A Semantic Digital Twinning Approach for the Management of Road Distress Data

Luca Bertolini[1], Inga Maria Giorgadze[2], Faridaddin Vahdatikhaki[2] and Fabrizio D'Amico[1]

[1]Department of Civil, Computer Science and Aeronautical Technologies Engineering, Roma Tre University, Italy (luca.bertolini, fabrizio.damico)@uniroma3.it
[2]Department of Construction Management & Engineering, University of Twente, The Netherlands (i.m.giorgadze, f.vahdatikhaki)@utwente.nl

Keywords: Pavement failure, Digital Twin.

Abstract

This paper addresses the challenge of managing and maintaining road pavements effectively throughout their lifecycle. Emphasizing the need for a cohesive data collection and handling approach, the paper highlights the vulnerability caused by the absence of a structured information system, resulting in reworks, information loss, and misinterpretation of collected data. A unified data handling and collection structure, introducing the concept of a pavement Digital Twin through a standardized data structure is presented. The Digital Twin aims to integrate information regarding pavement failures, with the future aim of predicting deterioration and facilitating informed decision-making in management and maintenance interventions. Complexities in pavement failures, categorized into surface and subsurface modes, prompt the necessity for a reliable classification and representation system. The proposed methodology introduces a grid of cells for surface failures and three-dimensional voxels for subsurface failures, providing a structured approach for data integration throughout the pavement's lifecycle. An approach for the semantic representation of pavement distresses is also presented. The proposed methodology stresses the importance of scalability and flexibility in data storage, forming the basis for a comprehensive Digital Twin of road pavements. Finally, through the use of multiple pavement failure datasets, the methodology is shown to have a high potential in providing a structure for comprehensive management of road data.

1. Introduction

Management and maintenance of civil infrastructures, particularly transportation infrastructures, are among the highest priorities for stakeholders and public administrations. For all the parties involved throughout the lifecycle of infrastructure, collection and management of data about assets' condition is critical for the development of strategies to reduce critical failures.

This data is required to monitor road conditions and ensure that contractors and infrastructure management agencies can be notified and held accountable for maintenance and repairs, should deteriorations arise within a road project's guarantee period. To continuously improve design, building, and maintenance procedures, information aids clients and contractors in better evaluating the long-term effects of decisions made throughout the pavement life.

Currently, data collection and handling procedures over the course of a road lifecycle can vary a lot, especially between different asset owners and contractors. The lack of a structured information system and information fragmentation make pavement management vulnerable to significant rework, information loss, assessment errors, and misinterpretation of the collected data.

In this context, Building Information Modeling (BIM) aims at integrating lifecycle information into three-dimensional digital models of analyzed assets of civil infrastructures (Costin et al., 2018) (Bradley et al., 2016). However, standardized procedures to produce, integrate, represent and upkeep data in BIM are still lacking in the road sector. This represents a challenge for the industry, as efficient methods to exchange data between the different players and disciplines that take part in a road project throughout its lifecycle are not available. Several research studies have delved into the use of BIM for infrastructure data collection and management, for different phases of their lifecycle and survey techniques (Bertolini et al., 2023) (Bosurgi et al., 2022), but a cohesive approach to information structuring is still lacking.

Starting from the design phase of the road pavement, through construction and management, a cohesive approach to data handling must be developed. Indeed, contractors need to be kept updated regarding infrastructure conditions to schedule maintenance activities. A correct data managing process can ensure that information can be used to predict pavement failures and effectively assign resources to handle distresses that may arise during the road's operation phase. Moreover, a structured information handling methodology can be used to link those failures to design, construction or maintenance activities, optimizing the entire lifecycle assessment.

On-site surveys are usually carried out to monitor the operational condition of transportation infrastructure. Non-destructive testing techniques, such as Ground Penetrating Radar (GPR), LiDAR (Laser Imaging Detection and Ranging), Laser Profilometer, etc., have been ever more utilized to carry out these inspections (Gagliardi et al., 2023) (Xiong et al., 2023) (Elseicy et al., 2022). However, these surveys are typically carried out independently by multiple operators over a long time. The subsequent output analyses typically take place separately as well. To guarantee that the evaluations are conducted more effectively, one methodology that keeps and updates this data throughout time and permits an integrated study of all available datasets would be essential. Moreover, pavement condition information is currently collected and handled inconsistently and without connection to its entire lifecycle, making it hard to trace pavement failures to their root causes. By providing a framework to store information regarding pavement failures that can allow the correlation between these elements and the different phases of the lifecycle, a system capable of executing pavement deterioration predictions, unveiling degradation mechanisms and making informed decisions regarding management and maintenance interventions could be obtained.

Furthermore, defining a unique structure for data handling and collection is an important step towards the concept of a pavement Digital Twin. In the context of this research, Digital Twins can be defined as models that contain real-time dynamic

data, provided by the possible implementation of sensors and IoT (Internet of Things). A Digital Twin of road pavements can only be achieved if a standardized data structure is defined, allowing information exchange and interoperability between the different operators and processes that take part in the lifecycle management to happen efficiently and dynamically.
Based on the presented issues, the primary aim of this paper is to provide a high-level structure for the semantic representation of road distress data. Moreover, a process to collect and register failure data that can produce a structured database of pavement distresses, which can be accessible and updateable, is another objective of the research hereby presented.

2. Failure Modes

Pavement failure modes can be categorized into two main groups: surface and subsurface failure modes (Figure 1). Surface failure modes affect the surface layer of pavements and are exposed for most of their extension to the open air (i.e., cracks, potholes, etc.). They can be detected by visual inspection or more advanced methods such as LiDAR (Bosurgi et al., 2022). Subsurface failures mainly develop inside the pavement layers (i.e., concealed cracks, mixture segregation, etc.) and are not apparently exposed to the open air, but they can still cause severe damage to the road and pose a threat to its correct operation. The development of subsurface and surface failure modes is often connected, as damage to the pavement generally is not limited to the surface. Moreover, some types of subsurface distresses can develop and become surface distresses after a certain period (i.e., concealed cracks develop to become surface cracks as multiple loads are applied on the pavement). These types of pavement distresses can't be detected by visual on-site inspection, as more advanced instruments are necessary to identify them. GPR has been proven effective for the detection of pavement layer configuration and the identification of subsurface failure modes (Liang et al., 2022) (Li et al., 2021) (Gao et al., 2020) (Benedetto et al., 2017).

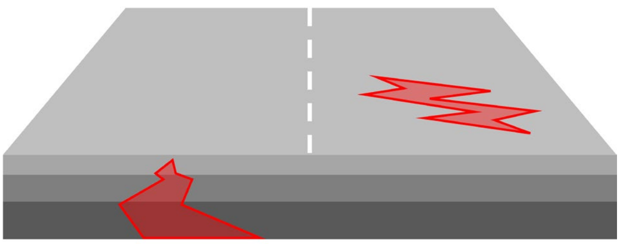

Figure 1. Schematic representation of surface and subsurface failures in a road pavement.

This first categorization allows us to define the starting point of the proposed methodology, as it is necessary to define a universal reference point in relation to which the failures' data is to be stored. While defining these elements, it's necessary to consider that for the subsurface distresses the three-dimensional aspects could be more prominent than for the surface failures. Indeed, information regarding the depth at which the distress is detected, its volume and configuration are all relevant data that must be collected for these kind of pavement failures. Moreover, regarding the surface failures, their position in relation to the pavement is only a matter of geographic coordinates, as they are expected to be positioned on its surface. Therefore, for the surface failures a two-dimensional element can act as the main reference point to store the related information, while for the subsurface distresses a three-dimensional element is necessary. However, these reference points must be connected, to allow for the comparison and correlation between these different failures, for a correct analysis of the pavement conditions.

3. Methodology

The proposed methodology is based on the decomposition of the pavement layers into a regular grid of cells. These elements can be the reference for the integration of data related to surface failures. To consider the three-dimensional aspects of subsurface failures, a three-dimensional reference must then be determined. Voxels were chosen as the elements to be used to collect information regarding these types of failures. A voxel, short for "volumetric pixel", is a three-dimensional pixel or the smallest unit of a three-dimensional grid representing a volumetric space. In contrast to traditional pixels in two-dimensional images, which have width and height, voxels also have depth, making them the reference elements for representing 3D information and objects. Each voxel can contain information about a specific point in the 3D space, consequently enabling it to store data related to subsurface failures.
Thus, a number of voxels corresponding to the different pavement layers is assigned to each cell of the previously mentioned grid (Figure 2). Each of these voxels can then contain information regarding the failures present in the depth of the pavement and relate them to their corresponding cell. Cells and voxels are the building blocks necessary to create a Digital Twin of the road pavement. By placing them alongside each other, the complete structure of the pavement can be reproduced in digital form. As they are objects capable of storing information, they can also reproduce the data related to each part of the pavement, allowing for the analysis necessary for the lifecycle assessment of the pavement itself.

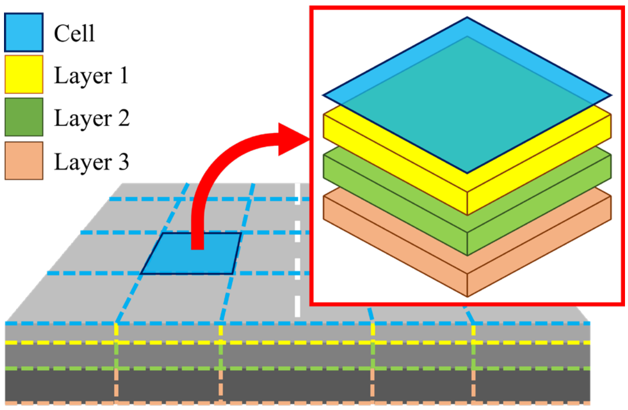

Figure 2. Cells and voxels.

3.1 Failure Mode Data Integration

The first step in the process is the identification of new failure data. As a pavement distress is detected, it is necessary to verify its connection to previous inspection outputs, in order to ensure that the dynamic aspects of failures over time are properly showcased. Since data coming from on-site surveys could contain errors related to the position of the detected failure, a correction process is then applied. This ensures that distresses detected over time are compared only if they represent the same failure modes. Conversely, if two failures identified in different surveys don't correspond to the same element, the comparison between the corresponding data shouldn't be carried out. The correction process begins by identifying the most similar failure from the previous data registration within a radius of 4 meters

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W5-2024-49-2024

around the newly detected distress. Then, the amount of error in terms of positioning offset is determined. This step is repeated for all the detected failures that fall within the previously mentioned area. As the last failure is identified and the corresponding error detected, the next step is to calculate the average of all data errors and consequently apply the necessary corrections. In particular, the applied corrections are weighted based on the previously available data, as old and already corrected information is considered more reliable and therefore more accurate. The process to handle and integrate failure data into the proposed reference grid is showcased in Figure 3.

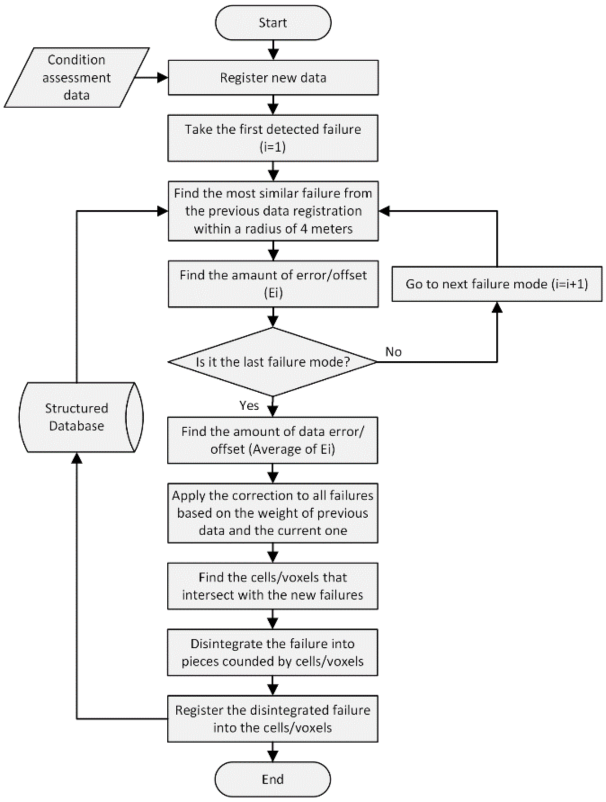

Figure 3. Data integration process.

Once a surface pavement failure is detected and the correction process is completed, the corresponding information (such as shape, dimensions, etc.) is traced and georeferenced onto the grid. Then, the failure is divided between the intersected cells so that each one of them contains the part of the distress that resides in its area. The failure information is disintegrated between the different cells accordingly so that they can be the reference for the later analysis of the pavement. Likewise, subsurface failures can be represented in 3D and placed in voxels. The distresses are then disintegrated into the voxels they intersect with so that these elements can store the related data and transmit it to the cell to which they belong. A schematic representation of this process is shown in Figure 4, showing how a surface and a subsurface failure are divided in the corresponding 2D and 3D grids.

Once this process is complete, there will be a collection of cells and voxels containing failure data. Each one of these elements can then be selected and the information stored in it analyzed (Figure 5). By applying this process, a structured database of pavement failures is obtained. This can be updated over time by repeating the previously described process as new condition assessment data is provided by surveys carried out on-site. Such a database can provide useful and up-to-date information to asset owners and contractors to carry out data driven decisions regarding maintenance interventions and management activities. The data present in each cell and voxel can then be aggregated for larger sections of the pavement, allowing to scale up the methodology in size and form. Therefore, starting from a lane segment, entire stretches of road could be analyzed, as they would be containing all the information stored in the corresponding cells.

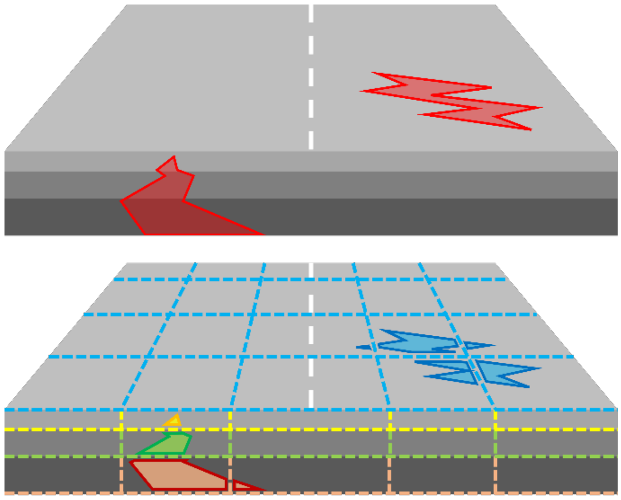

Figure 4. Failure modes disaggregated into the grids formed by the cells and voxels.

Moreover, other than describing the conditions of different road sections, the proposed methodology also allows to inquire the database regarding specific layers of the pavement. Indeed, by aggregating data contained in multiple voxels that are part of the same layer, an analysis of the different failures contained in that pavement layer can be carried out. A diagram in Figure 6, in the following page, shows the layering of information as described. While a cell can contain multiple voxels, a voxel can be part of only one cell, and the same rule applies for cells and the road pavement. Therefore, a cell contains all the information related to each of its voxels, while a voxel only contains information directly integrated into it, while still being able to make connections to data contained in other elements that are part of the same pavement. By continuously updating and integrating information into these objects, it becomes possible to trace the evolution of each failure over time, which can allow for complex analysis of the pavement's conditions.

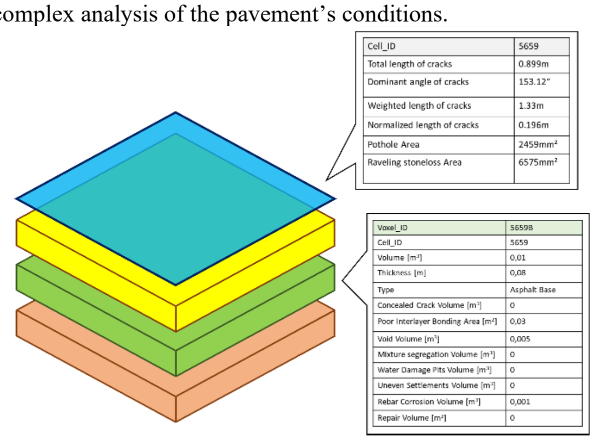

Figure 5. Cells and voxels information can be accessed to analyze the failures that are affecting that portion of the pavement.

The cells and voxels not only work as a reference point for the integration of failure modes data but also must contain all other information related to the lifecycle of the pavement. For instance, information regarding construction processes and pavement design can be integrated into the different cells. Each one of these objects will then contain data such as the compaction efficiency of each layer, the type of mixture used, the temperature of the asphalt mixture during construction. As multiple types of pavement data are stored in cells and voxels, correlations between failures and other aspects of the asset's lifecycle can then be executed. Moreover, latent information can be extracted, performing predictions regarding future degradations or correlations between the appearance of failures and the design and construction processes. This way, a standardized method for storing and representing pavement data can be achieved, allowing better data exchange and interoperability in the industry.

4. Case Study

Multiple pavement failures datasets were used to define a case study to test the proposed methodology. The first dataset was obtained using a Laser Crack Measuring System, capable of distinguishing different kinds of surface failures. This instrument scans the surface of the pavement and detects inconsistencies that can be related to specific types of pavement distresses. Figure 7 shows a portion of the dataset, with different color-coded types of distresses (e.g. cracks, potholes, patches, etc.). The failures are then categorized in relation to the kind of distress they represent and are attributed information regarding their geometric configuration and their position in relation to the coordinate system used for the survey in question.

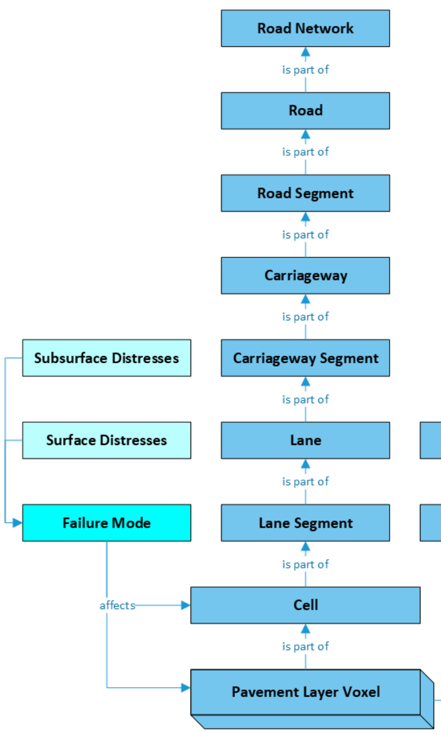

Figure 6. Data structure for pavement failures.

The correction process then takes place, confronting the failures included in the current dataset with other distresses obtained from previous surveys. As this phase is concluded, it is possible to proceed with the superimposition of the grid over the multiple failures, disintegrating them and integrating their information into the different cells comprising said grid.

Moreover, GPR surveys were carried out over different stretches of road, providing multiple datasets that can be used to identify subsurface pavement failures. In the presented case study, the Hi-Mod 200-600 ground-coupled pulsed GPR system manufactured by IDS Georadar S.p.A. was employed. The system is equipped with a multifrequency antenna working at the central frequencies of 200 MHz and 600 MHz. The acquisition parameters are shown in Table 1. As the instrument is carried over the road, it emits electromagnetic impulses towards the ground.

Parameter	Value
Frequency	600 MHz
Time window	60 ns
Horizontal resolution	0.05 m
Samples	512

Table 1. Margin settings for A4 size paper

As these impulses come in contact with the elements comprising the pavement, they are reflected back to the GPR. By determining the time passed between the signal emission and the received reflection, and the reflected signal intensity, it is possible to determine the configuration of the pavement layers and highlight possible pavement distresses (Liang et al., 2022) (Gao et al., 2020). To optimize the signal-to-noise ratio and increase the dependability of the interpretation phase, the dataset obtained from the survey underwent standard processing procedures such as zero cutting, band pass filtering, migration, and background noise removal. Significant geometric details about the configuration of the pavement layers were found through the interpretation of the processed data.

The process of converting time to depth involved measuring the propagation velocity (v) of electromagnetic waves using the formula:

$$v = c/\sqrt{\varepsilon_r} \quad (1)$$

where: c = speed of light in a vacuum
ε_r = dielectric constant
v = propagation velocity

The dielectric constant (ε_r) was then computed using the hyperbola fitting method (Jol, 2008), yielding values of 6 and 8 for the layers of hot-mix asphalt (HMA) and concrete, respectively.

In this context, tomographies of the pavement allow to visualize spots of inhomogeneities in reflective behavior at progressive depths, which can be related to subsurface failures (Figure 8). Using multiple techniques, it's then possible to determine their geometrical characteristics and the type of distress that they represent. As for the surface distresses, the dataset is then corrected, and the failures disaggregated between the different voxels that they intersect.

The databases utilized in this case study are comprised by real data obtained from on-site surveys. The surface failures dataset and the subsurface failure dataset are the result of inspections carried out on different infrastructures. Therefore, their combined use in the case study hereby presented is only meant to demonstrate the feasibility of the methodology in handling the available information and allowing simultaneous analysis of both datasets.

The resulting analysis allows to inquire the information contained in the cells and voxels, while also confronting the two available datasets. In fact, it is possible to identify the cells that are affected by a certain type of surface failure while also containing pavement voxels that are affected by subsurface failures. For instance, the cells that are affected by surface cracks while also residing over a portion of pavement that has a certain percentage of volume affected by concealed cracks can be identified. By performing these analyses, it would be possible to study the correlation between different types of pavement failures, to allow for more efficient pavement management strategies. Moreover, by scaling up the analysis, entire stretches of road can be examined in the same manner. Furthermore, as more information is implemented through this methodology, such as design and construction data, it becomes possible to carry out analyses linking pavement failures with different phases of the pavement lifecycle.

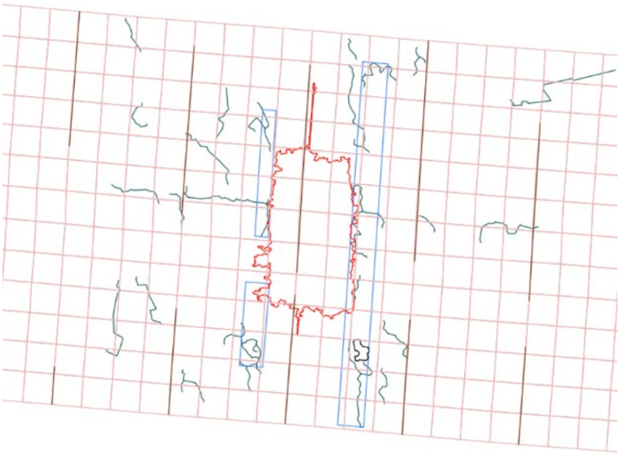

Figure 7. Surface pavement failures.

This first application represents an initial step towards a finalized implementation of a semantic modeling procedure for pavement failures. The proposed methodology is then shown to have high potential for providing a system to store and handle pavement failure data. Moreover, the method could provide a structure for the comprehensive management of road condition data.

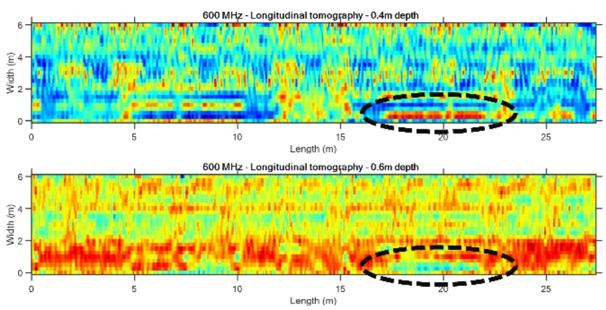

Figure 8. GPR tomography at different depths in the pavement, and possible pavement distresses.

5. Conclusions and Future Developments

In summary, the importance of effective management and maintenance of road pavements throughout their lifecycle, is emphasized. The need for a cohesive approach to data collection and handling is also highlighted, addressing the current vulnerability in the form of a lack of a structured information system. This absence results in rework, information loss, assessment errors, and misinterpretation of collected data.

The development of a unified data handling and collection structure is a necessary step to an optimized management system for pavements. In this context, Digital Twins, which can be facilitated by a standardized methodology, aim to integrate information on pavement failures, predict deterioration, identify degradation mechanisms, and support informed decision-making regarding management and maintenance interventions.

The complexities of pavement distresses, categorized into surface and subsurface failures, underscore the necessity for a reliable classification and representation system. The proposed methodology introduces a regular grid of cells for two-dimensional surface failures data collection, and voxels for three-dimensional subsurface failures. This structured approach allows for the integration of data related to pavement failures, providing a reference point for analysis throughout the pavement's lifecycle.

The methodology emphasizes the importance of scalability and flexibility in data storage. Cells and voxels can be scaled up to represent entire stretches of roads, allowing for a more generalized analysis of the pavement's conditions. By using these elements as reference points for data integration, a Digital Twin of the road pavement can be achieved. To test the proposed method, multiple pavement failure datasets are used, and the described processes applied to them. By combining surface and subsurface failure data, a comprehensive representation of the pavement can be achieved. Indeed, the methodology is shown to allow the combined analysis of multiple failures datasets, with the possibility to expand the current investigation to lifecycle data as well. In conclusion, the critical need for a standardized and integrated approach to data handling in road pavement management is highlighted. The proposed methodology provides a systematic framework for representing and analyzing pavement failures, facilitating data exchange, interoperability, and informed decision-making throughout the pavement's lifecycle.

Moreover, some further reflections need to be addressed regarding the proposed methodology presented in this paper. As the way to structure and integrate data is shown, what information should be integrated into the cells and voxels is still to be defined. Furthermore, semantic modeling of pavement information must be a priority to effectively create a system capable of fully representing the information associated with its lifecycle. Developing ontologies related to the various aspects of a pavement management systems, starting from the design phase and through the construction and maintenance phases, while also representing all data associated with its materials, failure modes and other elements is a priority to allow the development a functioning pavement Digital Twin. In this context, existing standards for digital models of pavements and roads should be studied to define how they would need to be developed to integrate the proposed data structure. By utilizing the same data handling and structuring procedures, interoperability between different standards can be optimized. This would allow for improved management and maintenance of road assets, making it possible for the different players involved not to be bound to a particular standard or software in their activities.

Moreover, rules for upscaling and downscaling information between different elements, such as multiple cells information scaled up to represent a portion of pavement, must be defined and tested to ensure that data loss can be avoided, and the resulting analysis can accurately represent the pavement's conditions.

Data obtained from on-site surveys can pose a challenge to the correct application of the proposed methodology. Indeed, the quality and scale of the information provided by inspections can vary greatly between different operators and instruments. A

standardization of the requirements for good quality data should be defined to ensure that the process of data integration in the Digital Twin environment can correctly proceed.
Lastly, to assure that a correct data handling process is achieved throughout the entire lifecycle of a road pavement, it is important that all the players involved in its management procedures define how the available information be exchanged between each other. Data ownership, its access and the related limitations are arguments that need to be addressed if the industry is to proceed in the use of Digital Twins to manage projects in the construction industry.

References

Benedetto, A., Tosti, F., Bianchini Ciampoli, L., & D'Amico, F. (2017). An overview of ground-penetrating radar signal processing techniques for road inspections. *Signal Processing*, *132*, 201–209. doi:10.1016/j.sigpro.2016.05.016

Bertolini, L., D'Amico, F., Napolitano, A., Bianchini Ciampoli, L., Gagliardi, V., & Romer Diezmos Manalo, J. (2023). A BIM-based approach for pavement monitoring integrating data from non-Destructive Testing methods (NDTs). *Infrastructures*, *8*(5), 81. doi:10.3390/infrastructures8050081

Bosurgi, G., Pellegrino, O., & Sollazzo, G. (2022). Pavement condition information modelling in an I-BIM environment. *International Journal of Pavement Engineering*, *23*(13), 4803–4818. doi:10.1080/10298436.2021.1978442

Bosurgi, Gaetano, Sollazzo, G., Bongiorno, N., & Pellegrino, O. (2020). Automatic crack detection results using a novel device for survey and analysis of road pavement condition. In *Lecture Notes in Civil Engineering* (pp. 431–440). Cham: Springer International Publishing.

Bradley, A., Li, H., Lark, R., & Dunn, S. (2016). BIM for infrastructure: An overall review and constructor perspective. *Automation in Construction*, *71*, 139–152. doi:10.1016/j.autcon.2016.08.019

Costin, A., Adibfar, A., Hu, H., & Chen, S. S. (2018). Building Information Modeling (BIM) for transportation infrastructure – Literature review, applications, challenges, and recommendations. *Automation in Construction*, *94*, 257–281. doi:10.1016/j.autcon.2018.07.001

Elseicy, A., Alonso-Díaz, A., Solla, M., Rasol, M., & Santos-Assunçao, S. (2022). Combined use of GPR and other NDTs for road pavement assessment: An overview. *Remote Sensing*, *14*(17), 4336. doi:10.3390/rs14174336

Gagliardi, V., Tosti, F., Bianchini Ciampoli, L., Battagliere, M. L., D'Amato, L., Alani, A. M., & Benedetto, A. (2023). Satellite remote sensing and non-destructive testing methods for transport infrastructure monitoring: Advances, challenges and perspectives. *Remote Sensing*, *15*(2), 418. doi:10.3390/rs15020418

Gao, J., Yuan, D., Tong, Z., Yang, J., & Yu, D. (2020). Autonomous pavement distress detection using ground penetrating radar and region-based deep learning. *Measurement: Journal of the International Measurement Confederation*, *164*(108077), 108077. doi:10.1016/j.measurement.2020.108077

Jol, H.M. (2008) Ground penetrating radar theory and applications. *Elsevier*, Amsterdam, Netherlands. Available online: https://www.sciencedirect.com/book/9780444533487/ground-penetrating-radar-theory-and-applications

Li, H., Li, N., Wu, R., Wang, H., Gui, Z., & Song, D. (2021). GPR-RCNN: An algorithm of subsurface defect detection for airport runway based on GPR. *IEEE Robotics and Automation Letters*, *6*(2), 3001–3008. doi:10.1109/lra.2021.3062599

Liang, X., Yu, X., Chen, C., Jin, Y., & Huang, J. (2022). Automatic classification of pavement distress using 3D ground-penetrating radar and deep convolutional neural network. *IEEE Transactions on Intelligent Transportation Systems: A Publication of the IEEE Intelligent Transportation Systems Council*, *23*(11), 22269–22277. doi:10.1109/tits.2022.3197712

Xiong, C., Yu, J., & Zhang, X. (2023). Use of NDT systems to investigate pavement reconstruction needs and improve maintenance treatment decision-making. *International Journal of Pavement Engineering*, *24*(1), 1–15. doi:10.1080/10298436.2021.2011872