EdinBurg Peer Reviewed Journals and Books Publishers
Journal of Information and Technology
Vol. 5||Issue 1||pp 16-24||March||2025
Email: info@edinburgjournals.org|

Benchmarking Machine Learning Models for Landslide Susceptibility: A Study in the Ngororero Sector

Jean de Dieu Hagenimana[1*], Djuma Sumbiri[2]
[1,2]Computing and Information Sciences, University of Lay Adventists of Kigali, Rwanda
Corresponding Author Emails: jados2@gmail.com; sumbirdj@gmail.com

How to Cite: de Dieu Hagenimana, J., & Sumbiri, D. (2025). Benchmarking Machine Learning Models for Landslide Susceptibility: A Study in the Ngororero Sector. *Journal of Information and Technology, 5*(1), 16-24.

Abstract

This study evaluates the performance of six machine learning algorithms: Decision Trees (DT), Neural Networks (NN), Support Vector Machines (SVM), Random Forests (RF), Gradient Boosting Machines (GBM), and k-nearest Neighbors (k-NN) for landslide prediction in the Ngororero sector, Rwanda. Using Sentinel-2 satellite imagery, meteorological data, and topographical datasets from 2015, 2019, and 2023, the study incorporates critical features such as slope, rainfall, soil type, and vegetation cover. The findings indicate significant temporal and algorithmic variations in prediction performance. K-Nearest Neighbors and Random Forest consistently achieved high accuracies, with kNN showing a value of 84% in the 2019 dataset and more than 80% in other datasets. Random Forest demonstrated robust performance with a 78.98% accuracy in 2015. The research concludes that k-nearest Neighbors and Random Forest are optimal for predicting landslides in the Ngororero sector.

Keywords: *Landslide prediction, Machine learning algorithms, Feature selection, Rwanda*

Received: 20 February 2025 ***Revised: 06 March 2025*** ***Accepted: 08 March 2025***

1. Introduction

Landslides are significant natural hazards that affect lives and infrastructure, particularly in mountainous areas. Urbanization, deforestation, and climate change exacerbate landslide risks by destabilizing slopes(Guzzetti et al., 1995). Heavy rainfall is a critical trigger for landslides(Uwihirwe et al., 2020), particularly in regions like Rwanda, where steep slopes and high population density compound vulnerability and caused 95 deaths in Ngororeo District during May 2023 (Ayalew et al., 2005; Nwazelibe et al., 2023). Rwanda is in Central Africa, bordered by Uganda, Tanzania, the Democratic Republic of the Congo, and Burundi. It covers an area of 26,338 km2, spanning latitudes from 1°04’ to 2°51’ south and longitudes from 28°45’ to 31°15’ east. As per the 2022 census, Rwanda’s population is approximately 13,246,394(NISR, 2022). The country’s landscape is predominated by mountains, with an average altitude of around 1700 meters, Mt. Karisimbi being the highest peak at 4507 meters above sea level(BİZİMANA & SÖNMEZ, 2015). While the central plateau features volcanic mountains and rolling hills, the eastern region is much flatter, with elevations less than 1.5 km. The varied terrain contributes to Rwanda’s climate, which is significantly influenced by altitude. Average annual temperatures are around 18.5°C, and yearly rainfall is about 1,250 mm(Henninger, 2013). However, the Western Province faces more rain-induced network disruptions than other regions. The lowlands of the southwest, particularly the Bugarama plain, represent a part of the African Rift Valley's tectonic depression and sit at an altitude of 900 meters(Sumbiri et al., 2016). Rwanda's topography and population density affect its vulnerability to floods and landslides. A study by researchers (Panelli et al., 2024) indicated that the northern, western, and southern provinces are primarily exposed to landslide hazards. The significant landslide hazard factors influence elevation, slope, rainfall, and poor land management. Rutsiro, Ngororero, and Karongi districts in Rwanda's Western Province have the highest landslide concentration. The relative importance of conditioning factors indicates that geology, rainfall, distance to the road, slope, and NDVI factors played a crucial role in landslides in the studied area. Twenty-nine of these landslides dammed rivers, resulting in high losses of fertile soils and infection of fresh waters with chemicals used to cultivate those lands; the types of landslides in Rwanda are influenced by different factors such as soil properties, slope steepness, undercutting of the slope, and external triggering factors such as rainfall and wind. A study by authors (Claude et al., 2020) indicated that "the residents are unaware of landslide causal factors due to low levels of education and training". Landslides in Rwanda, particularly in the Ngororero sector, are exacerbated by its hilly terrain and heavy seasonal rainfall, and no study has been conducted focusing on the landslide susceptibility in the area. This research addresses this gap by leveraging geospatial data from the Ngororero sector and evaluating machine learning models like Decision Trees, Random Forests, k-NN, SVM, and Gradient-Boosting Machines for landslide prediction. The unique approach of combining theoretical insights with empirical validation contributes to the growing knowledge of landslide forecasting, emphasizing scalable and effective solutions for regions with limited computational resources.

Integrating Sentinel-2 satellite imagery and geospatial data offers an unprecedented opportunity to assess landslide susceptibility more accurately(Montello & Rossi, 2024). Machine learning, as a subset of artificial intelligence, provides advanced capabilities for predictive analytics, enabling researchers to identify patterns and predict future occurrences based on historical data(Witten et al., 2011).

2. Literature Review

Landslide prediction remains a complex yet critical endeavor, particularly in regions with challenging topographies like Rwanda. Previous studies highlight the effectiveness of statistical and machine learning models in assessing landslide susceptibility. For instance, authors(Hong et al., 2016) utilized models like Evidential Belief Function and Random Forest to achieve notable prediction accuracy in Lianhua County, China. Similarly, (Pourghasemi & Rahmati, 2018) demonstrated the robust performance of Random Forest and Boosted Regression Trees in Iran, achieving accuracies of 83.7% and 80.7%, respectively. While ensemble methods and profound learning advancements have shown global potential, their application in resource-constrained environments such as Rwanda remains underexplored.

A recent study by Zeng et al. (2023) in Dazhou Town, China, introduced an ensemble framework to improve landslide susceptibility predictions. The stacking-based RF-XGBoost model demonstrated superior accuracy, outperforming standalone models like XGBoost, and was the most effective approach for identifying landslide-prone areas. In Rwanda, Nahayo et al. (2019) used Geographic Information Systems (GIS) and statistical methods to map landslide hazards, identifying factors like elevation and slope as key predictors. The resulting map categorized areas into five hazard classes ranging from very low to very high, with northern, western, and southern provinces identified as high-risk areas due to factors like elevation, slope, and poor land management. In a study on predicting landslides in the Ngororero District of Rwanda using Random Forest (RF) and Logistic Regression (LR), Kuradusenge et al. (2020) demonstrated that incorporating antecedent rainfall data significantly improved landslide predictions using Random Forest and Logistic Regression models(Kuradusenge et al., 2020).

Researchers(Nahayo et al., 2018) used Geographic Information Systems (GIS) and statistical methods to map landslide hazards, identifying factors like elevation and slope as key predictors. The resulting map categorized areas into five hazard classes ranging from very low to very high, with northern, western, and southern provinces identified as high-risk areas due to factors like elevation, slope, and poor land management.

3. Methodology

3.1 Study Area and Data Sources

The study focuses on the Ngororero sector in the Western Province of Rwanda. The region faces challenges related to environmental sustainability, including deforestation, soil erosion, and vulnerability to natural disasters such as landslides and flooding. The sector's varying elevation and slope create distinct microclimates and ecological zones, which are essential to understanding the environmental and socioeconomic dynamics of the area. The region's complex terrain, ranging from steep slopes to valleys, makes it highly susceptible to landslides. The location is predominantly rural, with agriculture being the main economic activity. This region has a history of landslide occurrences due to its hilly terrain and seasonal heavy rainfall (Kuradusenge et al., 2020; Sumbiri et al., 2016). It provides an ideal case study for examining predictive analytics in a landslide-prone area.

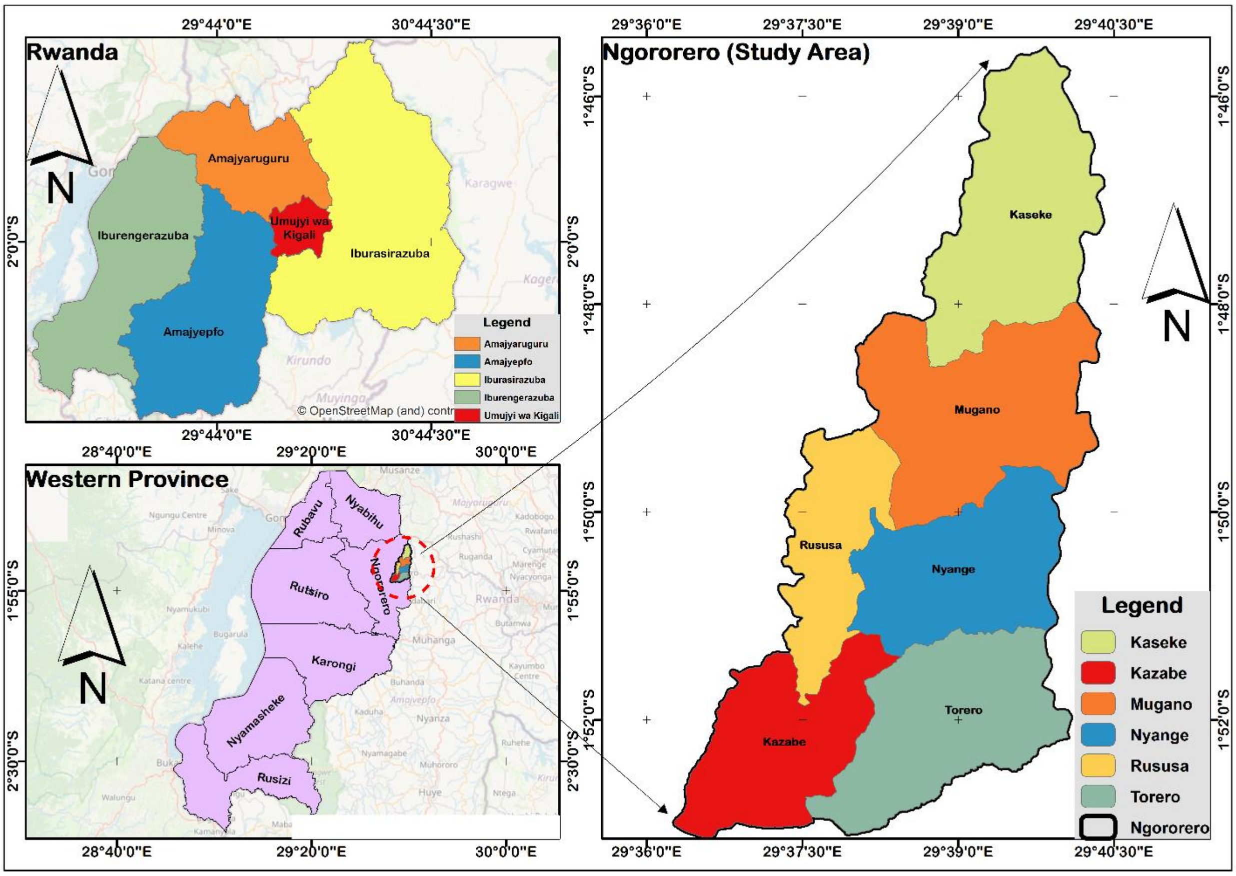

Figure 1: Study Area

The following data sources were utilized:

• Satellite imagery from Sentinel-2 for vegetation indices and land cover.

• Meteorological data from Meteo Rwanda detailing rainfall intensity and variability.

• Topographical data from ALOS Global DEM for elevation and slope gradients.

• Geological soil composition and texture data from Rwanda’s Ministry of Environment.

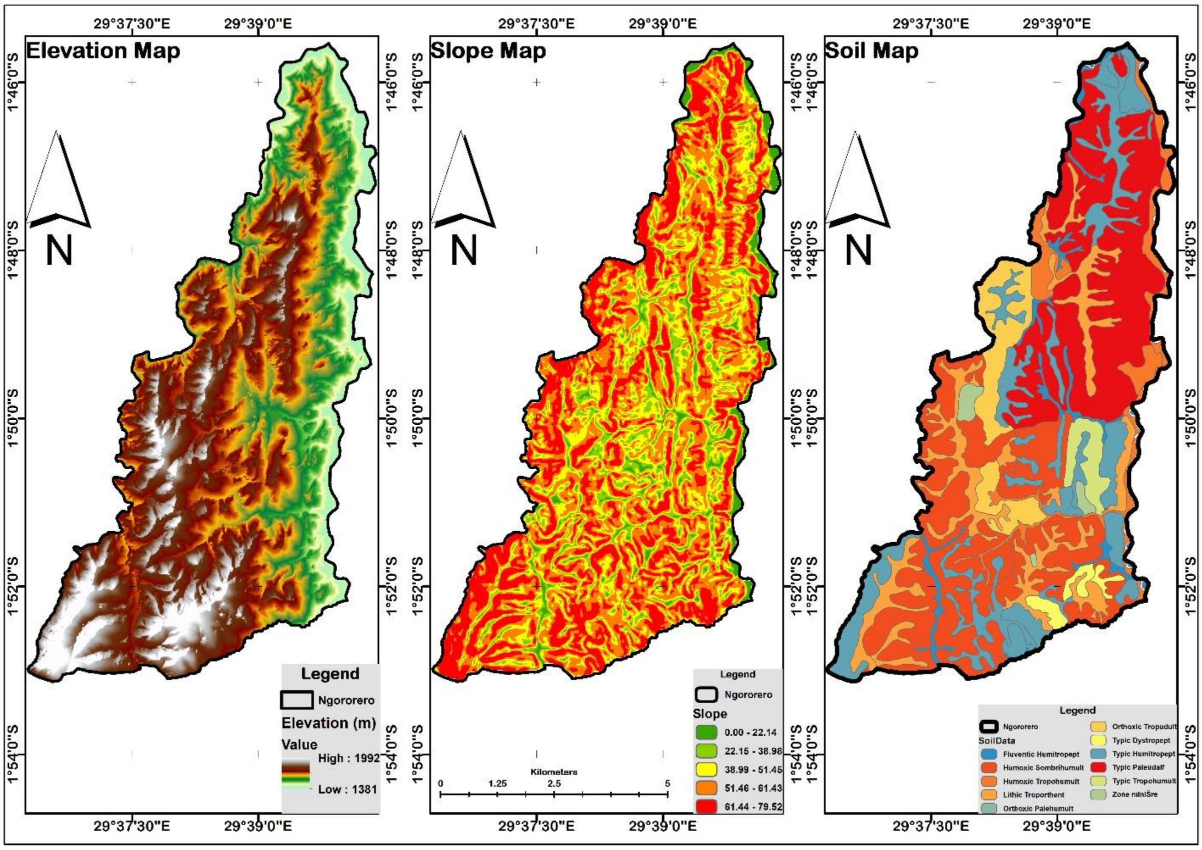

Figure 2: Rainfall map across 2015, 2019, and 2023

EdinBurg Peer Reviewed Journals and Books Publishers
Journal of Information and Technology
Vol. 5||Issue 1||pp 16-24||March||2025
Email: info@edinburgjournals.org|

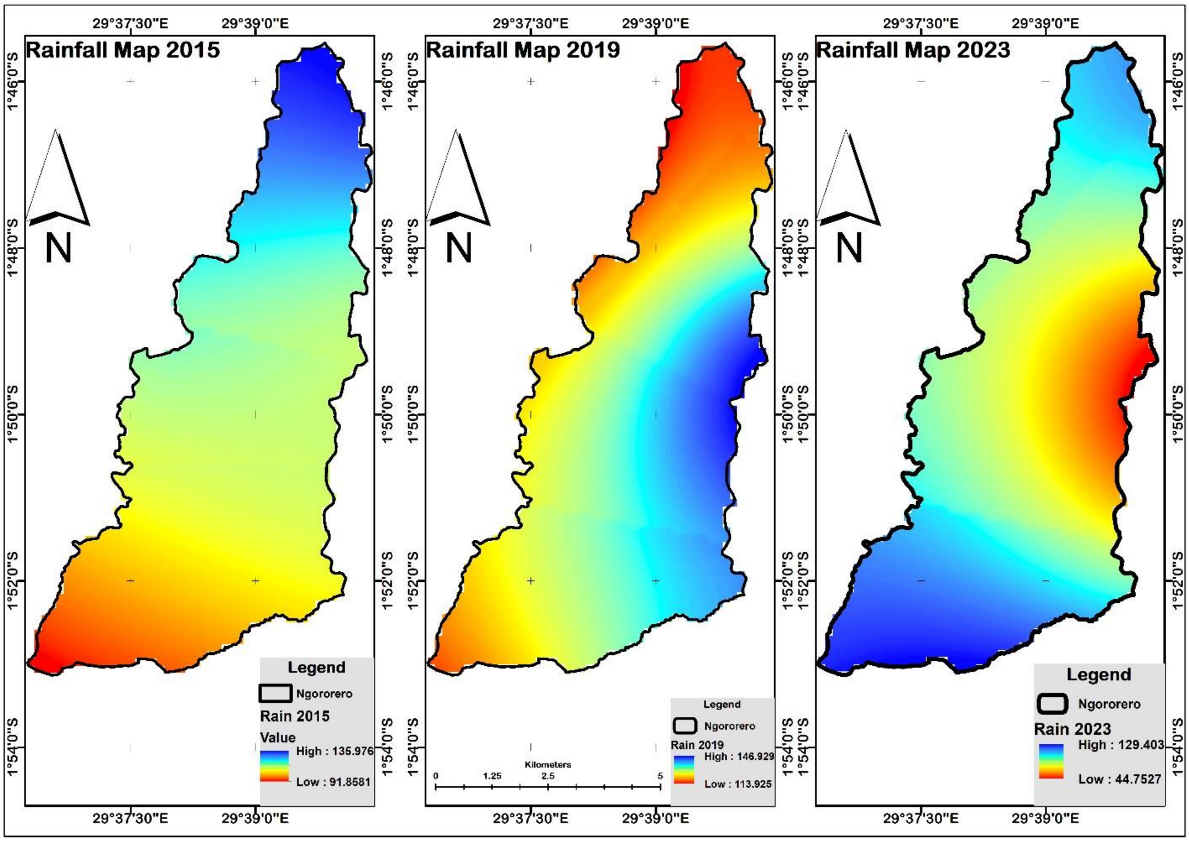

Figure 3: Elevation, Slope, and Soil map

3.2 Data Preprocessing

Data preprocessing included handling missing values using mean imputation for numerical features and mode imputation for categorical ones. Feature scaling was applied to ensure uniformity, particularly for variables such as rainfall and elevation. The key steps included:
1. Normalization: Scaled features to a 0–1 range for models sensitive to input magnitudes, such as SVM.

2. Feature selection: Prioritized slope, elevation, NDVI, and rainfall as key predictors based on domain expertise.

3. Class balancing: Used SMOTE to address the imbalance between landslide and non-landslide instances, enhancing model sensitivity. The preprocessed datasets covered 2015, 2019, and 2023, providing a temporal dimension for evaluating algorithmic performance.

3.3 Training and Validation

The datasets were split into training (70%) and testing (30%) subsets, with five-fold cross-validation ensuring robustness.

Grid search optimization was employed to fine-tune hyperparameters such as the depth of decision trees, the number of hidden layers in neural networks, and SVM kernel types. The SMOTE technique was employed to address class imbalance in the dataset.

4. Results and Discussion

The study compared the performance of Decision Tree, Neural Network, Support Vector Machine (SVM), K-Nearest Neighbors, Random Forest, and Gradient Boosting Machine (GBM) models across datasets from 2015, 2019, and 2023. Results showed that Neural Network, GBM, and Random Forest achieved consistently high accuracy, with Neural Network and GBM outperforming other models regarding precision, recall, and ROC-AUC. For instance, Neural Network and GBM achieved peak accuracies of over 99% in the 2023 dataset, demonstrating their adaptability and precision in predicting landslide-prone areas. Decision Tree and SVM showed vital accuracy but were less robust in handling the 2023 dataset, which presented greater complexity.

Model Performance

A detailed evaluation revealed the following insights:

- ✓K-Nearest Neighbors (k-NN) was the best-performing model, achieving 84.61% accuracy in 2019.
- ✓Random Forest also performed well, with high accuracy and AUC scores.
- ✓Gradient Boosting Machine (GBM) had the lowest performance, especially in 2015
- ✓Decision Trees showed high recall (80.52% in 2023), making them useful for landslide detection.

Impact of Temporal Variability

- ✓Model accuracy improved over time due to stronger correlations between rainfall, slope, and landslides.

Rainfall's correlation with landslides increased from 0.57 (2015) to 0.87 (2023).

- ✓Data quality enhancements (normalization, oversampling) improved model performance.

Table 1: Year-by-Year Model Comparisons

Metric	***Best Model (2015)***	***Best Model (2019)***	***Best Model (2023)***
Accuracy	k-NN (80.85%)	k-NN (84.61%)	k-NN (80.56%)
Precision	k-NN (81.35%)	k-NN (84.77%)	Decision Tree(84.35%)
Recall	k-NN (80.61%)	k-NN (80.61%)	k-NN (80.56%)
F1-score	k-NN (80.94%)	k-NN(84.34%)	k-NN (80.74%),Decision Tree(80.74%)
ROC-AUC	k-NN (89%)	k-NN (92%)	k-NN(90%)

5. Conclusion

The study compared the performance of Decision Tree, Neural Network, Support Vector Machine (SVM), K-Nearest Neighbors, Random Forest, and Gradient Boosting Machine (GBM) models across datasets from 2015, 2019, and 2023. Results showed that K-Nearest Neighbors and Random Forest achieved consistently high accuracy and outperformed in precision, recall, F1-Score, and ROC-AUC. For instance, K-Nearest Neighbors and Random Forest achieved peak accuracies of over 80% and 78% across all datasets, demonstrating their adaptability and precision in predicting landslide-prone areas

6. Recommendations

This research recommends adopting K-Nearest Neighbors and Random Forest models to enhance landslide prediction accuracy and model robustness, as they consistently demonstrated high accuracy and generalization across datasets.

Given the importance of accurate and comprehensive data, it is recommended that more detailed datasets like high-resolution satellite imagery be integrated. Including real-time data sources such as weather updates and geospatial information would further enhance the predictive capabilities of the models.

Neural Networks can be improved through careful parameter tuning and data pre-processing in areas where more complex relationships exist. Although GBM was less reliable in this study, it may benefit from further optimization, particularly in handling complex and high-dimensional datasets.

ORCID

Jean de Dieu HAGENIMANA https://orcid.org/0000-0003-0360-4806

References

Ayalew, L., Yamagishi, H., Marui, H., & Kanno, T. (2005). Landslides in Sado Island of Japan: Part II. GIS-based susceptibility mapping with comparisons of results from two methods and verifications. *Engineering Geology*, *81*(4), 432–445. https://doi.org/10.1016/J.ENGGEO.2005.08.004

BİZİMANA, H., & SÖNMEZ, O. (2015). Landslide Occurrences in The Hilly Areas of Rwanda, Their Causes and Protection Measures. *Disaster Science and Engineering*, *1*(1), 1–7. https://dergipark.org.tr/en/pub/dse/issue/34556/381756

Claude, M. J., Martin, N. V., Abias, M., Francoise, M., Johnson, U., Tonny, K., & Martine, U. (2020). Mapping Landslide Susceptibility and Analyzing Its Impact on Community Livelihoods in Gakenke District, Northern Rwanda. *Journal of Geoscience and Environment Protection*, *08*(05), 41–55. https://doi.org/10.4236/gep.2020.85003

Guzzetti, F., Carrara, A., Cardinali, M., & Reichenbach, P. (1995). Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology*, *13*(6), 1995.

Henninger, S. M. (2013). Does the global warming modify the local Rwandan climate? *Natural Science*, *05*(01), 124–129. https://doi.org/10.4236/ns.2013.51a019

Hong, H., Pourghasemi, H. R., & Pourtaghi, Z. S. (2016). Landslide susceptibility assessment in Lianhua County (China): A comparison between a random forest data mining technique and bivariate and multivariate statistical models. *Geomorphology*, *259*, 105–118.

https://doi.org/10.1016/J.GEOMORPH.2016.02.012

Kuradusenge, M., Kumaran, S., & Zennaro, M. (2020). Rainfall-induced landslide prediction using machine learning models: The case of ngororero district, rwanda. *International Journal of Environmental Research and Public Health*, *17*(11), 1–20. https://doi.org/10.3390/ijerph17114147

Montello, F., & Rossi, C. (2024). *LANDSLIDE MAPPING FROM SENTINEL-2 IMAGERY THROUGH CHANGE DETECTION*.

Nahayo, L., Li, L., & Mupenzi, C. (2018). Spatial Distribution of Landslides Vulnerability for the Risk Management in Ngororero District of Rwanda. *Eajst*, *8*(2), 1–14.

NISR. (2022). Fifth Population and Housing Census - 2022. *National Institute of Statistics of Rwanda*.

Nwazelibe, V. E., Egbueri, J. C., Unigwe, C. O., Agbasi, J. C., Ayejoto, D. A., & Abba, S. I. (2023). GIS-based landslide susceptibility mapping of Western Rwanda: an integrated artificial neural network, frequency ratio, and Shannon entropy approach. *Environmental Earth Sciences*, *82*(19), 1–23. https://doi.org/10.1007/s12665-023-11134-4

Pourghasemi, H. R., & Rahmati, O. (2018). Prediction of the landslide susceptibility: Which algorithm, which precision? *CATENA*, *162*, 177–192. https://doi.org/10.1016/J.CATENA.2017.11.022

Sumbiri, D., Afullo, T. J. O., & Alonge, A. (2016). Rain attenuation prediction for terrestrial links at microwave and millimeter bands over Rwanda. *Progress in Electromagnetic Research Symposium (PIERS)*, 4233–4236. https://doi.org/10.1109/PIERS.2016.7735592

Uwihirwe, J., Hrachowitz, M., & Bogaard, T. A. (2020). Landslide precipitation thresholds in Rwanda. *Landslides*, *17*(10), 2469–2481. https://doi.org/10.1007/s10346-020-01457-9

Witten, I. H., Frank, E., & Hall, M. A. (2011). Data Mining. In *Data Mining*. https://doi.org/10.1016/C2009-0-19715-5