Uncertainty in AI-Based Terrain Traversability Prediction from Incomplete Geospatial Data

Marek Wyszyński [1], Krzysztof Pokonieczny [1]

[1] MUT, Faculty of Civil Engeering and Geodesy, Military University of Technology, Warsaw, Poland – marek.wyszynski@wat.edu.pl

Keywords: GeoAI, terrain traversability, passability mapping, geospatial data quality, uncertainty, artificial neural networks.

Abstract

Terrain traversability mapping is a geospatial problem with direct relevance to off-road mobility assessment, route planning, crisis response, and spatial decision support. Although machine learning has become an increasingly attractive complement to rule-based GIS procedures, model quality in this domain is still commonly summarized by aggregate error statistics alone. This paper examines uncertainty in AI-based terrain traversability prediction and argues that global accuracy is not sufficient for evaluating the usefulness of map products. The study estimates an index of passability (IOP) from structured geospatial attributes aggregated to 100 m x 100 m grid cells. The learning dataset contains 236,617 records described by 115 non-empty features derived from a larger attribute structure affected by missing values. A compact artificial neural network was used as a regression model and achieved good overall performance, with BIAS = 0.0024, MAE = 0.0097, and RMSE = 0.015 on a held-out test subset. However, spatial inspection of the results shows that prediction error is not randomly distributed. The largest deviations occur in terrain contexts that are absent, simplified, or weakly represented in the reference geodata, including mining areas and wetlands. The paper therefore frames uncertainty in traversability mapping as a compound outcome of model behaviour, thematic incompleteness of source data, and limited spatiotemporal representativeness of the training sample. By shifting attention from model novelty to data-driven reliability, the study contributes a more transparent perspective to GeoAI research in the open geospatial domain.

1. Introduction

Terrain traversability is a long-established topic in spatial analysis; however, it remains methodologically demanding because passability is not an intrinsic property of land cover alone. Rather, it arises from the combined influence of topography, surface cover, hydrological conditions, soils, linear obstacles, local objects, and the capabilities of the moving platform. Consequently, the same terrain unit may be considered traversable under one set of environmental conditions and constrained under another. Traversability assessment is therefore most appropriately treated as a multi-factor geospatial problem rather than as a single-class classification task (K. Pokonieczny, 2018; Meng et al., 2018).

Conventional GIS-based approaches estimate passability by decomposing terrain into surface, linear, and point features and subsequently combining them into a synthetic index. This logic remains valuable because it preserves interpretability and can be implemented with standard spatial operations. Its limitations are equally clear: the procedure is labor-intensive, often depends on expert-defined rules, and is not well suited to representing non-linear interactions among many variables (Hubáček et al., 2015; Kim and Choi, 2012; Pokonieczny and Wyszyński, 2016; Polanco et al., 1997; Rybansky et al., 2015).

A key methodological step in this transition is the conversion of heterogeneous vector data into a consistent spatial representation suitable for quantitative analysis. Terrain objects differ not only in thematic meaning, but also in geometry: polygon features are commonly represented by their area within an analytical unit, linear features by their length, and point objects by their frequency. When these measures are aggregated to a regular grid and supplemented with terrain-morphology variables, each cell can be expressed as a numerical feature vector. This representation preserves the spatial context of the source data while enabling the application of statistical and machine-learning techniques to continuous passability modelling.

Once geospatial features have been transformed into a structured attribute space, machine-learning methods provide a practical complement because they can estimate complex relationships between terrain characteristics and a continuous passability coefficient.

Despite the increasing use of GeoAI in terrain assessment, comparatively little attention has been paid to the relationship between global predictive performance and local spatial reliability. Most evaluations emphasise summary statistics calculated over the complete test dataset, whereas the geographic distribution of residuals is rarely treated as a primary analytical result. This omission is significant because prediction errors may be concentrated in specific terrain categories, geographic contexts, or poorly represented combinations of features. The present study addresses this gap by combining conventional regression metrics with spatially explicit residual analysis and by examining how incomplete geospatial reference data affect the reliability of the resulting passability surface.

At the same time, machine learning introduces a distinct set of methodological risks. A model may achieve good average scores while still producing spatially misleading predictions in locations where the training data do not adequately represent the actual terrain process (Campbell et al., 1997; Pokonieczny, 2018; Rada et al., 2020). This issue is particularly important for map-based products because end users interpret spatial patterns rather than loss values. A map with acceptable mean error may nevertheless mislead route planning or terrain interpretation if uncertainty is spatially clustered. Overall accuracy and operational usefulness should therefore be regarded as related, but not equivalent, criteria.

This paper addresses that problem by analysing uncertainty in AI-based terrain traversability prediction. The objectives are twofold. First, it presents a practical workflow for estimating a continuous passability coefficient from structured geospatial attributes using an artificial neural network. Second, it examines where the resulting errors occur and what they reveal about the underlying data. The central argument is that uncertainty in terrain traversability prediction cannot be explained solely by model architecture; it also reflects incomplete reference geodata, uneven representation of extreme cases, and limited spatiotemporal diversity in the learning sample.

2. Study area, data, and target variable

The analysis uses a study area in northeastern Poland derived from VMAP Level 2. For modelling purposes, the area was partitioned into regular basic fields with a spatial resolution of 100 m x 100 m (Figure 1). This regular grid served as the analytical support for both feature generation and target assignment. The method follows earlier work in which terrain passability is represented at the level of a basic field and estimated from the composition of features located within it.

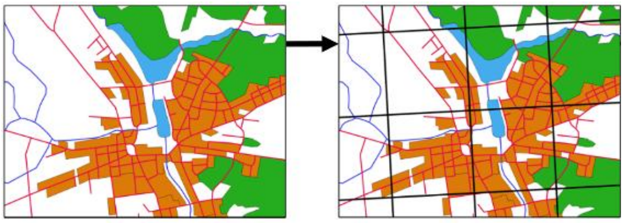

Figure 1. Spatial data partitioning into regular basic fields.

Within each grid cell, the vector database was transformed into a structured set of numerical descriptors. Surface features such as forests, lakes, built-up areas, and wetlands were represented by area within the cell. Linear features such as roads, railways, rivers, and clearways were represented by feature length. Point features such as buildings or farm objects were represented by counts. Each cell was also described by a terrain-morphology variable derived from slope.

This conversion from a discrete vector model to a continuous analytical feature space is a key methodological step because it allows a conventional geospatial database to be used as input for statistical learning while preserving the spatial meaning of the original classes.

The full source structure contained 227 parameters. Because many attributes were empty as a consequence of missing thematic content in the reference data, only 115 non-empty parameters were retained for modelling. The resulting learning table comprised 236,617 records. Exploratory analysis showed that the distribution of the target variable, the IOP coefficient, was not normal and that low-value cases were relatively rare in comparison with mid-range values above 0.5. This imbalance is relevant for regression performance because sparsely represented ranges are less likely to be learned with equal fidelity.

Correlation analysis confirmed that the passability coefficient was associated with several geospatial descriptors. The strongest reported relationships listed in Table 1, Figure 2 were observed for forest area (0.840), park area (0.496), lake area (0.352), swamp area (0.292), clearway length (0.269), a depth-related hydrological feature (0.227), clay soil (0.205), slope-related information (0.183), and road length (0.172). These values are plausible from a terrain-analysis perspective and support the assumption that passability can be estimated from structured geospatial features.

The target variable was a continuous passability coefficient in the range 0-1. It was assigned to each cell through an expert-informed terrain assessment procedure used in prior work.

Layer name	Correlation
foresta_af	0.840
parka_aft	0.496
lakea_aft	0.352
swampa_aft	0.292
clearwl_lf	0.269
depthcl_lf	0.227
soil_glin	0.205
dtedl1	0.183
roadl_lft	0.172
z_slowtogo	0.146
soil_pias	0.137
z_notogo	0.133
sanctrya_a	0.131
farmp_pft	0.127
treesp_pft	0.121

Table 1. Layers with strongest correlation.

In the present article, that coefficient is treated as the reference signal that the machine-learning model attempts to approximate from the explanatory variables. The task is therefore formulated as supervised regression rather than classification.

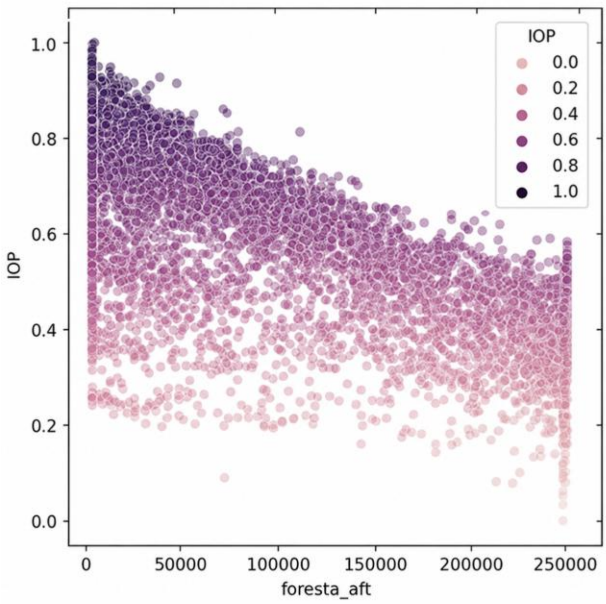

Figure 2. Relationship between IOP and forests area.

3. Methods

The learning workflow was implemented in an open-source Python environment using GeoPandas, pandas, NumPy, scikit-learn, and TensorFlow/Keras. Geospatial input was read from columnar storage, empty columns were removed, remaining missing values were filled with zero, and explanatory variables were normalized with MinMaxScaler prior to model fitting. For the final evaluation reported here, 20% of the observations were reserved as a held-out test subset.

A compact artificial neural network was selected for the predictive model. The network consists of an input layer, one hidden layer with 57 neurons, and a single-neuron output layer. The hidden layer uses the ReLU activation function and L2 regularization with a factor of 0.001. A dropout layer with probability 0.5 was applied to reduce the risk of overfitting. The output layer also uses ReLU and L2 regularization. The model contains 6,670 trainable parameters, which makes it lightweight in comparison with contemporary deep-learning architectures and appropriate for tabular geospatial regression.

Model training used the Adam optimizer with an inverse-time decay learning-rate schedule. Early stopping and model checkpointing were applied in order to stabilize optimisation and retain the best-performing model state. This training configuration reflects a balance between simplicity and regularization: the objective was not to maximize architectural complexity, but to obtain a transparent baseline that would make the effects of data quality easier to interpret.

Model performance was evaluated on the held-out test subset using BIAS, mean absolute error (MAE), and root mean squared error (RMSE). In addition to these aggregate statistics, predicted values were reconnected to the original grid-cell geometries to enable spatially explicit validation. Residuals were mapped and compared with the reference passability surface in order to identify geographically concentrated patterns of overestimation and underestimation. Areas exhibiting the largest discrepancies were subsequently examined against the underlying source data to determine whether the errors were associated with rare terrain conditions, incomplete feature vectors, or missing thematic information. This combined statistical and spatial evaluation was used to distinguish general model error from uncertainty inherited from the geospatial reference database.

The methodological decision to keep the network simple is deliberate. The article does not claim that neural networks are inherently superior to other regression approaches; rather, it uses a compact ANN as a practical instrument for testing how well structured geospatial attributes can approximate a continuous passability signal and for identifying the conditions under which that approximation becomes unreliable.

4. Results

The trained model achieved good aggregate performance on the held-out test data. The reported statistics are BIAS = 0.0024, MAE = 0.0097, and RMSE = 0.015. Training and validation curves (Figure 3) indicate stable convergence, with no clear evidence of severe overfitting or underfitting.

The comparison of the reference and predicted value distributions provides an additional perspective on model behaviour. The model reproduces the dominant central part of the IOP distribution reasonably well, particularly for the ranges that are most strongly represented in the training sample.

However, the predicted distribution is slightly more concentrated than the reference distribution, which indicates a tendency to smooth rare or extreme observations towards more frequently occurring values. This behaviour is characteristic of regression models trained on imbalanced data and helps explain why the aggregate statistics remain favourable while the least represented passability classes exhibit larger deviations.

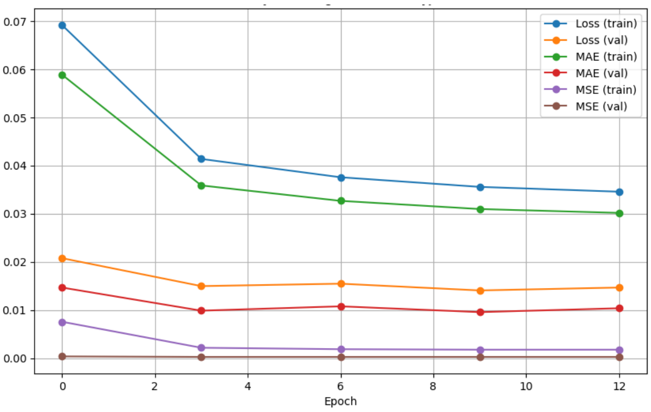

Figure 3. Training and validation curve.

At first sight, these values suggest that the model provides an adequate approximation of the reference passability coefficient. (Figure 4).

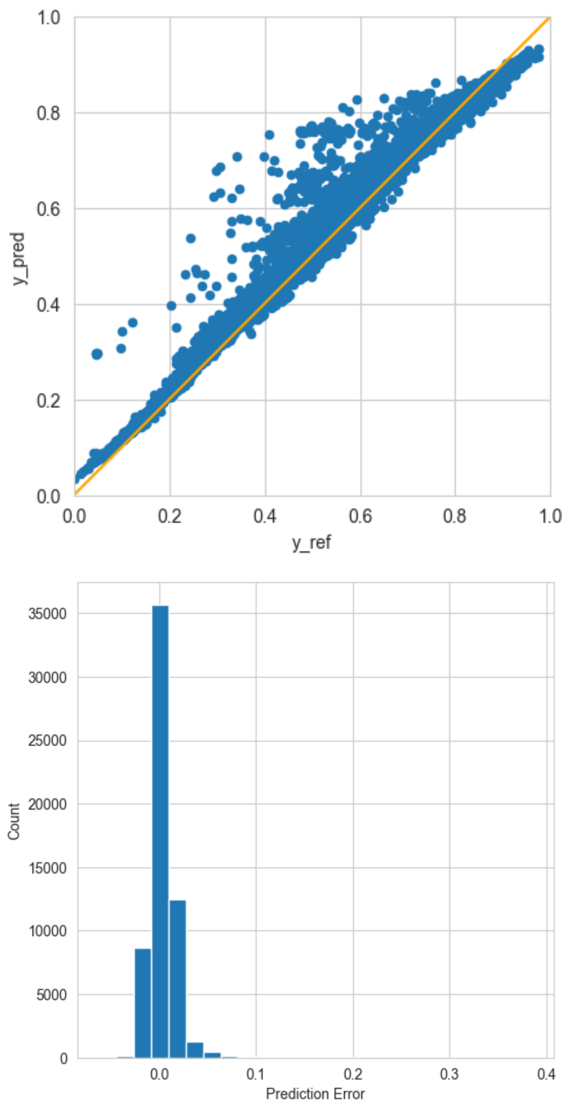

Figure 4. Training results, prediction error.

https://doi.org/10.5194/isprs-archives-L-4-W1-2026-341-2026

A more detailed pattern emerges when the predicted values are compared with the reference values across the full coefficient range. The relationship departs from the ideal prediction line at the lower end of the range and becomes less stable near the upper boundary. Very low IOP values are predicted with lower accuracy, while predictions for very high values also show greater deviation from the reference data. In both cases, the model tends to shift estimates towards the central, more strongly represented part of the distribution. This pattern is consistent with the imbalance identified during exploratory data analysis, where extreme coefficient ranges were represented by relatively few observations (Figure 5).

Visual comparison of the reference and predicted maps shows that the model preserves the broad spatial pattern of terrain traversability across most of the study area. Large forested, agricultural, and relatively open terrain units are generally reproduced consistently, and the principal zones of higher and lower passability remain visible in the prediction.

The model therefore captures the dominant landscape-scale relationships represented in the training data. Differences become more pronounced near local boundaries and in grid cells containing unusual combinations of terrain features, suggesting that broad spatial structures are reconstructed more reliably than rare or highly localised conditions.

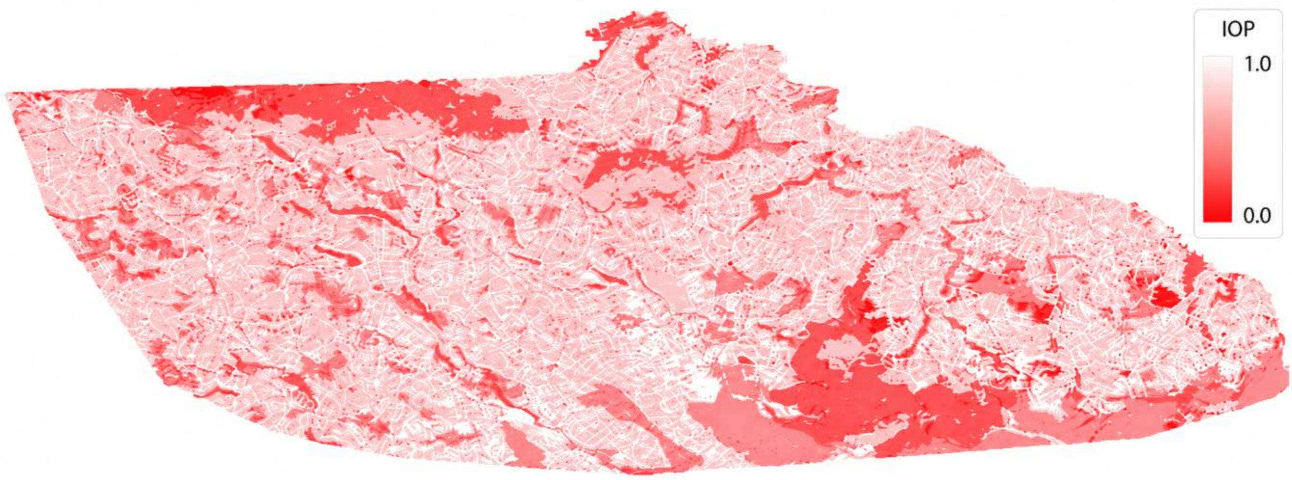

Figure 5. Passability map based on the ANN model.

The residual distribution is broadly centred around zero, and the proportion of large absolute errors remains limited, which is consistent with the reported BIAS, MAE, and RMSE values. Nevertheless, the residuals are not spatially random. Their geographic distribution reveals local clusters of higher error magnitude, indicating that uncertainty is structured by terrain context rather than being caused by statistical noise alone. The spatial pattern of residuals therefore provides information that cannot be obtained from global evaluation metrics.

The spatial residual analysis further indicates that error magnitude is related to the completeness of the feature vectors describing individual grid cells. Cells represented by commonly occurring and consistently encoded terrain categories generally produce smaller prediction errors. By contrast, cells containing rare objects, uncommon combinations of features, or incompletely described terrain conditions are more likely to form local clusters of overestimation or underestimation. This relationship confirms that prediction reliability varies spatially and depends not only on the learned model parameters, but also on the thematic completeness of the underlying geospatial database.

Two local cases are especially revealing. The largest discrepancy occurs in a mining area visualised in Figure 6. According to the project report, the error arose because the corresponding grid cells did not contain sufficiently detailed information describing the presence and spatial characteristics of the mine. Consequently, the feature representation supplied to the model did not adequately reflect the actual terrain conditions. The model generated a prediction that was consistent with the available input attributes but inconsistent with the real-world characteristics of the location.

A second important discrepancy was identified in a wetland area. In this case, missing or incomplete swamp-related information in the VMAP database contributed to a visible mismatch between the reference and predicted passability values. The model was therefore unable to account fully for the effect of wet ground conditions because the relevant thematic information was absent from the corresponding feature vectors. As in the mining example, the local prediction error resulted not only from model behaviour but also from an incomplete representation of the terrain in the source data.

These cases demonstrate that the model may achieve low average error while still producing locally misleading results. Increasing the complexity of the neural-network architecture would not necessarily resolve such discrepancies, because the missing terrain information is not available to the model during either training or prediction. Errors of this type originate upstream in the analytical workflow, during the creation of the geospatial reference database and the transformation of terrain objects into machine-learning features.

The results therefore justify a distinction between average predictive accuracy and spatial reliability. Aggregate statistics describe the overall numerical agreement between predicted and reference values, but they do not indicate whether errors are geographically concentrated or associated with specific terrain categories. For map-based decision support, this distinction is essential because a relatively small local error cluster may affect the interpretation of a critical terrain feature or the feasibility of a proposed route.

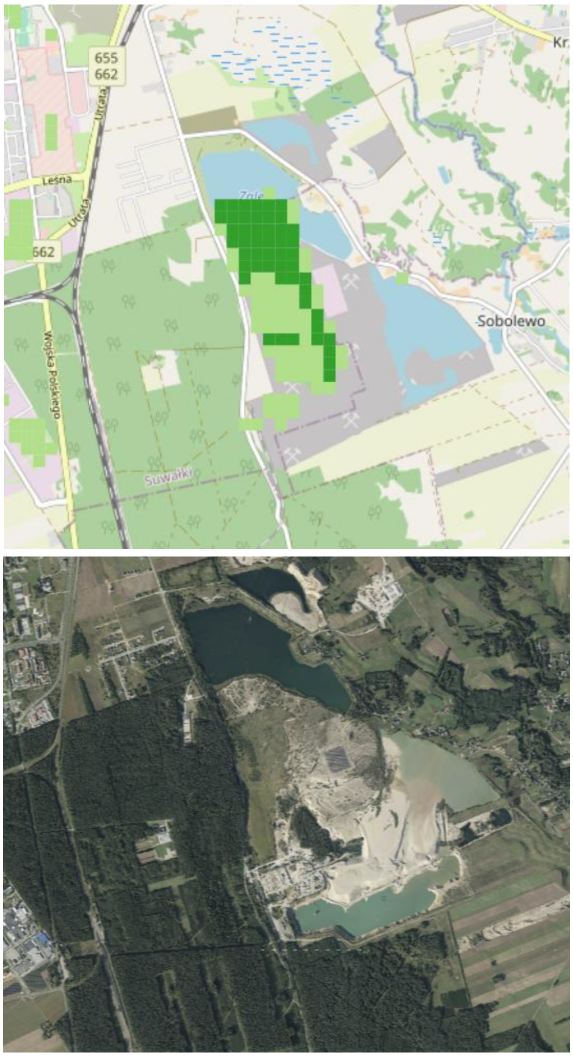

Figure 6. Model overestimation of passability in mine area. The authors’ passability layer presented over an OpenStreetMap basemap (© OpenStreetMap contributors; data available under the Open Database License, ODbL) and orthophotomap from the Polish National Geodetic and Cartographic Resource (PZGiK), provided by the Head Office of Geodesy and Cartography (GUGiK) via Geoportal.gov.pl;. imagery year: 2021 accessed on 12.2025. Source: GUGiK/PZGiK

Consequently, the resulting traversability surface should be interpreted together with information on feature availability, thematic completeness, and the representativeness of the training data. A machine-learning-derived geospatial product cannot be evaluated solely as a numerical approximation problem. If the source database omits, simplifies, or inconsistently encodes critical terrain categories, the model may generate spatially unreliable outputs even when global error measures remain low. Uncertainty should therefore be interpreted as a compound outcome of model behaviour, data imbalance, and geospatial data incompleteness.

5. Discussion

The results have three principal implications for GeoAI-based terrain analysis. First, they confirm that structured geospatial attributes can be used to estimate a continuous passability coefficient with relatively low average error. This finding is consistent with earlier work on terrain assessment and supports the continued use of machine-learning methods in traversability studies.

Second, the study demonstrates that uncertainty in geospatial machine learning is often inherited from the source data rather than produced exclusively by model design. This point is methodologically important because it shifts the analytical emphasis from architectural tuning to the quality, completeness, and semantic adequacy of the reference database. In other words, better models alone will not necessarily resolve uncertainty if the geospatial representation of terrain remains incomplete.

Third, the study highlights a representativeness problem that extends beyond thematic completeness. The training sample is geographically limited and lacks explicit temporal labels, even though passability is inherently sensitive to seasonality and changing environmental conditions. A model that performs well within the study area may therefore not retain the same level of accuracy when transferred to another region or to similar terrain under different seasonal conditions.

Fourth, the results indicate that model evaluation should combine conventional statistical metrics with spatially explicit residual analysis. Aggregate measures such as MAE and RMSE are useful for summarising overall performance, but they may conceal geographically concentrated errors associated with specific terrain classes or missing source information. Mapping the residuals made it possible to identify clusters of overestimation and underestimation and to relate them directly to deficiencies in the underlying geospatial database. This suggests that spatial diagnostics should be treated as a standard component of GeoAI model validation whenever the output is intended to support map-based interpretation or decision-making.

Fifth, the observed error patterns have practical implications for the deployment of traversability models in operational environments. Predictions should not be interpreted as equally reliable across the entire study area, particularly where rare terrain conditions, incomplete thematic attributes, or out-of-distribution combinations of features occur. Future implementations should therefore complement the predicted passability layer with indicators of data completeness, model confidence, or local prediction reliability. Such additional information would allow users to distinguish between areas supported by representative training data and locations where the model output should be treated with greater caution.

These observations reinforce the relevance of the paper to the FOSS4G Academic Track. The contribution is not a generic machine-learning exercise, but an open geospatial study concerned with the reliability of derived map products. By emphasising the interaction among data quality, spatial analysis, and open-source implementation, the work is aligned with broader discussions in GeoAI about transparency, reproducibility, and operational trustworthiness.

6. Conclusions

Terrain traversability can be estimated from structured geospatial attributes using a compact artificial neural network, and the reported aggregate statistics indicate that the approach is viable for regression-based passability modelling. However, the study also shows that acceptable global error values do not guarantee spatial reliability. The most informative errors occur in terrain contexts that are absent, simplified, or weakly represented in the reference geodata.

A further conclusion concerns the transferability of the proposed approach. Because the model was trained within a geographically limited study area and on a dataset without explicit temporal information, its performance cannot be assumed to remain unchanged in other regions or under different seasonal conditions.

Future research should therefore include cross-region validation, more balanced representation of extreme passability values, and the integration of temporally variable factors such as precipitation, soil moisture, and freeze–thaw conditions. These extensions would help distinguish uncertainty caused by local data limitations from uncertainty resulting from genuine environmental variability.

The practical implication is straightforward. In GeoAI applications for traversability mapping, model evaluation should extend beyond BIAS, MAE, and RMSE to include spatially explicit error analysis and critical review of the thematic completeness of the source database. Without that broader perspective, apparently accurate models may still produce locally misleading outputs in the very places where reliable terrain information is most needed.

References

Campbell, L., Lotwin, A., DeRico, M.M.G., Ray, C., 1997. The use of artificial intelligence in military simulations. In: Proceedings of the 1997 IEEE International Conference on Systems, Man, and Cybernetics: Computational Cybernetics and Simulation, Orlando, FL, USA, 12–15 October 1997, Vol. 3, pp. 2607–2612. doi: 10.1109/ICSMC.1997.635328.

Hubáček, M., Rybansky, M., Cibulova, K., Brenova, M., Ceplova, L., 2015. Mapping the passability of soils for vehicle movement. *KVÜÕA Toimetised*, 21, 5–18.

Kim, T.J., Choi, K., 2012. GIS for transportation. In: Kresse, W., Danko, D.M. (eds.), *Springer Handbook of Geographic Information*. Springer, Berlin, Heidelberg, pp. 503–521. doi: 10.1007/978-3-540-72680-7_26.

Meng, X., Cao, Z., Liang, S., Pang, L., Wang, S., Zhou, C., 2018. A terrain description method for traversability analysis based on elevation grid map. *International Journal of Advanced Robotic Systems*, 15(1), 1729881417751530. doi: 10.1177/1729881417751530.

Pokonieczny, K., 2018. Use of a multilayer perceptron to automate terrain assessment for the needs of the armed forces. *ISPRS International Journal of Geo-Information*, 7(11), 430. doi: 10.3390/ijgi7110430.

Pokonieczny, K., Wyszyński, M., 2016. Automation of the terrain assessment classification due to passability for the needs of crisis managment. In: Proceedings of the 16th International Multidisciplinary Scientific GeoConference SGEM 2016, Book 2, Vol. 3, pp. 119–126. doi: 10.5593/SGEM2016/B23/S11.016.

Polanco, X., François, C., Keim, J.-P., 1998. Artificial neural network technology for the classification and cartography of scientific and technical information. *Scientometrics*, 41(1–2), 69–82. doi: 10.1007/BF02457968.

Rada, J., Rybansky, M., Dohnal, F., 2020. Influence of quality of remote sensing data on vegetation passability by terrain vehicles. *ISPRS International Journal of Geo-Information*, 9(11), 684. doi: 10.3390/ijgi9110684.

Rybansky, M., Hofmann, A., Hubacek, M., Kovarik, V., Talhofer, V., 2015. Modelling of cross-country transport in raster format. *Environmental Earth Sciences*, 74, 7049–7058. doi: 10.1007/s12665-015-4759-y.