## Landslide25_31

# Data-Driven Landslide Susceptibility Analysis in the Ukrainian Carpathians: A Machine Learning Approach

***R. S. Ajin** *(Department of Earth Sciences, University of Florence (UNIFI))*, **K. Hadiatska** *(Department of General and Historical Geology, Institute of Geology, Taras Shevchenko National University of Kyiv)*, **O. Ivanik** *(Department of General and Historical Geology, Institute of Geology, Taras Shevchenko National University of Kyiv; Department of Geosciences, University of Cincinnati)*, **D. Kravchenko** *(Department of General and Historical Geology, Institute of Geology, Taras Shevchenko National University of Kyiv)*, **E. Petrushenko** *(Department of General and Historical Geology, Institute of Geology, Taras Shevchenko National University of Kyiv)*, **S. Segoni** *(Department of Earth Sciences, University of Florence (UNIFI))*

***SUMMARY***

The Carpathian region of Ukraine is notably prone to landslides because of its complex geology, rugged terrain, and relatively high rainfall. This investigation presents the first implementation of a machine learning (ML) method—specifically, the eXtreme Gradient Boosting (XGBoost) algorithm—for landslide susceptibility mapping in the Transcarpathian region of Ukraine. Utilizing a landslide inventory comprising 697 recorded landslides, the model integrated ten predisposing factors selected through Variance Inflation Factor (VIF) analysis to mitigate multicollinearity. The model's efficacy was measured using the Receiver Operating Characteristic (ROC) curve, which yielded an Area Under the Curve (AUC) score of 0.71. Although the model did not demonstrate high predictive performance, this first-of-its-kind ML application marks a critical step towards improved geohazard management in Ukraine.

## Introduction

The Carpathian region of Ukraine is notably prone to landslides due to its complicated geology, steep gradients, abundant rainfall, and anthropogenic factors, which have led to substantial loss of residential areas, infrastructure, and ecological systems (Shtohryn & Kasiynchuk, 2024). Susceptibility mapping is vital for the evaluation of potential hazards and for guiding decisions related to land use. While various studies have created susceptibility maps for this region employing traditional heuristic and statistical techniques (Ivanik et al., 2022; Hadiatska et al., 2023), machine learning (ML) approaches—despite their rising global usage and impressive predictive performance (Liu et al., 2021; Ajin et al., 2024)—have yet to be utilized in this region. This research presents an ML-oriented approach to produce a data-driven susceptibility map of the Transcarpathian region, utilizing historical landslide records together with relevant geological and environmental parameters, with the intention of facilitating improved geohazard management.

## Study area

Positioned on the southwestern slopes of the Carpathian Mountains (Fig. 1a), the Transcarpathian region (Fig. 1b) is mainly drained by the Tisza River along with its tributaries (Shtohryn et al., 2020). Geologically, it is characterized by two significant tectonic units: the Carpathian fold system and the Transcarpathian inner depression (Shtohryn et al., 2020; Kasiyanchuk & Shtohryn, 2021). The stratification of these units, along with associated erosion processes, influences landslide activity (Ivanik et al., 2019). This region exhibits significant humidity, with yearly precipitation varying from 600 to 1100 mm, and is notably vulnerable to landslides, largely due to the elevated groundwater levels and the frequent occurrence of intense, erratic rainfall (Shtohryn et al., 2020).

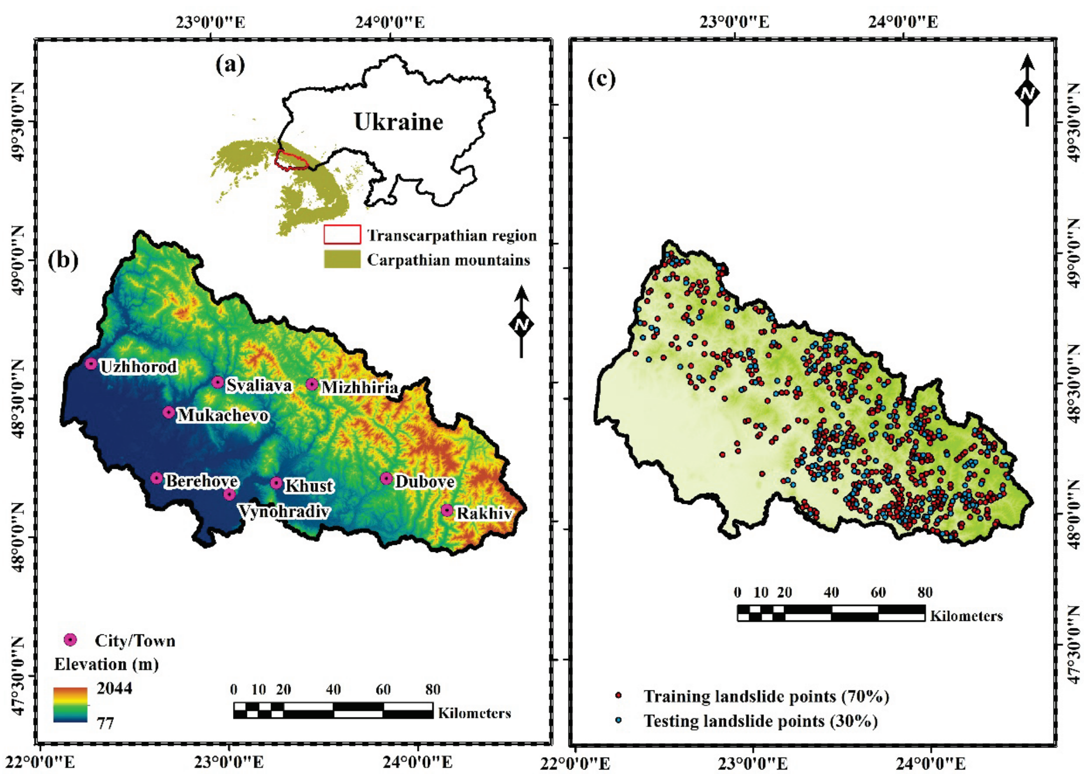


***Figure 1*** *(a) Location of the Carpathian Mountains in Ukraine, (b) Transcarpathian region, and (c) Landslide inventory with training and testing points.*

**Method**

The inventory comprises 697 landslides, which were randomly split into 488 (70%) training points and 209 (30%) validation points (Fig. 1c). To ensure balance in the dataset for modeling, an equal number of non-landslide points was utilized. Ten predisposing factors were selected (Fig. 2). The variance inflation factor (VIF) values of these factors were computed to exclude multicollinearity. The eXtreme Gradient Boosting (XGBoost) algorithm was employed for modeling, implemented on the Kaggle platform (https://www.kaggle.com/). XGBoost is an ML algorithm that integrates decision trees based on a technique known as gradient boosting. It offers multiple advantages over other algorithms, such as the capacity to handle missing data, effective parallel processing, and its effectiveness with large and complex datasets (Tarwidi et al., 2023). The model's performance was evaluated through the Receiver Operating Characteristic (ROC) curve.

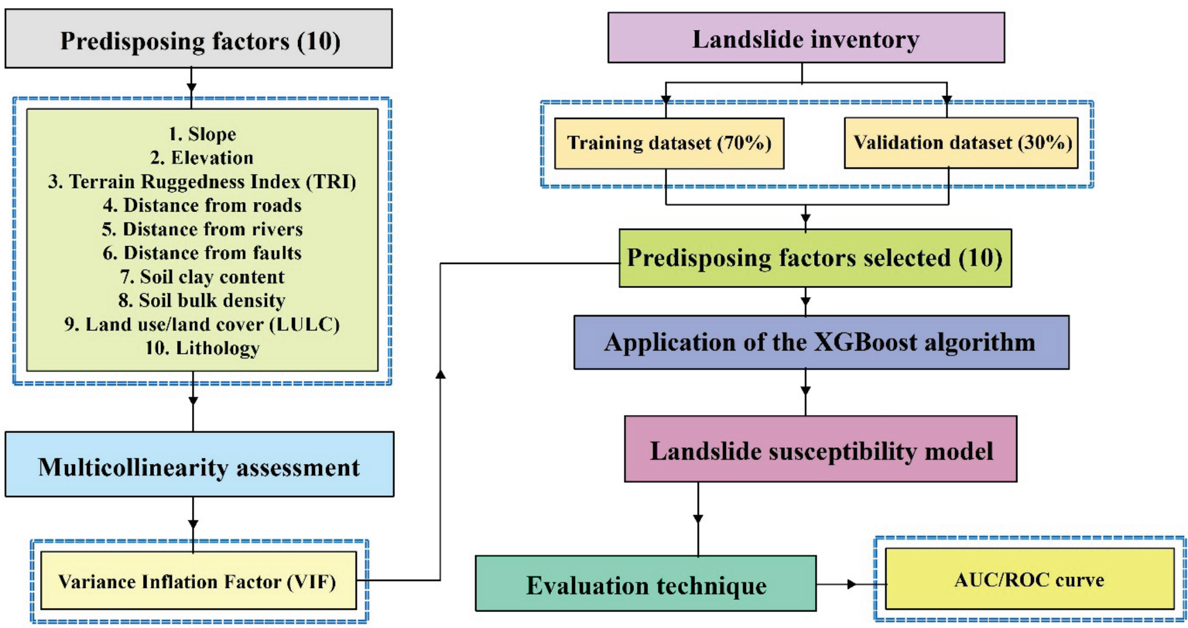


***Figure 2*** *Flowchart illustrating the susceptibility modeling process.*

**Results**

The analysis of multicollinearity affirmed that all 10 factors are suitable for modeling. Figure 3 displays the resulting susceptibility map. The regions identified as having a very high susceptibility are mainly situated in the Flysch Carpathians areas, which illustrate the rugged landscape and geological conditions that are likely to induce landslides. On the other hand, the lowland regions near the Tisza River predominantly belong to the low to very low susceptibility, which corresponds to their gentle topography and more stable geomorphological characteristics. The model's performance, as evaluated by the ROC curve, achieved a fair Area Under the Curve (AUC) score of 0.71 (White et al., 2023), as illustrated in Figure 4.

**Discussion and conclusions**

This study marks the first application of an ML algorithm—specifically, XGBoost—for landslide susceptibility modeling in Ukraine. The model achieved an AUC of 0.71, which is relatively modest compared to similar studies conducted globally (Can et al., 2021; Khan et al., 2025). One possible explanation is the limitation in the quality of the input data; to address this, further research is underway to ameliorate susceptibility assessments through the integration of upgraded environmental datasets and more homogeneous, updated landslide inventories. Despite these challenges, the study represents a relevant step towards landslide risk management in the Carpathians.

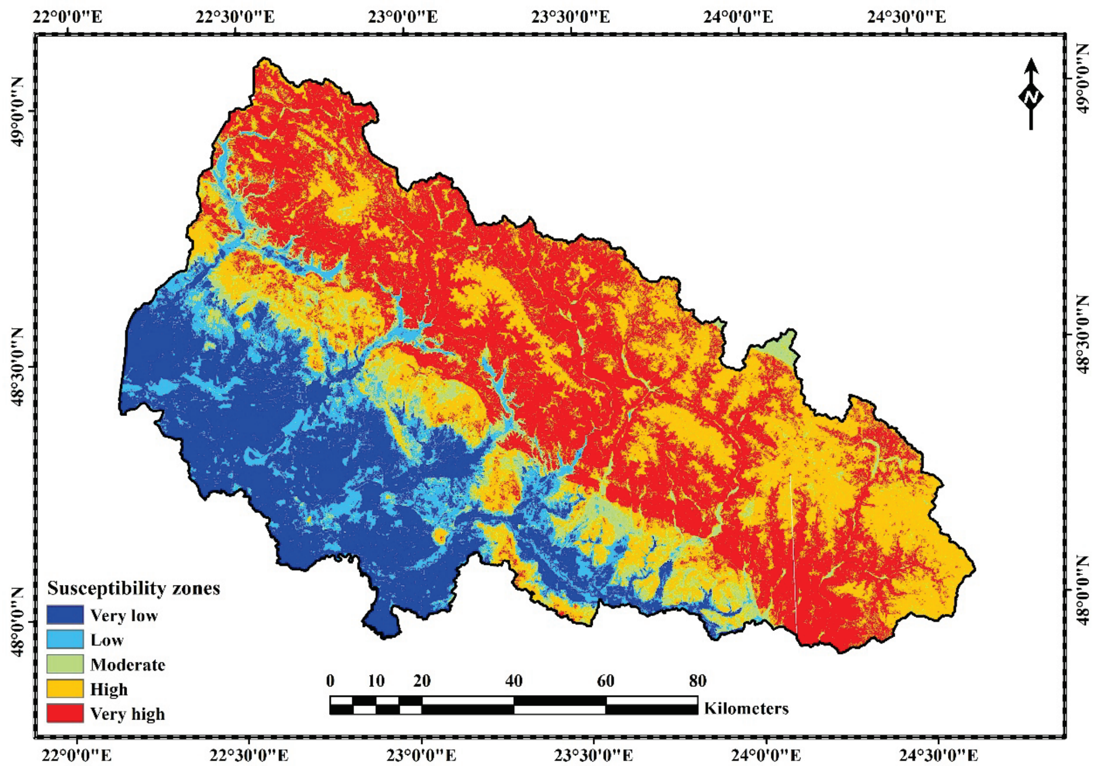


***Figure 3*** *Landslide susceptibility map of the Transcarpathian region, Ukraine.*

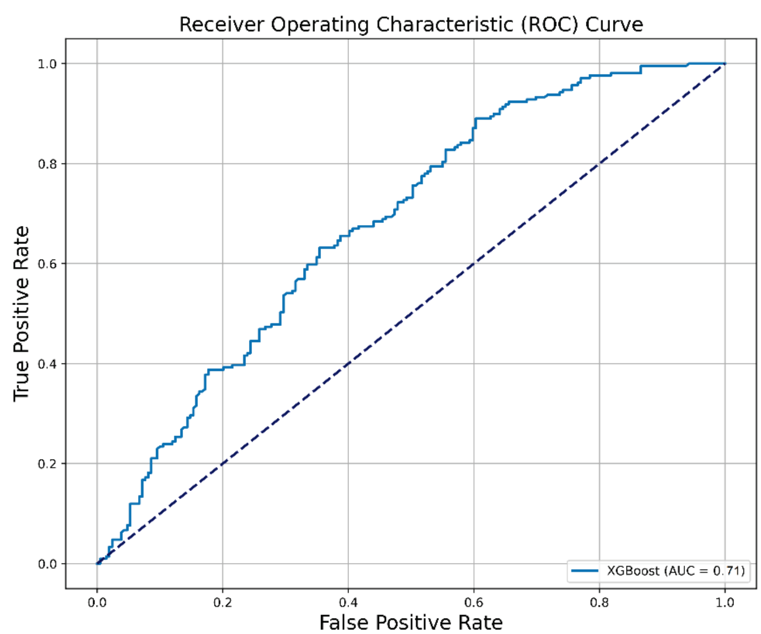


***Figure 4*** *ROC curve with corresponding Area Under the Curve (AUC) score.*

**Acknowledgements**

This research has been done in the framework of the project "Risk assessment of landslide hazards and impact on communities" as part of the Fulbright U.S. Scholar Program sponsored by the U.S. Department of State, and the EU Project LOC3G (MARIE SKLODOWSKA-CURIE ACTIONS (MSCA), grant No 101129729. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Fulbright Program or the U. S. Government.

**References**

Ajin, R.S., Segoni, S. & Fanti, R. (2024). Optimization of SVR and CatBoost models using metaheuristic algorithms to assess landslide susceptibility. *Scientific Reports, 14*, 24851. https://doi.org/10.1038/s41598-024-72663-x

Can, R., Kocaman, S., & Gokceoglu, C. (2021). A comprehensive assessment of XGBoost algorithm for landslide susceptibility mapping in the upper basin of Ataturk dam, Turkey. *Applied Sciences*, *11*(11), 4993. https://doi.org/10.3390/app11114993

Hadiatska, K., Ivanik, O., Tustanovska, L., Kravchenko, D., & Vilkhovyi, R. (2023). *Regional Landslide Hazard Assessment Using Susceptibility Mapping: a Case Study on the Middle Dnieper Area, Ukraine*. In *Proceedings of the Fourth EAGE Workshop on Assessment of Landslide Hazards and Impact on Communities* (pp. 1–5). European Association of Geoscientists & Engineers. https://doi.org/10.3997/2214-4609.2023500045

Ivanik, O., Menshov, O., Bondar, K., Vyzhva, S., Khomenko, R., Hadiatska, K., Kravchenko, D., & Tustanovska, L. (2022). Integrated approach to modelling and assessing the landslide hazards at the regional and local scale in Kyiv urbanized area, Ukraine. *Modeling Earth Systems and Environment*, *8*(4), 5479–5491. https://doi.org/10.1007/s40808-022-01447-x

Ivanik, O., Shevchuk, V., Kravchenko, D., Yanchenko, V., Shpyrko, S., & Gadiatska, K. (2019). Geological and geomorphological factors of natural hazards in Ukrainian Carpathians. *Journal of Ecological Engineering*, *20*(4), 177–186. https://doi.org/10.12911/22998993/102964

Kasiyanchuk, D., & Shtohryn, L. (2021). Assessment of the ecological risks of landslide damages in the Carpathian Region. *Grassroots Journal of Natural Resources 4*(3), 52-61. https://doi.org/10.33002/nr2581.6853.040306

Khan, D., Akram, W., & Ullah, S. (2025). Enhancing landslide susceptibility predictions with XGBoost and SHAP: a data-driven explainable AI method. *Geocarto International, 40*(1), 2514725. https://doi.org/10.1080/10106049.2025.2514725

Liu, S., Yin, K., Zhou, C., Gui, L., Liang, X., Lin, W., & Zhao, B. (2021). Susceptibility assessment for landslide initiated along power transmission lines. *Remote Sensing, 13*(24), 5068. https://doi.org/10.3390/rs13245068

Shtohryn, L., Kasiyanchuk, D., & Kuzmenko, E. (2020). The problem of long-term prediction of landslide processes within the Transcarpathian Inner Depression of the Carpathian Region of Ukraine. *Carpathian Journal of Earth and Environmental Sciences*, *15*(1), 157–166. https://doi.org/10.26471/cjees/2020/015/118

Shtohryn, L. V., & Kasiynchuk, D. V. (2024). Analysis of natural and man-made factors of landslide development in the Carpathian region using GIS. *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu*, *5*, 93–98. https://doi.org/10.33271/nvngu/2024-5/093

Tarwidi, D., Pudjaprasetya, S. R., Adytia, D., & Apri, M. (2023). An optimized XGBoost-based machine learning method for predicting wave run-up on a sloping beach. *MethodsX*, *10*, 102119. https://doi.org/10.1016/j.mex.2023.102119

White, N., Parsons, R., Collins, G., & Barnett, A. (2023). Evidence of questionable research practices in clinical prediction models. *BMC Medicine*, *21*(1), 339. https://doi.org/10.1186/s12916-023-03048-6