Proceedings of the 10th World Congress on New Technologies (NewTech'24)
Barcelona, Spain - August 25-27, 2024
Paper No. ICCEIA 127
DOI: 10.11159/icceia24.127

New Insights into Road Cavity Detection from GPR Data

Ahmed Elseicy[1], Mercedes Solla[1], Pedro Arias[1]
[1]CINTECX, GEOTECH research group, Universidade de Vigo
Campus As Lagoas - Marcosende, Vigo, Spain
ahmedmossadibrahim.elseicy@uvigo.es; merchisolla@uvigo.es; parias@uvigo.es

Abstract – Maintaining the integrity of transportation infrastructure is critical for resilience and safety. Subsurface changes, along with climate change and aging infrastructure, can all contribute to the development of sinkholes, a critical concern for infrastructure. However, early detection is possible through characterization of the factors that influence sinkhole formation. Ground-penetrating radar (GPR) is a practical tool for non-destructive subsurface monitoring and early detection of sinkholes. Nevertheless, conventional GPR evaluation relies heavily on subjective analysis. Deep learning (DL) techniques can automate and improve GPR data analysis, especially for large amounts of collected data. Despite the success of DL in the field of computer vision, limited data availability prevents its widespread application in GPR surveys. In this paper, an overview of GPR applications for cavity detection in transportation infrastructure is discussed, highlighting key findings and limitations. It also explores data preparation techniques, including synthetic data generation and data augmentation, to facilitate the automation of cavity detection from GPR data using DL approaches.

***Keywords*:** Transportation infrastructure, Sinkholes, GPR, Pavement, Cavities, Deep learning, Data generation

1 Introduction

A key factor in assuring the resilience and safety of transportation infrastructure is the integrity of the ground they are built upon. Any weaknesses or deficiencies in the ground (e.g. inadequate drainage or fluctuating water table) can lead to serious structural problems, such as sinkholes, pavement failures, and bridge collapses, which pose significant risks to public safety and the environment. Indeed, the incidence of sinkhole collapses affecting transportation infrastructure has been on the rise in recent years, primarily due to the impact of climate change on foundations soils. The increasing frequency and intensity of extreme weather events, such as drought and heavy rainfall, are accelerating pavement deterioration and exacerbating soil instability [1]. These superficial distresses (cracking and rutting) create pathways for water ingress, leading to the washing away of soil materials. Moreover, the alternating cycles of wetting and drying can have detrimental effects on certain types of soils. In clay soils, these cycles induce swelling and shrinkage, causing soil movement and instability. Similarly, freezing and thawing cycles can weaken sandy silts, making them more prone to erosion and structural failure.

Sinkhole formation could be detectable from the surface due to ground depressions, but in many cases the pavement may appear intact despite underlying structural weaknesses. The structural integrity of pavement is crucial for supporting traffic loads safely. Within this framework, geophysical subsurface monitoring plays a crucial role in assessing the real condition of the subground and detecting early signs of sinkhole formation that may not be visible through conventional visual inspection methods. Among these geophysical methods, ground-penetrating radar (GPR) stands out as a valuable tool due to its non-destructive and non-invasive nature. In the context of transportation infrastructure, GPR is widely used for evaluating both flexible and rigid pavements, as well as the underlying subgrade [2, 3]. GPR provides valuable insights into the subsurface conditions for predicting sinkhole formation [4], such as cavities, settlements, and moisture damage. However, its effectiveness relies heavily on the signal processing methods employed and the expertise of who interprets. Additional development is therefore much needed to implement feasible processing and interpretational techniques that minimize subjectivity and maximize the quality and accuracy of the results obtained [5]. Recently, deep learning (DL) techniques and their application in signal processing and object detection became state-of-the-art (SoA), due to their remarkable detection speed and accuracy compared to conventional image processing techniques. DL applications have the potential to significantly reduce analysis subjectivity by automatically detecting hidden patterns in complex data. While DL algorithms have been widely adopted in computer vision tasks such as image classification and segmentation, their application to GPR data analysis is relatively new and still underexplored. The main reason is the limited availability of labeled data for training purposes. Furthermore, when dealing with cavity detection, the complexity arises from the great variety of features (or signatures) representing a sinkhole or cavity. This diversity necessitates a larger number of annotated images for effective model training. To overcome this limitation, synthetic data (numerical modelling) and data augmentation methods are used for the creation of diverse and realistic datasets, providing valuable training samples and improving model generalization [6].

Following the above motivation and current SoA, this work presents (i) a review of the application of the GPR method to detect cavities in linear infrastructure, highlighting its main findings and limitations, and (ii) a discussion on the data preparation techniques used to automate cavity detection from GPR data. Future perspectives on the application are also commented.

2 GPR applied to road cavity detection

The GPR method has shown remarkable capabilities in detecting cavities in road infrastructure, along with other aspects related sinkhole formation. Table 1 describes some notable contributions found in the published literature:

Table 1: Examples of published works related to GPR cavity detection.

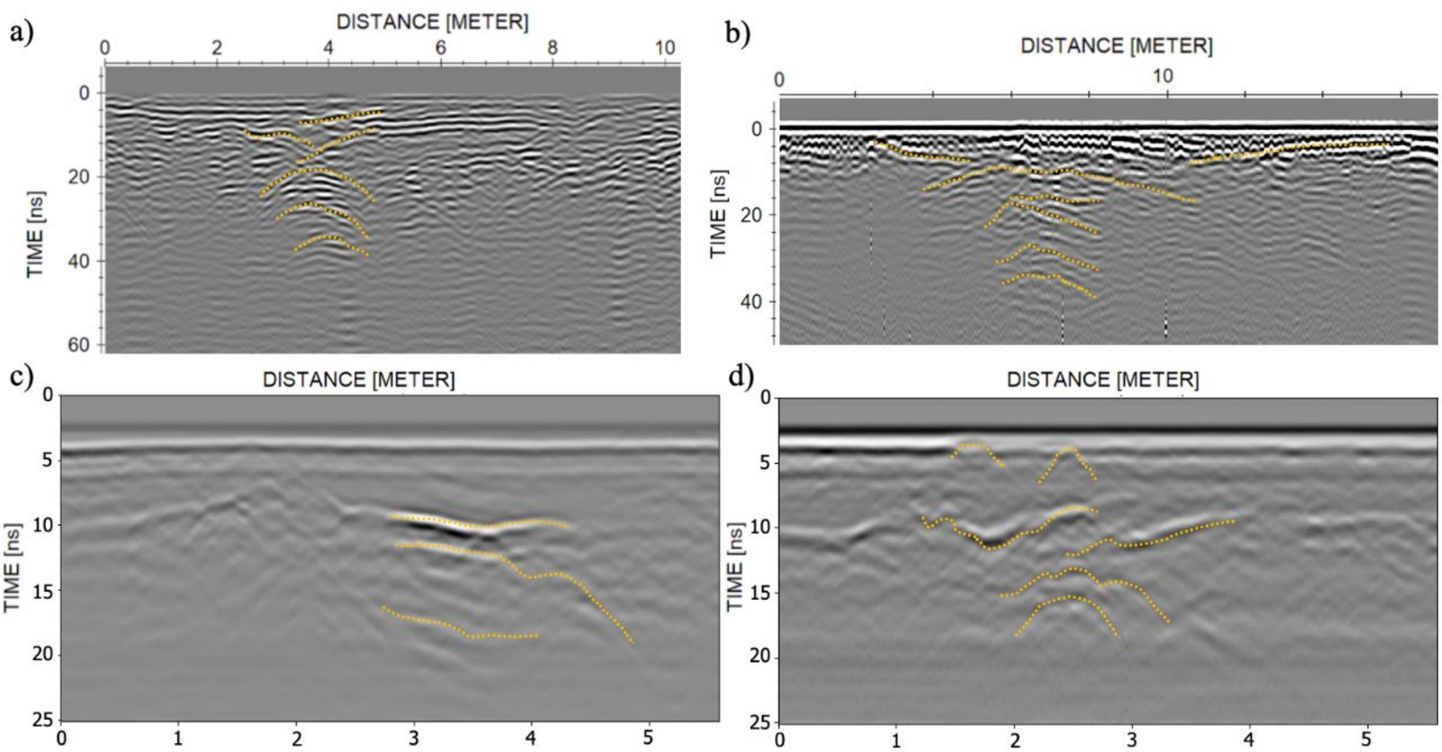
Objective	Antenna frequency	Findings	Limitations	Reference
To investigate a sinkhole in an urban area	180 MHz	GPR identified a concealed sinkhole probably caused by sagging	GPR was affected by the presence of highly conductive anthropogenic deposits and landscaped areas. Complementary NDT (INSAR, LiDAR, ERT) was used to overcome this limitation	[7]
To detect sinkholes in urban areas	400 MHz	GPR detected surface breaks, sinkholes, and down-dipping layers (sinking) due to forming sinkholes	The penetration depth of GPR was limited. Another limitation is to assess urban areas with presence of targets (e.g. buildings). Complementary NDT was used: InSAR and reflection seismic	[8]
To detect road deterioration and possible causes	500 & 800 MHz	A sinkhole was identified at 2 m deep, without superficial signs	Data interpretation improved when using higher antenna frequencies. Complementary NDT used: IRT, RGB, TLS	[9]
To detect cavities and galleries	400 & 200 MHz	GPR detected cavities and galleries at depths up to 3 m	The penetration depth of GPR was limited. Complementary NDT used: ERT	[10]
To identify forming sinkholes and karstic features	600 & 200 MHz	GPR detected subsidence, fractured rock and cavities	Complementary NDT used: refraction Seismic (and geological aspects) to assess potential sinkhole geohazards	[11]
To detect a sinkhole into karst-related systems	100 MHz	GPR detected reflectivity anomalies associated with karst features (e.g., weathering, fracturation) with reliable results up to 10-m depth	Not accurate enough to reveal the karstic depression due to the thickness of soils infilling, depth of bedrock, and presence of potential geological factors that could attenuated the signal and limited the resolution. Complementary NDT used: refraction seismic and ERT	[12]
To monitor a high-risk sinkhole in an urban area	100 & 200 MHz	GPR provided information on the internal structure of the sinkhole and the subsidence mechanisms	Signal noise produced by artificial elements (e.g., walls of the buildings' basement, pipes, cables). The GPR profiles acquired perpendicularly to the deformation structures offer more information on the subsidence structures than those with significant obliquity. The dip of the reflections is affected by the topographic correction. Complementary data: high-precision levelling	[13]
To characterize karst aquifer	100 MHz	GPR detected potential targets with elongated shape typical for karstic channels	Strong influence of anthropic structures (e.g., pipes) that could lead to misinterpretation (not karstic voids). Complementary NDT used to give	[14]

systems under urban areas			confidence in interpretation: ERT, reflection seismic and InSAR	

As observed in Table 1, integrating GPR with complementary geophysical methods is a common approach to overcome overcome technical limitations and misinterpretation. Combining multiple techniques, such as electrical resistivity tomography (ERT), seismic and InSAR provides more reliable information at different scales and resolutions.

Sinkholes and subsidence mechanisms are complex structures that generate complex and random reflections patterns, making difficult the GPR data interpretation. Figure 1 presents some reflections patterns produced by cavities or sinkholes mechanisms, showing the variety of GPR signatures associated to these kinds of features. In this context, the use of more sophisticated tools, such as artificial intelligence (AI), is highly valued to assist in the interpretation process of GPR data. AI methods, including DL models, can analyse large volumes of data and automatically extract characteristic patterns of sinkhole formation and subsurface features indicative of subsidence.

Fig. 1: Examples of GPR signatures for cavities and sinkholes mechanisms: 500 MHz data (a and b) and 900 MHz data (c and d).



3 Data augmentation methods for automated cavity detection

The pattern of the cavities can be complex, with many signal variations depending on their size, shape, the filling material and the surrounding soil. In addition, the pattern can be accompanied by other clutter and artifacts with stronger reflections, making it a non-dominant feature in the profile and, hence, difficult to interpret. Traditional GPR analysis often involves manual interpretation, which can be subjective and time-consuming, making it impractical for preparing large datasets with ground truth.

In the past few years, deep learning models, particularly Convolutional Neural Networks (CNNs), have received more attention in the research field [15]. Those models can automate the detection process thanks to their capabilities for automatic feature extraction and handling large GPR datasets. Therefore, they offer a faster and more consistent solution. However, the training process requires a large amount of labelled data, which is a challenging task. Recently, more research has focused on using supplementary synthetic data from numerical simulations to increase the size of the dataset. Still, generating realistic cavity samples is computationally intensive, and one model generates one sample. Another approach is to use generative adversarial networks (GANs) for data augmentation, which can increase the amount of data used for training [16]. Table 2 lists examples of related deep learning-based methods that used different approaches to synthetic data generation and data augmentation. While the collected real data can be limited in size, data simulation and augmentation can increase the overall size of the training dataset to be able to train a robust model. It's important to note that deep learning models often require thousands of images to train properly, making these data augmentation techniques particularly valuable.

Table 2: Examples of published works related to GPR cavity detection using deep learning with synthetic data incorporation.

Objective	Antenna Frequency	Collected field data	Synthetic data approach	Training data from augmentation and simulation	Reference
To detect cavities and cracks from B-scan data	200 MHz 400 MHz 900 MHz	Cavity samples: 408 Crack samples: 397 Total Images: 763	Gain compensation, station spacing, and radar signal mapping	Up to 4376 images	[17]
To detect voids from B-scan data	N/A	20 images	SinGAN gprMax	400 using SinGAN 100 images using gprMax	[18]
Recognition of pavement distress from A-scan data	500 MHz 900 MHz 1.6 GHz	500 traces	gprMax	281 traces using 900 MHz antenna and 600 traces using 1.6 GHz antenna	[19]
To detect subsurface voids using 3D data	Multi-channel radar system	Void samples: 88	gprMax	35 void samples	[20]

3.1 GPR data simulations

GPR simulations are the most common method to generate GPR data as they can generate realistic GPR scenarios based on a specific antenna frequency and with high details. Moreover, pavement layer compositions and soil properties can be controlled. Also, the shape, size, depth, and material of the cavity and surrounding loose soil can be adapted based on the purpose of the simulation. Simulations are also giving the ability to generate 3D data by simulating parallel profiles [21]. For deep learning purposes, the simulation output can be interpreted and annotated easily as the ground truth is known. Moreover, it can be used for training or testing purposes as they are not driven from the existing data.

Figure 2 shows an example for a simulation using the gprMax package [22]. Several cavities with irregular shapes have been positioned at different depths, and the synthetic data was generated by considering an antenna frequency of 1 GHz.

The simulation process can be computationally demanding and, usually, one model results in only one image during dataset generation. To add more variety to the generated data, a simple image fusion can be used to add a background from from clutter-free data to the simulation to augment more samples. Figure 3 shows an example of the image fusion between between simulated 400 MHz cavity with a background of real data. This approach can be further expanded to copy and paste cavities over different real data backgrounds.

Fig. 2: Examples of GPR simulation for shallow cavities using 1 GHz antenna: The model (left) and the corresponding simulation (right).

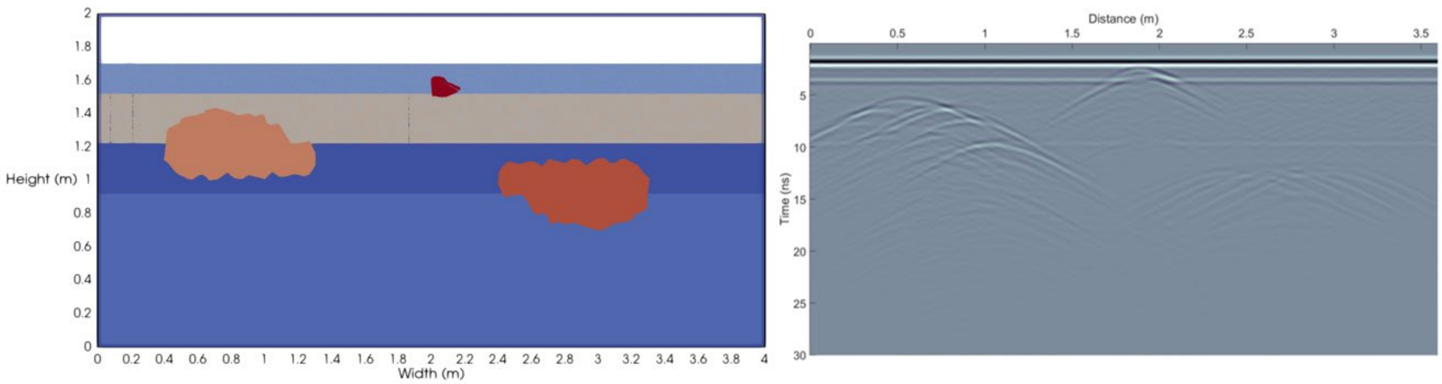
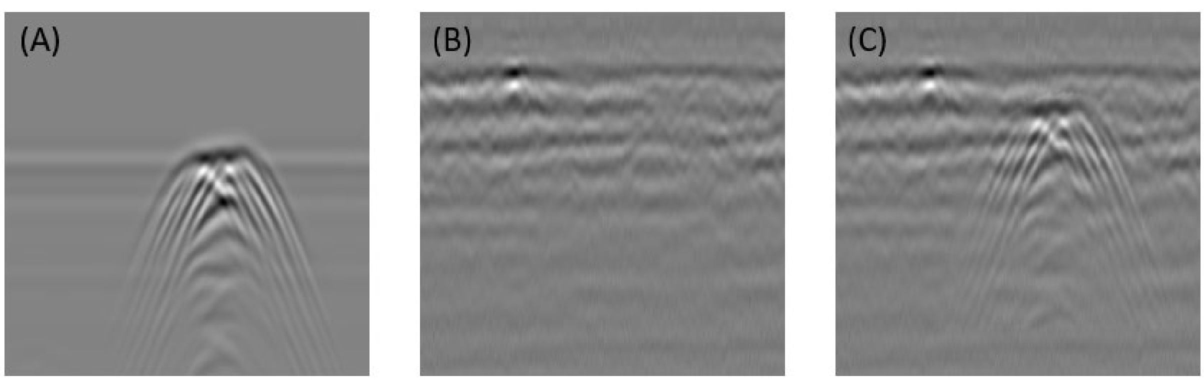


Fig. 3: Example for a basic image fusion to add clutter for a simulated data: A) simulated cavity using a 400 MHz antenna; B) real GPR data sample, and C) overlay between both images.



3.2 Traditional data augmentation

A common approach for data augmentation is to apply transformations to the existing real-world dataset, which can help the model generalize better to unseen situations. Operations such as translation and rotation are applied to the B-scan images to shift the objects' position and have more diversity in the cavity examples. Other operations can be image flipping, noise injection, and brightness and contrast adjustments, which can be similar to applying different gain functions. The augmented data can be further checked with image similarity methods such as Simple Linear Iterative Clustering Phash (SLIC-Phash) [23] to eliminate the similar or repeated augmented images which will improve the quality of the final training dataset. The selection of the data augmentation method depends on the problem under investigation and the target object of interest.

3.3 Generative adversarial networks

Generative adversarial networks (GANs) are a recent deep learning approach used for data augmentation. The basic concept involves training two neural networks in competition: a generator that attempts to create realistic data and a discriminator that tries to distinguish between real and generated data. While approaches like CycleGAN can be effective for generating GPR data [24], they often require large, diverse datasets to achieve optimal results. Training GANs can be problematic with cavities data as they are limited and difficult to obtain. Single-image techniques like SinGAN may be more suitable for cavity data augmentation as they require only a single image to train the model. Figure 4 shows an example of cavity data generated using SinGAN [25]. While it generates new data, the output often strongly resembles the original image. In addition, the input image should be chosen carefully to ensure that the cavity pattern is clear and is the dominant feature in the image. Another approach, ExSinGAN [26], allows control over semantics, structure, and texture within the input data, enabling customization of the generated output. Figure 5 demonstrates data generated by this method, illustrating the impact of changing different parameters. While GAN training can be resource-intensive, it can generate theoretically unlimited data, unlike the traditional simulation methods, which generate only one output. However, the output of the GAN models needs to be interpreted as there is no corresponding ground truth.

Fig. 4: Examples of GPR data augmentation using SinGAN: Input image (left) and the corresponding generated data (right).

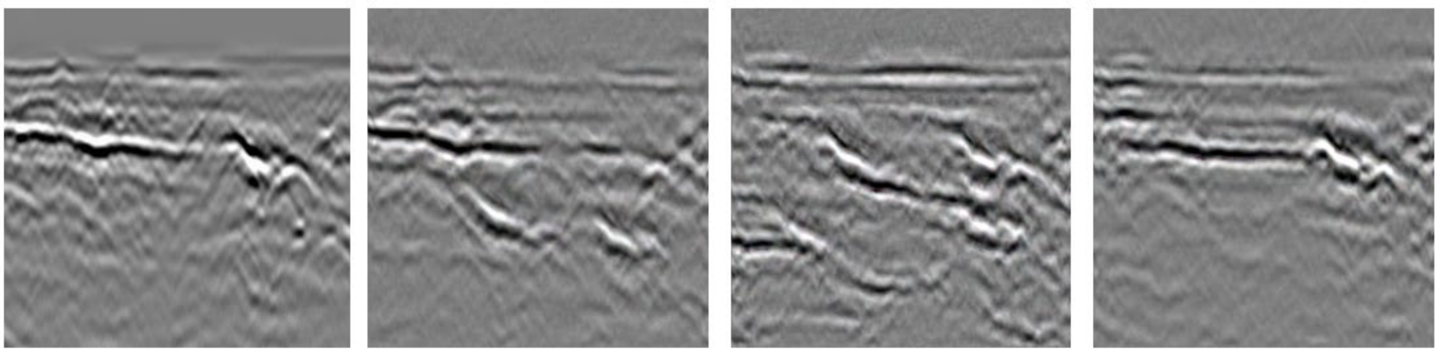
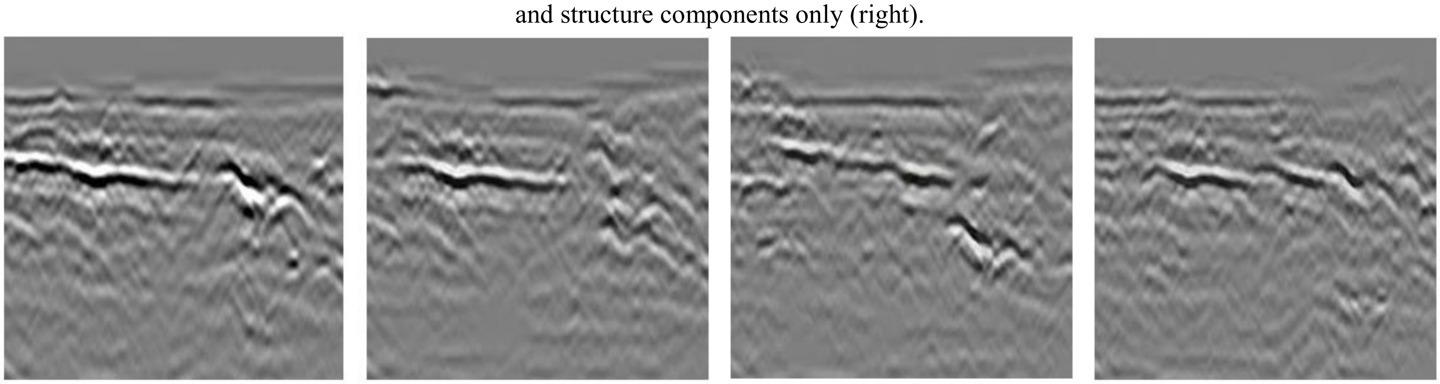


Fig. 5: Examples of GPR data augmentation using ExSinGAN: Input image (left) and the corresponding generated data using texture and structure components only (right).



4 Conclusion and further challenges

The GPR method has proven to be highly effective in detecting cavities in road infrastructure and assessing various aspects related to sinkhole formation. To overcome its inherent limitations, GPR is often complemented with other geophysical methods such as ERT and InSAR to obtain a more comprehensive understanding of subsurface conditions and improve the accuracy in detection. However, conducting geophysical prospection often requires careful planning, data acquisition, processing, and interpretation, which can take considerable time and resources.

Deep learning methods can automate cavity detection on a network level during regular inspection surveys, but the need for large training datasets remains a challenge. This paper briefly reviewed data augmentation and synthetic generation methods, highlighting their potential. A balanced approach that combines GPR data simulation and augmentation can provide diverse and realistic datasets needed for resource-efficient tasks. Our preliminary results will contribute to building a cavity database for training a robust object detection model, such as YOLOv9, which is currently considered SoA for object detection. The preliminary experiments suggested that having a training dataset between 3000 to 4000 images using the discussed methods is a good starting point to train SoA cavity detection model.

To sum up, intelligent monitoring plays a crucial role in ensuring the safety and longevity of transportation infrastructure. The early detection of any changes or deterioration in the ground leads to prevent potential hazards. By prioritizing the integrity of the ground, transportation authorities can enhance the resilience and longevity of infrastructure, ultimately ensuring the safety and well-being of the traveling public. Moreover, implementing proactive maintenance strategies allows them to optimize investments and resource allocation effectively, thus prolonging the lifespan of transportation assets.

Acknowledgements

Authors acknowledge the funding received from the STRADAR project (TED2021-130183B-I00), funded by MCIN/AEI/10.13039/501100011033 and by the "European Union NextGenerationEU/PRTR". M. Solla acknowledges the grant RYC2019–026604–I, funded by MCIN/AEI/10.13039/501100011033 and by "ESF Investing in your future". This work presents the preliminary results of the research stay of A. Elseicy in the Belgian Road Research Centre (BRRC), who acknowledges all the support and mentoring provided by Dr. Carl Van Geem, Dr. Colette Gregoire, Dr. Audrey Van der Wielen, Yves Pollet, and Ali Yeganeh, as well as some of the data that makes this research possible.

References

[1] H. Jeong, H. Kim, K. Kim and H. Kim, "Prediction of Flexible Pavement Deterioration in Relation to Climate Change Using Fuzzy Logic," *Journal of Infrastructure Systems*, vol. 23, no. 4, 04017008, 2017.

[2] M. Solla, V. Pérez-Gracia and S. Fontul, "A review of GPR application on transport infrastructure: troubleshooting and best practices," *Remote Sensing*, vol. 13, no. 4, 672, 2021.

[3] A. Elseicy, A. Alonso-Díaz, M. Solla, M. Rasol and S. Santos-Assunçao, "Combined use of GPR and other NDTs for road pavement assessment: an overview," *Remote Sensing*, vol. 14, 4336, 2022.

[4] A. Ronen, M. Ezersky, A. Beck, B. Gatenio and R.B. Simhayov, "Use of GPR method for prediction of sinkholes formation along the Dead Sea Shores, Israel," *Geomorphology*, vol. 328, pp. 28-43, 2019.

[5] M. Solla and N. Fernández, "GPR analysis to detect subsidence: A case study on a loaded reinforced concrete pavement," *Int. J. Pavement Engineering*, 2022.

[6] D. Reichman, L.M. Collins and J.M. Malof, "Some good practices for applying convolutional neural networks to buried threat detection in GPR," *in Proceedings of the 9th Int. Workshop on Adv. Ground Penetrating Radar*, 2017.

[7] D. Carbonel, V. Rodríguez-Tribaldos, F. Gutiérrez, J.P. Galve, J. Guerrero, M. Zarroca, C. Roqué, R. Linares, J.P. McCalpin and E. Acosta, "Investigating a damaging buried sinkhole cluster in an urban area (Zaragoza city, NE Spain) integrating multiple techniques: Geomorphological surveys, DInSAR, DEMs, GPR, ERT, and trenching," *Geomorphology*, vol. 229, pp. 3–16, 2015.

[8] A. Busetti, C. Calligaris, E. Forte, G. Areggi, A. Mocnik and L. Zini, "Non-Invasive Methodological Approach to Detect and Characterize High-Risk Sinkholes in Urban Cover Evaporite Karst: Integrated Reflection Seismics, PS-InSAR, Leveling, 3D-GPR and Ancillary Data. A NE Italian Case Study," *Remote Sens.*, vol. 12, 3814, 2020.

[9] S. Lagüela, M. Solla, I. Puente and F.J. Prego, "Joint use of GPR, IRT and TLS techniques for the integral damage detection in paving," *Constr. Build. Mater.*, vol. 174, pp. 749–760, 2018.

[10] D. Gómez-Ortiz and T. Martín-Crespo, "Assessing the risk of subsidence of a sinkhole collapse using ground penetrating radar and electrical resistivity tomography," *Eng. Geol.,* vol. 149-150, pp. 1–12, 2012.

[11] L. De Giorgi and G. Leucci, "Detection of Hazardous Cavities Below a Road Using Combined Geophysical Methods," *Surv. Geophys.*, vol. 35, pp. 1003–1021, 2014.

[12] M. Cueto, J. Olona, G. Fernández-Viejo, L. Pando and C. López-Fernández, "Karst-induced sinkhole detection using an integrated geophysical survey: A case study along the Riyadh Metro Line 3 (Saudi Arabia)," *Near Surf. Geophys.*, vol. 16, pp. 270–281, 2018.

[13] J. Sevil, F. Gutiérrez, C. Carnicer, D. Carbonel, G. Desir, Á. García-Arnay and J. Guerrero, “Characterizing and monitoring a high-risk sinkhole in an urban area underlain by salt through non-invasive methods: Detailed mapping, high-precision leveling and GPR,” *Eng. Geol.*, vol. 272, 105641, 2020.
[14] R. Martel, P. Castellazzi, E. Gloaguen, L. Trépanier and J. Garfias, “ERT, GPR, InSAR and tracer tests to characterize karst aquifer systems under urban areas: the case of Quebec City,” *Geomorphology*, vol. 310, pp. 45-56, 2018.
[15] M.-S. Kang, N. Kim, S. B. Im, J.-J. Lee, and Y.-K. An, “3D GPR Image-based UcNet for Enhancing Underground Cavity Detectability,” *Remote Sens.*, vol. 11, no. 21, p. 2545, Oct. 2019, doi: 10.3390/RS11212545.
[16] X. Zhang, L. Han, M. Robinson, and A. Gallagher, “A Gans-Based Deep Learning Framework for Automatic Subsurface Object Recognition from Ground Penetrating Radar Data,” IEEE Access, vol. 9, pp. 39009–39018, 2021, doi: 10.1109/ACCESS.2021.3064205.
[17] G. Yue, C. Liu, Y. Li, Y. Du, and S. Guo, “GPR Data Augmentation Methods by Incorporating Domain Knowledge,” *Appl. Sci.*, vol. 12, no. 21, p. 10896, Oct. 2022, doi: 10.3390/app122110896.
[18] G. Chen, X. Bai, G. Wang, L. Wang, X. Luo, M. Ji, P. Feng, Y. Zhang, “Subsurface Voids Detection from Limited Ground Penetrating Radar Data Using Generative Adversarial Network and YOLOV5,” in International Geoscience and Remote Sensing Symposium (IGARSS), Oct. 2021, pp. 8600–8603. doi: 10.1109/igarss47720.2021.9554954.
[19] J. Xu, J. Zhang, and W. Sun, “Recognition of the typical distress in concrete pavement based on gpr and 1d-cnn,” *Remote Sens.*, vol. 13, no. 12, p. 2375, Jun. 2021, doi: 10.3390/rs13122375.
[20] T. Yamaguchi, T. Mizutani, M. Kimiro, and T. Hirano, “Detecting Subsurface Voids from GPR Images by 3-D Convolutional Neural Network using 2-D Finite Difference Time Domain Method,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, pp. 1–1, 2022, doi: 10.1109/jstars.2022.3165660.
[21] H. Huang, X. Hao, L. Pei, J. Ding, Y. Hu, and W. Li, “Automated detection of through-cracks in pavement using three-instantaneous attributes fusion and Swin Transformer network,” *Autom. Constr.*, vol. 158, no. August 2023, p. 105179, 2024, doi: 10.1016/j.autcon.2023.105179.
[22] C. Warren, A. Giannopoulos, and I. Giannakis, “gprMax: Open source software to simulate electromagnetic wave propagation for Ground Penetrating Radar,” *Comput. Phys. Commun.*, vol. 209, pp. 163–170, Dec. 2016, doi: 10.1016/j.cpc.2016.08.020.
[23] N. Wang, Z. Zhang, H. Hu, B. Li, and J. Lei, “Underground Defects Detection Based on GPR by Fusing Simple Linear Iterative Clustering Phash (SLIC-Phash) and Convolutional Block Attention Module (CBAM)-YOLOv8,” *IEEE Access*, vol. 12, no. January, pp. 1–1, 2024, doi: 10.1109/access.2024.3365959.
[24] H. Wang, S. Ouyang, Q. Liu, K. Liao, and L. Zhou, “Buried target detection method for ground penetrating radar based on deep learning,” *J. Appl. Remote Sens.*, vol. 16, no. 01, p. 018503, Jan. 2022, doi: 10.1117/1.jrs.16.018503.
[25] T. R. Shaham, T. Dekel, and T. Michaeli, “SinGAN: Learning a Generative Model from a Single Natural Image,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2019-October, pp. 4569–4579, May 2019, doi: 10.1109/ICCV.2019.00467.
[26] Z. Zhang, C. Han, and T. Guo, “ExSinGAN: Learning an Explainable Generative Model from a Single Image,” 32nd Br. Mach. Vis. Conf. BMVC 2021, May 2021, Accessed: Mar. 31, 2024. [Online]. Available: https://arxiv.org/abs/2105.07350v2