

DOCTORAL THESIS SUMMARY

Contributions to the Study of Land Use and Land Cover in Flood Susceptibility Assessment Using Aerospace Remote Sensing Methods

Author: Eng. **Ioana PAVEL (VIZIREANU)**

Email: vizireanu.yoana@gmail.com, *phone:* +40 746 957 328

PhD Supervisor: Col.(r) Prof. Univ. Dr. Eng. **Dan-Gabriel RĂDUCANU**

This doctoral thesis addresses the assessment of flood susceptibility at a local scale through the integration of high-resolution geospatial data and modern machine learning methods, with the aim of analysing hydrogeomorphological processes in rural environments. The research is framed within the context of increasing extreme hydrometeorological events and accelerated land use and land cover transformations, which directly influence infiltration, surface runoff, and water accumulation processes, making the identification and quantification of flood susceptibility essential.

The study area is Mihăești commune, Argeș County, located within the catchment of the Târgului River, a representative area for the geomorphological and functional conditions specific to the sub-Carpathian sector of Argeș County. The selection of the study area reflects its morphological variability, the heterogeneous structure of land use, and the presence of sectors prone to local water accumulation, enabling a detailed investigation of the relationships between environmental factors and flood susceptibility.

The main objective of the research is the development and application of an integrated methodology for local-scale flood susceptibility assessment, based on high-resolution remote sensing data and machine learning algorithms, combined with methods for interpreting predictive models. **Specific objectives** include the compilation of a multi-sensor geospatial database, the derivation of relevant morphometric and thematic variables, the definition of the dependent variable through automated hydrogeomorphological methods, the implementation and comparison of classification algorithms, the evaluation of predictive performance, the analysis of model interpretability, independent spatial validation, and the integration of results into a built environment analysis.

The thesis builds on the limitations of approaches based on generalised altimetric models or medium-resolution data, which do not adequately capture the fine spatial heterogeneity of the factors governing water accumulation processes. In this context, an integrated set of high-resolution geospatial data is used, including airborne LiDAR data (mean point density of approximately 8 points/m²), a LiDAR-derived digital terrain model at 1 m resolution, Sentinel-2 multispectral imagery (10 m), high-resolution orthophotos (0.2 m), and thematic vector data.

Through the integration of these sources, a multi-sensor geospatial database comprising 8,982,391 spatial units is developed, each characterised by a coherent set of explanatory variables, including elevation, slope, curvature, aspect, distance to the drainage network, land use and land cover classes, and the NDVI index. The dependent variable is defined through Bluespot hydrogeomorphological analysis, used for the automated identification of topographic depressions with water ponding potential based on the micro-relief derived from the digital terrain model. The susceptible areas thus identified account for 5.53% of the analysed surface, reflecting a significant class imbalance and justifying the use of appropriate modelling techniques.

In the modelling stage, three classification algorithms are implemented and compared: Random Forest, XGBoost, and logistic regression. Model training is performed on a stratified sample of 1,000,000 observations, while the final evaluation is conducted on an independent test set of 300,000 observations. The class imbalance, characterised by an approximate ratio of 1:17, is addressed using the SMOTE technique, and hyperparameter optimisation is performed through cross-validation. The robustness of the results is assessed using two data splitting strategies: 70–30, in which 70% of the data is used for training and 30% for testing, and 80–10–10, in which 80% is used for training, 10% for validation, and 10% for testing. The comparison of these strategies confirms the consistency of the results and the stability of model performance, with observed differences being minor and without significant impact on overall performance.

The evaluation of predictive performance demonstrates the superiority of the Random Forest model, which achieves high values across all statistical indicators (AUC–ROC = 0.868; accuracy = 83.9%; precision = 20.5%; recall = 66.5%; F1 score = 0.313; MCC = 0.306), confirming its capacity to model complex non-linear relationships between environmental variables and flood susceptibility. Based on the optimal model, a flood susceptibility map classified into five classes is generated, highlighting the predominance of low susceptibility levels and the concentration of high susceptibility classes in low-lying areas and in proximity to the drainage network. These results confirm that the integration of high-resolution LiDAR data with ensemble models, namely Random Forest and XGBoost, provides a robust and interpretable framework for local-scale flood susceptibility assessment.

A distinctive aspect of the research is the coupling of the high spatial detail of the input data with the requirement for explainability of machine learning results. The proposed framework aims to generate a predictive susceptibility map accompanied by the interpretation of the relationships between environmental variables and susceptibility, thereby enhancing the practical relevance of the results.

Model interpretability analysis represents a central component of the research and is conducted using the SHAP method, which enables the quantification of each explanatory variable's contribution to the model output and the characterisation of relationships between the analysed factors and susceptibility. This approach allows the results to be interpreted in physical-geographical terms, facilitating the understanding of the mechanisms governing water accumulation. The results highlight the dominant role of elevation, slope, land use and land cover classes, and distance to the drainage network, as indicated by variable importance assessment methods (SHAP, Gini importance, and permutation importance). In addition, SHAP analysis enables the identification of critical variable thresholds relevant for delineating areas of high susceptibility.

Model validation is performed through independent spatial methods, including concentric buffer zone analysis relative to the drainage network and a North–South comparative analysis, which demonstrate the geographical coherence of the results and the model's capacity to reflect realistic spatial distributions of susceptibility. These methods go beyond classical statistical evaluation, providing insight into model behaviour in relation to the spatial structure of the analysed phenomenon. The integration of the results into an applied analysis of the built environment highlights the spatial distribution of buildings across susceptibility classes, enabling the identification of sectors with high exposure to water accumulation.

The main contributions of the thesis consist in the development of an integrated methodological framework for local-scale flood susceptibility assessment, based on the combined use of high-resolution remote sensing data and machine learning methods. A first element of novelty lies in the systematic use of LiDAR data for metric-resolution micro-relief characterisation, enabling the detailed identification of topographic depressions and subtle slope variations relevant to water accumulation. At the same time, the integration of sub-metric

resolution land use and land cover information leads to a significantly improved representation of surface conditions compared to approaches based on generalised datasets.

An important contribution consists in the comparative evaluation of machine learning algorithms within a unified framework, allowing the assessment of their performance under the conditions of a complex and imbalanced dataset. The analysis highlights the advantages of the Random Forest model in capturing non-linear relationships between environmental variables and susceptibility. In addition, the use of interpretability techniques, particularly the SHAP method, represents a significant contribution by enabling the explanation of model outputs and their linkage to real physical processes.

The thesis further contributes through the integration of independent spatial validation, which enables verification of the geographical coherence of the results. The assessment of the influence of dataset configuration highlights the model's stability and the impact of this configuration on the obtained results. The integration of the results into an applied framework, through reference to the built environment, completes the methodological approach and demonstrates the practical utility of the developed model.

Through these elements, the thesis proposes a coherent, reproducible, and adaptable framework applicable to flood susceptibility analysis in similar contexts characterised by high spatial variability and the availability of high-resolution data.

Future research directions involve extending the methodological framework through the testing of additional machine learning algorithms, such as Support Vector Machine (SVM) approaches or convolutional neural network (CNN)-based models, in order to evaluate their capability to handle datasets characterised by pronounced class imbalance. Furthermore, future studies may include the use of additional evaluation indicators, such as the area under the Precision–Recall curve (PR-AUC), complementary to the AUC–ROC metric, as well as the identification of an optimal classification threshold for improving the balance between recall and precision.

Overall, the study contributes to the advancement of modern flood susceptibility assessment methods through the integration of high-resolution geospatial data with advanced modelling and interpretation techniques. The proposed methodological framework has direct applicability in supporting spatial planning processes and in identifying areas with high predisposition to flooding, particularly in contexts characterised by high spatial variability.