SPATIALLY AWARE LANDSLIDE SUSCEPTIBILITY PREDICTION USING A GEOGRAPHICAL RANDOM FOREST APPROACH

A. Teke[1*], T. Kavzoglu[1]

[1]GTU, Department of Geomatics Engineering, 41400 Gebze Kocaeli, Turkey - (a.teke2020, kavzoglu)@gtu.edu.tr

KEY WORDS: Landslide Susceptibility, Geographical Random Forest, Machine Learning, Mapping Unit, Spatial Autocorrelation.

ABSTRACT:

Landslide susceptibility prediction practices have been increasingly reliant on non-geographically-oriented (i.e., aspatial) machine learning algorithms. While these approaches have exhibited increasing success, they have often faced criticism for their limited consideration of spatial autocorrelations and local variations across geographical space, thereby neglecting the concept of spatial non-stationarity. To fulfill the research gap, this work applies a geographical random forest (GRF) approach, contrasting it with the conventional random forest (RF) algorithm. To this end, the study area, encompassing the Lake Sapanca Basin and its surroundings, was subdivided into 4,452 slope-based mapping units. The effectiveness of both predictive models was then measured by using overall accuracy (OA) and area under the curve (AUC). The results revealed that the GRF (OA = 80.82% and AUC = 85.22%) outperformed the RF algorithm (OA = 75.34% and AUC = 82.50%) by approximately 5% in OA, and demonstrated a 3% improvement in AUC score. The Wilcoxon signed-rank test confirmed significant differences (95% level) between the predictions of both models. The slope parameter emerged as the globally most influential factor, but local interpretations disclosed notable variations in the importance of causative factors contingent upon location. For instance, the curvature parameter was the most important geospatial covariate in around one-third (34.23%) of the slope units, mostly concentrated in the northernmost zones of the study area. On the other hand, elevation was the most important factor for 14.67% of the slope units primarily located in the southern region.

1. INTRODUCTION

Landslide occurrences are complex natural events characterized by non-linear behaviours and multifaceted mechanisms, making it challenging to pinpoint their exact triggers (Kavzoglu et al., 2021). They result in severe material damage, substantial financial losses, loss of human lives, and disrupt the Earth's natural equilibrium by altering its surface (Kutlug Sahin et al., 2017). To mitigate these impacts, crafting landslide susceptibility maps emerges as a crucial strategy. These maps can depict the spatial arrangement of landslide and non-landslide areas and are widely acknowledged as a crucial tool and cornerstone in contemporary literature.

A plethora of approaches has been introduced to predict landslide susceptibility since the 1970s with expert-based notes taken on paper (Brabb et al., 1972). Then, the potent alternatives emerged as heuristic or knowledge-driven models based on the subjective judgment of the assessor. Subsequently, the spotlights firmly turn to the data-driven models, initially in a bivariate context and later evolving into multivariate approaches. Presently, the literature is heavily populated by machine learning algorithms, such as support vector machines (Kavzoglu et al., 2014), decision trees (Arabameri et al., 2021), random forest (Yilmaz et al., 2022), LightGBM (Sun et al., 2023), CatBoost (Ye et al., 2022) and XGBoost (Kavzoglu and Teke, 2022a), offering more flexible alternatives to multivariate statistical tools and significantly improving predictive capabilities.

The current landscape of the literature has been increasingly reliant on non-geographically oriented (i.e., aspatial) machine learning algorithms, indicating a virtual monopoly in the field. While these approaches have exhibited increasing success, they have often faced criticism for their limited consideration of spatial autocorrelations and local variations across geographical space (Chalkias et al., 2020), thereby neglecting the concept of spatial non-stationarity. These problems can also result in generalized models that inadequately capture the nuances of different regions. When these spatial elements are ignored, models might make predictions that do not seamlessly align with real-world spatial constraints. Until now, a series of studies has used geographically weighted models, enabling the incorporation of spatial context into the analysis of landslide susceptibility, such as Geographically Weighted Regression (GWR) (Chalkias et al., 2020), Geographically Weighted Logistic Regression (GWLR) (Gu et al., 2022), Geographical Random Forest (GRF) (Quevedo et al., 2022), and Geographically Weighted Artificial Neural Networks (GWANN) (Zhao et al., 2024). However, these studies mostly used the grid cells (i.e., pixels) as the main mapping unit. Even though spatial resolution may be prioritized in these studies, the units of grid cells might prove too localized to effectively capture unstable conditions, particularly when predicting larger-scale slide failures. Additionally, the landslide susceptibility maps generated through pixel-based methods often consist of cells at metric resolutions, each assigned a specific susceptibility value, lacking spatial coherence or connectivity constraints among neighbouring pixels within an individual slope (Martinello et al., 2021). Consequently, selecting a suitable method for the delineation of terrain surfaces has become another problematic topic especially when dealing with models accounting spatial heterogeneity. Even though researchers mostly use the pixels or grids as default mapping units to segment the landslide-prone zones, they may pose certain drawbacks in some cases. These include susceptibility to spectral-mixing, where diverse surface characteristics within a single pixel can confound the interpretation of landslide-prone areas. Additionally, the computational intensity associated with these methods can significantly slow down analyses and increase resource requirements. Also, as spatial resolution increases, individual pixels may cease to adequately represent the defining traits of classification targets.

Motivated by the research gaps mentioned above, this study employs a geographical random forest (GRF) approach using slope-based mapping units, aiming to address the constraints presented by traditional aspatial landslide susceptibility modeling methods and grid mapping units. This work also contrasts the GRF approach with the conventional random forest (RF) algorithm to compensate for these limitations effectively.

2. STUDY AREA AND DATASET

The study area, encompassing the Lake Sapanca Basin and its surroundings, is situated within a tectonic depression (Figure 1). Spanning approximately 16 kilometres in length and reaching a maximum width of 6 kilometres, the lake displays an elongated shape and is distinguished by its clear warm waters, exhibiting oligotrophic and monomictic properties (Duru, 2017; Temiz et al., 2022). The topography of the Lake Sapanca Basin is characterized by fault-controlled mountain ranges, forming a complex arrangement of horsts and grabens. Tectonic activity in the region has resulted in various faults and folds, notably the North Anatolian Fault situated north of the basin, a highly active fault responsible for significant earthquakes in the area's history.

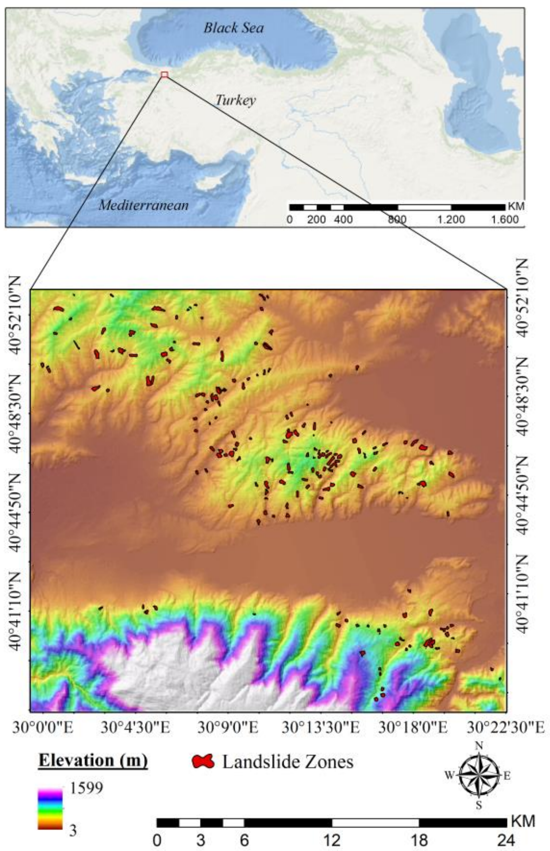

Figure 1. Location map of the study area and landslide zones.

The expansion of human settlements and population growth in the Lake Sapanca Basin has significantly changed the land, making it more prone to erosion (Ikiel, 2022). This increased urban and rural development has strained the basin's capacity to support life, causing harm to the natural environment. Over the last thirty years, settlements have extended beyond their limits due to industrial growth, leading to higher population density. Consequently, this has resulted in the loss of vegetation in the study area, elevating the risk of erosion.

In the literature, a guiding ideology integral to creating landslide susceptibility maps is that past landslide occurrences in specific regions are indicative of potential future events in similar areas. Consequently, landslide inventory maps hold significant importance in subsequent stages as they furnish vital details like the type, precise location, and extent of landslides. This research utilized a historical database of landslides sourced from the General Directorate of Mineral Research and Exploration (GDMRE) of Turkey, which is part of the "Turkey Landslide Inventory Map" project. The main aim was to identify where landslides occurred across the country and create a comprehensive digital database on a national level. To achieve this, areas prone to landslides were marked on a detailed map using remote sensing methods such as aerial imagery and on-site surveys, at a scale of 1: 25,000. This resulting inventory is accessible through a dedicated web portal that complies with legal regulations (http://yerbilimleri.mta.gov.tr/home.aspx). The identified landslide zones were then transformed into a raster format, resulting in 9,181 pixels or grid cells. Approximately 827 hectares of land within the basin experienced landslides, varying in size from 2363 m^2 to 139,499 m^2 per individual landslide area. Each pixel or sample in the inventory represents a ground surface area of 30x30 meters.

Major Factors	Covariates	Source	Scale/Resolution
Topographical	Aspect	SRTM - DEM	30 m
	CI		
	Curvature		
	Elevation		
	Plan Curvature		
	Profile Curvature		
	Slope		
	TPI		
	TRI		
	Valley Depth		
Hydrological	TWI	SRTM - DEM	30 m
Geology	Lithology	GDMRE	1:100,000
Environmental	LULC	ESRI Sentinel-2 LULC Time Series	10 m

Table 1. Data source and scale/resolution of geospatial covariates.

Expanding on the prior studies and analyzing the main features of the basin, 13 geospatial covariates (i.e., landslide-related parameters) were initially chosen (Table 1). Most of these factors, such as aspect, convergence index (CI), curvature, elevation, plan curvature, profile curvature, slope, topographic position index (TPI), topographic roughness index (TRI), topographic wetness index (TWI), and valley depth, were derived from the digital elevation model of the Shuttle Radar Topography Mission, with a spatial resolution of 30 meters. The lithology map at a scale of 1:100,000 was obtained from GDMRE. Additionally, the land use land cover (LULC) map was acquired from the ESRI Sentinel-2 10m Land Use/Land Cover Time Series, derived from ESA Sentinel-2 imagery at a resolution of 10 meters, available globally.

Aside from collecting landslide samples, identifying non-landslide instances is another crucial step in landslide susceptibility mapping practices. This study adopts the methodology proposed by Gómez and Kavzoglu (2005). The approach is based on selecting non-landslide samples exclusively from areas that are completely free of landslides, mirroring how landslide instances are gathered from high-risk regions. This method operates on the premise that terrains with slopes less than 5% and river channels are unlikely to experience landslides. To ensure a balanced dataset and avoid biases in the analysis, an equivalent number of non-landslide pixels were collected following the procedure, matching the total count of landslide cases.

3. METHODOLOGY

The methodological design of this work comprises six primary steps. Initially, the focus is on gathering landslide-related parameters and collecting samples from both landslide and non-landslide areas. These factors are then amalgamated into a multi-layer image composite to distinguish between landslide and non-landslide regions. Subsequently, the slope units within the basin were computed with the LaGriSU toolpack. For the construction of predictive models, 70% of the samples were utilized to train both GRF and RF models, while the remaining data was reserved for accuracy evaluation. The third step involves analyzing potential correlations among geospatial covariates within the database through multicollinearity analysis. Following this, the GRF and RF algorithms were employed to generate landslide susceptibility maps, and a comprehensive examination of both the global and local interpretations of the GRF algorithm was conducted to ascertain the dominant influence governing landslide activities within the basin. The performance of the susceptibility maps was assessed using four widely recognized accuracy criteria. Finally, any potential disparities in accuracy were evaluated using the Wilcoxon sign-ranked test.

3.1 Slope-based Mapping Units

Segmentation of landscape into terrain units is an important concern for discussion in the literature. Until now, a series of mapping units, such as grid cells, aspect units, unique-condition units, small watershed units, topographic units, and administrative divisions, has been used to subdivide the landslide-susceptible area under investigation. Ideally, the goal is to maximize homogeneity within each mapping unit or among the identities forming a mapping unit while permitting maximum heterogeneity among different units. In line with this aim, slope units delineated by drainage and divide lines align well with these expectations. Unlike their counterparts such as unique-condition units or pixels, slope units are closely associated with the geomorphological and hydrological processes by reflecting the physical features and characteristics of the terrain (Alvioli et al., 2016).

The study area initially consisted of 771x493 grid cells. The slope units were delineated using LaGriSU (Landslide Grid and Slope Units) QGIS toolpack v 0.2 embedded within the QGIS platform (Althuwaynee, 2021). Through the integration of SAGA and GDAL libraries within QGIS's hydrological analysis module, this tool creates catchment basins in positive relief, based on the original DEM data. The toolpack requires only landslide locations with either point or polygon features and DEM. Once applied the framework, a total of 4452 slope units were extracted.

3.2 Random Forest (RF)

Random Forest (RF) (Breiman, 2001), a member of the ensemble machine learning family, stands as a well-established model rooted in individual decision trees (DTs). One of the central problems encountered when applying DTs is the propensity for overfitting—wherein the model excessively intricacies itself to fit the training data, resulting in a compromise in generalization when confronted with test data. The RF method adeptly addresses this concern by embracing an ensemble learning framework, crafting a forest of decision trees via bootstrap aggregation, widely recognized as the bagging approach.

The advantage of RF lies in its ability to curtail overfitting by constructing multiple decision trees and amalgamating their outputs. Each tree, generated using the DT algorithm, embarks on a recursive process. It segments the data into subsets based on selected features, striving at every step to maximize information gain or minimize impurity. This recursive partitioning process forms a multitude of diverse decision trees, each capturing unique aspects of the dataset. When predictions are to be made using the RF algorithm, the collective wisdom of all trees within the forest is harnessed. For classification tasks, the final prediction is determined by the majority vote among the constituent trees. Conversely, in regression tasks, the amalgamation of predictions yields the mean value, ensuring a comprehensive and balanced outcome.

3.3 Geographical Random Forest (GRF)

The RF algorithm has long been used across various domains (Colkesen et al., 2023; Kavzoglu and Teke, 2022b; Tonbul et al., 2022). However, its conventional application lacks spatial awareness, potentially neglecting the complex spatially heterogeneous nature of phenomena such as landslides and the variability present in remotely sensed data. To tackle this limitation, Georganos et al., (2021) introduced the GRF as an extension of RF encompassing the spatial considerations. The working principle of GRF is analogous to the GWR. Specifically, it creates localized RF models in which a number of nearby observations are grouped for each geographic location. Hence, this can enhance the RF models by calibrating them locally, and boost their adaptability to diverse spatial contexts. Every data point is linked to a submodel focusing solely on neighbouring observations within a defined geographic area termed the 'neighborhood' or 'core'. The bandwidth, representing the distance between a data point and its kernel, dictates the scope of these subpatterns. Typically, two kinds of kernels are employed: 'adaptive' and 'hard'. While the former is determined by the n nearest neighbours, the latter utilizes a circular radius as its bandwidth.

3.4 Accuracy Assessment Metrics

To accurately assess the effectiveness of predictive models in identifying landslide-prone areas, evaluating the generated maps is crucial. In our study, four important accuracy assessment metrics - area, under curve (AUC), accuracy (OA), sensitivity, and specificity – were considered to measure the model's performance and prediction capabilities. AUC is a commonly used metric in landslide susceptibility mapping, as it evaluates the model's ability to distinguish between different thresholds while balancing sensitivity and specificity. AUC values range from 0 to 1, with higher values indicating better discriminatory power in identifying landslide-prone areas compared to non-landslide areas. The higher the AUC, the better the performance of the models at separating between the landslide and non-landslide samples. The OA is another widely used metric measuring the degree of agreement between predicted and observed samples of landslides and non-landslides. It offers a comprehensive assessment of how well a model can predict outcomes. Sensitivity, also referred to as the true positive rate, reflects the model's ability to accurately recognize areas affected by landslides. It indicates the proportion of true landslide occurrences that the model correctly identifies, showcasing its accuracy. Similarly, specificity, or the true negative rate, demonstrates how well the model can correctly distinguish non-landslide areas.

3.5 Statistical Significance

It is essential to utilize both statistical significance testing and accuracy assessment metrics in order to accurately compare the performance of landslide susceptibility maps generated by different machine learning algorithms. The application of the Wilcoxon signed-rank test, a non-parametric statistical test, was integral in determining the discrepancies between the predictions of various models in this study.

The test relies on a null hypothesis, assuming that there is no difference in the performances of the models by exploring p-values and z-values to assess the likelihood of either rejecting or accepting this hypothesis. Different significance levels (e.g., %1 or 5%) can be used to refute or accept the hypothesis. When the p-value falls below the specified threshold of 0.05 and the z-value exceeds the critical margins of -1.96 or +1.96, a significant difference in model performances becomes evident, resulting in the clear rejection of the null hypothesis. Conversely, if the p-value is at or above 0.05 and the z-value stays within the range of -1.96 to +1.96, the null hypothesis remains valid, indicating an absence of substantial variation between the models being examined.

3.6 Programming Language and Reproducibility

Once obtaining the slope units in the QGIS platform using the LaGriSU toolpack, all analysis was carried out in R Studio with the following main libraries: *tidyverse* for data manipulation and handling, *sf* and *BAMMtools* for importing and analyzing the geospatial covariates, *randomForest* for the application of regular RF, and *SpatialML* for GRF. To provide reproducibility and increase the transparency of the study, the main script of the code was shared in the following GitHub repository: (https://github.com/tekeali/SpatiallyAwareLSM).

4. RESULTS & DISCUSSION

In the initial phase of this study, a multicollinearity test was conducted whether there exists a potential issue undermining the goodness-of-fit of the artificially intelligent models. The test involved the computation of the variance inflation factor (VIF) and tolerance (TOL) for each covariate, as presented in Table 2. Multicollinearity is a common problem in modeling when predictor variables are highly correlated, as it can affect the accuracy of the individual effects of landslide-related parameters and compromise the reliability of the model's predictions. To identify potential multicollinearity among the geospatial covariates, two indicators are commonly used: tolerance (TOL) and variance inflation factor (VIF). Generally, a VIF score above 10 or a TOL score below 0.1 indicates significant multicollinearity. In our study, the VIF scores ranged from 1.10 to 5.31, suggesting varying degrees of interdependence among the geospatial covariates, while the TOL scores ranged from 0.19 to 0.91. Despite the presence of some multicollinearity, our results suggest that it may not significantly impact the accuracy of the model's outcomes.

Covariates	Tolerance	VIF
Aspect	0.91	1.10
CI	0.29	3.48
Curvature	0.20	5.05
Elevation	0.27	3,75
Lithology	0.79	1.26
LULC	0.83	1.20
Plan Curvature	0.41	2.41
Profile Curvature	0.29	3.40
Slope	0.23	4.42
TPI	0.19	5.31
TRI	0.28	3.56
TWI	0.23	4.34
Valley Depth	0.33	3.04

Table 2. Multicollinearity test for geospatial covariates.

Following the initial analyses, a series of accuracy assessment metrics, including area under curve (AUC), overall accuracy (OA), sensitivity, and specificity, were used to measure the predictive performances of both algorithms. The results revealed significant differences in performance between the two artificially intelligent models. The GRF demonstrated a significantly higher overall accuracy score of 80.82% compared to RF's 75.34%. Additionally, the GRF showed superior discriminatory power with an AUC value of 85.22%, surpassing RF's AUC of 82.50%. When it came to correctly identifying positive cases, the GRF also outperformed RF with a sensitivity rate of 80.49% compared to RF's 74.42%. Similarly, GRF showed higher proficiency in specificity with a score of 81.25%, surpassing RF's 76.67%. The differences in performance could be attributed to the unique mechanisms of each algorithm. GRF, specifically engineered to incorporate spatial elements, likely benefits from this spatial context, enhancing its predictive accuracy. By integrating this important aspect of the landslide susceptibility mapping practices, the

GRF might excel in identifying localized patterns or relationships within the data that RF, devoid of geographic considerations, might overlook. Consequently, the inclusion of geographical factors in GRF might have contributed to its superior performance metrics across accuracy, AUC, sensitivity, and specificity when compared to RF. These results underscore the significance of feature relevance and the impact of incorporating contextual information, such as spatiality, in improving the predictive power of machine learning models.

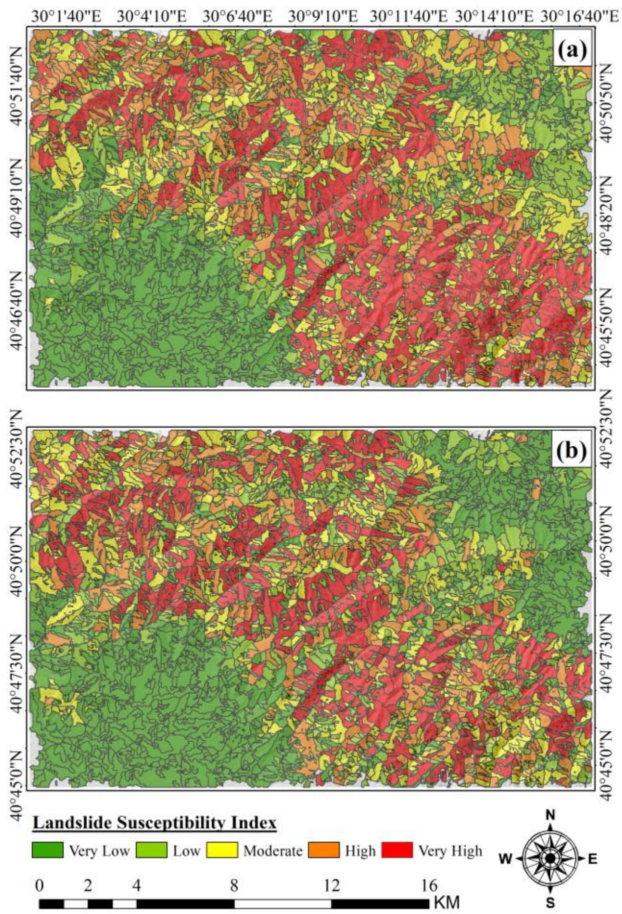

Figure 2. Landslide susceptibility maps with slope-based mapping units produced by (a) GRF and (b) RF.

The produced landslide susceptibility maps were discretized into five distinct categories—very low, low, moderate, high, and very high—through the application of the Jenks classification approach (as illustrated in Figure 2). The areas characterized by high and very high landslide susceptibility were primarily clustered in the north-western and south-eastern sectors of the Lake Sapanca Basin, diverging from expectations of concentration along the lakeshore areas. This divergence can be ascribed to the distinctive topography prevalent in the northwestern sectors, characterized by steeper slopes and elevated terrain in contrast to the relatively flatter lakeshore regions. Such topographical disparities contribute to increased soil saturation and diminished frictional resistance, thereby amplifying the susceptibility to landslides. On the other hand, the distribution of landslide susceptibility zones varied across landslide susceptibility maps, displaying some inconsistencies. The GRF model exhibited notably increased areas categorized as very high and high landslide susceptibility zones, predominantly concentrated in the north-central region of the area under study.

The application of the Wilcoxon signed-ranked test to ascertain the statistical significance in predictive performances between the two algorithms yielded compelling results. The test confirmed a statistically significant difference between the predictions generated by these models at a 95% significance level. This statistical significance suggests that the variation in performance metrics, such as accuracy, AUC, sensitivity, and specificity, observed between GRF and RF is not merely due to chance but holds substantive significance. Consequently, these findings reinforce the notion that GRF and RF indeed produce predictions with significantly divergent levels of accuracy and effectiveness in discerning patterns within the dataset. The validation of this statistical significance bolsters the confidence in the superiority of one model over the other, providing substantial evidence to support the selection or preference for one algorithm based on its demonstrated performance.

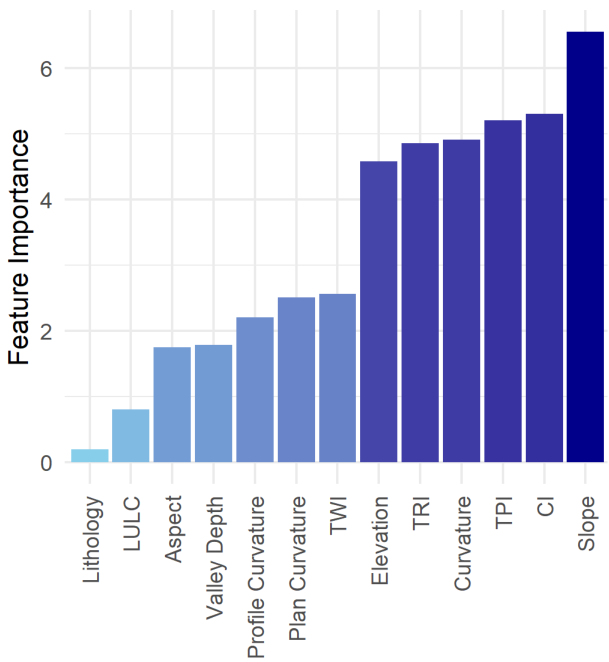

Figure 3. Global explanation of geospatial covariates with GRF.

Both the global and local interpretation of the GRF algorithm was also made (Figure 3). The results showed that the slope parameter emerged as the globally most influential factor, but local interpretations disclosed notable variations in the importance of causative factors contingent upon location. For example, the curvature parameter was the most important geospatial covariate in approximately one-third (34.23%) of the slope units, mostly located in the northernmost zones of the basin. On the other hand, elevation was the most important landslide-related parameter for 14.67% of the slope units primarily located in the southern sector of the study area. These results underscore the importance of comprehending the spatial patterns and site-specific characteristics inherent in landslide conditioning factors linked to landslide occurrences. Understanding the intricate spatial dynamics affecting landslide susceptibility can greatly strengthen the predictive and preventive capabilities for mitigating landslide incidents.

While many studies have utilized a variety of mapping units such as grid cells and pixel-based approaches, this study diverges from conventional methods and instead focuses on slope units and incorporates the spatially weighted GRF method. This approach aims to promote homogeneity within each unit while also allowing for diversity among different units, as the hydrological and geomorphological conditions within natural landscapes are closely linked to these units. For instance, Chowdhuri et al. (2020) experimented with the evidential belief function (EBF), GWR, and RF method and their ensemble forms (RF-EBF and RF-GWR). RF-EBF and RF-GWR models showed prediction capabilities of 91.8% and 89.9%, respectively. Likewise, Zhao et al. (2024) used artificial neural networks, support vector machines, RF, GWR, and GWANN to predict landslide susceptibility of Yichang City, Hubei Province. The GWANN model exhibited the strongest performance (AUC = 0.788) among all models tested. It was closely followed by the ML model ANN (AUC = 0.771), SVM (AUC = 0.754), and RF (AUC = 0.759). However, the GWR model displayed the lowest performance (AUC = 0.738) among the models evaluated. Furthermore, the authors highlighted that employing grid cells with a 90-meter spatial resolution presents challenges in storing the spatial distance matrix and factor weight matrix for extensive study areas due to limitations in hardware memory and computing power. They proposed leveraging distributed computing techniques to process extensive datasets, thereby enhancing the model's predictive capabilities at a local scale. Despite the obtained satisfying performances, these studies mainly adopted the pixel-based mapping unit, which might limit the nuanced understanding of spatial dynamics of the landslide phenomena.

In the literature, there are several tools available for generating slope units, such as r.*slopeunits* (Alvioli et al., 2016) and the recently developed Slope Unit Maker (Woodard et al., 2023). r.*slopeunits* has been utilized and tested across various studies (e.g., Aguilera et al., 2022; Camilo et al., 2017; Schlögel et al., 2018), establishing a certain level of trust and reliability within the literature. On the other hand, LaGriSU may require an evaluation of its stability, interoperability, and robustness given its relatively limited adoption in landslide susceptibility mapping practices. Therefore, future studies may consider the integration of segmentation quality metrics to understand the effectiveness of this particular tool.

Another issue worth discussing here is the adopted terrain segmentation methodology, which can significantly influence the landslide prediction process. This decision plays a critical role as it affects how features are represented, how sensitive the model is, how efficiently it runs, and the overall quality of the segmentation. The segmentation scale directly determines the level of detail versus generalization captured in terrain attributes, necessitating an exploration of the trade-offs between granularity and oversimplification. In the context of this study, the exclusive reliance on slope units as the primary mapping units holds numerous advantages, particularly in capturing terrain-related variables relevant to landslide prediction. However, other potentially valuable mapping units might have been overlooked. Alternative segmentation methodologies, such as those rooted in diverse geomorphological and hydrological features (such as grid units, catchment areas, and aspect-based units), or the adoption of object-oriented image segmentation (known as OBIA or GEOBIA), could provide a more holistic representation of terrain characteristics impacting landslides. These different units might capture distinct aspects of the landscape that, when integrated, could enhance the quality and accuracy of landslide susceptibility maps. Recognizing this, our future studies aim to incorporate a broader spectrum of mapping units derived from varied segmentation methodologies.

Beyond the chosen mapping unit for landslide susceptibility prediction practices, the process of down-sampling and up-sampling may introduce inconsistencies, which is a matter evident in numerous studies within the literature. In our endeavour, various sources were tapped to procure geospatial covariates, some bearing distinct spatial resolutions. For instance, DEM derivatives maintained a spatial resolution of 30 meters, while the LULC map derived from Sentinel-2 boasted a finer resolution of 10 meters. To ensure a more uniform dataset for analysis, a deliberate choice was made in favor of a 30-meter spatial resolution. This decision aimed to mitigate virtual resolution increases and curtail potential distortions or biases stemming from resolution disparities. By adopting this approach, our study sought to create a more consistent and reliable dataset, thereby reducing the potential distortions or biases stemming from resolution variations. The study produced slope units from these pixels, strategically introducing a terrain-related variable that exhibits greater constancy across diverse spatial scales.

5. CONCLUSIONS

This study aims to improve the accuracy of the produced landslide susceptibility maps by applying the GRF approach integrated with the slope-based mapping units, thereby alleviating the limitations posed by non-geographically oriented approaches and grid mapping units. Furthermore, it endeavours to assess the comparative effectiveness of this approach against the conventional RF algorithm, intending to establish a more coherent and precise predictive model for areas prone to landslides. The culmination of this research leads to several key conclusions:

Firstly, the GRF, specifically designed to integrate spatial information into the prediction stage, demonstrated a superior performance over the regular RF model, showcasing approximately a 5% and 3% increase in OA and AUC scores, respectively. Secondly, the application of the Wilcoxon signed-rank test validated a statistically significant difference between GRF and RF predictions at a 95% confidence level. This confirms that the differences seen in performance metrics among the models were not coincidental; rather, they indicate significant variations in accuracy and effectiveness. Finally, when analyzing the GRF algorithm on a global scale, it highlighted slope as the most impactful factor. However, on a local level, some variations emerged, showcasing different primary causative factors across various locations. Recognizing these spatial patterns and site-specific characteristics is paramount in effectively predicting and mitigating landslide incidents.

While acknowledging the plethora of geographically weighted models in existing literature, this study deliberately narrows its focus to specific themes to maintain a manageable scope. Notably, certain novel members of the geographically weighted model family, such as geographically weighted gradient boosting and geographically convolutional neural network weighted regression, were not integrated into this investigation. In future pursuits, we aspire to expand the horizons of this research by integrating these advanced models. By doing so, we seek to enhance and diversify the methodologies applied in spatially focused landslide susceptibility assessment. The inclusion of these sophisticated models holds promise in offering heightened precision, nuanced spatial insights, and a more comprehensive understanding of the intricate relationships between geospatial variables and landslide occurrences. This expansion aims to push the boundaries of knowledge in landslide susceptibility assessment, fostering a richer understanding of spatially varying influences and improving the predictive accuracy of such models.

REFERENCES

Aguilera, Q., Lombardo, L., Tanyas, H., Lipani, A., 2022: On the Prediction of Landslide Occurrences and Sizes via Hierarchical Neural Networks. *Stochastic Environmental Research and Risk Assessment* 36, 2031–2048.

Althuwaynee, O.F., 2021. LaGriSU v 0.2 (Landslide Grid and Slope Units): Binary (1,0) Samples Extraction Tool Pack for Landslides Mapping in QGIS. https://github.com/Althuwaynee/LaGriSU_Landslide-Grid-and-Slope-Units-QGIS_ToolPack.

Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., Guzzetti, F., 2016: Automatic Delineation of Geomorphological Slope Units with r.slopeunits v1.0 and Their Optimization for Landslide Susceptibility Modeling. *Geoscientific Model Development* 9, 3975–3991.

Arabameri, A., Chandra Pal, S., Rezaie, F., Chakrabortty, R., Saha, A., Blaschke, T., Di Napoli, M., Ghorbanzadeh, O., Thi Ngo, P.T., 2022: Decision Tree Based Ensemble Machine Learning Approaches for Landslide Susceptibility Mapping. *Geocarto International* 37, 4594-4627.

Brabb, E.E., Pampeyan, E.H., Bonilla, M.G., 1972: Landslide susceptibility in San Mateo County, California. US Geological Survey Miscellaneous Field Studies Map MF-360.

Breiman, L., 2001: Random Forests. *Machine Learning*. 45, 5–32.

Camilo, D.C., Lombardo, L., Mai, P.M., Dou, J., Huser, R., 2017: Handling High Predictor Dimensionality in Slope-Unit-Based Landslide Susceptibility Models through LASSO-Penalized Generalized Linear Model. *Environmental Modelling & Software* 97, 145–156.

Chalkias, C., Polykretis, C., Karymbalis, E., Soldati, M., Ghinoi, A., Ferentinou, M., 2020: Exploring Spatial Non-Stationarity in the Relationships between Landslide Susceptibility and Conditioning Factors: A Local Modeling Approach Using Geographically Weighted Regression. *Bulletin of Engineering Geology and the Environment* 79, 2799–2814.

Chowdhuri, I., Pal, S.C., Arabameri, A., Ngo, P.T.T., Chakrabortty, R., Malik, S., Das, B., Roy, P., 2020: Ensemble Approach to Develop Landslide Susceptibility Map in Landslide Dominated Sikkim Himalayan Region, India. *Environmental Earth Sciences* 79, 476.

Colkesen, I., Kavzoglu, T., Atesoglu, A., Tonbul, H., Ozturk, M.Y., 2023: Multi-Seasonal Evaluation of Hybrid Poplar (P. Deltoides) Plantations Using Worldview-3 Imagery and State-Of-The-Art Ensemble Learning Algorithms. *Advances in Space Research* 71, 3022–3044.

Duru, U., 2017: Shoreline Change Assessment Using Multi-Temporal Satellite Images: A Case Study of Lake Sapanca, NW Turkey. *Environmental Monitoring and Assessment* 189, 385.

Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E., Kalogirou, S., 2021: Geographical Random Forests: A Spatial Extension of the Random Forest Algorithm to Address Spatial Heterogeneity in Remote Sensing and Population Modelling. *Geocarto International* 36, 121–136.

Gomez, H., Kavzoglu, T., 2005: Assessment of Shallow Landslide Susceptibility Using Artificial Neural Networks in Jabonosa River Basin, Venezuela. *Engineering Geology* 78, 11-27.

Gu, T., Li, J., Wang, M., Duan, P., 2022: Landslide Susceptibility Assessment in Zhenxiong County of China Based on Geographically Weighted Logistic Regression Model. *Geocarto International* 37, 4952–4973.

Ikiel, C., 2022: Erosion Risk Analysis in Sapanca Lake Basin. *Journal of Human Sciences* 19, 625–640.

Kavzoglu, T., Sahin, E.K., Colkesen, I., 2014: Landslide Susceptibility Mapping Using GIS-Based Multi-Criteria Decision Analysis, Support Vector Machines, and Logistic Regression. *Landslides* 11, 425–439.

Kavzoglu, T., Teke, A., Yilmaz, E.O., 2021: Shared Blocks-Based Ensemble Deep Learning for Shallow Landslide Susceptibility Mapping. *Remote Sensing* 13, 4776.

Kavzoglu, T., Teke, A., 2022a: Advanced Hyperparameter Optimization for Improved Spatial Prediction of Shallow Landslides Using Extreme Gradient Boosting (XGBoost). *Bulletin of Engineering Geology and the Environment* 81, 201.

Kavzoglu, T., Teke, A., 2022b: Predictive Performances of Ensemble Machine Learning Algorithms in Landslide Susceptibility Mapping Using Random Forest, Extreme Gradient Boosting (XGBoost) and Natural Gradient Boosting (NGBoost). *Arabian Journal for Science and Engineering* 47, 7367–7385.

Kutlug Sahin, E., Ipbuker, C., Kavzoglu, T., 2017: Investigation of Automatic Feature Weighting Methods (Fisher, Chi-square and Relief-F) for Landslide Susceptibility Mapping. *Geocarto International* 32, 956-977.

Martinello, C., Cappadonia, C., Conoscenti, C., Agnesi, V., Rotigliano, E., 2021: Optimal Slope Units Partitioning in Landslide Susceptibility Mapping. *Journal of Maps* 17, 152–162.

Quevedo, R.P., Maciel, D.A., Uehara, T.D.T., Vojtek, M., Rennó, C.D., Pradhan, B., Vojteková, J., Pham, Q.B., 2022: Consideration of Spatial Heterogeneity in Landslide Susceptibility Mapping Using Geographical Random Forest Model. *Geocarto International* 37, 8190–8213.

Schlögel, R., Marchesini, I., Alvioli, M., Reichenbach, P., Rossi, M., Malet, J.-P., 2018: Optimizing Landslide Susceptibility Zonation: Effects of DEM Spatial Resolution and Slope Unit Delineation on Logistic Regression Models. *Geomorphology* 301, 10–20.

Sun, D., Wu, X., Wen, H., Gu, Q., 2023: A LightGBM-Based Landslide Susceptibility Model Considering the Uncertainty of Non-Landslide Samples. *Geomatics, Natural Hazards and Risk*

14, 2213807.

Temiz, T., Sonmez, O., Dogan, E., Oner, A., Opan, M., 2022: Evaluation of the Effects of Land-Use Change and Increasing Deforestation in the Sapanca Basin on Total Suspended Solids (TSS) Movement with Predictive Models. *Acta Geophysica* 70, 1331–1347.

Tonbul, H., Colkesen, I., Kavzoglu, T., 2022: Pixel- and Object-Based Ensemble Learning for Forest Burn Severity Using USGS FIREMON and Mediterranean Condition DNBRs in Aegean Ecosystem (Turkey). *Advances in Space Research* 69, 3609–3632.

Woodard, J.B., Mirus, B.B., Wood, N.J., Allstadt, K.E., Leshchinsky, B.A., Crawford, M.M., 2023: Slope Unit Maker (SUMak): An Efficient and Parameter-Free Algorithm for Delineating Slope Units to Improve Landslide Susceptibility Modeling. *Natural Hazards and Earth System Sciences Discussions* 2023, 1–18.

Ye, P., Yu, B., Chen, W., Liu, K., Ye, L., 2022: Rainfall-Induced Landslide Susceptibility Mapping Using Machine Learning Algorithms and Comparison of Their Performance in Hilly Area of Fujian Province, China. *Natural Hazards* 113, 965–995.

Yilmaz, E.O., Teke, A., Kavzoglu, T., 2022. Performance evaluation of depthwise separable CNN and random forest algorithms for landslide susceptibility prediction. *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, Kuala Lumpur, Malaysia, 5477–5480.

Zhao, Z., Xu, Z., Hu, C., Wang, K., Ding, X., 2024: Geographically Weighted Neural Network Considering Spatial Heterogeneity for Landslide Susceptibility Mapping: A Case Study of Yichang City, China. *CATENA* 234, 107590.