Proceedings of the 21st ICSMGE, Vienna, Austria, 14 – 19 June 2026. Pistrol, Adam & Schweiger (eds.)
Published by: ÖGG, Austrian Society for Geomechanics, Salzburg, Austria, ISBN 978-3-9503898-4-5
https://doi.org/10.53243/ICSMGE2026-1305

Interpretable Artificial Intelligence for Subsidence Mapping Using City-Scale Geotechnical and Infrastructure Features

Homin Park
Singapore University of Social Sciences, Singapore; hominpark@suss.edu.sg

Yongmin Kim, Yuan Shen Chua
University of Glasgow, United Kingdom; yongmin.kim@glasgow.ac.uk; 2942017C@student.gla.ac.uk

Junghwan Kim
The Seoul Institute, South Korea; imkim@si.re.kr

Taesik Kim
Hongik University, South Korea; taesik.kim@hongik.ac.kr

ABSTRACT: Ground subsidence, including the sudden formation of sinkholes, poses a serious threat to urban infrastructure, public safety, and long-term sustainability. Anticipating such events requires predictive models that are not only accurate but also interpretable and applicable in data-scarce environments. This study proposes a feature-driven framework for estimating subsidence risk using geotechnical and urban infrastructure data. We evaluate two approaches, namely XGBoost and TabNet, based on 22 features derived from geological strata, utility networks, and surface-level urban characteristics. To address the scarcity of labeled incident data, we implement a spatial partitioning strategy to simulate real-world deployment, and apply class weighting to mitigate extreme class imbalance. Experimental results show that both models achieve high recall across a range of classification thresholds. In particular, XGBoost offers significantly greater interpretability through its tree-based decision structure and explicit feature attribution. Feature importance analysis highlights the dominant influence of subsurface utility density, particularly from telecommunications and sewer systems, in identifying high-risk zones. Given the safety-critical nature of ground subsidence prediction, we conclude that XGBoost provides a more actionable and trustworthy basis for decision support in urban geohazard management.

KEYWORDS: Ground subsidence, machine learning, deep learning, interpretability.

1 INTRODUCTION

Ground subsidence, including the sudden formation of sinkholes, is a critical geohazard that threatens the safety, reliability, and sustainability of urban infrastructure. Often sudden and localized, such events can lead to severe damage to roads, buildings, pipelines, and underground utilities. More importantly, they pose direct risks to human life and public safety, especially in densely populated areas where even minor ground failures can result in injuries, fatalities, or large-scale evacuations. As cities continue to grow and expand their reliance on subterranean infrastructure, the need for accurate and scalable predictive tools has become increasingly urgent.

Recent advances in artificial intelligence have enabled the development of predictive models capable of assessing subsidence risks with increasing accuracy (Bilgilioğlu et al., 2025; Lee et al., 2023; Razavi-Termeh et al., 2025). Despite these advancements, existing approaches exhibit several limitations that hinder their practical deployment. First, deep learning models often function as black-box systems, offering limited interpretability. In domains such as subsidence prediction, where outputs can influence public safety interventions and infrastructure planning, this opacity reduces stakeholder trust and complicates expert validation. Second, many high-performing models rely on specialized input features, such as InSAR deformation metrics, which are not consistently available to municipal agencies or urban planners, particularly in data-constrained regions. Finally, the scarcity of labeled data presents a persistent challenge, as subsidence events are rare, leading to highly imbalanced datasets.

To address these challenges, we propose the use of eXtreme Gradient Boosting (XGBoost), a decision-tree-based ensemble model, for pixel-wise subsidence likelihood estimation (Chen and Guestrin, 2016). Unlike deep neural networks or attention-based architectures such as TabNet, XGBoost provides feature importance measures and enables inspection of individual decision trees, allowing domain experts to trace the rationale behind predictions. This interpretability makes it particularly suitable for high-stakes applications where human oversight is critical. Our framework leverages 22 geospatial and infrastructural features, including geological strata, land use, and subsurface utility characteristics, which are generally available to municipal authorities. By combining data efficiency, predictive performance, and clear interpretability, the proposed XGBoost-based approach offers a practical pathway toward reliable and interpretable geohazard prediction in urban environments.

2 METHODOLOGY

Our proposed framework, shown in Figure 1, includes three subcomponents, namely data preparation (describing input features and preparations done for model training), model architecture (detailing explored models), and model training (highlighting the training procedures employed).

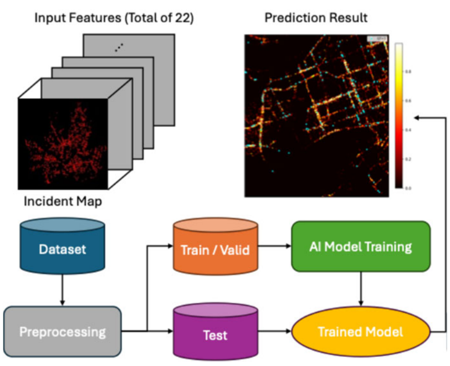

Figure 1. Overview of the proposed framework.

2.1 *Input Features and Data Preparation*

2.1.1 *Feature Description*

To support the pixel-wise prediction of subsidence risks, we compiled a feature set comprising 22 variables, as listed in Table 1. These features were derived from geotechnical investigations, utility records, and urban infrastructure databases. These features were selected based on their relevance to ground deformation and their general availability to municipal authorities or urban planners. Collectively, these features provide a comprehensive representation of subsurface conditions and anthropogenic influences that may contribute to local instability (Ji et al., 2024).

We first included a depth measurement of geological strata: clay, gravel, sand, silt, weathered rock, and the ground surface. These measurements were derived from borehole records and indicated as either layer thickness or depth to top. They reflect soil stratification, which governs settlement behavior, water infiltration, and redistribution of stress in the subsurface.

Subsurface utility infrastructure was characterized through two major groups of features. The first group consists of pressurized or gravity-driven pipeline networks, including gas, water supply, sewer, and district heating systems. For these networks, we primarily considered pipeline density, diameter, burial depth, slope, and installation year. These attributes capture both the physical footprint and the age of underground utilities, which influence subsidence risk through excavation history, leakage, long-term material degradation, and localized soil disturbance. The second group represents conduit-based networks, such as telecommunication and electrical systems, for which duct density, diameter, and installation year were included. Although individual ducts are smaller than major pipelines, their extensive coverage and frequent maintenance can contribute to shallow ground instability, especially in densely serviced areas.

By combining geological and infrastructural indicators, the selected features capture the interaction between in-situ soil conditions and human activities. The dual perspective improves the model's ability to identify areas prone to subsidence while ensuring that the inputs remain accessible in city scale management systems. This design supports predictive performance and interpretability, which are both critical for practical decision making in urban geohazard management.

2.1.2 *Feature Preparation for Model Input*

To prepare the selected features for predictive modeling, all 22 variables were spatially aligned and stacked into a three-dimensional tensor with dimensions corresponding to height, wide, and feature channels. Each feature layer represents a spatial grid of the study area, where each pixel contains the corresponding value of that feature. This unified representation facilitates a consistent input format for all models considered in this study. XGBoost and TabNet process each pixel independently, treating the 22 feature values at a given location as a tabular input vector.

To ensure numerical stability and balanced feature contribution, each feature layer was normalized prior to model training. We applied feature wise normalization to rescale each layer based on its own distribution, transforming the values to a consistent range. This step is particularly important given the heterogeneous nature of the input features, which include physical dimensions such as depth or diameter as well as density-based variables. Proper normalization prevents features with large absolute values from dominating the learning process and improves both convergence and generalization across models.

Table 1. Input features and their corresponding XGBoost-based importance scores. The top five features are highlighted in bold.

Category	Feature Name	Importance Score
Geological Stratigraphy (Depth)	*Clay*	0.025
	Gravel	0.026
	Sand	0.024
	Silt	0.025
	Weathered Rock	0.023
	Ground Surface	0.021
Gas Infrastructure	*Density*	0.021
	Diameter	0.022
	Installation Year	0.027
Heating System	*Diameter*	0.024
	Installation Year	0.015
Telecommunication Ducts	***Density***	**0.293**
	Diameter	**0.035**
	Installation Year	0.020
Electrical Power Ducts	***Density***	**0.035**
	Diameter	0.022
Sewer System	***Density***	**0.216**
	Burial Depth	0.020
	Slope	0.023
	Installation Year	0.020
Water Supply Infrastructure	*Density*	0.022
	Diameter	**0.038**

2.2 *Model Description*

2.2.1 *XGBoost*

To explore an interpretable and computationally efficient model, we implemented an eXtreme Gradient Boosting (XGBoost) classifier for pixel-wise prediction of ground instability. XGBoost is a high-performance ensemble method based on gradient-boosted decision trees, known for its robustness on structured tabular data. Unlike neural networks, which operate as black-box models, XGBoost offers explicit feature importance measures and decision logic, making it well-suited for applications where model transparency is critical.

The classifier was trained using the same 22 geotechnical and infrastructural features, and each pixel was treated as an independent instance described by its corresponding feature values. XGBoost builds an additive model by sequentially training decision trees that correct the residuals of previous ones, optimizing a regularized objective function that balances classification performance and model complexity. This iterative refinement allows the model to capture nonlinear relationships and feature interactions without extensive preprocessing or feature engineering.

In addition to achieving competitive predictive performance, the XGBoost model provides feature importance scores, which are used to evaluate the relative contribution of each input variable to the classification task. This built-in interpretability makes XGBoost particularly valuable in risk-sensitive geotechnical applications, where understanding the reasoning behind predictions is essential for decision-making and policy implementation.

2.2.2 *TabNet*

To establish a baseline model, we implemented TabNet, a deep learning architecture specifically designed for learning from tabular data while maintaining a degree of interpretability (Arik and Pfister, 2021). Unlike conventional fully connected networks, TabNet employs a sequential attention mechanism that selectively focuses on the most relevant features at each decision step. This feature selection process is implemented using sparse attention masks, which allow the model to capture complex, non-linear relationships between input variables without requiring extensive manual feature engineering.

In the context of subsidence prediction, TabNet is particularly appealing because it can automatically identify key geospatial and infrastructural features that contribute most to ground instability risk. The model processes each pixel's 22-dimensional feature vector through multiple decision steps, producing intermediate feature masks that highlight the variables influencing the prediction. These masks provide a form of model interpretability, as they reveal which features the network prioritized when assessing the likelihood of subsidence at a given location. Despite this interpretability advantage over conventional deep neural networks, TabNet does not offer a fully human-readable decision path like decision-tree-based models. Instead, its explanations rely on feature importance scores derived from attention distributions, which can guide expert interpretation but are less explicit than the rule-based structure of XGBoost.

3 RESULTS AND DISCUSSION

To assess the effectiveness and practical applicability of the proposed subsidence prediction framework, we conducted a comprehensive evaluation using real-world geotechnical and infrastructure data. At first, predictive performance was evaluated to quantify the ability of the models to identify locations with high subsidence risk. Lastly, interpretability was analyzed to demonstrate how feature-driven methods, particularly XGBoost, can provide actionable insights into the factors driving model predictions.

3.1 *Dataset Partitioning and Experimental Setup*

3.1.1 *Train and Test Strategy*

A major challenge in subsidence prediction is the scarcity of labeled data, as ground failures such as sinkholes are rare and highly localized. In our study, only a single city-scale incident map was available, which required a spatial partitioning approach to generate training, validation, and test samples. Rather than using random pixel sampling, which risks spatial information leakage and produces overoptimistic performance estimates, we divided the study area into a 2 by 2 grid of non-overlapping quadrants. Each quadrant functions as a self-contained spatial unit, ensuring that pixels used for training do not appear in the test set. During model evaluation, three quadrants are used for training, while the remaining quadrant is reserved for testing.

To prevent overfitting and ensure generalization, hyperparameters such as learning rate, regularization terms, and more were empirically adjusted. Both model outputs a probability score between 0 and 1 for each pixel, which is converted into a binary classification using a threshold. Note that the binary cross entropy loss was used for both models.

3.1.2 *Handling of Class Imbalance*

Ground subsidence events are inherently rare, resulting in a highly imbalanced dataset. In our study, the labeled dataset contained 1,247,113 negative pixels representing stable ground and only 1,472 positive pixels corresponding to documented subsidence events. This extreme imbalance presents a major challenge for predictive modeling because classifiers are biased toward the majority class and often underpredict rare events.

To mitigate this issue, XGBoost and TabNet were trained with class weighting, where the scale positive weight was set to the ratio of negative to positive samples. This ensures that each positive instance exerted proportionally greater influence during training and reducing the likelihood of false negatives.

Table 2. Performance with varying prediction threshold.

Threshold	Model	Precision	Recall	F1
0.7	*XGBoost*	0.014	0.502	0.028
	TabNet	0.010	0.865	0.020
0.5	*XGBoost*	0.008	0.888	0.017
	TabNet	0.006	0.936	0.013
0.2	*XGBoost*	0.001	1.000	0.002
	TabNet	0.004	0.981	0.008

3.1.3 *Performance Metrics*

Evaluating predictive performance in the context of ground subsidence requires metrics that reflect the high class imbalance and the safety-critical nature of the task. Conventional accuracy is not informative in this setting because the overwhelming prevalence of negative samples would allow a trivial model that predicts all negatives to achieve near-perfect accuracy while completely failing to identify rare subsidence events.

Instead, we adopt precision, recall, and F1-score as our primary evaluation metrics. Precision measures the proportion of predicted positives that correspond to actual subsidence events, reflecting the model's ability to avoid false alarms that could lead to unnecessary interventions or resource allocation. Recall quantifies the proportion of actual subsidence events that are correctly identified, which is crucial in safety-critical applications where missed detections can result in significant infrastructure damage or threats to public safety. F1-score, defined as the harmonic mean of precision and recall, provides a single measure that balances the trade-off between these two competing objectives.

3.2 *Predictive Performance Analysis*

The predictive performance of XGBoost and TabNet was evaluated with varying decision thresholds as shown in Table 2. Across all thresholds, the models exhibit high recall but very low precision, a characteristic that reflects the inherent challenges of this application. For instance, at a threshold of 0.5, XGBoost achieves a recall of 0.888 while maintaining a precision of only 0.008. TabNet shows a similar trend, reaching a recall of 0.936 with a precision of 0.006. Consequently, the F1 scores remain low across all settings. At a high prediction threshold of 0.7, both models generate conservative positive predictions, yielding low recall but slightly higher precision. XGBoost achieves 0.014 precision with 0.502 recall, while TabNet reaches 0.010 precision with 0.865 recall. Lowering the threshold to 0.2 drives recall near saturation but reduces precision to near zero.

This behavior is primarily explained by the severe class imbalance and the nature of the incident map. Furthermore, the incident map may not capture all future events or areas with high latent risk, as regions that are truly vulnerable but have not yet failed are labeled as negative. In such cases, models that identify a broader set of potentially hazardous areas will naturally achieve high recall but low precision, as many predicted positives may correspond to areas that have not yet experienced visible incidents.

From a practical decision-making perspective, this imbalance in model behavior is not necessarily detrimental. In safety-critical applications, recall is often prioritized over precision, as missing a high-risk location could lead to catastrophic consequences. The low precision simply indicates that some predicted risk areas require field validation or further monitoring. Urban safety managers and geotechnical engineers can leverage these model outputs, as shown in Figure 2, to prioritize site inspections, deploy additional sensors, or schedule preventive maintenance, ensuring that resources are focused on the high risk regions.

Overall, TabNet demonstrates consistently higher recall, suggesting stronger sensitivity to positive cases, whereas XGBoost exhibits slightly better precision at higher thresholds, which may be advantageous in risk-sensitive decision-making where minimizing false positives is critical. These results underscore the importance of threshold tuning in operational deployment and suggest that additional techniques, such as cost-sensitive learning or anomaly ranking, may be required to balance detection sensitivity with actionable precision.

3.3 *Interpretability and Feature Importance from XGBoost*

One of the key advantages of using XGBoost for subsidence prediction lies in its intrinsic interpretability. Table 1 presents the input features and their corresponding XGBoost-based importance scores, with the top five features highlighted in bold. The results indicate that subsurface utility density features dominate the predictive landscape. This finding aligns with engineering intuition, as areas with dense underground utilities are more susceptible to ground disturbances due to excavation history, leakage, and long-term soil weakening. In contrast, geological stratigraphy features, while relevant, exhibit lower importance scores, suggesting that anthropogenic factors play a stronger role in explaining the observed subsidence events.

To further illustrate model interpretability, Figure 3 shows the SHAP plot for a specific instance (i.e., a pixel), which quantifies each feature's contribution to model outputs at the instance level (Lundberg and Lee, 2017). The horizontal spread of each feature reflects the magnitude and direction of its influence. Consistent with the global importance scores, telecommunication duct density and sewer line density exert the largest positive contributions to subsidence likelihood predictions. Meanwhile, geological features such as sand, silt, and gravel layers primarily provide stabilizing effects or minor contributions. This dual perspective of global importance and local attribution provides domain experts with actionable insights: high-risk predictions are often linked to regions with concentrated underground utilities, which can guide both preventive monitoring and maintenance planning.

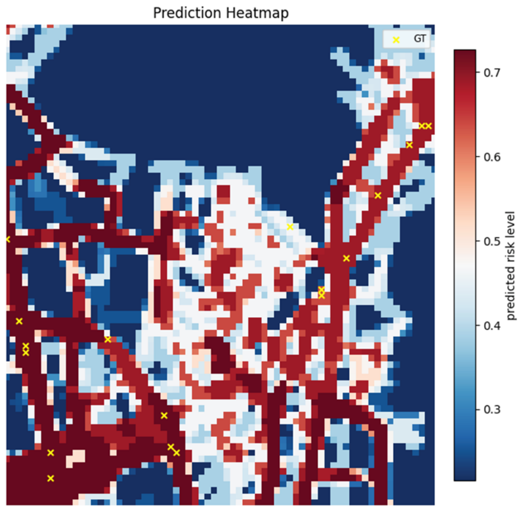

Figure 2. XGBoost prediction heatmap.

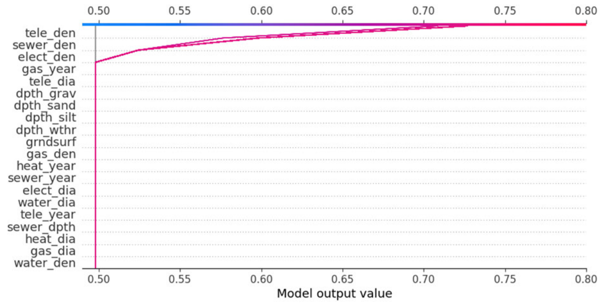

Figure 3. XGBoost instance level interpretability.

Feature attribution analysis confirms that subsurface utility infrastructure, particularly telecommunications and sewer networks, plays a critical role in subsidence risk. While TabNet delivers competitive performance, its limited transparency makes it less suitable for applications that require expert validation and public accountability. Given the safety-critical nature of subsidence prediction, we recommend XGBoost as a more practical tool for supporting early risk identification and resource prioritization in urban geohazard management. Future work may explore hybrid approaches or the integration of temporal and sensor-based data to enhance model robustness and operational relevance.

4 CONCLUSIONS

This study presents an interpretable and scalable framework for ground subsidence prediction using city-scale geotechnical and infrastructure data. We compare the performance of XGBoost and TabNet across highly imbalanced datasets and spatially partitioned evaluation settings. Both models achieve high recall, demonstrating strong potential for identifying hazardous areas in data-scarce urban environments. In particular, XGBoost offers a distinct advantage in interpretability, allowing users to inspect decision rules and quantify feature contributions directly.

5 REFERENCES

Bilgilioğlu, S.S., Gezgin, C., Iban, M.C., Bilgilioğlu, H., Gündüz, H.I. and Arslan, Ş., 2025. Explainable Sinkhole Susceptibility Mapping Using Machine-Learning-Based SHAP: Quantifying and Comparing the Effects of Contributing Factors in Konya, Türkiye. Applied Sciences, 15(6), p.3139.

Lee, S., Kang, J. and Kim, J., 2023. Machine learning-based predictive model of ground subsidence risk using characteristics of underground pipelines in urban areas. IEEE Access, 11, pp.69326-69336.

Razavi-Termeh, S.V., Sadeghi-Niaraki, A., Ali, F., Pirasteh, S. and Choi, S.M., 2025. Enhancing land subsidence susceptibility mapping using deep tabular learning optimization with metaheuristic algorithms. Gondwana Research.

Chen, T. and Guestrin, C., 2016, August. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

Arik, S.Ö. and Pfister, T., 2021, May. Tabnet: Attentive interpretable tabular learning. In Proceedings of the AAAI conference on artificial intelligence (Vol. 35, No. 8, pp. 6679-6687).

Lundberg, S.M. and Lee, S.I., 2017. A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

Ji, Y.M., Kim, J.H., Lee, J.W., and Choi, H. Development of Ground Subsidence Risk Assessment Model Using Analytic Hierarchy Process Method, Journla of Korean Geotechnical Society, 40(6), 57-65.