LANDSLIDE SUSCEPTIBILITY MAPPING USING A SPATIAL CV RANDOM FOREST APPROACH: A CASE STUDY OF BLAGOEVGRAD PROVINCE, SOUTHWEST BULGARIA

Rosen Iliev, Boyko Ranguelov, Garo Mardirossian, Lachezar Filchev

Space Research and Technology Institute – Bulgarian Academy of Sciences
e-mail: ilievrosen@space.bas.bg, branguelov@gmail.com, garo.mardirossian@gmail.com, lachezarhf@space.bas.bg

***Keywords**: Blagoevgrad Province; landslide susceptibility; Random Forest; machine learning; geomorphometry; digital elevation model; spatial cross-validation*

***Abstract:** This study presents a Landslide Susceptibility Model (LSM) for Blagoevgrad Province, Bulgaria, developed through an integrated geomorphometric and machine learning framework within a Python-based QGIS environment. The approach combines a landslide inventory of 108 mapped events with ten conditioning factors derived from remote sensing data and the EU-DEM v1.1 (25 m) dataset. A Random Forest classifier was used to estimate spatial landslide susceptibility probabilities. The model achieved a mean AUC of 0.908, PR-AUC of 0.796, and Brier score of 0.124, indicating strong discrimination and relatively good probabilistic prediction performance. Spatial cross-validation showed stable model performance across folds, while the low mean residual spatial autocorrelation (Moran's I = 0.037) indicates limited remaining spatial structure in the model residuals. The resulting susceptibility patterns show a clear spatial correspondence with the distribution of known landslides and the main geomorphological characteristics of the study area. The proposed LSM can support regional-scale landslide susceptibility assessment, environmental planning, and risk management in mountainous areas.*

КАРТОГРАФИРАНЕ НА ПОДАТЛИВОСТТА КЪМ СВЛАЧИЩА ЧРЕЗ ПОДХОД С ПРОСТРАНСТВЕНА КРЪСТОСАНА ВАЛИДАЦИЯ И СЛУЧАЙНА ГОРА (RANDOM FOREST): ИЗСЛЕДВАНЕ НА ПРИМЕРА НА ОБЛАСТ БЛАГОЕВГРАД, ЮГОЗАПАДНА БЪЛГАРИЯ

Росен Илиев, Бойко Рангелов, Гаро Мардиросян, Лъчезар Филчев

Институт за космически изследвания и технологии – Българска академия на науките
e-mail: ilievrosen@space.bas.bg, branguelov@gmail.com, garo.mardirossian@gmail.com, lachezarhf@space.bas.bg

***Ключови думи:** Област Благоевград; свлачищна податливост; случайна гора; машинно обучение; геоморфометрия; цифров модел на релефа; пространствена кръстосана валидация*

***Резюме:** Настоящото изследване представя модел за картографиране на податливостта към свлачища (Landslide Susceptibility Model, LSM) за област Благоевград, България, посредством интегриран подход, съчетаващ геоморфометричен анализ и машинно обучение в базирана на Python среда в QGIS. Методологичната рамка интегрира инвентар от 108 картографирани свлачищни явления, с десет условни фактора, извлечени от данни от дистанционни изследвания и цифровия модел на релефа EU-DEM v1.1 с пространствена резолюция 25 m. За моделиране на пространственото разпределение на вероятността за свлачищна податливост е приложен класификаторът Random Forest. Полученият модел достига средна стойност на площта под ROC кривата (AUC) от 0,908, площ под кривата Precision–Recall (PR-AUC) от 0,796 и Brier score от 0,124. Тези показатели свидетелстват за висока дискриминационна способност на модела и относително добра точност на вероятностните му оценки. Резултатите от пространствената кръстосана валидация показват относително стабилна производителност на модела между отделните пространствени подмножества. Допълнителният анализ на остатъчната пространствена автокорелация показва ниска средна стойност на Moran's I (0,037), което предполага ограничена остатъчна пространствена структура в прогнозните грешки на модела. Пространственото*

разпределение на моделираната податливост към свлачища показва ясно съответствие с локализацията на регистрираните свлачища, както и с основните геоморфоложки характеристики на изследваната територия. Разработеният модел за картографиране на свлачищната податливост предоставя приложима рамка за регионална оценка на пространственото разпределение на потенциално податливи към свлачища територии и може да подпомогне процесите на териториално планиране, както и управлението на свлачищния риск в планински райони. Предложеният подход предоставя методична рамка за пространствено характеризиране на податливостта към свлачища, в унисон с целите на проекта „Космически и наземни технологии за изследване на каскадни мултирискови събития – модели, приложения, прекурсори“ (SPARC), и създава основа за интегрирането на резултатите в анализа на каскадни мултирискови събития.

Introduction

Landslide Susceptibility Models (LSMs) quantify the spatial probability of landslide occurrence as a function of environmental conditions and are widely used in hazard assessment, land-use planning, and risk mitigation. Methodologically, LSMs establish relationships between landslide inventories and conditioning factors derived primarily from Digital Elevation Models (DEMs) and remote sensing, including slope, aspect, curvature, lithology, and land cover. Machine learning (ML) methods, particularly Random Forest (RF), are increasingly applied because of their ability to capture nonlinear interactions among conditioning variables and model complex susceptibility patterns (Gautam et al., 2021; Joshi et al., 2025; Sun et al., 2023; Wang et al., 2023; Zhou et al., 2025).

In Bulgaria, GIS- and DEM-based approaches have demonstrated their applicability to regional landslide and natural-hazard assessment. Ivanova (2023) investigated water-related natural hazards, including landslides, in the Rhodope Mountains, while Tcherkezova et al. (2023) developed landslide and rockfall susceptibility models for the Chepelarska River Basin using remote sensing datasets, including DEM and CORINE Land Cover. Recent regional assessments of Bulgaria (Berov et al., 2025; Ivanov et al., 2022) provide generalized susceptibility patterns; however, their spatial scale may not adequately represent local variability in terrain stability.

Against this background, the present study develops a Landslide Susceptibility Model for Blagoevgrad Province by integrating ten conditioning factors derived from DEM and remote sensing data with an inventory of 108 mapped landslides of different types. A Random Forest framework is employed to capture nonlinear relationships between terrain and environmental characteristics and observed landslide occurrence. The resulting LSM aims to provide higher spatial detail and a data-driven representation of susceptibility, supporting regional land-use planning and landslide risk mitigation in the heterogeneous mountainous environment of southwestern Bulgaria.

Study area

Blagoevgrad Province, southwestern Bulgaria, covers 6,449.5 km² and occupies a geomorphologically and environmentally diverse transitional zone between continental and Mediterranean climatic influences. The region encompasses the major mountain systems of Pirin, Rila, and the Western Rhodopes, together with deeply incised valleys of the Struma, Strumeshnitsa, and Mesta rivers. The pronounced relief energy, steep slopes, and complex drainage network provide favourable conditions for gravity-driven mass movements.

The geological setting is highly heterogeneous, ranging from resistant crystalline basement rocks to weathered sedimentary formations, resulting in substantial spatial variability in slope strength, weathering, and erosion susceptibility. Seasonal and high-intensity precipitation events represent important landslide-triggering processes. In addition, deforestation, road construction, and urban expansion locally modify surface drainage and vegetation cover, potentially increasing slope instability.

The combined effects of complex topography, heterogeneous lithology, climatic forcing, and anthropogenic disturbance make Blagoevgrad Province an appropriate setting for regional landslide susceptibility modelling. The pronounced intra-regional variability further requires a spatially explicit approach capable of resolving differences in terrain and environmental conditions.

Methods and data

Landslide Susceptibility Modeling Framework

The Landslide Susceptibility Model (LSM) for Blagoevgrad Province was developed using a fully automated Python-based geoprocessing workflow implemented in QGIS. The workflow constitutes an original software framework for automated landslide susceptibility mapping, with the complete source code openly available on GitHub and archived in Zenodo (Iliev, 2026). The workflow applies GDAL-based preprocessing to harmonize all input datasets in terms of spatial resolution,

coordinate reference system, and spatial extent, ensuring a consistent predictor dataset for machine learning analysis.

The model uses ten spatially distributed conditioning factors derived from the Copernicus EU-DEM (25 m), representing complementary characteristics of terrain morphology, hydrological processes, and surface complexity.

The EU-DEM served as the primary elevation source from which a series of geomorphometric derivatives was generated. The final predictor set consisted of:

✓ EU-DEM (EEA, 2019), representing absolute elevation and regional geomorphometric context. Elevation controls broad climatic, hydrological, and geomorphic gradients and provides a first-order descriptor of terrain energy.

✓ Flow Accumulation, describing the potential convergence of surface runoff and concentration of drainage pathways that may contribute to slope failure.

✓ Slope Gradient, representing local terrain inclination and controlling the gravitational forces acting on slope materials.

✓ Slope Fractal, representing local slope variation and abrupt changes in slope morphology. The variable was calculated using a variogram-based fractal approach following the methodology originally proposed by Mark and Aronson (1984).

✓ Lacunarity (Plotnick et al., 1993), expressing the spatial heterogeneity and fragmentation of terrain patterns.

✓ Elevation Anomaly, representing the relative elevation of each cell with respect to its surrounding neighborhood and highlighting local topographic prominence or depressions.

✓ Distance to Roads, derived from the OpenStreetMap (OSM) road network, representing the Euclidean distance from each raster cell to the nearest mapped road. This predictor represents the spatial association between landslide susceptibility and areas of potential anthropogenic terrain modification.

✓ Distance to Rivers, derived from the OpenStreetMap (OSM) hydrographic network, representing the Euclidean distance to the nearest mapped river or stream. Proximity to river channels represents the spatial association with fluvial erosion and local hydrological conditions that may contribute to slope instability.

✓ Aspect, describing the compass orientation of slopes. Aspect influences the spatial distribution of solar radiation, evapotranspiration, and soil moisture, thereby affecting weathering processes, vegetation cover, and slope stability.

✓ Land Cover (EEA, 2023), obtained from the Copernicus Land Monitoring Service at 10 m spatial resolution, representing the dominant land-cover class for each raster cell. Land cover influences slope stability through its effects on vegetation reinforcement, infiltration capacity, and human land-use activities, thereby providing information on both natural and anthropogenic controls of landslide susceptibility.

All raster layers were resampled to a common spatial resolution of 25 m and aligned to the same coordinate reference system and spatial grid prior to model development.

The predictor set comprised conventional terrain and environmental variables, supplemented by fractal and anomaly-based descriptors to capture local geomorphological variability. An iterative selection procedure based on Random Forest feature importance, Variance Inflation Factor (VIF), and spatial cross-validation reduced the initial variable set to ten non-redundant predictors.

The modelling dataset included 108 landslide observations and 216 spatially constrained pseudo-absence samples. Negative samples were excluded within 75 m of known landslides, corresponding to approximately three 25 m raster cells, yielding a 1:2 class ratio.

A Random Forest classifier with 250 trees, maximum depth of 10, and minimum leaf size of four was used. Class imbalance was addressed using "balanced_subsample", and isotonic calibration was applied through "CalibratedClassifierCV". Spatial validation used five-fold "StratifiedGroupKFold" with 250 m spatial blocks, ensuring that observations from the same block were not shared between training and validation sets while maintaining class balance.

Model performance was assessed using ROC-AUC, PR-AUC, Brier score, sensitivity, specificity, and Matthews Correlation Coefficient (MCC), with a 0.5 probability threshold for the latter three metrics. Residual spatial autocorrelation was evaluated using Moran's I with an eight-nearest-neighbour structure. Cross-validated AUC uncertainty was estimated from 200 spatial bootstrap resamples of pooled out-of-fold predictions, with up to 500 observations per resample. A final calibrated Random Forest was subsequently fitted to the complete dataset.

The resulting calibrated probabilities were used to generate the susceptibility map. Raw and smoothed probability surfaces were produced, with the latter used for visualization. Continuous susceptibility values were classified into five relative classes using the 20th, 40th, 60th, and 80th percentiles.

Data and GIS software

DEM and Land Cover Data

The primary elevation dataset was EU-DEM v1.1, provided by the Copernicus Land Monitoring Service at 25 m spatial resolution and developed by the European Environment Agency (EEA) from SRTM and ASTER GDEM data. It was used to derive all terrain-related predictors (EEA, 2019). Land-cover information was obtained from Copernicus Land Cover 2023 at 10 m resolution and incorporated as an environmental predictor representing vegetation, surface cover, and anthropogenic land use (European Environment Agency, 2023).

Inventory data

For the training of the landslide susceptibility model (LSM) in Blagoevgrad Province, an inventory of documented landslides was compiled from an online catalog provided by Geoprotection Pernik, available through the Ministry of Regional Development and Public Works of the Republic of Bulgaria (https://gz-pernik.mrrb.government.bg/landslide/). The dataset contains the geographic coordinates of 108 recorded landslide events and was used as the primary source of positive samples for model calibration and validation.

GIS software

All spatial preprocessing, raster processing, feature integration, model training, and prediction were performed in the open-source QGIS environment, integrated with Python libraries for spatial, statistical, and machine learning analysis.

Results and discussion

The statistical reliability and predictive performance of the LSM for Blagoevgrad Province are presented in Table 1 and Table 2 and Figure 1 and Figure 2, followed by the corresponding analysis and discussion.

Table 1. Performance and Validation Metrics of the Landslide Susceptibility Model

Metric	Mean	Standard Deviation	Minimum	Maximum
AUC	0.908	0.025	0.869	0.937
PR-AUC	0.796	0.086	0.670	0.918
Brier Score	0.124	0.016	0.108	0.149
Sensitivity	0.647	0.062	0.588	0.737
Specificity	0.894	0.024	0.848	0.913
Matthews Correlation Coefficient (MCC)	0.561	0.037	0.524	0.630
Moran's I	0.037	0.049	−0.040	0.090

Table 2. Multicollinearity (Variance Inflation Factor – VIF) and Predictor Importance Analysis of the Landslide Susceptibility Model

Predictor	Importance (%)	VIF
Distance to Roads	21.27	1.21
EU-DEM	12.90	1.20
Slope Fractal	12.18	1.10
Slope Gradient	10.91	1.07
Distance to Rivers	9.79	1.27
Elevation Anomaly	9.10	1.06
Lacunarity	6.46	1.59
Flow Accumulation	6.16	1.04
Aspect	5.89	1.02
Land Cover	5.35	1.13

The spatially cross-validated Random Forest showed strong and relatively stable predictive performance (Table 1). Mean ROC-AUC was 0.908 (SD = 0.025; range = 0.869–0.937), while PR-AUC was 0.796 (SD = 0.086; range = 0.670–0.918), indicating good discrimination with greater fold-to-fold variability in precision–recall performance. The mean Brier score of 0.124 (SD = 0.016; range = 0.108–0.149) indicates relatively low probabilistic prediction error, although it does not alone demonstrate full calibration. At a 0.5 threshold, mean sensitivity and specificity were 0.647 and 0.894,

respectively, while MCC was 0.561 (range = 0.524–0.630), indicating moderate classification agreement. Residual Moran's I was low (mean = 0.037; range = −0.040–0.090), suggesting limited remaining spatial structure in the residuals and reducing concern that model evaluation was dominated by strong residual spatial dependence.

Predictor importance (Table 2) was dominated by anthropogenic, topographic, morphometric, and hydrological variables. Distance to Roads was the most influential predictor (21.27%), followed by EU-DEM (12.90%), Slope Fractal (12.18%), Slope Gradient (10.91%), Distance to Rivers (9.79%), and Elevation Anomaly (9.10%). The four principal terrain variables (EU-DEM, Slope Fractal, Slope Gradient, and Elevation Anomaly) accounted for 45.09% of total importance. Distance to Rivers and Flow Accumulation contributed 9.79% and 6.16%, respectively, while Lacunarity, Aspect, and Land Cover contributed 6.46%, 5.89%, and 5.35%.

The high importance of Distance to Roads may reflect the influence of road-related modifications to slope geometry and drainage, although it may also partly reflect accessibility and the spatial distribution of officially documented landslides. The contribution of terrain and hydrological variables indicates that topographic morphology and drainage conditions provide substantial information for distinguishing susceptibility levels.

Multicollinearity was low across all predictors, with VIF values of 1.02–1.59 and a maximum of 1.59 for Lacunarity (Table 2), indicating no substantial variance inflation.

The model demonstrated strong discrimination and comparatively stable performance under spatial cross-validation. The lower sensitivity (0.647) should be interpreted in relation to the available inventory and the 1:2 positive-to-negative sampling ratio, as the limited number and spatial distribution of positive observations may affect sensitivity estimation across validation folds.

The susceptibility maps (Fig. 1 and Fig. 2) show a spatially coherent pattern, with the highest predicted susceptibility concentrated along the Struma, Mesta, and Strumeshnitsa valleys and their transitions with mountainous terrain, including the Blagoevgrad–Simitli, Sandanski–Petrich, and Gotse Delchev kettles. Pronounced hotspots occur in the Kresna Gorge, the Pirin and Slavyanka Mountains, northern Belasitsa, the Ograzhden massif, and the southwestern margins of the Western Rhodopes. Lower susceptibility generally characterizes the interior parts of the major mountain massifs.

The spatial pattern is consistent with the dominant contribution of terrain-related predictors and the hydrological variables Distance to Rivers and Flow Accumulation, while the concentration of susceptibility along transport corridors corresponds to the high importance of Distance to Roads. Overall, the LSM provides a spatially coherent representation of susceptibility across the province, with elevated values concentrated mainly in valley–mountain transition zones and areas of complex terrain.

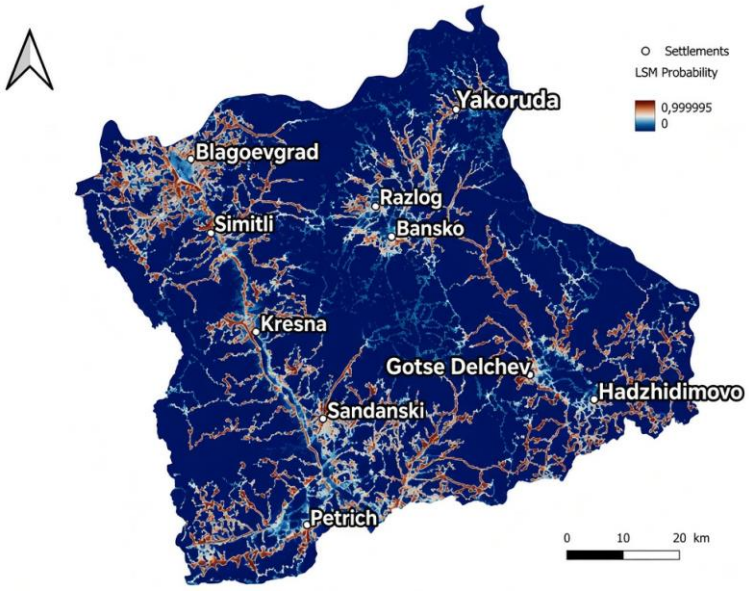


Fig. 1. Landslide Susceptibility Map of Blagoevgrad Province

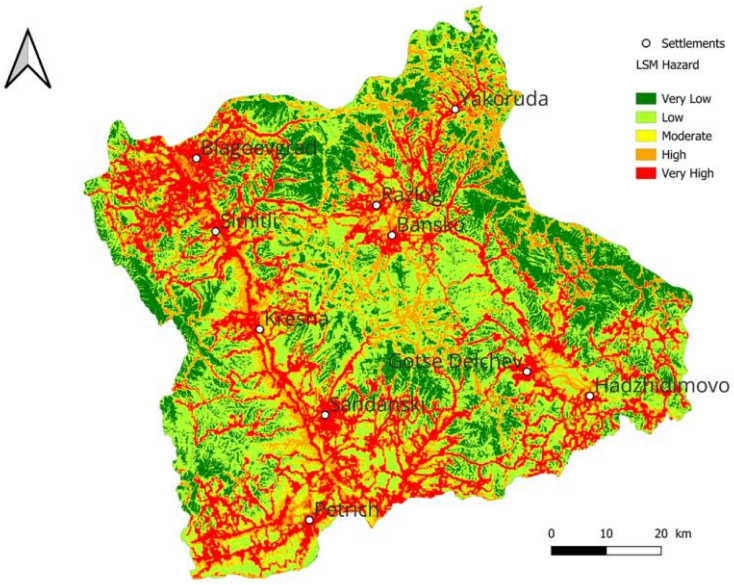


Fig. 2. Reclassified Landslide Susceptibility Map of Blagoevgrad Province Using Quantile Classification

Conclusion

This study developed a spatially explicit Landslide Susceptibility Model for Blagoevgrad Province using DEM-derived and environmental predictors within a Python-based QGIS framework. The calibrated Random Forest model, evaluated through spatial cross-validation, demonstrated strong and relatively stable predictive performance. Susceptibility was spatially differentiated, with higher probabilities concentrated mainly in valley–mountain transition zones and areas influenced by complex terrain and proximity to roads and rivers. The open-source, reproducible workflow integrates landslide inventory data, automated geospatial processing, spatial validation, and probability calibration, providing a transferable framework for regional landslide susceptibility assessment and supporting land-use planning, environmental assessment, and risk management in mountainous regions.

Funding and Acknowledgements

This study was nationally co-funded by the project “Space and Ground Technologies for Studying Cascading Multi-hazard Events - models, Applications, Precursors (SPARC)”, supported under Co-funding Contract No. KP-06-COST / 23 / dated 18 August 2026, concluded between the Space Research and Technology Institute at the Bulgarian Academy of Sciences (SRTI–BAS) and the Bulgarian National Science Fund (BNSF) within the framework of BNSF-1.009 – “Procedure for Providing National Co-funding for the Participation of Bulgarian Research Teams in Established Actions in Science and Technology.”

References:

1. Berov, B., Gadzhev, G., Ivanov, P., Dobrev, N., Ivanov, V., Nikolova, N., Toteva, A., Krastanov, M., Nankin, R., Frantzova, A. “Assessment of Landslide Susceptibility in Bulgaria, in the Context of Climate Change, Using the Mora and Vahrson Method.” Geologica Balcanica, 54(1), 2025, pp. 3–16. DOI: 10.52321/GeolBalc.54.1.3.
2. European Environment Agency. “EU-DEM v1.1 – Pan-European Digital Elevation Model (25 m).” Copernicus Land Monitoring Service, 2019. https://www.copernicus.eu/en.
3. European Environment Agency. “Copernicus Land Cover 2023 – High Resolution Land Cover (10 m).” Copernicus Land Monitoring Service, 2023. https://land.copernicus.eu/.
4. Gautam, P., Kubota, T., Sapkota, L. M., et al. “Landslide Susceptibility Mapping with GIS in High Mountain Area of Nepal: A Comparison of Four Methods.” Environmental Earth Sciences, 80, 2021, 359. DOI: 10.1007/s12665-021-09650-2.
5. Iliev, R. “Spatial Cross-Validated Random Forest for Landslide Susceptibility Mapping (Version V1.0.0).” Computer software, Zenodo, 2026. DOI: 10.5281/zenodo.22685563.
6. Ivanov, P., Dobrev, N., Berov, B., Frantzova, A., Krastanov, M., Nankin, R. “Landslide Risk for the Territory of Bulgaria by Administrative Districts.” Geologica Balcanica, 51(3), 2022, pp. 21–28. DOI: 10.52321/GeolBalc.51.3.21.

7. Ivanova, E. “Water-Related Natural Hazard Assessment: A GIS-Based Methodology for the Rhodope Mountain Range in Bulgaria.” In Proceedings of the Nineteenth International Scientific Conference Space, Ecology, Safety (SES23), Sofia, Bulgaria, 2023, pp. 291–296.
8. Joshi, B. R., Bhandary, N. P., Acharya, I. P., KC, N., Bhandari, C. “Landslide Susceptibility Mapping Optimization for Improved Risk Assessment Using Multicollinearity Analysis and Machine Learning Technique.” Applied Sciences, 15(22), 2025, 12152. DOI: 10.3390/app152212152.
9. Mark, D. M., Aronson, P. B. “Scale-Dependent Fractal Dimensions of Topographic Surfaces: An Empirical Investigation, with Applications in Geomorphology and Computer Mapping.” Mathematical Geology, 16(8), 1984, pp. 671–683. DOI: 10.1007/BF01033029.
10. Ministry of Regional Development and Public Works. “Landslide Register.” Geoprotection Pernik, n.d. https://gz-pernik.mrrb.government.bg/landslide/.
11. Plotnick, R. E., Gardner, R. H., O’Neill, R. V. “Lacunarity Indices as Measures of Landscape Texture.” Landscape Ecology, 8(3), 1993, pp. 201–211. DOI: 10.1007/BF00125351.
12. Sun, D., Gu, Q., Wen, H., Xu, J., Zhang, Y., Shi, S., Xue, M., Zhou, X. “Assessment of Landslide Susceptibility along Mountain Highways Based on Different Machine Learning Algorithms and Mapping Units by Hybrid Factors Screening and Sample Optimization.” Gondwana Research, 123, 2023, pp. 89–106. DOI: 10.1016/j.gr.2022.07.013.
13. Tcherkezova, E., Zareva, E., Yordanov, N. “GIS-Based Landslide and Rockfall Susceptibility Zoning in Chepelarska River Basin (Western Rhodope Mountains, Bulgaria).” Engineering Geology and Hydrogeology, 37, 2023, pp. 117–132. DOI: 10.52321/igh.37.1.117.
14. Wang, X., Du, A., Hu, F., Liu, Z., Zhang, X., Wang, L., Guo, H. “Landslide Susceptibility Evaluation Based on Active Deformation and Graph Convolutional Network Algorithm.” Frontiers in Earth Science, 11, 2023, 1132722. DOI: 10.3389/feart.2023.1132722.
15. Zhou, C., Gan, L., Cao, Y., Wang, Y., Segoni, S., Shi, X., Motagh, M., Singh, R. P. “Landslide Susceptibility Assessment of the Wanzhou District: Merging Landslide Susceptibility Modelling (LSM) with InSAR-Derived Ground Deformation Map.” International Journal of Applied Earth Observation and Geoinformation, 136, 2025, 104365. DOI: 10.1016/j.jag.2025.104365.