Review | Received 10 July 2024; Accepted 2 September 2024; Published 7 September 2024
https://doi.org/10.55092/sc20240010

Review of multimodal data and their applications for road maintenance

Shirin Malihi[1,*], Lilia Potseluyko[1,*], Alwyn Mathew[1], Hamidreza Alavi[1], Varun Kumar Reja[1], Yuandong Pan[1], Leonardo Binni[1, 2], Guangming Wang[1], Xiang Wang[1] and Ioannis Brilakis[1]

[1] Department of Engineering, University of Cambridge, Cambridge CB3 0FA, UK
[2] Polytechnic University of Marche, Ancona, Italy

* Correspondence authors ; E-mail: sm2852@cam.ac.uk, lp625@cam.ac.uk.

Abstract: The application of multimodal data in road maintenance has attracted considerable attention due to its potential to enhance decision-making processes and improve infrastructure resilience. This paper provides a comprehensive review of the utilisation of various modalities of multimodal data, including LiDAR, RGB images, thermal images, ground-penetrating radar (GPR), text, audio, and some others for road maintenance tasks. The research methodology thoroughly examines existing literature, categorising data modalities and analysing their respective applications. The paper discusses the integration and fusion of multimodal data, spatial and temporal analysis techniques, decision support systems, strategies for resilience and adaptability and information requirements in for road maintenance. It also explores data structures for integration into digital twin, advanced methodologies for sensor fusion, integration of new sensors and data types and multimodal sensors into road maintenance. This comprehensive review underscores the significance of multimodal data in enhancing the efficiency and effectiveness of road maintenance activities and identifies gaps in the automatic fusion of different modalities in the context of road asset management.

Keywords: maintenance; LiDAR; camera; RGB images; thermal images; GPR; Digital Twin; text; audio; GPS; IMU; game engines; decision support system; data structure; data fusion; road asset; pavement; subsurface; multimodal sensors; information requirements; vibration; spectroscopy; robotics; transformer; GPT; AI; NeRF; large language models (LLM)

1. Introduction and background

1.1. Definitions

This paper presents a comprehensive review of the employment of multimodal data in the context of road maintenance procedures. Multimodal data refers to data that is collected from multiple modalities. These sources can include various sensors, devices, or systems that capture different types of information. In the context of road maintenance, multimodal data *traditionally* include data from sources such as:

- Traffic cameras: Capturing visual information about traffic flow, road conditions, and incidents.
- GPS devices: Providing location data for vehicles and assets on the road network.
- Accelerometers and gyroscopes: Recording motion and orientation data for vehicles and infrastructure.
- Weather stations: Gathering meteorological data such as temperature, precipitation, and

wind speed.
- Road surface sensors: Measuring parameters such as pavement condition, friction, and temperature.
- Vehicle sensors: Monitoring vehicle performance, fuel consumption, and emissions.
- LiDAR (Light Detection and Ranging): Utilising laser scanning technology to measure distances and create detailed 3D representations of road surfaces and surroundings.
- Textual records (mainly from inspection and maintenance reports): including a description of road assets, inspection specifications, identified issues, defects information, maintenance priority, required maintenance activities, completed tasks, and so on.

The paper examines the merits and drawbacks of conventional modalities traditionally employed in road maintenance practices. Additionally, it delves into modalities utilised in other domains that have yet to be thoroughly investigated in the context of road maintenance. Furthermore, an investigation into user requirements for traffic inspections is conducted to ascertain the necessary data for making informed decisions regarding repair processes.

Road maintenance refers to the regular upkeep and repair of roads, and motorways and other transportation infrastructure to ensure their safe and efficient operation. It includes a range of activities aimed at preserving the condition and functionality of road networks, such as repairing potholes, resurfacing pavements, cleaning drainage systems, replacing signage, and maintaining roadside vegetation.

1.2. Problem statement

Roads are cost-effective to build and expand (£8 to £24 millions/mile) when compared to rail (£50 to £400+ millons/mile)[1, 2] . Over time, this has led to an enormous 246,700 miles of motorways and roads in the UK carrying over 83% (328 billion vehicle miles) of all passenger miles travelled and 79% of all domestic freight [3]. Roads are responsible for 69% of all transport Greenhouse Gas Emissions (GHG), the highest of all UK industries. Roads are increasingly safer, but still responsible for 1,784 deaths and 25,511 serious injuries just in 2019. Motorways and A roads are only 13% of the road network's total length but carry 66% of the total road traffic on a shoestring budget, receiving only 15% of the total public expenditure on all transport modes. This lack of investment is a leading cause of steady annual traffic delay increases, with the average delay estimated at 47.3 seconds/vehicle/mile in 2018. This means that 4.3 billion labour hrs/year are wasted in traffic. These numbers highlight the importance of efficient road network expansion, maintenance, and repair. Interventions that are carried out before they are really needed or after they could have been easily corrected can waste resources and unnecessarily increase GHG and accident rates. Unfortunately, this is too often the case, with (i) 3% of all 'A' roads, 6% of 'B' and 'C' roads, and 16% of unclassified roads categorised as needing substantial maintenance, (ii) repair/maintenance road closures alone costing the UK £26.2 m/year, and (iii) the annual carriageway maintenance budget shortfall per authority ballooning to £4.1 million in 2018/19[4]. Years of underfunding within both revenue and capital budgets have led to insufficient cyclical maintenance of roads that, along with the climate change impact and increased traffic volumes, particularly on the minor road network, have accelerated further structural decline.

We contend that this cycle will persist as long as roads remain inadequately documented and monitored, leading to reactive maintenance practices. However, with recent advancements in multimodal data processing, there is no justification for perpetuating this trend. Future road data acquisition processes have the potential to become more intelligent, delivering real-time information and insights to maintenance teams and data analysts. This enhancement enables better application of the capital invested in road maintenance.

1.3. Research significance

The significance of this research lies in its potential to address persistent challenges in road maintenance practices. By highlighting the impact of inadequate documentation and reactive maintenance, the study underscores the urgent need for innovative solutions. Multimodal data fusion plays a significant role in the realm of road asset management. Fusing multiple modalities will result in a comprehensive understanding and assessment of road surface conditions, as each modality senses the environment differently. The exploration of recent advancements in multimodal data processing offers a promising avenue for improving road infrastructure management. By harnessing real-time information and insights, stakeholders can make more informed decisions, leading to optimised resource allocation and enhanced road maintenance outcomes. Ultimately, this research has the potential to contribute to the development of smarter and more efficient road maintenance practices, resulting in safer and more sustainable transportation networks.

1.4. Research structure

This paper begins with an Introduction and Background section, which encompasses definitions, problem statements, and research significance. Gaps in the literature subsection focuses on summarising literature on multimodal data in the context of roads. Following this, the Research Methodology section is presented. Subsequently, the main focus of the research is delved into the multimodal data user requirements, classification, fusion, and real-time integration for the road maintenance and operation section. The Discussion section follows, where the findings and implications of the research are discussed in detail. Finally, the Conclusion section provides a summary and concluding remarks for the research.

This research was conducted by a team of academic experts with diverse expertise in mobile mapping hardware design, photogrammetry, computer graphics, digital twin integration, data modelling, and machine learning. Our goal is to create a real-time, continuously updated digital twin of the UK Road Network using multimodal data. As a first step, we plan to investigate and align the multidisciplinary advancements in multimodal data fusion and its integration with digital twin technology.

1.5. State of research

From the investigated research statistics, there are several notable gaps in the review of multimodal data for road inspections:

Limited coverage of specific road asset types: While there is substantial research on LiDAR-based mobile mapping for road maintenance assessments, the coverage of specific road asset types varies. For instance, there is a significant focus on pavement conditions, traffic signs, and vegetation, but other critical assets such as drainage systems, lampposts and retro-reflectivity quality of road assets receive comparatively less attention[5–7]. Section 3.6.7 Data modalities and their applications for road assets maintenance investigates the capability of each modality to provide the necessary level of information for specified road assets. Table 3 summarises the outcomes of the literature review of each modality and its applicability for each road type asset investigation.

Sparse integration of GPR and thermal camera data: Although there are some instances of integrating ground penetrating radar (GPR) and thermal camera data with LiDAR for road maintenance inspections, the number of studies exploring this combination remains relatively low. This indicates a gap in understanding the potential benefits and challenges of integrating multiple modalities for comprehensive road inspections[8–10]. Sections 3.6.3 Thermal images and Section 3.6.4 GPR explore existing research of these modalities, their benefits and challenges.

Limited research on automation of data fusion with operational data: such as text reports data and routine inspection data. Despite the importance of operational data for road maintenance, there are relatively few studies focusing on it's fusion with LiDAR and GPR data. This gap suggests a need for more research in this area to enhance road maintenance management and infrastructure planning [11–13].

Challenges in Multimodal Data Fusion: While there are numerous studies on multimodal data fusion for road inspections, gaps exist in the comprehensive integration of different data modalities. The limited coverage of specific fusion approaches highlights the need for more diverse and robust fusion methodologies tailored to different road assets.

The review of the research statistics has shown insufficient attention in the literature towards data fusion and multimodal data. In the rest of this paper, we delve into details of the existing research while comparing the described outputs with inspection information requirements and tailored inspection needs of each asset type. Sections 3.8. Advanced sensor fusion and Section 3.9. More sensor data fusion of the paper particularly focus on innovations in multimodal data sensing fusion methodologies.

1.6. Research aim and objectives

We aim *to explore the state-of-the-art in multimodal data fusion and integration with digital twins*. This involves investigating the latest techniques and methodologies for combining various types of data to create a cohesive and comprehensive digital representation of the road network. By understanding the current advancements and challenges in this field, we can identify the most effective approaches for data fusion, address potential integration issues, and lay a strong foundation for developing a robust and reliable digital twin that can significantly enhance road maintenance and operational efficiency. By collaborating on this paper with a multidisciplinary research team, we aim to assemble the puzzle pieces to fully leverage the strengths of each discipline. This *multidisciplinary approach* is crucial as it allows us to integrate diverse perspectives and expertise, ensuring that our solutions are both innovative and practical, ultimately *leading to a more accurate and efficient pipeline where each step is seamlessly interoperable with the others.*

Objectives

- Objective 1 - To conduct a comprehensive review of the utilisation of various modalities of multimodal data, including LiDAR, RGB images, thermal images, ground-penetrating radar (GPR), text, and sound, for road maintenance.
  Approach: This involves thoroughly examining existing literature, classifying data modalities, analysing their respective applications, and identifying gaps in the existing research.
- Objective 2 - To explore integration and fusion techniques of multimodal data, spatial and temporal analysis methods, decision support systems, and strategies for resilience and adaptability in the context of road maintenance.
  Approach: This includes discussing data structures for integration into digital twins, advanced sensor fusion methodologies, and future directions for incorporating additional sensor data to enhance road maintenance and operation activities.
- Objective 3 - To understand the additional values that each modality can bring to DT in the context of road maintenance.
  Approach: This involves identifying the unique contributions of each data modality (LiDAR, RGB images, thermal images, GPR, text, and sound) to the enhancement of Digital Twins.

1.7. State of practice

There are many industry tools for working with mobile mapping and multimodal data, which can be grouped by the functionalities they perform.

Data Collection

Industry-leading companies producing equipment for traffic-speed mobile mapping include Trimble (Trimble MX9 [14]), Leica Geosystems (Leica Pegasus [15]), Topcon (OP-S3 [16]), REIGL (VMX-2HA [17]), and NavVis (NavVis VLX [18]). These systems allow for road inspection at traffic speed without the need for road closures, offering high precision and dense point clouds. These companies also provide additional software for pre-processing the data. However, a drawback is that the hardware often designed on LiDAR and RGB data requires additional equipment to collect other modalities like GPR and thermal data.

The Traffic Speed Deflectometer (TSD) by Greenwood Engineering[19] provides multimodal data collection for pavement, including deflectometer data, GPR, pavement temperature, point clouds, and continuous RGB orthomosaic of the pavement.

A significant drawback of industry state-of-the-art tools for data collection is their high cost and the need for planning and preparation prior to surveys, which limits the frequency of usage. These tools are commonly used for pre-construction scanning, road repair schemes, and yearly analytical surveys, unsuitable for daily/weekly operational maintenance activities. Hence, there is a significant gap in the hardware market in identifying low-cost solutions that would be affordable to use in daily and weekly surveys. In addition to this, the industry data collection survey do not privide real-time integration of multimodal data. Commonly these steps will be completed after additional data processing.

Geographic Information System Software.

Several Geographic Information System (GIS) software solutions are capable of working with multimodal data, integrating various data types such as LiDAR, RGB, GPR, thermal imaging and others. As some data modalities are best represented in 2D, GIS Engines are powerful tools for combining 2D and 3D data. Many industry-leading vendors include GIS functionality; however, in this category, Esri ArcGIS [20] and QGIS [21] represent a substantial share of the market.

QGIS is an open-source software, which means it's freely available to use and modify. This makes it accessible to a wide range of users and encourages collaboration and innovation. QGIS and ArcGIS support various data formats, including vector, raster, and 3D data. They can handle different types of spatial data, making it suitable for integrating and analysing multimodal datasets. QGIS has a plugin architecture that allows users to extend its functionality. There are numerous plugins available for tasks such as data processing, analysis, and visualization, enabling users to customize QGIS according to their specific needs. QGIS and ArcGIS can leverage cloud computing services for processing large datasets or performing computationally intensive geospatial analyses. Users can set up virtual machines (EC2 instances) on AWS and install QGIS on them to run processing tasks in the cloud. The drawback of both QGIS and ArcGIS is latency and rendering delays when working with large point cloud datasets.

BIM Software

Integration of multimodal data continues its steady growth into BIM-related applications. To name a few: Autodesk's Infraworks [22], Civil3D [23], Navisworks [24], Recap [25], Trimble's TBC [26], RealWorks [27], Quadri [28], and Bentley's Microstation [29], OpenRoads [30]. The primary purpose of BIM applications is to support construction planning and design, so the integration of point clouds is commonly required for landscape study, quantity takeoff, and other critical design and planning tasks.

While Building Information Modelling (BIM) models are valuable tools for design, construction and facility management, using them for recording real-time state of roadscan pose certain challenges. Firstly, many road construction models are created in CAD formats containing road alignments and coordinates of point assets and lack comprehensive 3D representation. Secondly, BIM models are typically designed for the pre-construction and planning phases, updating them in real-time to reflect changes in the physical environment might not be feasible due to model parsing and asset referencing issues. BIM Models, Ifc files commonly have limited support for texture mapping and no parameters available to reflect damage and changes with the road assets.

Road Assets Feature Extraction Tools

Feature extraction tools, including *TopoDOT* [31], *Ordinance Survey* [32], and *Orbit3DM* [33], are integral in processing and analyzing data from mobile mapping and other survey technologies. Specifically designed for extracting significant insights from point cloud data, these software solutions enable various applications in infrastructure management, construction, and geographic information systems (GIS).

TopoDOT is a leading software solution for extracting features from LiDAR point clouds working as a plugin for Bentley Systems products like Microstation and OpenRoads. It is best used when exported in a dedicated format from pre-processing software like TBC. TopoDOT extracts road alignments, road markings lines, and road grid lines. It aligns RGB images by references. Feature extraction tools considerably reduce workload compared to manual processing and information extraction, however, the process is yet a long way away from reaching full automation. The extracted results from these tools are presented in CAD formats, which would require further processing to generate a textured 3D reconstruction model.

Classification and Segmentation.

TBC (Trimble Business Center) is a comprehensive software package designed for surveyors and geospatial professionals. It provides tools for processing, analyzing, and managing geospatial data, including point clouds. Seamlessly integrates with other Trimble products, allowing for a smooth workflow. Trimble produces modalities of LiDAR and RGB, besides GNSS and IMU for location and rotation. Laser profilers, projected and perspective cameras capture data from road assets. Trimble Business Center handles different processing including filtering, pre-processing, registration, classification and vectorisation[34]. Terrasolid [35] ecosystem provides users with Terra Scan, Terra Modeller, Terra Match, Terra Photo. They support different formats, classification routines of point cloud data, filtering and vectorisation. Meanwhile, rectification and orthomosaicing are a part of image-to-image adjustments using ground control points in Terra Photo. TIN and mesh, coloured shaded surfaces and triangle nets, contour lines, slope directions and textured surfaces are visual functions for using these modalities [36].

Photogrammetry Generation

Photogrammetry involves the process of reconstructing 3D models from images. The software analyses the overlapping images or scan data, identifying common features and triangulating their positions in 3D space. This process results in a detailed 3D representation of the photographed or scanned object or scene. Photogrammetry generation software encompasses tools like Metashape (formerly Agisoft Photoscan), RealityCapture, 3D Survey, and ContextCapture [37–40].

These photogrammetry generation software packages each offer unique strengths and are tailored to different industry needs. Metashape and RealityCapture are known for their high accuracy and speed, making them suitable for a wide range of applications. 3D Survey provides a user-friendly approach specifically for surveyors, while ContextCapture excels in infrastructure and urban planning with its detailed and scalable models. For mobile mapping datasets Agisoft Metashape currently holds leading position. However, with the best results so far the overlap and density of the mobile mapping data produce mesh with gaps and low resolution, which makes it insufficient for traffic officers and data analytics to use for decision-making support.

Game Engines

Game engines like Unity [41] and Unreal Engine[42] have gained popularity for their ability to handle complex datasets and create immersive virtual environments [43]. Integrating multimodal data with game engines presents opportunities for various industries, including simulation, training, urban planning, and entertainment. Unity offers versatile tools for importing and processing multimodal data, including point clouds, GIS data, photogrammetry models, and sensor data. Unity includes SEnsorSDK and SystemGraph which manages graphical features such as rendering, lighting, meshing, texturing, and shading. 3DVEM solves transformations, alignment, orientation and registration of 3D data including LiDAR and image-derived point clouds [44] . Unreal Engine excels in handling large-scale environments and high-fidelity graphics, making it suitable for processing and visualising multimodal data. Its Blueprint visual scripting system enables rapid prototyping and iteration of data-driven interactions and simulations. Geometry Scripting and Procedural Content Generation Graph(PCG) integrate well with point clouds creating procedural(dynamic) textures. Unreal Engine's Datasmith plugin facilitates the import of CAD data, GIS data, point clouds, and other formats with the preservation of metadata and hierarchy.

The review of the state of practice reveals several gaps and limitations in current approaches to working with multimodal data:

- Limited integration with real-time operations: Many industry tools for data collection, such as those for traffic-speed mobile mapping, are designed for pre-construction scanning or yearly analytical surveys rather than operational maintenance activities. This limits their ability to provide real-time updates for digital twins or support frequent maintenance activities.
- Inadequate support for texture mapping and damage information: BIM models, commonly used for design and construction planning, lack comprehensive 3D representation for real-time road state recording. They often have limited support for texture mapping and lack parameters to reflect damage and changes with road assets.
- Manual Processing requirements: Despite advancements in feature extraction tools like TopoDOT, the process is still far from fully automated. While these tools considerably reduce workload compared to manual processing, further processing is often required to generate textured 3D reconstruction models.
- Insufficient resolution and gaps in meshes: Photogrammetry generation software packages produce mesh with gaps and low resolution when processing mobile mapping data. This makes the data insufficient for decision-making support by traffic officers and data analysts.
- Overall, there is a need for further research and development to address these gaps and limitations in current practices for working with multimodal data. This includes improving real-time integration capabilities, automating processing workflows, enhancing support for texture mapping and damage reflection, and increasing the resolution and accuracy of photogrammetry-generated meshes. Additionally, dynamic integration capabilities in game engines need to be expanded to better support real-time operations and decision-making processes.

1.8. Gaps in the literature

The field of multimodal data fusion for road maintenance and operation is rapidly evolving, driven by continuous advancements in data collection hardware and analytical techniques. Consequently, one significant gap in the existing literature is the need for regular reviews to keep pace with these innovations. Many previous publications may not reflect the current state-of-the-art due to the frequent introduction of new sensors, improved data processing algorithms, and enhanced integration methodologies. This highlights the importance of periodically reassessing and updating the body of knowledge in this domain.

A critical limitation of existing research is the tendency to focus on specific assets or individual data modalities. For instance, numerous studies concentrate exclusively on pavement condition assessment using a single type of data, such as LiDAR or thermal imaging, without considering the synergistic benefits of integrating multiple data sources. This narrow focus limits the comprehensiveness of the solutions proposed and their applicability to the broader spectrum of road maintenance and operational needs.

Furthermore, many papers lack a holistic review and integration with Digital Twin (DT) technologies. While DTs offer a promising framework for real-time, dynamic representation of physical assets, their potential remains underexplored in the context of road infrastructure management. There is a noticeable deficiency in the literature that addresses the integration of multimodal data with DTs, which is essential for creating a cohesive and actionable digital representation of the road network.

In summary, the following gaps have been identified in the current literature:

1. Need for regular reviews: Due to rapid advancements in data collection hardware, existing literature may become outdated quickly, necessitating frequent reassessment.
2. Limited scope on specific assets or modalities: Many studies focus narrowly on specific road assets or single data modalities, missing the opportunity to leverage the full potential of multimodal data fusion.
3. Lack of integration with DTs: There is a paucity of research on the integration of multimodal data with Digital Twin technologies, which is crucial for a comprehensive approach to road maintenance and operation.

Addressing these gaps is essential for advancing the field and harnessing the full potential of multimodal data fusion and Digital Twins in enhancing the efficiency and effectiveness of road maintenance and operation.

2. Research methodology

The methodology used for this paper is shown in Figure 1. The authors used a qualitative approach for this review. This included expert interviews, a review of available literature, and other relevant documentation.

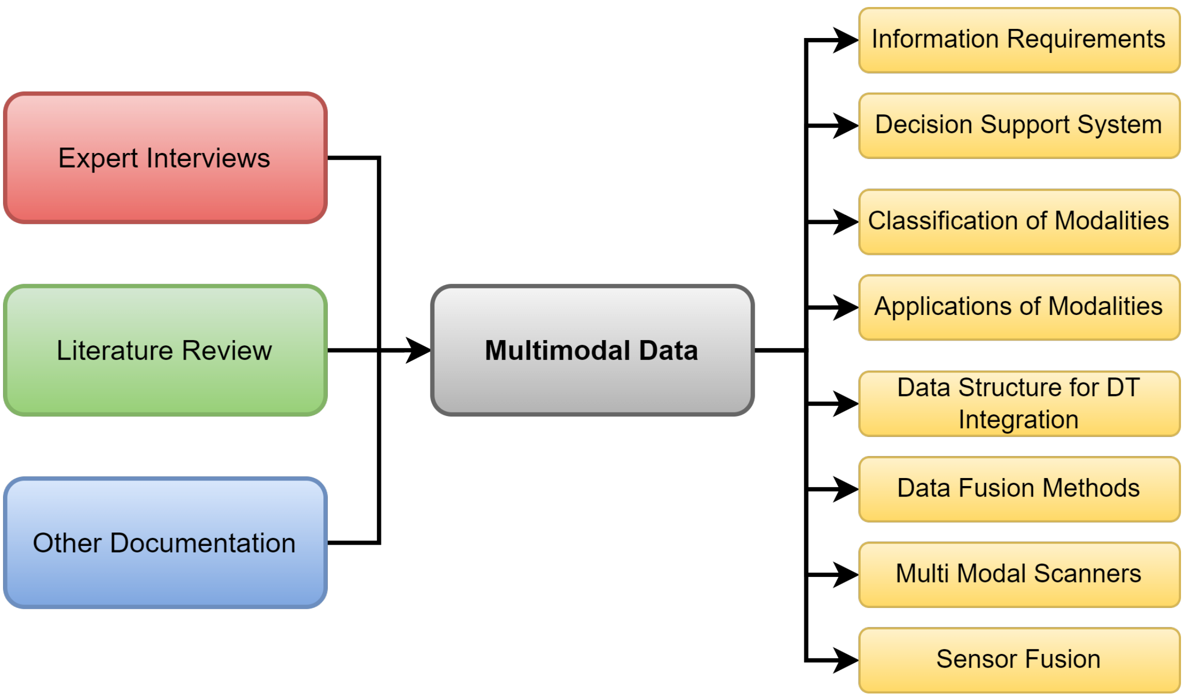

Figure 1. Research methodology.

The expert interviews were conducted systematically with industry experts with technical knowledge about multimodal data and its applications. A set of questions were asked in an unstructured way, so that all the aspects related to information requirements, decision support systems, classification of modalities, application of modalities, data structure for DT integration and Data fusion methods were covered. The information collected through these interviews was then analysed and utilised for this review.

The literature review identified the relevant literature from two comprehensive databases, Scopus and Web of Science. The Keywords search used were "Multimodal Data + Roads" and "Multimodal Data + Maintenance + Roads". A limited set of identified literature was analysed and used for this review.

Other documentation consisted of reviewing road datasets with multiple data modalities. Technical documents available online and in GitHub repositories were also utilised and reviewed to identify the data fusion methods.

3. State of the art for multi-modal data

The advent of multimodal data in road maintenance represents a significant leap in enhancing the efficiency, resilience, and adaptability of road infrastructure systems. Leveraging diverse data sources such as sensor networks, remote sensing technologies, and advanced imaging techniques, the integration and fusion of multimodal data provide comprehensive insights into road conditions, traffic patterns, and environmental factors. This section delves into the state-of-the-art methodologies for utilising multimodal data, emphasising the importance of information requirements, integration techniques, spatial and temporal analysis, decision support systems, and the role of resilience and adaptability in road maintenance. By harnessing these advanced data-driven approaches, road authorities can optimise asset management, predictive maintenance, and overall operational strategies, ensuring sustainable and robust road networks.

3.1. Information requirements

Establishing information requirements for designing digital twins of various road assets necessitates a comprehensive understanding of the specific data needed to accurately represent and simulate these assets. This involves identifying key attributes and parameters such as geometric dimensions, material properties, and real-time operational data [45]. For different road assets, such as pavements, bridges, and signage, unique sets of information must be gathered, including structural health, traffic loads, environmental conditions, and maintenance history. Furthermore, integrating IoT sensors and leveraging data from existing infrastructure management systems can enhance the fidelity and functionality of the digital twins [46]. Ensuring data interoperability, accuracy, and consistency is crucial, as is the need for a scalable data architecture to accommodate future enhancements and expansions of the digital twin models [47]. This foundational information framework supports more efficient asset management, predictive maintenance, and informed decision-making in road infrastructure projects. The theoretical aspects underlying the information requirements for leveraging multimodal data in road maintenance are discussed in the following subsections.

3.2. Integration and fusion of multimodal data

Effective utilisation of multimodal data in road maintenance necessitates clearly understanding the requirements, which encompass various aspects that ensure road networks' smooth functioning and upkeep. The integration combines data from various sources to create a unified view of road conditions and operational parameters [48]. At the same time, fusion techniques aim to extract meaningful insights by merging complementary information from different modalities [49]. Using machine learning methods and ontological frameworks play crucial roles in leveraging multimodal data into road maintenance decision-making.

3.3. Spatial and temporal analysis

Multimodal data in road maintenance exhibit spatial characteristics that influence the dynamics of road networks and operational processes. Spatial analysis involves examining the geographic location and condition of road assets, traffic patterns, and environmental factors to identify spatially-dependent trends and patterns. Temporal analysis of this data focuses on understanding how road conditions, traffic flow, and operational parameters vary over time, enabling predictive modelling and trend analysis. Techniques such as spatial data analytics, time-series analysis, and geospatial visualisation tools empower road authorities to extract valuable insights from multimodal data and make informed decisions regarding maintenance scheduling, traffic management strategies, and infrastructure investments.

3.4. Decisions support systems

Decision support systems (DSS) are pivotal in leveraging multimodal data for road maintenance by providing analytical tools and decision-making frameworks to support planning, monitoring, and control activities. DSS integrates multimodal data sources with computational models, optimisation algorithms, and visualisation techniques to facilitate data-driven decision-making at various levels of the road management hierarchy. From strategic planning and asset management to tactical routing and incident response, DSS empower road authorities to assess alternative scenarios, evaluate trade-offs, and optimise resource allocation based on real-time and predictive insights derived from multimodal data.

3.5. Resilience and adaptability

The resilience and adaptability of road maintenance systems depend on their ability to anticipate and respond effectively to disruptions, uncertainties, and changing conditions. Multimodal data is a foundational resource for enhancing system resilience by enabling proactive risk management, contingency planning, and adaptive decision-making in response to unforeseen events such as natural disasters, accidents, or infrastructure failures. By harnessing multimodal data streams, road authorities can enhance the robustness and agility of their operational processes, thereby minimising disruptions, mitigating risks, and maintaining the functionality and safety of road networks under diverse operating conditions.

Table 1 breaks down these and provides a detailed overview of each information requirement, highlighting its significance and relevance in leveraging multimodal data for road maintenance.

Table 1. A detailed overview of each information requirement in the context of leveraging multimodal data for road maintenance.

Aspect	Information Requirement	Description
Integration and Fusion of Multimodal Data	Data from sensor networks	Sensor networks provide real-time data on various aspects such as traffic flow, road surface conditions, and weather parameters, facilitating continuous monitoring and assessment of road conditions.
	Remote sensing data	Remote sensing technologies, such as camera imagery, satellite imagery, RADAR and LiDAR offer a comprehensive view of road networks and surrounding areas, enabling the assessment of road infrastructure, and environmental factors without any destructive effect or contacting surfaces of road infrastructures.

Table 1. *Cont.*

Aspect	Information Requirement	Description
	Traffic monitoring data	Traffic monitoring systems, including loop detectors, RFID readers, and GPS tracking systems, capture real-time data on traffic flow, congestion levels, and vehicle movements, enabling traffic management and optimisation of transportation systems.
Spatial and Temporal Analysis	Geographic distribution of road assets	Understanding the spatial distribution of road assets such as bridges, tunnels, and signage systems facilitates spatial analysis for identifying spatially-dependent trends, optimising asset management strategies, and prioritising maintenance activities based on geographical considerations.
	Traffic patterns	Analyzing traffic patterns, including origin-destination flows, peak hours, and route preferences, enables road authorities to optimize traffic management strategies, identify congestion hotspots, and improve the overall efficiency of transportation networks.
	Environmental factors	Incorporating environmental factors such as weather conditions, air quality, and terrain characteristics into spatial analysis enables road authorities to assess the impact of environmental factors on road conditions, traffic flow, and infrastructure resilience, informing decision-making and risk management efforts.
	Temporal variations in road conditions and traffic flow	Monitoring temporal variations in road conditions, traffic flow, and congestion levels over different time intervals (e.g., daily, seasonal) facilitates trend analysis, predictive modelling, and the identification of recurring patterns, enabling proactive planning and adaptive decision-making in response to changing conditions.
Decision Support Systems	Geographic distribution of road assets	Integrating computational models such as traffic flow simulations, pavement deterioration models, and optimisation algorithms into decision support systems enables predictive modelling, scenario analysis, and optimisation of road maintenance activities.
	Optimisation algorithms	Employing optimisation algorithms such as genetic algorithms, particle swarm optimisation, and linear programming techniques facilitates resource allocation, route optimisation, and scheduling of maintenance activities, enhancing operational efficiency and cost-effectiveness.
	Visualisation techniques	Utilising visualisation techniques such as geographic information systems (GIS), dashboards, and interactive maps enables the representation of complex data sets and analytical results in a visually intuitive manner, facilitating data exploration, decision-making, and communication with stakeholders.
	Real-time and predictive insights from multimodal data	Leveraging real-time and predictive insights derived from multimodal data sources enables decision-makers to monitor ongoing events, anticipate future trends, and proactively respond to emerging issues, enhancing the responsiveness and effectiveness of road maintenance activities.
Resilience and Adaptability	Proactive risk management	Implementing proactive risk management strategies, such as hazard identification, vulnerability assessment, and risk mitigation planning, enables road authorities to anticipate potential threats and vulnerabilities, reducing the likelihood and impact of adverse events on road networks.

Table 1. *Cont.*

Aspect	Information Requirement	Description
	Contingency planning	Developing contingency plans and response protocols for various scenarios, including natural disasters, accidents, and infrastructure failures, facilitates rapid response and recovery efforts, ensuring minimal disruption and maintaining the functionality of road networks under adverse conditions.
	Adaptive decision-making	Embracing adaptive decision-making processes that enable flexibility, agility, and responsiveness to changing conditions empowers road authorities to adjust strategies, allocate resources, and implement interventions dynamically, enhancing the resilience and adaptability of road maintenance systems.

Table 2 provides a comprehensive overview of different modalities, data types and formats and their applications in road maintenance.

Table 2. Technologies and capabilities in multimodal data for road maintenance.

Technology	Data type and formats	Capabilities
LiDAR	3D point clouds (LAS/LAZ, PLY, XYZ, OBJ, ASC/CSV, and BIN)	• Mapping road surfaces and infrastructure features • Detection of terrain topography
Camera	gray images (PEG, PNG, GIF, TIFF, BMP, and SVG) - videos (MP4, AVI, MOV, WMV, and MKV) - color images (JPEG, PNG, and TIFF)	• Visual inspection of road conditions • Traffic monitoring • Incident detection • Lane marking detection • Anomaly detection • Pavement condition assessment
Thermal Imaging	Infrared radiation (JPEG, PNG, or TIFF, FLIR's radiometric JPEG format (.RJPEG))	• Detection of temperature variations • Identification of thermal anomalies • Detection of defects in pavements and infrastructure
GPR (Ground Penetrating Radar)	Electromagnetic pulses (SEG-Y, GPRMAX, DT1, DZT, and RADAN)	• Non-destructive testing of subsurface layers • Detection of buried utilities and voids • Assessment of pavement structure and thickness
IMU (Inertial Measurement Unit)	Linear and angular motion sensors (CSV, JSON, or proprietary binary formats)	• Real-time data on vehicle dynamics • Integration with GPS for precise vehicle localisation • Motion tracking
GPS	Satellite signals (GPX (GPS Exchange Format)), NMEA (National Marine Electronics Association) 0183, KML (Keyhole Markup Language))	• Precise geographic positioning • Vehicle tracking • Route optimisation

Table 2. *Cont.*

Technology	Data type and formats	Capabilities
Radar	Radio waves (HDF5, NetCDF, CSV)	• Object detection and tracking • Measurement of distance, speed, and direction • Supplemental data for traffic monitoring and collision avoidance
Weather Sensors	Meteorological parameters	• Real-time data on weather conditions • Proactive response to weather-related hazards • Integration with road condition monitoring systems
Wireless Communication Systems	Data exchange between vehicles and infrastructure	• Vehicle-to-vehicle (V2V) communication • Vehicle-to-infrastructure (V2I) communication • Real-time traffic information • Cooperative collision avoidance • ITS applications
Roadway Sensors	Embedded pavement sensors	• Vehicle presence detection • Traffic counting • Speed measurement • Traffic flow monitoring
Environmental Sensors	Environmental parameters	• Monitoring of air quality, noise levels, and pollutant concentrations (Air Quality, Noise, etc.) • Assessment of environmental impacts • Identification of pollution hotspots

3.6. Classification and applications of data modalities

The digitisation process of physical assets is complex. Modern data collection technologies, such as laser scanners and cameras, have greatly enhanced the efficiency of capturing geometric information that reflects the as-is state of facilities in terms of Digital Twins [50]. GPS and IMU observations as positional modalities play important roles in the fusion of other modalities [51]. Integration of GPS and IMU measures the transition and rotation of sensors. Furthermore misalignment and boresight are of important parameters which these positional sensors can solve for relational transformation between modalities [52]. Other modalities are categorized and investigated as following:

3.6.1 LiDAR

The practical benefits of LiDAR technology in infrastructure management are undeniable. Its effectiveness in detecting urban objects and facilitating road asset inventory is a testament to its versatility [53]. By augmenting LiDAR with other sensors like cameras, GPR, and IMU, mapping capabilities can be enhanced, leading to more precise road defect detection and improved traffic safety measures [54]. LiDAR-equipped UAVs have emerged as invaluable assets in surface defect detection and maintenance initiatives, offering a holistic approach to upkeep infrastructure[55]. Using adjacency analysis in virtual twining of man-made objects is one of the state-of the -art usages of LiDAR [56].

Collaborating LiDAR with GNSS technology has paved the way for developing accurate pothole detection systems and revolutionizing road maintenance practices [57]. While LiDAR excels in providing accurate scans of road infrastructure, incorporating cameras is often deemed necessary to capture crucial road surface textures, thereby augmenting the efficacy of road inspection processes. In addition, LiDAR point clouds are also used in vegetation segmentation and classification on roads. By analysing point clouds to identify point-based and neighbourhood-based features, removing planar surfaces, and applying a random forest classifier followed by a rectangularity-based region growing algorithm, vegetation points can be effectively segmented and classified into linear objects [58].

However, spread of laser pulses allows penetration through vegetation and sensing surface data beneath the canopy. Moreover, highly accurate and Large areas can be scanned in a short time compared with terrestrial methods. However, It needs much larger investment into equipment. So although economical when used on large scales, it can be expensive for capturing data in smaller areas. Meanwhile typically LiDAR data is not colourised making it difficult to interpret without overlaying RGB photos.

3.6.2 RGB images

In contrast to laser scanning, which allows the capture the environment in 3D, imaging usually produces a 2D visual representation, i.e., images. An image sensor receives light that is focused by an optical lens and transfers information to a digital signal. There are advantages of using image sensors to capture the environment compared with using laser sensors. The first one is that the capturing device is usually less expensive. A huge amount of different camera options can be selected, and even modern cell phones nowadays also have good camera lenses. The second advantage is that it requires much less professional training to use cameras than to use a laser scanner, which makes it possible for all stakeholders of the facility to capture the current status of the asset. However, 2D images contain only 2D information, which requires image processing technologies to map 2D virtual information to 3D spatial space [59, 60].

For road maintenance, particularly in identifying defects, RGB images can efficiently identify various road surface defects, such as cracks and potholes, which can deteriorate road quality and compromise safety. Advanced image processing and deep learning techniques are integral to these systems, enhancing detection accuracy and reducing high costs and inefficiencies associated with manual inspections [54]. In [61], defects on asphalt pavement are detected by convolutional neural networks, utilising image data collected by a mobile mapping system. [62] presents a dataset that is composed of 45,788 images captured with a high-resolution industrial camera for pavement distress detection and classification. In [63], the authors propose an approach to detect pavement distress in images collected by a drone with a high-resolution camera.

RGB-D cameras, also called depth cameras, work by capturing both RGB images as well as a depth map of the scene, which is achieved by the use of structured light or time-of-flight techniques. While structured light projects a known pattern of light onto the scene and analyses the deformation of the pattern to compute depth, time-of-flight emits a pulse of light and measures the time it takes for the pulse to return. RGB-D cameras are used in a variety of applications, including computer vision, robotics, and augmented reality. They provide a more comprehensive understanding of the environment than traditional RGB cameras alone, allowing for more accurate perception and interaction with the world. Various road defects including patching, cracks, and potholes are detected in data collected by RGBD sensors and unsupervised approach [64].

RGB images are a reliable source for detecting road furniture particularly traffic signs. Many types of defects related to traffic signs could be detected only in the RGB modality. This includes graffiti, dirt, vegetation coverage, faded damaged text and directions. Both RGB and point cloud can be used to detect traffic sign tilt, deformation and missing elements. There are several experiments have been conducted in the literature to localise and 3D reconstruct traffic signs [65–67], however the majority of the scholarly literature focuses on sign detection and recognition for autonomous driving needs [68–70].

Debris and Illegal objects on the road and its sides could also be detected only using RGB modality as it provides necessary context for road maintenance staff and AI detection algorithms. Automation of road safekeeping by clearing debris and illegal object detection has been explored in the literature [71].

3.6.3 Thermal images

Thermal imaging, or infrared thermography (IRT), utilises infrared cameras to capture the heat distribution of objects or areas. This technique operates on the principle that all objects above absolute zero emit radiation, detectable by IRT cameras in the infrared spectrum, similar to how standard cameras capture visible light. As an object's temperature increases, so does the radiation it emits, allowing thermal images to display temperature variations in different colours.

IRT has proven highly effective for detecting non-visible damage on roads by identifying variations in heat signatures indicative of subsurface anomalies such as voids, moisture accumulation, and delamination within pavement layers. Recent studies have advanced this application; for instance, automated methods for detecting sub-pavement voids using IRT have been proposed [72–74]. [73] demonstrated that enhancements such as Principal Component Thermography(PCT). Sparse PCT could significantly improve IRT's detection capabilities. Further, PCT analysis has been utilised to increase the accuracy of damage detection [74]. However, IRT's effectiveness can be compromised by environmental factors such as water presence, shadows, or direct sunlight, necessitating a standardised operating protocol to optimise performance. Additionally, [75] developed a numerical model to predict thermal contrasts in concrete roads, which could forecast subsurface delamination under various environmental conditions. This model was validated with experimental results from an actual concrete block, showcasing its utility as a predictive tool for thermal contrast assessment.

IRT has also been used for road surface damage detection and quality assessment. In [76] and [77], the authors applied IRT to detect pavement defects including chipped slab corners, reflective cracks, local deterioration with scaling, longitudinal cracks, crack networks, grout joining paving blocks, and joint deformations. In [78], IRT is used to detect voids above damaged culverts and drainage pipes and quantify the dimension and severity of the defects. In [79], delamination in RC bridge decks is detected by IRT for rapid bridge inspection. In summary, IRT is a non-intrusive method that does not require surface contact or alteration, preserving the integrity of the structure being inspected. However, external factors such as sunlight, wind, and temperature can influence the thermal readings and affect the accuracy of results.

Infrared radiation cannot go through water or glass. Meanwhile thermal cameras cannot identify individuals because infrared radiation does not create detailed enough images. However, because they generate very-high-contrast images, thermal cameras are just as effective at night as they are during the day and deliver high performance in all weather and air conditions.

3.6.4 GPR

Ground Penetrating Radar (GPR) is a modality of sensing data which can investigate the shallow subsurface of the ground, roads, railways and bridges. It uses the electromagnetic waves for subsurface measurements. The electromagnetic wave is radiated and travels through the subsurface material until it hits an object or surface that has different electrical and magnetic characteristics; it scatters back the wave, so it will be detected by the receiver antenna [8, 80].

The sensitivity of the frequency spectrum of GPR to the typology of materials was investigated in [81] for bituminous mixes, granular and cement-treated materials. It addresses the usage of GPR in monitoring asphalt pavement and determining the structural condition of the pavement. A GPR equipment proper for surveying is horn antennas, which can operate at traffic speed, with frequencies ranging from 1 to 2.5 GHz, for penetration depths of approximately 0.4 m to 1 m, respectively. Another type of antenna is dipole antenna which was developed for use in geological survey, and can be implemented for use in contact with the surface, with the best range of frequencies from 400 MHz to 2.5 GHz. Array Multichannel antennas include a large number of antennas recording simultaneously to enable faster data collection. Usually, GPR antennas are mounted on a vehicle, alongside a GNSS receiver and a distance measurement indicator for geo-tagging and measuring travelled distance [9].

A trace or A-scan is recorded by a single antenna. It is in the time domain and after conversion is referenced in distance or depth. B-scan is a 2D slice which can be visualized by radargram. It contains multiple A-scans in a row. C-scan is created as 3D data by stacking several B-scans and a depth slice of this data. Using an array antenna, the 3d data can be directly generated [82].

The relationship between the velocity of the wave and material properties is the basis for using GPR to investigate the subsurfaces because the velocity is different between materials with different electrical properties, and a signal passed through two different materials will have two different travel time [83].

On the other hand, the deterioration and distress under the surface cannot be examined accurately using traditional methods such as hammer sounding, chain dragging, and test pits. According to [84] GPR was employed as a non-destructive method for routine subsurface inspections, especially in transport infrastructures. This technique can be applied on-site to assess flexible pavements, detect anomalies that could indicate damage in the airport runways, assess the track conditions of railways, inspect vertical structures of retaining walls, detect unknown geometries in the interior of bridges, detect moisture damage and delamination in asphalt pavement [85], and to measure thickness and detect damages of lining layers of tunnels.

GPR is a viable method to detect and determine drainage pipes because they rarely contain metal. However clay-rich soils attenuate GPR signals, since drains are generally linear segments, it leads to a detectable pattern. In a study by the Ohio State University plastic pipes were detectable on single profiles as distinct hyperbolic responses[86].

The advantages of this modality lie in its ability to offer non-invasiveness, and versatility across different materials. However, it comes with limitations due to depth restrictions and challenges in interpreting complex subsurface conditions.

3.6.5 Text

National inventory databases typically contain textual records, including inspection and maintenance reports, which are invaluable for evaluating the progression of every road asset condition and determining necessary maintenance measures. These reports, compiled during each assessment, are strictly connected to the operator's knowledge. Therefore, reports inherently entail subjectivity, potentially leading to human errors and inconsistencies.

Information included within textual reports from road inspections and maintenance activities is rich and multifaceted, encompassing data such as road type, inspection specifications, identified issues, severity of defects, and their respective locations, maintenance priority, required maintenance activities, and so on. Furthermore, the reports may not be restricted to a single asset only (e.g., pavement), but can include information concerning any element forming the road infrastructure (e.g., barriers, signs, and so on). The details are articulated in natural language by often using more than one concept term to refer to the same object and usually describe multiple instances of the same type of deficiency. Currently, these textual resources are used indirectly to support condition assessment and predictive maintenance, but are not used in a quantitative capacity. Consequently, there exists a demand for automated techniques in text information extraction and data fusion to both transform unorganised report data into structured datasets and fuse them with other data modalities (e.g., visual data) for advanced quantitative analysis.

Text data fusion with other data modalities has been extensively studied in computer science and machine learning fields [49, 87–90]. Previous research has primarily examined textual and visual data fusion in various domains such as image captioning [91], question answering [92], image retrieval[93], clinical prediction models[94], and image annotation [95]. However, its application to infrastructure asset condition assessment and predictive maintenance remains challenging due to the complexity and disorder of report data [96].

In the infrastructure domain, extensive research has delved into Information Extraction (IE), the process of distilling crucial details from unstructured textual data like inspection reports [97]. Automated information extraction methods have been devised to aid bridge condition assessments [98–101]. Techniques such as semi-supervised conditional random fields (CFR) [97], deep learning frameworks [96, 99], and hybrid data fusion have been explored [98, 101, 102]. These methods aim to extract localised condition information, synthesise inspection narratives, normalised named entities and concepts, correlate information to overall condition ratings, and integrate visual and text data. Besides, the work proposed by Momtaz *et al.* [96] takes a step forward in terms of multimodal data fusion in the infrastructure domain, proposing a framework for predicting condition ratings of bridge components by fusing textual and visual data. Their framework aims to reduce the uncertainty associated with manual condition assessments, thus minimising human involvement.

Despite these advancements, multimodal data fusion including text sources in the road infrastructure domain remains an unexplored area in research. The advantages from digitally structuring textual records from any road infrastructure inspection and maintenance activities primarily include overcoming the subjectivity, interpretability, and fragmentation of the information contained in them, as well as improving the accuracy of condition ratings in the case of pavements or other structural assets. Structuring textual information from disorganized reports also generates a side benefit, though no less important, concerning the opportunity to merge textual information with other data modalities. Such developments, as the one proposed by [96], benefit from access to different modalities as complementary sources of information that capture the reality of facts and eventual problems from different perspectives, and therefore to provide a more robust and consistent documentation of the physical asset. This opens the way for automated quantitative analyses of road assets conditions, including predictive capabilities.

3.6.6 Audio

The use of audio data for pavements evaluation is an emerging field that utilises the sound generated by the interaction between tires and road surfaces. This approach has the potential to complement existing road monitoring systems by providing real-time insights into road conditions. Several studies have demonstrated the potential of using audio data for road surface analysis. Evaluating the road surface comprises of determining its condition (e.g., dry, wet, and moist) and characteristics (e.g., texture, roughness, and cracking, thickness, layer conditions) [103].

In general, most of the research focused on using audio signals from microphone sensors embedded in the tires [103–108]. Ganji *et al.* [103] designed a custom equipment setup for collecting tire-road interaction noise, focusing on the macrotexture characteristics of the pavement. This method distinguishes surfaces with closely related macrotexture properties by processing the collected audio signals. Their approach highlights the effectiveness of using audio data to evaluate various pavement characteristics, such as texture and roughness.

Gagliardi *et al.* [107] have developed an embedded system designed for real-time road surface classification. This system uses acoustic data recorded inside the tire cavity, ensuring insulation from external noise. The collected audio data are processed into Mel spectrograms, which are then classified using a convolutional neural network (CNN) to determine the road's condition. The CNN is capable of distinguishing between good quality roads, damaged roads, silence, and unknown conditions with an accuracy of 90-93% depending on the model's quantisation. This system represents a significant advancement by providing a low-cost, low-power solution that can be deployed widely for continuous road health monitoring. Abdic *et al.* [108] have introduced a deep learning approach utilising recurrent neural networks (RNNs) to detect wet road surfaces from the audio of tire-surface interactions. Their system achieves an unweighted average recall (UAR) of 93.2%, effectively identifying wet conditions across various vehicle speeds, including when vehicles are stationary. This capability is crucial for enhancing safety by providing timely alerts about hazardous road conditions. On the other hand, few researches address the integration between audio data and other modalities, such as accelerometers to improve the road condition assessment [109–111] and GPS to provide a map visualization of the extracted conditions [111–113].

The use of audio data in road maintenance offers several benefits. The implementation of audio-based monitoring systems can be more cost-effective compared to traditional methods that require extensive hardware and manual inspections. The hardware required, such as microphones and embedded systems, is relatively inexpensive and easy to install. Audio data allows for real-time monitoring of road conditions processing and classifying road conditions on the fly, providing immediate feedback that is critical for timely maintenance actions. Unlike visual inspection systems that can be affected by lighting conditions and weather, audio-based systems can operate effectively in various environments. The placement of microphones inside the tyre cavity, as demonstrated in Gagliardi *et al.* [107], provides insulation from external noise, ensuring reliable data collection. Furthermore, combining audio data with other modalities, such as visual data, GPS and vibration sensors, could enhance the accuracy and reliability of road condition assessments. Multimodal approaches could leverage the strengths of each data type, providing a more comprehensive understanding of road health.

In this direction, Saeed *et al.* [114] proposed audio-image data fusion for classifying the loose gravel condition. In this work, they extracted spectrogram and roboflow from these modalities and a vgg-16 based segmentation method was applied for feature detection and afterwards feature-level fusion. The aim is to aid in adapting gravel road maintenance to reduce the environmental impact and enhance safety.

3.6.7 Data modalities and their applications for road assets maintenance

This section provides a summary of the different data modalities used in road maintenance applications reviewed in the previous sections. For this purpose, Table 3 is provided. Based on each data source analysed in the previous sections (i.e., LiDAR, RGB images, Thermal images, GPR, Text, and Audio), the different types of applications found in the literature that aim to support the maintenance of equally different types of road components (including pavements and other road assets such as markings, furniture, traffic signs, and so on) are categorised. In addition, the main references reviewed for each type of application are given. From the table, although the literature's inclination toward defect and damage detection applications on road pavement emerges strongly, the quantity and heterogeneity of potential applications of different data modalities on road infrastructure maintenance is nevertheless evident.

Table 3. Summary of the various data sources and their applications for different highway components.

Data Modality	Road Asset Type	Applications	Papers
LiDAR	Pavement	Defect and damage detection on pavement and assessing pavement distress	[54, 55, 57, 115]
	Road marking	Road marking defect detection	[116]
	Road furniture	Road asset inventory	[53]
	Vegetation	Road vegetation segmentation and classification	[58]
RGB images	Pavement	Defect and damage detection on pavement	[61 –64]
	Traffic signs	Defects e.g. graffiti	[65–70]
	Traffic signs	Removal of illegal objects e.g. illegal signs	[117, 118]
	Pavement	Automation of debris cleaning via robotics	[71]
Thermal images	Pavement	Detecting non-visible damages within pavement layers	[72–77]
GPR	Culverts, drainage	Defects above damaged culverts and drainage pipes	[78]
	Bridge deck	Bridge inspection (delimitation in RC bridge decks)	[79]
	Pavement	Subsurface inspections: deterioration and distress under the surface	[8, 9, 77, 83 –85]
	Pavement	Infrastructure mapping	[82, 84]
	Retaining walls	Subsurface inspections	[84]
	Vertical structure of bridges	Subsurface inspections	[84]
	Lining layers of tunnels	Subsurface inspections	[84]
	Drainage	Asset detection	[86]
Text	Bridges	Report's information extraction for asset condition assessment	[96–101]
	Bridges	Report's information extraction for deterioration prediction	[102]
Audio	Pavement	Evaluating road surface condition (e.g., dry, wet, and moist) and characteristics (e.g., texture and roughness)	[103]
	Loose gravel road pavement	Evaluating road surface condition and maintenance	[114]

3.7. Data structures for digital twin integration

The integration of advanced computational tools in civil engineering has revolutionised infrastructure management. Building Information Modeling (BIM) and Digital Twin technologies play a pivotal role in this transformation [119, 120]. BIM aids in managing buildings and infrastructure throughout their lifespan by creating digital representations [121]. Digital Twin technology takes BIM a step further by producing real-time virtual replicas of physical assets, which are updated with real-time data, enabling continuous monitoring and predictive maintenance [122, 123]. Digital twins are capable of managing large-scale infrastructure, such as traffic for roads, by processing data from sensors to make immediate assessments and predictions for maintenance [124]. The successful demonstration of digital twin technology in road maintenance has been highlighted in several case studies, outlining its practical benefits and implementation strategies, as summarized in Table 4. It is evident that digital twin technology accompanies a wave of change that arises from the road maintenance and management fields that have been discussed in these case studies. The technical solutions involved the amalgamation of several traffic modes as well as the use of the forecasts from analytics on the basis of the available data sources. Due to these methods, the reliability of the machine in the act of route maintenance is significantly improved and the autonomy the machine is given in the decision-making process can be safely ensured. In this section, we explore the fundamental data structures and their roles in integrating digital twins, focusing on the types of data, data models, and data management strategies required for the effective deployment of digital twins in road maintenance.

Table 4. Case studies of applying Digital Twin technology for road maintenance.

Case Study	Location	Overview	Implementation	Benefits
Singapore's Intelligent Transport System [125]	Singapore	Advanced ITS incorporating digital twin technology for road infrastructure management.	• Multimodal Data Integration: Traffic cameras, sensors, weather data, GPS data. • Predictive Maintenance: Real-time data and historical trends. • Simulation and Scenario Testing: Impact of heavy rainfall, etc.	• Improved maintenance efficiency. • Reduced traffic disruptions. • Cost savings from preventive maintenance.
UK's National Roads Telecommunications Services (NRTS) [126]	UK	Digital twins for managing road network maintenance.	• Sensor Networks: Real-time data on road conditions, traffic flow, and environmental factors. • Data Analytics: Identify patterns and predict maintenance needs. • Visualisation Tools: Visual representation of road network.	• Enhanced decision-making. • Increased safety and reduced accidents. • Optimized resource allocation.

Table 4. *Cont.*

Case Study	Location	Overview	Implementation	Benefits
California Department of Transportation (Caltrans) [127]	California, USA	Enhanced highway maintenance with digital twin technology.	• Real-Time Monitoring: IoT sensors, connected vehicles. • Maintenance Planning: Real-time data, predictive analytics. • Coordination and Communication: Unified digital platform.	• Improved accuracy in identifying maintenance needs. • Enhanced efficiency and effectiveness. • Reduced downtime and inconvenience for users.

3.7.1 Data types for Digital Twin integration

Creating a comprehensive digital twin requires integrating various types of data, ensuring a holistic representation of road infrastructure. This integration involves combining geospatial, structural, sensor, maintenance, traffic, and environmental data to provide a detailed and dynamic model. Each data type contributes uniquely, enhancing the fidelity and utility of the digital twin for applications like maintenance and real-time monitoring. The diverse data sets are crucial for accurate analysis and decision-making, as shown in Table 5.

Table 5. Data types for Digital Twin integration.

Data Type	Description
Geospatial Data	This includes geographic information system (GIS) data, topographic maps, and spatial coordinates defining the physical location and layout of road infrastructure.
Structural Data	This includes detailed specifications of road components, such as pavement layers, bridge dimensions, and materials used, providing a foundation for structural analysis.
Sensor Data	This data is gathered from various sensors such as LiDAR, thermal cameras, GPS, and GPR. These sensors provide spatio-temporal data on road conditions, traffic flow, and environmental factors.
Maintenance and Inspection Data	This includes textual records from inspection reports, maintenance logs, and defect databases that document the condition, repairs, and maintenance activities carried out over time.
Traffic Data	This includes traffic flow statistics, congestion patterns, and accident records, which are crucial for modeling and simulating traffic scenarios.
Environmental Data	This includes weather conditions, temperature, precipitation, and other environmental factors that impact road conditions and maintenance needs.

Digital twins rely on real-time data from various sensors, which monitor aspects such as structural health, traffic flow, and environmental conditions. This data is integrated using APIs and stored in relational databases such as PostgreSQL. Ensuring data integrity and providing seamless access to both historical and real-time data is critical for ongoing road maintenance decisions.

Integrating GIS and BIM is essential for creating a cohesive digital twin environment. GIS provides spatial context by mapping the physical location of assets, while BIM offers detailed 3D models of infrastructure components such as bridges and tunnels. The alignment of geographic and geometric data ensures accurate asset management and analysis, crucial for effective road maintenance.

PCD obtained from 3D scanning technologies like LiDAR represents the surface geometry of assets. Technologies like WebGL, Three.js, or Potree enable the efficient rendering and manipulation of large datasets in web browsers. This integration supports detailed inspections and maintenance planning by providing precise 3D representations of road surfaces and structures.

3.7.2 Data interoperability and standards

The interoperability between different data sources and formats is essential for effective digital twin integration. Common data formats like JSON, XML, and CSV are used for data exchange. Adhering to industry standards, such as IFC (Industry Foundation Classes) for BIM and OGC (Open Geospatial Consortium) standards for GIS data, ensures compatibility and facilitates seamless data integration.

APIs play a critical role in integrating various data sources into the digital twin platform. They provide a standardized way to access and manipulate data, enabling the integration of sensor data, GIS, BIM, and other information systems. The use of RESTful APIs ensures that data can be accessed and updated in real-time, supporting dynamic interactions with the digital twin.

3.7.3 Data security and management

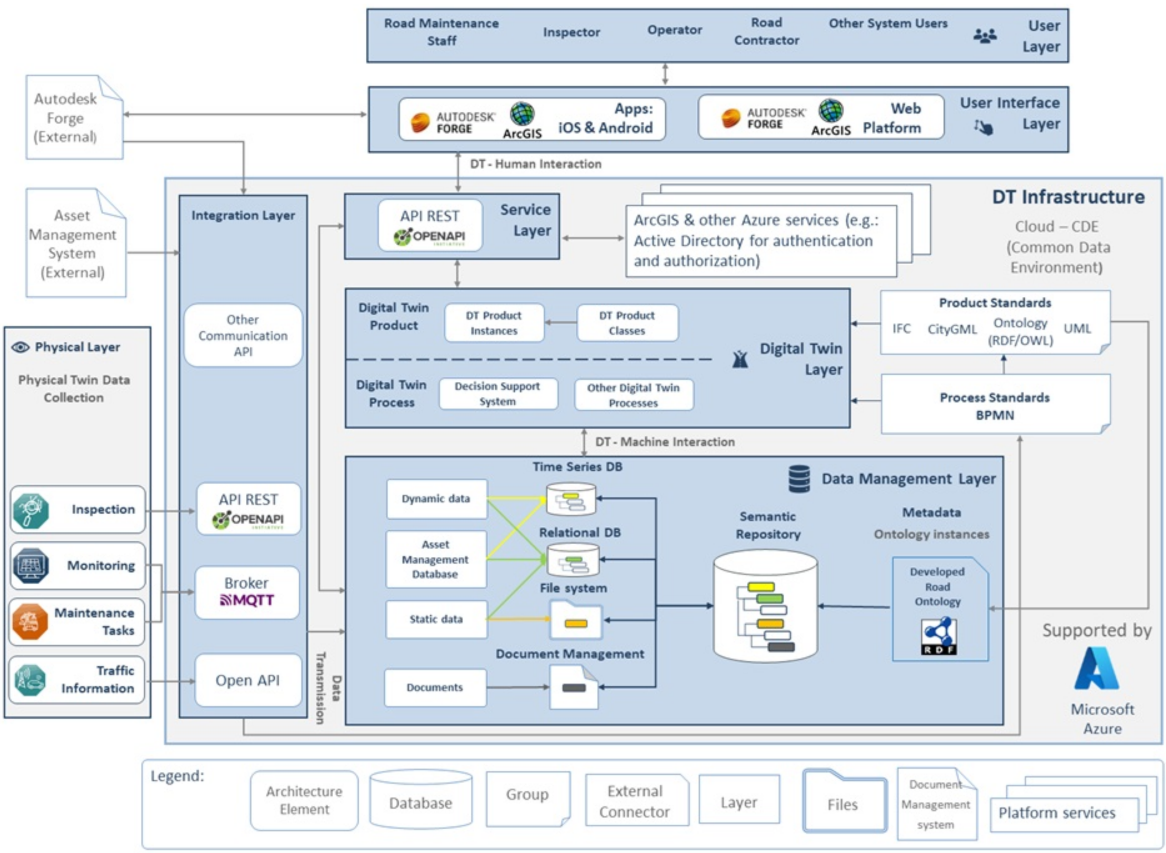

Figure 2. Digital Twin architecture for roads.

It is essential to ensure the security and confidentiality of data in digital twin integration. Strong security measures, such as encryption, access control, and secure communication protocols, should be implemented to allow only authorised users to access sensitive data, thus maintaining the security and reliability of the digital twin platform.

The data management layer plays a crucial role in decision-making within the digital twin architecture. It defines the types of data and the repositories for storing data from inspection, monitoring, maintenance, and traffic sources. The identified data types include Metadata, Static data, Dynamic data, Asset Management data, and Documents. The digital twin processes different types of data from various sources with different natures and formats, requiring optimised technologies for each type. The repositories for project data include: 1- Semantic Repository for metadata, 2- Time Series DB for dynamic data, 3- Relational DB for static, dynamic, and asset management data 4- File System for static data in files, 5- Document management for documents as shown in Figure 2 [128].

The Broker links sensor data and monitoring details, managing inputs from various sensors. This integration ensures that real-time monitoring data is seamlessly incorporated into the digital twin, facilitating decision-making for road maintenance and operations. Effective data management practices are crucial for maintaining the accuracy and usability of the digital twin. This includes regular data validation, cleaning, and updating processes to ensure that data remains relevant and reliable. Implementing a structured data management framework helps in organizing and storing data efficiently, facilitating easy access and analysis.

3.8. Advanced methodologies for fusion of different modalities

Different sensors typically bring complementary information about real-world sampling. For example, RGB cameras provide visual texture information, while LiDAR brings explicit 3D geometric information. Therefore, the development of automated fusion techniques for different sensors greatly assists in enhancing the automation level of road asset maintenance and management. Most recent studies are data-driven methods. These methods mainly focus on weighted fusion based on encoding different sensor information and learning weights to allocate different weights to different information, thus aggregating different information [129]. Recently, the development of some new fusion technologies has brought new possibilities for the maintenance and management of road assets, as follows.

Fusion technologies based on Transformer attention mechanisms [130] have rapidly developed in the field of computer vision [131] and are expected to be widely used for multimodal information fusion in road asset management [132, 133]. Transformers were initially used in natural language processing, as the understanding of language relies on the varying weights of different vocabulary and their positions in understanding the entire sentence. The key of the Transformer is the self-attention mechanism, which can learn the weights for each input feature, enabling the model to better understand the importance of each feature based on contextual information. In the case of multi-sensor data fusion, the self-attention mechanism can effectively align and weigh data from different sensors, thereby extracting more meaningful features. This joint learning of attention to global information and local multi-information input provides possibilities for the fusion of various information on roads. Transformer technology is expected to combine scene context when fusing different sensor information on roads, making judgments on road defects more accurate. Additionally, traditional sequence models, such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) [134], require sequential processing of data, while the Transformer processes input data in parallel. This characteristic allows it to efficiently handle large-scale data and simultaneously fuse data from multiple sensors. At the same time, by stacking multiple layers of Transformer encoders, the model can progressively extract and fuse deep features from multi-sensor data. Transformer models can be stacked in multiple layers and trained on a large scale [135]. Therefore, Transformer models hold promise for large-scale fusion of various sensor information in the field of road traffic.

Technologies based on consistent 3D spatial modelling, such as Neural Radiance Field

(NeRF) [136] and 3D Gaussians splatting [137], are rapidly developing in the field of robotics [73, 138]. NeRF uses neural networks to implicitly represent 3D scenes. The network maps 3D coordinates (x, y, z) and viewing directions to colour and density values. In other words, NeRF does not store explicit geometric information (such as vertices and faces) but encodes the appearance and structure of the scene into the weights of the neural network. In 3D Gaussians splatting [137], 3D Gaussians are used to represent complex geometry distributions in 3D space. Each Gaussian component represents a cluster of points with its own mean and covariance. These components can capture features such as the density and colour of different regions in 3D space. Since both of these encodings store information in 3D space, NeRF and 3D Gaussians can store not only geometric structural features but also other features, such as semantics, in 3D space when establishing a unified 3D spatial model [139, 140]. Their mathematical frameworks allow for flexible and scalable representations of complex 3D structures and features. This is very similar to the concept of digital twins in the civil engineering field, which aims to achieve 3D modelling of physical assets while extracting and detecting information such as defects in 3D space, thus achieving richer and more accurate defect geometric modelling and defect type recognition.

Recently, the application of large models like chatGPT in multiple domains has sparked a wave of interest in the application of generative models [141]. Generative models can incorporate learned knowledge into neural networks and generate the information needed in new scenarios based on conditional inputs. Thanks to the efficient and large-scale utilisation of existing experience by generative models, their applications are very broad. Recently, GPT-4o introduced image processing capabilities, enabling the model to handle and understand both text and images, thereby expanding its range of applications. For example, when asked which objects in an image are dynamic, ChatGPT can deduce the correct answer based on its existing knowledge and output it in text form [142]. In addition, OpenAI began offering fine-tuning capabilities starting with the GPT-3.5 and GPT-4 versions, allowing users to customize and fine-tune the user's own models. This feature is available through the OpenAI API, enabling users to adjust existing pre-trained models based on specific tasks and data, making the models more accurate and effective in certain application scenarios. Therefore, for problems such as road defect detection and classification, it is possible to train an expert generative model that, when fed road defect images, can detect and classify lane defects using large generative models.

Different sensors reflect different physical properties; therefore, more detailed physical modelling is a more interpretable way to integrate multimodal information. For example, in computer graphics, material properties define how light interacts with a surface, geometric properties affect the reflection of light, and texture properties provide surface detail and colour. Through techniques like ray tracing and image rendering, these characteristics interact with each other to collectively determine the visual appearance of the final rendered effect [143]. Since the real world is very complex and the physical mechanisms linking different modal information are also complex and variable, deep learning is often used to assist in areas where physical mechanisms are unclear. However, it is more robust and interpretable to integrate physical mechanisms. Therefore, it is a trend to compile deep learning and physical mechanisms together and achieve the fusion development of physical mechanism modelling and learning-based modelling.

Table 6. Advanced methodologies for fusion of different sensors.

Advanced methodologies	**Technology origin and characteristics**	**Applications on fusion of different sensors**	**Papers**
Transformer attention mechanisms [130]	Initially used in natural language processing to understand the entire sentence.	This technology can fuse global contextual information with local multi-information input, providing possibilities for the fusion of various information on roads.	[130–133]
Neural Radiance Field (NeRF) [136] and 3D Gaussians splatting [137]	Consistent 3D spatial modelling, rapidly developing in the field of computer graphics, computer vision, and robotics.	This technology can store semantics and other features in 3D space similar to the concept of digital twins.	[73, 136–139]
Large generative models like chatGPT	Generative models can incorporate learned knowledge into neural networks and generate the information needed in new scenarios based on conditional inputs.	ChatGPT 4.0 allows each user to fine-tune their own GPT with their own data. Therefore, for problems such as road defect detection and classification, it is possible to train an expert generative model that, when fed road defect images, can detect and classify lane defects using large generative models.	[141, 142]
Detailed physical modelling to connect the relationship between different sensors	Since the real world is very complex and the physical mechanisms linking different modal information are also complex and variable, deep learning is often used to assist in areas where physical mechanisms are unclear, achieving the fusion development of physical mechanism modelling and learning-based modelling.	More detailed physical modelling is a more interpretable way to integrate multimodal information. For example, in computer graphics, through techniques like ray tracing and image rendering, geometric and material information can be linked to RGB images, thus connecting the relationship between RGB images and LiDAR 3D geometry.	[143]

3.9. More sensor data for fusion

Robots can use tactile sensing to identify objects [144]. Thus, tactile sensors hold promise for achieving more accurate and detailed defect modelling through measurements of pressure and deformation with the ground. Vehicles can be seen as robots. Tyre sensors make the vehicles tactile. The tyres of the vehicles interact with the road surfaces directly. The roughness and the friction of the road surface will have an influence on the tyres by the physical tyre-road interaction. Optical tactile sensors by simply introducing light into the surface of a tyre could make a possible to recognise the extent of road defects through the deformation of the tyre's contact with the ground. For example, the long-gauge fibre Bragg gratings have been used for measuring the deformation of the tyre [145]. Mutli-fibre Bragg gratings were installed on the inner face of the tyre to measure the longitudinal deformation of the tyre [146]. Recent research has achieved the identification of material strain with high sensitivity by analysing the backscattered light spectra in a deformed optical fibre [147]. Integrating sensors into the tyres provides additional road surface monitoring abilities for the dynamic sensing of the road surface conditions by the interaction of the tyre and the road surface [148] in a contact way.

For example, the road friction coefficient can be measured by the tyre sensors in real time. The advantage of tactile sensors lies in their ability to measure road surfaces through direct contact, distinguishing them from non-contact sensors such as LiDAR, RGB cameras, and other similar devices. This direct contact method provides unique road surface information that non-contact sensors cannot obtain. Additionally, these tactile sensors can be seamlessly integrated into existing vehicles with minimal modifications, such as by embedding them into the tires. However, there are some drawbacks to consider. The cost of certain tactile sensors, particularly fiber optic sensors, can be high. Furthermore, the maintenance and durability of sensors embedded in tires present significant challenges that need to be addressed to ensure long-term reliability and performance.

As vehicles travel on a road, different defects cause distinct sounds when the vehicle contacts the defective road surface. Recent research [103] has outlined the effectiveness of previous machine-learning methods in pavement monitoring. With the rapid development of speech recognition analysis based on deep neural network [72], enhancing pavement monitoring through sound is also a viable research direction. The advantages of these sensors include their non-invasive nature and cost-effectiveness. However, environmental noise could pose a significant challenge.

Vibration signals obtained with accelerometer modules in the vehicles have been used for road surface monitoring. Recently, road surface profiles have been reconstructed with vibration signals from the accelerometers on the vehicles and the quarter-car model [149]. Ten electric vehicles were used for road profile monitoring on the highways and urban roads and obtained good road profile matching with the results from laser profilometers. Because the accelerometers were installed in the vehicles, this measurement can be implemented across all weather conditions.

Smartphones integrated with accelerometer modules, GPS modules, gyroscopes and other modules. Expensive road surface monitoring equipment restricts their accessibility. Road surface monitoring with data from smartphones becomes an alternative road surface data collection solution. The anomaly pavement conditions have been investigated by the data from smartphones. For example, pothole detection with the accelerometer modules in smartphones has been investigated in Latvia [150] and many other countries [151, 152]. The platforms can be the vehicles [153] and the bicycles [154]. Road condition monitoring with smartphones is used as a supplement to road surface monitoring.

Vibration-based sensors installed in vehicles can provide information on road surface conditions across various weather conditions and are cost-effective. They are particularly effective for monitoring road defects and assessing rough road profiles. However, their sensitivity is typically insufficient for detecting small defects and capturing detailed information about road surfaces.

Spectroscopy allows distinguishing the types of road surface conditions by measuring the materials' spectra. Dry, wet, and snowy road surfaces show different absorption and scattering properties. By measuring the reflectance of the materials, the types of road surfaces and road surface conditions can be obtained. After analysing the reflectance spectra from the road surfaces, the dry, moist, wet, frosty, icy, and snowy road surfaces were classified [155].

Spectroscopy can not only show the materials' types but also show other materials' properties, for example, the thickness of the water on the road surfaces. The thicknesses of the water on the asphalt road surfaces were determined with spectroscopy. A halogen light was used for the illumination of the asphalt road surfaces with 1 mm, 2 mm, 3 mm and 4 mm water depths [156]. By analysing the reflected light intensities, the depth of the waters were determined. The wavelengths chosen for the water depth measurements were 1310 nm, 1490 nm and 1690 nm.

The friction coefficients of asphalt pavement can be obtained with spectroscopy. Carmon and Ben-Dor [157] estimated the friction coefficients with the reflection of the light from

400 nm to 2500 nm wavelength range and showed the potential of using the reflectance spectroscopy to analyse the friction coefficients of the asphalt pavement quantitatively. A table for more sensor data for fusion is shown in Table 7.

Spectroscopy-based methods are also non-destructive methods and have the advantages to analyse material's spectral properties with the spectroscopy. The spectroscopy equipment generally has a high cost especially the photon detectors used for infrared wavelengths range which are employed for distinguishing materials.

Table 7. More sensor data for fusion.

More sensor data	Technology origin and characteristics	Applications on fusion of different sensors	Papers
Contact-surface data	To reconstruct the information of the surfaces, for example, the profile of the surfaces and friction of the surfaces in a contact approach.	The texture information, the friction information, the profile information of the road surfaces supplement the data fusion	[145, 146, 148]
Vibration data	Vibration signals can be measured with the sound sensors, the accelerometers, the gyroscopes to measure the pavement-relevant information.	The additional pavement-relevant data measured with vibration data supplement road surface data fusion, for example the road profile detection and the anomaly pavement condition detection.	[103, 149–154]
Spectral data	Spectroscopy is used to identify the materials' properties by measuring and analyzing the spectral properties.	Additional data for example the road surface status (dry, moist, wet,icy, etc.), water depths on the road surfaces, etc. provide additional road surface information for fusion.	[155–157]

3.10. Multimodal scanners

Road inspections employ scanners with various sensors, including LiDAR, cameras (color, thermal, near-infrared, etc.), and GNSS-inertial systems [158, 159]. The quality of these sensors typically correlates with their price, with higher-priced sensors generally offering superior performance. Scanners used for road inspections are primarily classified into two types: static and mobile. Static scanners remain stationary while scanning, while mobile scanners move and record data simultaneously. Despite the growing interest in mobile scanners within the inspection industry, static scanners remain the preferred choice for most companies due to their ease of use and long-standing reliability. However, the emergence of advanced mobile scanners, albeit costly, has the potential to reduce inspection times significantly. While static scanning technology has demonstrated stability, mobile scanners are still gaining widespread user acceptance.

One of the most compelling features of mobile mapping is its versatility. Unlike terrestrial laser scanning, which is limited to specific locations, mobile mapping can be conducted in diverse environments. By utilising advanced sensors and LiDAR technology mounted on vehicles, drones, or individuals, it provides a comprehensive and seamless view of the surroundings, even in road networks, complex terrains, and previously inaccessible areas. They typically incorporate mechanical LiDAR for precise 3D point cloud generation, and color cameras to enrich the point cloud with color information [160]. The choice of camera type, quality, and resolution is tailored to the specific use cases of road inspection, ensuring the highest level of performance. Additionally, GNSS-inertial systems serve as crucial components of scanners, acting as reference points for fusing sensor data to create a comprehensive point cloud. However, high-accuracy GNSS-inertial systems remain relatively expensive, contributing to the overall cost of advanced scanning technologies. While commonly employed for road inspection, the conventional sensor suite predominantly provides surface-level information about the pavement, often overlooking underlying issues that cause many defects. In contrast, ground-penetrating radar (GPR) sensors integrated with GNSS-based systems [161] offer a non-invasive means of collecting layer-by-layer underground pavement data. Although GPR sensors have effectively revealed subsurface conditions, their slower data collection process may necessitate road closures, limiting their widespread use within road networks.

The market now features advanced mapping systems developed by leading hardware vendors that can quickly produce high-quality colored point cloud data. These devices are usually embedded with LiDAR sensors, cameras, and GNSS-initial systems. Notable examples include the NavVis VLX [162], Leica Pegasus TRK [163], and Trimble MX9 [34]. These multimodal scanners efficiently capture data on various aspects of road infrastructure, including road surfaces, markings, pavement cracks [164], and detailed information on traffic signs, encompassing their type, position, and placement [165]. The future of road surface condition inspection lies in using autonomous vehicles equipped with mobile scanners [166].

4. Discussion

- As each modality can sense the environment differently, multimodal data fusion has been catching attention increasingly in road asset reconstruction and maintenance. LiDAR covers different surfaces and generates a regular point cloud, while RGB images are rich in edge and corner features. Thermal images capture infrared characteristics of the objects and provide detective features for vegetation, waterbodies and a range of stresses on pavement. Thermal imaging has shown its effectiveness in detecting non-visible damage on roads. As an additional data modality that provides the possibility of subsurface information extraction, it can be used in predictive maintenance. However, there are still problems in applying thermal imaging for such purposes in practice. Firstly, using thermal imaging requires more expertise and staff training for operating the data collection device. Secondly, unlike experiments in research labs, the lack of subsurface ground truth in collected datasets for roads in operation limits the interpretation of the data.
  GPR senses unseen areas underground and delivers object responses as a consequence of the disorder of waves, which can detect underground assets and anomalies. This modality suffers from the same problem as thermal images, namely, lack of subsurface ground.
  Text fusion catches attention in order to transform unorganized report data into structured datasets and fuse them with other data modalities (e.g., visual data) for advanced quantitative analysis. Despite the advancements related to text information extraction, the fusion of text with other data sources in the road infrastructure domain remains an unexplored area of research. Emerging audio-based pavement evaluation methods show promise for surface analysis. However, still little research deals with the integration of such data with other modalities.

- Implementation of multimodal data is not without its challenges. Data interoperability and consistency remain significant hurdles, as different technologies and data sources often use varied formats and standards. Developing a scalable data architecture that can accommodate the diverse and growing volume of data is essential for the successful deployment of digital twins in road maintenance. Furthermore, the reliance on advanced machine learning algorithms and ontological frameworks for data fusion and analysis necessitates robust computational infrastructure and expertise. Despite these challenges, the benefits of leveraging multimodal data far outweigh the difficulties, as it leads to more informed decision-making, optimised resource allocation, and enhanced resilience of road infrastructure. By addressing these challenges through standardised data protocols and investing in advanced analytical capabilities, the potential of multimodal data integration in transforming road maintenance practices can be fully realised.

- The integration of digital twins in road maintenance is transforming infrastructure management by making use of comprehensive data structures. Digital twins combine various types of data, such as geospatial, structural, sensor, maintenance, traffic, and environmental data, to enable predictive maintenance and improve decision support systems through advanced spatial and temporal analysis. To ensure seamless data integration, digital twins must be interoperable with standards like IFC and OGC and utilize APIs for real-time data access. Strong data security and effective management practices are crucial for maintaining data integrity and reliability. Although there are challenges in handling diverse data types, the advancement of techniques shows promise in further enhancing digital twin for more efficient and effective infrastructure management.

- In addition to traditional sensors such as LiDAR and RGB cameras used for road monitoring, tactile sensors, fibre optic sensors, spectral sensors, and more sensors are being used for road monitoring. These sensors provide more types of data that can be used for data fusion in road maintenance. For example, tactile sensors can provide details of road textures, fibre optic sensors provide information on the interaction between tyres and road surfaces, and spectral sensors expand the spectral range of RGB to distinguish the material properties on road surfaces. The application of these sensors for road monitoring can become a supplement to the data fusion for road maintenance. However, the challenges for these new sensors are that integrating these advanced sensors with existing systems requires significant modifications not only need to update the hardware, but also need to develop corresponding software for data processing. Developing algorithms for data fusion from the new sensors to the data from Lidar, RGB camera and other sensors a challenging task. The cost of the new sensors also needs to be considered for real in-site monitoring tasks. Some of the sensors may demonstrate feasibility in laboratory settings, transitioning to real-world applications introduces further complexities such as durability, environmental noise resistance. Despite the challenges, new sensor technologies show new potentials for the road maintenance by providing additional complementary multimodal data.

- Currently, there have been numerous road maintenance methods based on multimodal data. However, the wave of AI development continues to bring new methodologies and research perspectives for multimodal data-based road maintenance from different perspectives. These include the complementary integration of characteristics between different sensors with transformer-based attention learning, interactive modelling of 3D spatial features in 3D geometry learning, incorporation of expert knowledge based on generative models, and modelling multimodal data fusion combining AI and physical mechanisms for road maintenance, etc. These innovative research methodologies and AI technologies will elevate the intelligence of road maintenance to a new level in the future.

- Despite the presence of strong players like Leica, Faro, and Trimble in the multi-model scanner market, the current sensor modalities are primarily limited to laser and colour images. While this is a solid foundation, numerous other sensor modalities could significantly enhance road mapping capabilities. Both hardware and software limitations of these scanners still require improvement. Fast laser scanners are still relatively expensive for many road maintenance contractors. Although current systems can accurately register laser and image data, integrating other heterogeneous sensor data remains challenging. Thermal imaging and GPR are highly sought after for road inspection, yet collecting and registering these data types is more complex. However, the rapid advancements in sensor technology, computing power, and market demand indicate a promising future for multi-sensor road inspection scanners.

- This article explores how various data modalities can significantly enhance the development of road digital twins beyond current practices. Integrating these diverse modalities will require a concerted effort to unify and standardize the processes by which we gather, analyze, and curate road data. This involves rethinking the entire pipeline from the ground up—from the sensor combinations used in road inspection scanners to the methods we employ to process and store this data for downstream applications. Inspection scanners should be modular, allowing them to accommodate a wide range of sensors, such as LiDAR, cameras, and others, tailored to specific needs. Furthermore, advanced registration algorithms are needed to align data from different sensors, even when captured at different times of day or under varying environmental conditions. While AI models that normalize weather conditions in recorded data exist, they require significant refinement to be deployment-ready. Additionally, technologies that lag in speed and accuracy, such as GPR, must be prioritized in this multimodal approach. Finally, textual data should be recorded in a machine-standardized format, enabling large language models (LLMs) to digest and effectively provide insights into this data. By addressing these challenges, we can move toward a more robust and comprehensive approach to road maintenance, leveraging the full potential of multimodal data.

5. Conclusion

The integration of multimodal data in road maintenance opens new avenues for improving the accuracy and efficiency of infrastructure management. One of the key advantages lies in the ability to merge data from various sources to create a comprehensive and real-time understanding of road conditions. In this paper, we addressed challenges in road maintenance that multimodal data can tackle. We delved into multi-modal data applications and capabilities for road maintenance including conventional modalities of LiDAR, RGB images, GPR, thermal images and unconventional modalities of texts, audio, and outputs of unconventional sensors. Fusion of different modalities and AI-based strategies for integration of this data in different levels of data and models, besides commercialised solutions for multimodal processing, were enumerated. Furthermore, each modality was analysed from the viewpoint of data types, formats and capabilities in road maintenance.

Key attributes and parameters such as geometric dimensions and material properties were investigated to fulfil the information requirements for data fusion, spatial and temporal analyses, decision support systems, and resilience and adaptability. Data security, management, interoperability and standards for road maintenance were other important issues that this paper highlights. Integration of spatial and temporal, structural, traffic, and environmental data were discussed to enable the predictive maintenance and improve decision support systems. Novel fusion methods, new data types and modalities and multimodal scanners were other aspects of interest which cope with multimodal data in the ream of road maintenance.

The fusion of data types enables a more nuanced analysis of road assets, allowing for detailed assessments that were previously unattainable with single-mode data. Additionally, the integration of real-time traffic data and environmental conditions facilitates more dynamic and responsive maintenance strategies, ensuring that road networks remain operational under varying conditions.

What's more, the review of the research demonstrates a low coverage of specific road assets such as drainage systems and lampposts. Moreover, modalities of GPR and thermal images, along with text and audio, have rarely been integrated into road assets management. Our research explores the fact that multimodal data fusion needs attention to reach a comprehensive and automatic integration with digital twins and bring its values to it in the context of road maintenance. Advancements in multimodal data capture and process enable road construction, monitoring and maintenance more intelligent and near real-time.

Acknowledgments

Postdoc researchers are supported by the European Union's Horizon 2020 OMICRON project under agreement No. 955269, the European Union's AEGIR project under agreement No. 101079961, the Digital Roads Prosperity Partnership, supported by the Engineering and Physical Sciences Research Council (EPSRC) [EP/V056441/1], and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101034337.

Conflicts of interests

Authors declare no conflicts of interest.

Authors' contribution

I. Brilakis was invited by the journal to submit a contribution. I. Brilakis, S.Malihi and L. Potseluyko devised the main conceptual idea. S.Malihi and L. Potseluyko designed the study. I. Brilakis, L. Potseluyko and S.Malihi contributed section 1. V.K. Reja contributed section 2 and subsections 3.1,3.2,3.3, 3.4, 3.5. Y. Pan, A. Mathew, G. Wang, L. Potseluyko, L. Binni and S. Malihi contributed subsection 3.6. H. Alavi contributed subsection 3.7. G. Wang, X. Wang and A. Mathew contributed subsections 3.8, 3.9 and 3.10 respectively. S. Malihi, Y. Pan, A. Mathew, X. Wang, H. Alavi, L. Binni, G. Wang, and V.K. Reja contributed to section 4. S.Malihi, L. Potseluyko and I. Brilakis reviewed the manuscript.

References

[1] Department for Levelling Up H, Communities. Getting Great Britain building again: Speeding up infrastructure delivery. 2023. Available: https://www.gov.uk/government/publications/getting-great-britain-building-again-speeding-up-infrastructure-delivery/getting-great-britain-building-again-speeding-up-infrastructure-delivery (accessed on 7 June 2024).

[2] TransportXtra. Costs of Road Building Explained. 2024. Available: https://www.transportxtra.com/publications/local-transport-today/features/6693/costs-of-road-building-explained (accessed on 15 June 2024).

[3] Nico Kronberg SW. Transport Statistics Great Britain 2019. *Dep. Transp.* 2019 .

[4] DfT. Annual Local Authority Road Maintenance Survey 2020. *Dep. Transp.* 2020 .

[5] Zhai J, Sun Z, Huyan J, Yang H, Li W. Automatic pavement crack detection using multimodal features fusion deep neural network. *INT. J. PAVEMENT ENG.* 2023, 24(2):2086692.

[6] Zhong J, Zhu J, Huyan J, Ma T, Zhang W. Multi-scale feature fusion network for pixel-level pavement distress detection. *Autom. Constr.* 2022, 141:104436.

[7] Chen YL, Jahanshahi MR, Manjunatha P, Gan W, Abdelbarr M, *et al.* Inexpensive multimodal sensor fusion system for autonomous data acquisition of road surface conditions. *IEEE Sens. J.* 2016, 16(21):7731–7743.

[8] Persico R. In *Introduction to ground penetrating radar: inverse scattering and data*

processing, 1st, ed., Hoboken: John Wiley & Sons, 2014.

[9] Hopman V, Beuving E. Repeatability, reproducibility and accuracy of GPR measurements. In *Bearing Capacity Of Roads Volume 1*, Boca Raton: CRC Press, 2022, pp. 637–645.

[10] Pozzer S, De Souza MPV, Hena B, Hesam S, Rezayiye RK, *et al.* Effect of different imaging modalities on the performance of a CNN: An experimental study on damage segmentation in infrared, visible, and fused images of concrete structures. *NDT and E Int.* 2022, 132:102709.

[11] Shan M, Narula K, Worrall S, Wong YF, Perez JSB, *et al.* A Novel Probabilistic V2X Data Fusion Framework for Cooperative Perception. *arXiv* 2022, ArXiv:2203.16964.

[12] Gagolewski M. Data Fusion: Theory, Methods, and Applications. *arXiv* 2015 ArXiv:2208.01644.

[13] Song J, Zhao L, Skinner KA. LiRaFusion: Deep Adaptive LiDAR-Radar Fusion for 3D Object Detection. *arXiv* 2024, ArXiv:2402.11735.

[14] Trimble Geospatial. Trimble MX9 Mobile Mapping System. 2024. Available: https://geospatial.trimble.com/en/products/hardware/trimble-mx9 (accessed on 6 May 2024).

[15] Leica Geosystems. Leica Pegasus:Two Mobile Sensor Platform. 2024. Available: https://leica-geosystems.com/en/products/mobile-sensor-platforms/capture-platforms/pegasus-two (accessed on 17 May 2024).

[16] Topcon Mobile Mapping Solutions. Available: https://www.topconpositioning.com/mapping/mobile-mapping/optech-ops-3 (accessed on 22 March 2024).

[17] RIEGL VMX-2HA. Available: http://products.rieglusa.com/item/mobile-scanners/vmx-2ha-dual-scanner-mobile-laser-scanners/1103 (accessed on 22 June 2024).

[18] NavVis VLX. Available: https://www.navvis.com/navvis-vlx (accessed on 3 May 2024).

[19] Traffic Speed Deflectometer (TSD). Available: https://greenwood.dk/road/tsd/ (accessed on 5 April 2024).

[20] Esri. ArcGIS. 2024. Available https://www.esri.com/en-us/arcgis/products/arcgis-online/overview (accessed on 8 May 2024).

[21] QGIS Development Team. QGIS Geographic Information System. 2024. Available https://www.qgis.org/en/site/ (accessed on 21 March 2024).

[22] Autodesk. Infraworks. Available: https://www.autodesk.com/products/infraworks/overview (accessed on 26 May 2024).

[23] Autodesk. Civil3D. Available: https://www.autodesk.com/products/civil-3d/overview (accessed on 29 June 2024).

[24] Autodesk. Navisworks. Available: https://www.autodesk.com/products/navisworks/overview (accessed on 29 May 2024).

[25] Autodesk. Recap. Available: https://www.autodesk.com/products/recap/overview (accessed on 20 April 2024).

[26] Trimble. Trimble Business Center (TBC). Available: https://geospatial.trimble.com/products-and-solutions/trimble-business-center (accessed on 28 April 2024).

[27] Trimble. RealWorks. Available: https://geospatial.trimble.com/products-and-solutions/trimble-realworks (accessed on 17 May 2024).

[28] Trimble. Quadri. Available: https://geospatial.trimble.com/products-and-solutions/trimble-quadri (accessed on 14 March 2024).

[29] Bentley. Microstation. Available: https://www.bentley.com/en/products/microstation (accessed on 28 May 2024).

[30] Bentley. OpenRoads. Available: https://www.bentley.com/en/products/brands/openroa

ds (accessed on 29 May 2024).
[31] TopoDOT. Point Cloud Processing Software. 2024. Available: https://topodot.com/ (accessed on 10 June 2024).
[32] Ordnance-Survey. See the world in a new light. 2024. Available: https://www.ordnancesurvey.co.uk/ (accessed on 10 May 2024).
[33] Orbit3DM. 3DM FEATURE EXTRACTION. 2024. Available: https://orbitgt.com/products/3dm/ (accessed on 8 May 2024).
[34] Trimble. Trimble MX9. Available: https://geospatial.trimble.com/en/products/hardware/trimble-mx9 (accessed on 26 April 2024).
[35] Terrasolid Ltd. Terrasolid Software Suite. 2024. Available: https://www.terrasolid.com (accessed on 23 June 2024).
[36] Terra. TerraSolid. Available: https://terrasolid.com/ (accessed on 17 May 2024).
[37] Agisoft LLC. Metashape. 2024. Available: https://www.agisoft.com (accessed on 18 June 2024).
[38] Capturing Reality sro. RealityCapture. 2024. Available: https://www.capturingreality.com (accessed on 6 May 2024).
[39] 3Dsurvey. 3D Survey. 2024. Available: https://www.3dsurvey.si (accessed on 20 May 2024).
[40] Bentley Systems, Incorporated. iTwin Capture Modeler. 2024. Available: https://www.bentley.com/en/products/brands/contextcapture (accessed on 28 April 2024).
[41] Unity Technologies. Unity Documentation. 2024. Available: https://docs.unity.com.
[42] Epic Games 2024. Unreal Engine Documentation. Available: https://docs.unrealengine.com.
[43] Sindhu Pradeep M, Reja VK, Varghese K. ConXR: A Comparative Participatory Platform for Construction Progress Monitoring. *J. Inst. Eng. India Ser. A* 2024:1–11 .
[44] GIFLE. 3DVEM Software. Available: http://www.gifle.webs.upv.es/3dvem_en.html (accessed on 26 June 2024).
[45] Reja VK, Davletshina D, Yin M, Wei R, Adam QF, *et al.* A Digital Twin Based Approach to Control Overgrowth of Roadside Vegetation. In *Proceedings of the International Symposium on Automation and Robotics in Construction (IAARC)*. 2024 .
[46] Kober C, Adomat V, Ahanpanjeh M, Fette M, Wulfsberg JP. Digital twin fidelity requirements model for manufacturing. In *Proceedings of the Conference on Production Systems and Logistics: CPSL 2022*. 2022, pp. 595–611.
[47] Correia JB, Abel M, Becker K. Data management in digital twins: a systematic literature review. *Knowl. Inf. Syst.* 2023, 65(8):3165–3196.
[48] Esteva M, Xu W, Simone N, Gupta A, Jah M. Modeling data curation to scientific inquiry: A case study for multimodal data integration. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*. 2020, pp. 235–242.
[49] Lahat D, Adali T, Jutten C. Multimodal data fusion: an overview of methods, challenges, and prospects. *Proc. IEEE* 2015, 103(9):1449–1477.
[50] Davletshina D, Reja VK, Brilakis I. Automating Construction of Road Geometric Digital Twins Using Context and Location Aware Segmentation. *Available at SSRN 4767693* 2024 .
[51] Nafisi V, Beranvand S. Estimation of Total Electron Content Using Single Frequency Weighted Observations for Esfahan Province. *J. Surv. Eng.* 2005, 131(2):60–66.
[52] Malihi S, Maboudi M, Pourmomen M. MISALIGNMENT CALIBRATION OF ULTRACAM D AND XP. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* 2012, XXXIX-B1:577–579.

[53] Ma L, Li Y, Li J, Wang C, Wang R, *et al.* Mobile laser scanned point-clouds for road object detection and extraction: A review. *Remote Sens.* 2018, 10(10):1531.

[54] Zhou Y, Guo X, Hou F, Wu J. Review of intelligent road defects detection technology. *Sustainability* 2022, 14(10):6306.

[55] Bolourian N, Soltani MM, Albahria A, Hammad A. High level framework for bridge inspection using LiDAR-equipped UAV. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*. 2017 .

[56] Zavar H, Arefi H, Malihi S, Maboudi M. TOPOLOGY-AWARE 3D MODELLING OF INDOOR SPACES FROM POINT CLOUDS. *Int. arch. photogramm. remote sens. spat. inf. sci.* 2021, XLIII-B4-2021:267–274.

[57] Yu J, Jiang J, Fichera S, Paoletti P, Layzell L, *et al.* Road Surface Defect Detection–From Image-based to Non-image-based: A Survey. *arXiv* 2024, ArXiv:2402.04297.

[58] Lucas C, Bouten W, Koma Z, Kissling WD, Seijmonsbergen AC. Identification of linear vegetation elements in a rural landscape using LiDAR point clouds. *Remote Sens.* 2019, 11(3):292.

[59] Reja VK, Varghese K, Ha QP. Computer vision-based construction progress monitoring. *Autom. Constr.* 2022, 138:104245.

[60] Malihi S, Valadan Zoej MJ, Hahn M. Large-Scale Accurate Reconstruction of Buildings Employing Point Clouds Generated from UAV Imagery. *Remote Sens.* 2018, 10(7).

[61] Chun Pj, Yamane T, Tsuzuki Y. Automatic detection of cracks in asphalt pavement using deep learning to overcome weaknesses in images and GIS visualization. *Appl. Sci.* 2021, 11(3):892.

[62] Du Y, Pan N, Xu Z, Deng F, Kang H. Pavement distress detection and classification based on YOLO network. *Int. J. Pavement Eng.* 2020, (1):1–14.

[63] Zhu J, Zhong J, Ma T, Huang X, Zhang W, *et al.* Pavement distress detection using convolutional neural networks with images captured via UAV. *Autom. Constr.* 2022, (133-Jan.).

[64] Jahanshahi MR, Jazizadeh F, Masri SF, Becerik-Gerber B. Unsupervised approach for autonomous pavement-defect detection and quantification using an inexpensive depth sensor. *J. Comput. Civ. Eng.* 2013, 27(6):743–754.

[65] Zhou L, Wu G, Zuo Y, Chen X, Hu H. A Comprehensive Review of Vision-Based 3D Reconstruction Methods. *Sensors* 2024, 24(7).

[66] Musa A. Multi-view traffic sign localization with high absolute accuracy in real-time at the edge. In *Proceedings of the 30th international conference on advances in geographic information systems*. 2022, pp. 1–12.

[67] Wang M, Liu R, Yang J, Lu X, Yu J, *et al.* Traffic Sign Three-Dimensional Reconstruction Based on Point Clouds and Panoramic Images. *Photogramm. Rec.* 2022, 37(177):87–110.

[68] Li S, Wang S, Wang P. A Small Object Detection Algorithm for Traffic Signs Based on Improved YOLOv7. *Sensors* 2023, 23(16):7145.

[69] Zhu Y, Yan WQ. Traffic sign recognition based on deep learning. *Multimed. Tools Appl.* 2022, 81:17779–17791.

[70] Arasanipalai Mohan Aditya SM. Study of Traffic Sign Detection and Recognition Algorithms. *Int. J. Sci. Res.* 2016 .

[71] Almanzor E, Anvo NR, Thuruthel TG, Iida F. Autonomous detection and sorting of litter using deep learning and soft robotic grippers. *Front. Rob. AI* 2022, 9.

[72] Dhanjal AS, Singh W. A comprehensive survey on automatic speech recognition using neural networks. *Multimed. Tools Appl.* 2024, 83(8):23367–23412.

[73] Sucar E, Liu S, Ortiz J, Davison AJ. imap: Implicit mapping and positioning in real-time. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 6229–6238.

[74] Kulkarni NN, Dabetwar S, Benoit J, Yu T, Sabato A. Comparative analysis of infrared thermography processing techniques for roadways' sub-pavement voids detection. *NDT and E Int.* 2022, 129:102652.

[75] Rumbayan R, Washer GA. Modeling of environmental effects on thermal detection of subsurface damage in concrete. *Res. Nondestr. Eval.* 2014, 25(4):235–252.

[76] Gameliak I, Vyrozhemskyi V, Voloshyna I, Dmitriev M. Research of cement concrete pavement using thermal imaging method. In *The 4th Chinese–European Workshop on Functional Pavement Design, CEW 2016*. 2016 .

[77] Lagüela S, Carracelas MS, Puente I, Martínez FJP. Joint use of GPR, IRT and TLS techniques for the integral damage detection in paving. *Constr. Build. Mater.* 2018, 174.

[78] Sabato A, Yu T, Kulkarni NN, Dabetwar S. Detecting Subsurface Voids in Roadways Using UAS with Infrared Thermal Imaging. Tech. rep., Massachusetts. Dept. of Transportation. Office of Transportation Planning, 2022.

[79] Omar T, Nehdi ML, Zayed T. Infrared thermography model for automated detection of delamination in RC bridge decks. *Constr. Build. Mater.* 2018, 168:313–327.

[80] Yang X, Huang R, Meng Y, Liang J, Rong H, *et al.* Overview of the application of Ground-Penetrating Radar, Laser, infrared thermal Imaging, and ultrasonic in nondestructive testing of road surface. *Measurement* 2024, 224:113927.

[81] Pedret Rodés J, Martínez Reguero A, Pérez-Gracia V. GPR spectra for monitoring asphalt pavements. *Remote Sens.* 2020, 12(11):1749.

[82] Merkle D, Frey C, Reiterer A. Fusion of ground penetrating radar and laser scanning for infrastructure mapping. *J. Appl. Geod.* 2021, 15(1):31–45.

[83] Luo TX, Lai WW, Chang RK, Goodman D. GPR imaging criteria. *J. Appl. Geophys.* 2019, 165:37–48.

[84] Solla M, Pérez-Gracia V, Fontul S. A review of GPR application on transport infrastructures: Troubleshooting and best practices. *Remote Sens.* 2021, 13(4):672.

[85] Zhang J, Yang X, Li W, Zhang S, Jia Y. Automatic detection of moisture damages in asphalt pavements from GPR data with deep CNN and IRS method. *Autom. Constr.* 2020, 113:103119.

[86] Anonymous. Agricultural Drainage. 2024. Available: https://www.sensoft.ca/case-studies/drainage-tile-locator/ (accessed on 17 June 2024).

[87] Zhao F, Zhang C, Geng B. Deep Multimodal Data Fusion. *ACM Comput. Surv.* 2024 .

[88] Narayana P, Pednekar A, Krishnamoorthy A, Sone K, Basu S. HUSE: Hierarchical Universal Semantic Embeddings. *arXiv* 2019, ArXiv:1911.05978.

[89] Gallo I, Calefati A, Nawaz S, Janjua MK. Image and Encoded Text Fusion for Multi-Modal Classification. *arXiv* 2018, ArXiv:1810.02001.

[90] Nota G, Postiglione A, Carvello R. Text mining techniques for the management of predictive maintenance. *Procedia Comput. Sci.* 2022, 200:778–792.

[91] Bernardi R, Cakici R, Elliott D, Erdem A, Erdem E, *et al.* Automatic description generation from images: A survey of models, datasets, and evaluation measures. *J. Artif. Intell. Res.* 2016, 55:409–442.

[92] Antol S, Agrawal A, Lu J, Mitchell M, Batra D, *et al.* Vqa: Visual question answering. In *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 2425–2433.

[93] Piras L, Giacinto G. Information fusion in content based image retrieval: A

comprehensive overview. *Inf. Fusion* 2017, 37:50–60.

[94] Ye J, Hai J, Song J, Wang Z. Multimodal data hybrid fusion and natural language processing for clinical prediction models. *AMIA Jt. Summits Transl. Sci. Proc.* 2023, 2024:191–200.

[95] Zhu S, Li X, Shen S. Multimodal deep network learning-based image annotation. *Electron. Lett.* 2015, 51(12):905–906.

[96] Momtaz M, Li T, Harris DK, Lattanzi D. Multi-modal deep fusion for bridge condition assessment. *J. Infrastruct. Intell. Resilience* 2023, 2(4):100061.

[97] Liu K, El-Gohary N. Ontology-based semi-supervised conditional random fields for automated information extraction from bridge inspection reports. *Autom. Constr.* 2017, 81:313–327.

[98] Liu K, El-Gohary N. Similarity-based dependency parsing for extracting dependency relations from bridge inspection reports. In *Computing in Civil Engineering 2017*, 2017, pp. 316–323.

[99] Li T, Alipour M, Harris DK. Mapping textual descriptions to condition ratings to assist bridge inspection and condition assessment using hierarchical attention. *Autom. Constr.* 2021, 129:103801.

[100] Riveiro B, DeJong M, Conde B. Automated processing of large point clouds for structural health monitoring of masonry arch bridges. *Autom. Constr.* 2016, 72:258–268.

[101] Sun Z, Tang P, Shi Y, Xiong W. Visual-semantic alignments for automated interpretation of 3D Imagery data of high-pier bridges. In *ASCE International Conference on Computing in Civil Engineering 2019*. 2019, pp. 209–216.

[102] Liu K, El-Gohary N. Fusing data extracted from bridge inspection reports for enhanced data-driven bridge deterioration prediction: A hybrid data fusion method. *J. Comput. Civ. Eng.* 2020, 34(6):04020047.

[103] Ganji MR, Ghelmani A, Golroo A, Sheikhzadeh H. A brief review on the application of sound in pavement monitoring and comparison of tire/road noise processing methods for pavement macrotexture assessment. *Arch. Comput. Methods Eng.* 2021 28(4):2977–3000.

[104] Mednis A, Strazdins G, Liepins M, Gordjusins A, Selavo L. RoadMic: Road surface monitoring using vehicular sensor networks with microphones. In *Networked Digital Technologies: Second International Conference, NDT 2010*, Prague, Czech Republic, July 7–9, 2010, pp. 417–429.

[105] Alonso J, López J, Pavón I, Recuero M, Asensio C, *et al.* On-board wet road surface identification using tyre/road noise and support vector machines. *Appl. Acoust.* 2014, 76:407–415.

[106] Ibarra D, Ramírez-Mendoza R, Ibarra S. Characterization of the road surfaces in real time. *Appl. Acoust.* 2016, 105:93–98.

[107] Gagliardi A, Staderini V, Saponara S. An embedded system for acoustic data processing and AI-based real-time classification for road surface analysis. *IEEE Access* 2022, 10:63073–63084.

[108] Abdić I, Fridman L, Brown DE, Angell W, Reimer B, *et al.* Detecting road surface wetness from audio: A deep learning approach. In *2016 23rd International Conference on Pattern Recognition (ICPR)*. 2016, pp. 3458–3463.

[109] Masino J, Wohnhas B, Frey M, Gauterin F. Identification and Prediction of Road Features and their Contribution on Tire Road Noise. *WSEAS Trans. Syst. Control* 2017, 12.

[110] Johnsson R, Odelius J. Methods for road texture estimation using vehicle measurements. In *Proceedings of ISMA2012-USD2012*. 2012 .

[111] David J, Van Hauwermeiren W, Dekoninck L, De Pessemier T, Joseph W, *et al.* Rolling-noise-relevant classification of pavement based on opportunistic sound and vibration monitoring in cars. In *Proceedings of the 48th International Congress and Exposition on Noise Control Engineering*. 2019, pp. 2193:1–2193:7.

[112] Ramos-Romero C, Asensio C, Moreno R, de Arcas G. Urban road surface discrimination by tire-road noise analysis and data clustering. *Sensors* 2022, 22(24):9686.

[113] Ramos-Romero C, León-Ríos P, Al-Hadithi BM, Sigcha L, De Arcas G, *et al.* Identification and mapping of asphalt surface deterioration by tyre-pavement interaction noise measurement. *Measurement* 2019, 146:718–727.

[114] Saeed N, Alam M, Nyberg RG. A Multimodal Deep Learning Approach for Gravel Road Condition Evaluation through Image and Audio Integration. *Transp. Eng.* 2024, 16:100228.

[115] Yamaguchi T, Mizutani T. Quantitative road crack evaluation by a U-Net architecture using smartphone images and Lidar data. *Comput.-Aided Civ. Infrastruct. Eng.* 2024, 39(7):963–982.

[116] Zhu J, Bu T, Ma T, Huang X, Chen F. Raster-Based Point Cloud Mapping of Defective Road Marking: Toward Automated Road Inspection via Airborne LiDAR. *J. Transp. Eng.* 2024, 150(2):04024015.

[117] GovermentUK. Traffic Signs Manual - Chapter 1. 2018. Available: https://assets.publishing.service.gov.uk/media/5c419a1240f0b61704aec4d7/traffic-signs-manual-chapter-1.pdf (accessed on 13 June 2024).

[118] Ngeni FC, Mwakalonge JL, Comert G, Siuhi S, Vaidyan V. Monitoring of Illegal Removal of Road Barricades Using Intelligent Transportation Systems in Connected and Non-Connected Environments. *US Department of Transp.* 2022 .

[119] Alavi H, Forcada N, Bortolini R, Edwards DJ. Enhancing occupants' comfort through BIM-based probabilistic approach. *Autom. Constr.* 2021, 123:103528.

[120] Consilvio A, Hernández JS, Chen W, Brilakis I, Bartoccini L, *et al.* Towards a digital twin-based intelligent decision support for road maintenance. *Transp. Res. Procedia* 2023, 69:791–798.

[121] Alavi H, Bortolini R, Forcada N. BIM-based decision support for building condition assessment. *Autom. Constr.* 2022, 135:104117.

[122] Sacks R, Brilakis I, Pikas E, Xie HS, Girolami M. Construction with digital twin information systems. *Data-Centric Eng.* 2020, 1:e14.

[123] Alavi H. Building information modeling for facility managers. Ph.D. thesis, Universitat Politècnica de Catalunya, Barcelona, 2022.

[124] Zhang N, Alipour A. A two-level mixed-integer programming model for bridge replacement prioritization. *Comput.-Aided Civ. Infrastruct. Eng.* 2020, 35(2):116–133.

[125] Land Transport Authority of Singapore. Intelligent Transport Systems. 2024. Available: https://www.lta.gov.sg/content/ltagov/en/getting_around/driving_in_singapore/intelligent_transport_systems.html (accessed on 24 April 2024).

[126] AMG Systems. National Roads Telecommunications Services (NRTS) Case Study. 2024. Available: https://www.amgsystems.com/Files/Files/Applications%20and%20Case%20Studies/AMG_NRTS_CS.pdf (accessed on 24 May 2024).

[127] California Department of Transportation. California Department of Transportation (Caltrans). 2024. Available: https://dot.ca.gov/ (accessed on 24 June 2024).

[128] Chen W, Brilakis I. Developing Digital Twin Data Structure and Integrated Cloud Digital Twin Architecture for Roads. In *Computing in Civil Engineering 2023*, 2023, pp. 424–432.

[129] Xing J, Wu W, Cheng Q, Liu R. Traffic state estimation of urban road networks by

multi-source data fusion: Review and new insights. *Physica A Stat. Mech. Appl.* 2022, 595:127079.

[130] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, *et al.* Attention is all you need. *Adv. Neural Inf. Process. Syst.* 2017, 30.

[131] Khan S, Naseer M, Hayat M, Zamir SW, Khan FS, *et al.* Transformers in vision: A survey. *ACM Comput. Surv.* 2022, 54(10s):1–41.

[132] Lin Z, Wang H, Li S. Pavement anomaly detection based on transformer and self-supervised learning. *Autom. Constr.* 2022, 143:104544.

[133] Guo F, Qian Y, Liu J, Yu H. Pavement crack detection based on transformer network. *Autom. Constr.* 2023, 145:104646.

[134] Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D* 2020, 404:132306.

[135] Li Z, Wallace E, Shen S, Lin K, Keutzer K, *et al.* Train big, then compress: Rethinking model size for efficient training and inference of transformers. In *International Conference on machine learning*. 2020, pp. 5958–5968.

[136] Mildenhall B, Srinivasan PP, Tancik M, Barron JT, Ramamoorthi R, *et al.* Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 2021, 65(1):99–106.

[137] Kerbl B, Kopanas G, Leimkühler T, Drettakis G. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graphics* 2023, 42(4):1–14.

[138] Wang G, Pan L, Peng S, Liu S, Xu C, *et al.* NeRF in Robotics: A Survey. *arXiv* 2024, ArXiv:2405.01333.

[139] Zhu S, Wang G, Blum H, Liu J, Song L, *et al.* Sni-slam: Semantic neural implicit slam. *arXiv* 2023, ArXiv:2311.11016.

[140] Zhu S, Qin R, Wang G, Liu J, Wang H. Semgauss-slam: Dense semantic gaussian splatting slam. *arXiv* 2024, ArXiv:2403.07494.

[141] Wu T, He S, Liu J, Sun S, Liu K, *et al.* A brief overview of ChatGPT: The history, status quo and potential future development. *IEEE/CAA J. Autom. Sin.* 2023, 10(5):1122–1136.

[142] Lu J, Clark C, Zellers R, Mottaghi R, Kembhavi A. Unified-io: A unified model for vision, language, and multi-modal tasks. In *The Eleventh International Conference on Learning Representations*. 2022 .

[143] Kocsis P, Sitzmann V, Nießner M. Intrinsic Image Diffusion for Single-view Material Estimation. *arXiv* 2023, ArXiv:2312.12274.

[144] Pohtongkam S, Srinonchat J. Tactile object recognition for humanoid robots using new designed piezoresistive tactile sensor and dcnn. *Sensors* 2021, 21(18):6024.

[145] Zhao W, Zhang C, Zhang J. Continuous measurement of tire deformation using long-gauge strain sensors. *Mech. Syst. Signal Process.* 2020, 142:106782.

[146] Agliullin T, Gubaidullin R, Morozov O, Sahabutdinov AZ, Ivanov V. Tire strain measurement system based on addressed FBG-structures. In *2019 Systems of Signals Generating and Processing in the Field of on Board Communications*. 2019, pp. 1–5.

[147] Wang X, Xiao Y, Rans C, Benedictus R, Groves RM, *et al.* Enhanced Strain Measurement Sensitivity with Gold Nanoparticle-Doped Distributed Optical Fibre Sensing. *Struct. Control Health Monit.* 2024 .

[148] Maurya D, Khaleghian S, Sriramdas R, Kumar P, Kishore RA, *et al.* 3D printed graphene-based self-powered strain sensors for smart tires in autonomous vehicles. *Nat. Commun.* 2020, 11(1):5392.

[149] Skar A, Levenberg E. Road Profile Inversion from In-Vehicle Accelerometers. *J. Transp. Eng.* 2024, 150(1):04023040.

[150] Mednis A, Strazdins G, Zviedris R, Kanonirs G, Selavo L. Real time pothole detection using android smartphones with accelerometers. In *2011 International conference on distributed computing in sensor systems and workshops (DCOSS)*. 2011, pp. 1–6.

[151] Souza VM. Asphalt pavement classification using smartphone accelerometer and complexity invariant distance. *Eng. Appl. Artif. Intell.* 2018, 74:198–211.

[152] Shtayat A, Moridpour S, Best B. Using e-bikes and private cars in dynamic road pavement monitoring. *Int. J. Transp. Sci. Technol.* 2022, 11(1):132–143.

[153] Seraj F, Van Der Zwaag BJ, Dilo A, Luarasi T, Havinga P. RoADS: A road pavement monitoring system for anomaly detection using smart phones. In *International Workshop on Modeling Social Media*. 2014, pp. 128–146.

[154] Zang K, Shen J, Huang H, Wan M, Shi J. Assessing and mapping of road surface roughness based on GPS and accelerometer sensors on bicycle-mounted smartphones. *Sensors* 2018, 18(3):914.

[155] Casselgren J, Rosendahl S, Sjödahl M, Jonsson P. Road condition analysis using NIR illumination and compensating for surrounding light. *Opt. Lasers Eng.* 2016, 77:175–182.

[156] Casselgren J, Sjödahl M, LeBlanc J. Angular spectral response from covered asphalt. *Appl. Opt.* 2007, 46(20):4277–4288.

[157] Carmon N, Ben-Dor E. Rapid assessment of dynamic friction coefficient of asphalt pavement using reflectance spectroscopy. *IEEE Geosci. Remote Sens. Lett.* 2016, 13(5):721–724.

[158] Soilán M, Sánchez-Rodríguez A, del Río-Barral P, Perez-Collazo C, Arias P, *et al.* Review of laser scanning technologies and their applications for road and railway infrastructure monitoring. *Infrastructures* 2019, 4(4):58.

[159] Puente I, González-Jorge H, Martínez-Sánchez J, Arias P. Review of mobile mapping and surveying technologies. *Measurement* 2013, 46(7):2127–2145.

[160] Yu Y, Li J, Guan H, Wang C. Automated extraction of urban road facilities using mobile laser scanning data. *IEEE Trans. Intell. Transp. Syst.* 2015, 16(4):2167–2181.

[161] Rasol M, Pais JC, Pérez-Gracia V, Solla M, Fernandes FM, *et al.* GPR monitoring for road transport infrastructure: A systematic review and machine learning insights. *Constr. Build. Mater.* 2022, 324:126686.

[162] NavVis. NavVis VLX laser scanning systems. Available: https://www.navvis.com/vlx (accessed on 26 April 2024).

[163] Leica Geosystems AG. Leica Pegasus TRK. Available: https://leica-geosystems.com/en-gb/products/leica-pegasus-trk (accessed on 21 April 2024).

[164] Guan H, Li J, Yu Y, Chapman M, Wang C. Automated road information extraction from mobile laser scanning data. *IEEE Trans. Intell. Transp. Syst.* 2014, 16(1):194–205.

[165] Wen C, Li J, Luo H, Yu Y, Cai Z, *et al.* Spatial-related traffic sign inspection for inventory purposes using mobile laser scanning data. *IEEE Trans. Intell. Transp. Syst.* 2015, 17(1):27–37.

[166] Urano K, Hiroi K, Kato S, Komagata N, Kawaguchi N. Road Surface Condition Inspection Using a Laser Scanner Mounted on an Autonomous Driving Car. *2019 IEEE Int. Conf. Pervasive Comput. Commun.* 2019, pp. 826–831.