The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIX-B1-2026
XXV ISPRS Congress 2026 "From Imagery to Understanding", Commission I, 4–11 July 2026, Toronto, Canada

Developing an Urban Road Dataset: A Multi-Sensor Framework for DT and AI-Based Road Infrastructure Management

Vittorio Scolamiero [1-2], Piero Boccardo [2]

[1] Department of Civil, Building and Environmental Engineering (DICEA), Sapienza Università di Roma, Via Eudossiana, 18, 00184 Roma; vittorio.scolamiero@uniroma1.it

[2] Interuniversity Department of Regional and Urban Studies and Planning (DIST), Polytechnic of Torino, Viale Pier Andrea Mattioli, 39, 10125 Torino; vittorio.scolamiero@polito.it, piero.boccardo@polito.it

Keywords: Urban Digital Twin, Road Infrastructure Monitoring, MMS, ALS, LiDAR, BIM

Abstract

High-quality multi-sensor data are essential for advancing Digital Twins (DTs), AI-based road infrastructure monitoring, and automated urban modelling. Despite growing interest, publicly available datasets that simultaneously integrate mobile mapping systems (MMS), aerial LiDAR (ALS), imagery, BIM-derived assets, and detailed pavement defect annotations within a unified, DT-ready framework remain scarce. This paper presents the Turin Urban Road Dataset, a multi-sensor, multi-layer data framework consolidating research outputs developed within the Turin Digital Twin initiative. The dataset emerges from the integration of previously validated methodologies for pavement condition assessment, BIM-based asset modelling, and point cloud semantic classification, formalised here into a coherent and reusable resource for urban road infrastructure research. It combines high-density MMS LiDAR (1,100 pts/m² on pavement), ALS LiDAR (30–40 pts/m²), RGB/NIR and 360° panoramic imagery, georeferenced BIM entities, and over 7,000 manually classified pavement defects across three representative urban scenes totalling approximately 200 million annotated points. All data sources are harmonised in EPSG:25832 through a quality-controlled pipeline ensuring geometric alignment, semantic coherence, and metadata completeness, consistent with DT data-quality principles. Rather than a fully validated benchmark, the dataset represents a structured and reproducible foundation for future experimentation in semantic segmentation, AI-based defect detection, and DT development, bridging the gap between geospatial acquisition, semantic enrichment, and infrastructure-level decision support.

1. Introduction

Urban road infrastructure represents one of the most critical components of contemporary cities, directly influencing mobility, safety, maintenance costs, and sustainability objectives. The increasing adoption of DT frameworks for urban management has created new opportunities for monitoring, analysing, and predicting infrastructure conditions through data-driven approaches (Batty, 2018; Boje et al., 2020). In this context, the integration of multi-sensor survey technologies — such as MMS, ALS, and high-resolution imagery — enables the generation of dense and detailed 3D representations of urban environments, supporting applications such as semantic segmentation, pavement condition assessment, and asset-level infrastructure modelling. However, the value of these technologies is only fully realised when acquisition, semantic enrichment, and infrastructure representation are organised within a coherent, reproducible data framework that can serve as a foundation for both research and operational DT applications.

Despite these technological advances, the development of robust data-driven approaches for road infrastructure management remains strongly dependent on the availability of structured and domain-specific datasets. While several large-scale urban datasets have been introduced in recent years, such as Cityscapes (Cordts et al., 2016), Semantic3D (Hackel et al., 2017), and TUM2TWIN (Wysocki et al., 2026), Santiago Urban Dataset SUD (Gonzalez-Collazo et al., 2024), their primary focus is on general urban scene understanding rather than road infrastructure monitoring. Conversely, domain-specific datasets, such as Camhighways: the Cambridge highways dataset (d'Avigneau et al., 2025), are mainly oriented toward highway environments and do not fully capture the semantic and geometric complexity of dense urban road systems. In addition, recent studies on DT implementations highlight the importance of integrating geometric accuracy with consistent semantic structuring and reproducible data organization (Fuller et al., 2020). However, these principles are still only partially addressed in existing road-focused 3D datasets, which often lack explicit connections between survey data, infrastructure elements, and condition-related information.

To address this gap, this paper presents the Turin Urban Road Dataset, a multi-sensor, multi-layer data framework for urban road infrastructure monitoring and DT-oriented applications, developed within the Turin Digital Twin initiative (La Riccia et al., 2024). Rather than a dataset designed from scratch as a standalone benchmark, it represents the consolidation of research outputs accumulated across a series of methodological studies by the authors (Scolamiero et al., 2025; Scolamiero et al., 2026), covering pavement condition assessment, BIM-based asset modelling, and point cloud semantic classification. By formalising these outputs into a unified and structured resource, the present work aims to make the underlying data infrastructure available and reusable for the wider research community. The dataset integrates high-density MMS point clouds with ALS data to combine detailed street-level geometry with large-scale contextual information. Radiometric content is provided through nadir and 360° imagery, supporting the identification of surface conditions and urban features. A key component is the inclusion of more than 7,000 georeferenced pavement defects, described through geometric and semantic attributes, and spatially linked to BIM entities representing road infrastructure elements such as pavement surfaces, poles, lighting systems, and traffic signals. This integration enables a direct connection between observed surface degradation and the corresponding infrastructure components within a DT-compatible data model.

A subset of three representative urban scenes, totaling approximately 200 million manually labelled points, has been annotated using a consistent semantic classification scheme including ground, buildings, vegetation, vehicles, and vertical poles. This annotated subset provides a reference basis for future applications in semantic segmentation and infrastructure analysis.

While the dataset is still under active development, this contribution formalises its underlying structure, semantic organisation, and integration strategy, establishing the methodological and data foundation upon which future extensions and validation studies will build. By combining multi-sensor data, pavement condition information, and BIM-based representations within a unified and interoperable framework, the Turin Urban Road Dataset aims to provide a structured and reproducible resource at the intersection of urban road monitoring, geospatial data science, and operational DT systems.

1.1 Objectives

The primary goal of this research is to develop a structured, application-oriented dataset for urban road infrastructure that bridges high-resolution reality capture data with semantic modelling and DT representations. To this end, the work pursues four main objectives.

The first objective is the integration of multi-source geospatial data within a common spatial framework. MMS point clouds, ALS data, and image-based information are combined in a shared coordinate reference system (EPSG:25832), ensuring geometric consistency and enabling joint visualisation, interpretation, and semantic processing while preserving the native resolution and acquisition geometry of each source.

The second objective is the definition of a semantic classification scheme tailored to urban road environments. The ontology focuses on elements directly relevant to road infrastructure analysis — ground surfaces, buildings, vegetation, vehicles, and vertical poles — with poles further structured into subclasses including lighting systems, traffic signals, and road signs. Manual annotation of representative scenes using CloudCompare ensures consistency and usability for downstream segmentation and asset extraction tasks.

The third objective, and the core contribution of this work, is the integration of pavement condition information and BIM entities within the same data framework. More than 7,000 pavement defects are stored as georeferenced vector data enriched with geometric and semantic attributes, and spatially linked to BIM elements representing road infrastructure components. This establishes a direct connection between observed surface degradation and the corresponding infrastructure assets, supporting DT-oriented condition-aware modelling.

The fourth objective is the development of a structured resource that can support future research in urban road modelling, semantic segmentation, and infrastructure monitoring. A subset of three representative urban scenes (figure 1) is manually annotated to provide high-quality labelled data for future experimentation in AI-based classification, defect detection, and BIM integration, rather than constituting a fully validated benchmark in terms of model performance.

Overall, the proposed approach focuses on bridging the gap between raw geospatial data and semantically enriched infrastructure representations, providing a structured and reproducible foundation for future developments toward fully operational DT for urban road systems.

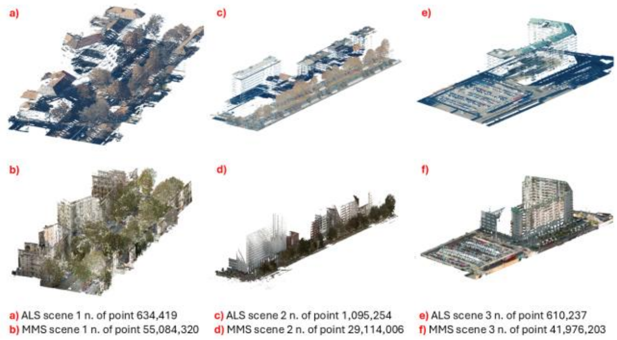

Figure 1. Subset of representative urban scenes included in the dataset. A) Dense built-up area; B) Vegetated corridor; C) Mixed residential-commercial area.

2. Related work

Recent years have seen the development of large-scale datasets supporting semantic segmentation and 3D urban understanding, driven by advances in computer vision and geospatial data acquisition. Early benchmarks such as Cityscapes focused on 2D image-based scene understanding, while subsequent datasets extended these concepts to 3D environments using LiDAR data. In this context, Semantic3D, SemanticKITTI (Behley et al., 2019), Paris-Lille-3D (Roynard et al., 2018), and Toronto-3D (Tan et al., 2020) have become widely adopted benchmarks for point cloud semantic segmentation, providing large-scale annotated datasets for urban classification tasks.

These datasets significantly advanced research in 3D semantic segmentation, particularly in the context of autonomous driving and general urban scene understanding. However, their primary focus remains on object-level classification rather than infrastructure-oriented analysis. In particular, road pavement condition, surface degradation, and the relationship between observed data and infrastructure assets are not explicitly addressed.

More recent efforts have explored the integration of multimodal data and DT concepts. The SUD combines handheld and mobile laser scanning acquisitions to provide detailed urban semantic labelling, with a focus on street-level elements. Similarly, TUM2TWIN introduces a multimodal dataset integrating LiDAR, imagery, and DT representations to support data-driven urban analysis. While these approaches move toward richer data integration, they are still primarily oriented toward general urban modelling rather than road infrastructure monitoring.

In parallel, research on road infrastructure has produced datasets and methodologies specifically targeting asset monitoring and roadway analysis. Among these, Camhighways: the Cambridge highways dataset, provides a structured dataset for highway environments, including semantic annotation of road elements and infrastructure components. This work demonstrates the potential of structured semantic classification for infrastructure management applications, particularly in controlled road scenarios.

However, highway-focused datasets differ significantly from dense urban environments, where higher object density, frequent occlusions, vegetation, and mixed-use infrastructure introduce additional semantic and geometric complexity. As a result, methods and datasets developed for highway contexts are not directly transferable to urban road systems.

Furthermore, recent studies on DT implementations emphasize the importance of semantic consistency, data traceability, and integration across heterogeneous data sources (Weil et al., 2023). Nevertheless, these principles are only partially reflected in existing datasets, which rarely provide explicit links between survey data, condition assessment, and structured infrastructure models.

Previous work by the authors has explored the integration of MMS data and pavement condition assessment within a BIM-based DT framework, demonstrating the potential of combining geometric and semantic information for urban road monitoring (Scolamiero et al., 2025 and Scolamiero et al., 2026). However, these approaches were primarily focused on methodology development rather than on the creation of a structured and reusable dataset. The present work builds upon these foundations by formalizing a comprehensive dataset designed to support reproducibility, future validation, and large-scale analysis.

Overall, despite the availability of large-scale urban benchmarks and emerging infrastructure-oriented datasets, there remains a lack of integrated resources specifically designed for urban road systems. In particular, existing datasets do not simultaneously combine high-density MMS and ALS point clouds, detailed pavement defect annotations, semantically structured urban assets, and BIM-linked infrastructure representations within a unified framework.

The Turin Urban Road Dataset is developed to address this gap by providing a multi-sensor, multi-layer dataset explicitly tailored to urban road infrastructure monitoring and DT applications. By integrating geometric data, semantic classification, pavement condition information, and BIM entities, the proposed dataset aims to support advanced analysis and decision-making processes for next-generation urban road management systems (figure 2).

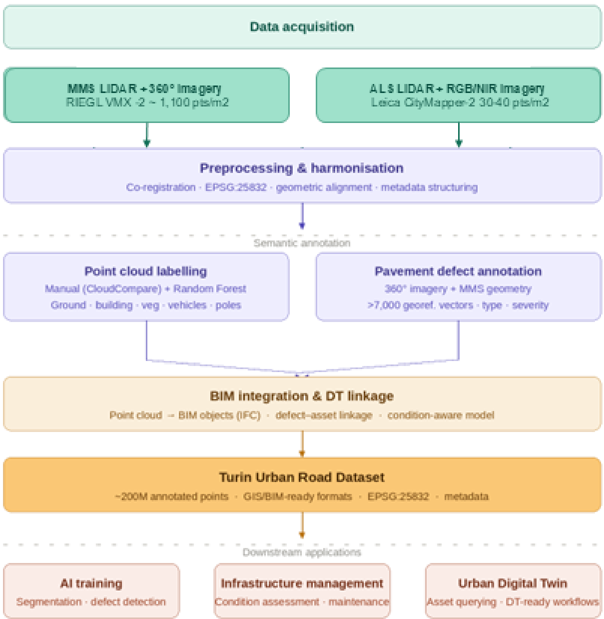

Figure 2. Overview of the Turin Urban Road Dataset development framework. The pipeline integrates two acquisition sources. Two parallel annotation workflows produce semantically labelled point clouds and a georeferenced pavement defect catalogue. These outputs are integrated through a BIM linkage layer that establishes explicit connections between survey data, semantic classification, and parametric infrastructure objects.

3. Dataset development framework

The dataset covers selected road segments in the city of Turin, representing heterogeneous urban morphologies including dense built-up areas, vegetated corridors, and mixed residential–commercial zones. The dataset integrates multiple complementary sources of information, combining ground-based observations, aerial measurements, and infrastructure models to support both geometric reconstruction and infrastructure monitoring tasks. In particular, the dataset includes MMS LiDAR-derived point clouds and panoramic RGB imagery, ALS point clouds together with RGB and NIR aerial imagery, BIM/IFC models describing road geometry and material attributes, and defect shapefiles representing both manually and automatically detected pavement anomalies.

Ground-level data were acquired in 2023 using a RIEGL VMX-2 mobile mapping platform (figure 3). The system integrates dual LiDAR scanners, GNSS positioning, and an inertial measurement unit (IMU), enabling the generation of highly accurate georeferenced point clouds. The acquisition produced dense point clouds with an average density of approximately 1,100 points/m² on the pavement surface, recorded at an acquisition speed of 80 km/h. Under differential GNSS correction, the system provides centimetre-level absolute positioning accuracy, while the resulting point cloud exhibits sufficient geometric fidelity to capture surface discontinuities and small-scale pavement defects. Simultaneously acquired 360° panoramic RGB imagery complements the LiDAR data and supports visual inspection and semantic labelling.

Aerial data were collected in 2022 using a Leica CityMapper-2 system (figure 4). The ALS dataset provides a nominal density of approximately 30–40 points/m² with a ground sampling distance (GSD) of about 5 cm, ensuring continuous coverage of the urban road corridor and the surrounding built environment. Nadir RGB and NIR images were acquired simultaneously and provide additional spectral information for contextual interpretation. Although ALS data do not provide sufficient spatial resolution for fine-scale pavement defect detection, they ensure large-scale geometric continuity and contextual integration within the urban DT framework.

Figure 3. The MMS RIEGL VMX-2 mounted on the vehicle for ground-based acquisition.

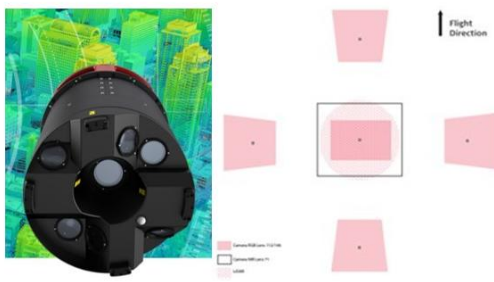

Figure 4. The airborne sensor Leica City-Mapper 2.

https://doi.org/10.5194/isprs-archives-XLIX-B1-2026-565-2026

3.1 Semantic ontology definition

The semantic ontology underlying the Turin Urban Road Dataset was designed to ensure structured, reproducible, and interoperable annotation of urban road environments. The annotation workflow employs hybrid manual–automatic labelling strategies supported by strict annotation guidelines and multi-annotator validation procedures.
Following these practices, the ontology was developed according to three guiding principles. First, infrastructure relevance was prioritized, meaning that semantic classes were defined to support road infrastructure monitoring rather than generic urban scene understanding. Second, hierarchical clarity was introduced to allow semantically related elements to be aggregated or separated depending on analytical needs. Finally, DT compatibility was considered to ensure that semantic entities can be directly linked to BIM objects and asset-level infrastructure representations.
Unlike generic urban segmentation datasets, the ontology therefore focuses on elements directly relevant to road management, including pavement surfaces, vegetation, and roadside infrastructure.
Two complementary annotation layers were implemented. The first concerns semantic labelling of point clouds, where high-density MMS and ALS point clouds were manually labelled according to a predefined schema including ground and road surfaces, buildings, vegetation, vehicles, vertical poles and unclassified elements. This classification supports scene-level understanding, AI-based semantic segmentation, and spatial filtering of dynamic or irrelevant objects. Manual labelling was conducted using CloudCompare, following explicit annotation guidelines designed to minimize ambiguity.
The second annotation layer focuses on pavement defects and operates independently from the general semantic segmentation. In this case, defects are digitized as georeferenced polygons enriched with geometric and semantic attributes, forming a condition-aware data layer suitable for infrastructure condition assessment, AI-based defect detection, and BIM-based infrastructure management. To ensure reproducibility, class definitions and annotation criteria were explicitly documented. The semantic schema aligns with ASPRS LAS classification standards whenever possible, extending classification codes only when domain-specific categories such as vehicles or poles are required. The ontology was also designed to remain scalable, allowing future refinements and the introduction of additional subclasses without affecting the existing hierarchical structure.

3.2 Manual labelling protocol

The semantic labelling of the point clouds was performed manually following a structured annotation protocol designed to ensure consistency and reproducibility across the dataset.
The annotation process was conducted directly on the point clouds using CloudCompare, which provides interactive tools for point selection, visualization, and class assignment. Unlike automated segmentation-based workflows, the labelling procedure was intentionally performed on the original point clouds in order to preserve the native geometric structure and point density of the dataset.
During the annotation process, the point clouds were visually inspected and labelled using a combination of manual selection tools and attribute-based visualization. Color information, point elevation, and intensity values were used to facilitate the identification of different urban elements. For example, height information helped distinguish between ground-level objects and elevated structures such as roofs or poles, while geometric patterns and spatial context supported the identification of vegetation layers and vertical urban elements.
Points were manually assigned to semantic classes through iterative selection and validation steps. CloudCompare tools such as polygonal selection, segmentation, and manual refinement were used to isolate groups of points belonging to specific objects or surfaces. This interactive workflow allowed annotators to progressively refine class boundaries and resolve ambiguities caused by occlusions, irregular sampling density, or overlapping structures.
The final semantic ontology includes the following classes: ground, vertical poles, static and dynamic vehicles, vegetation and building (figure 5, table 1-2). Semantic labels were stored using the LAS 1.4 classification framework to ensure compatibility with standard GIS, BIM, and machine learning processing pipelines.
To improve annotation reliability, the labelling process included iterative verification steps in which labelled scenes were reviewed and corrected when necessary. This procedure helped reduce subjective interpretation and ensured that the final dataset maintains a consistent semantic structure across different urban environments and acquisition conditions.

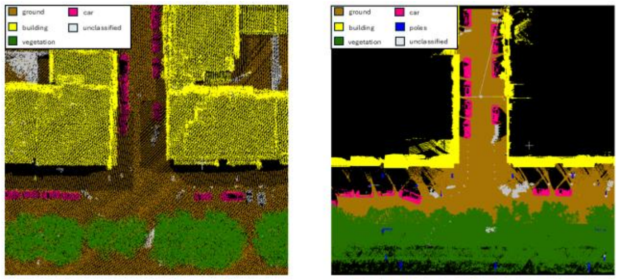

Figure 5. Top-down view of a manually labelled point cloud segment, illustrating the semantic annotation of urban elements.

ALS Classes scene 1	n. of point	%
ground	241,178	38.02
building	171,063	26.96
vegetation	197,826	31.18
car	10,929	1.72
ALS Classes scene 2		
ground	484,840	44.27
building	356,342	32.54
vegetation	208,648	19.05
car	8,031	0.73
ALS Classes scene 3		
ground	400,390	65.61
building	148,931	24.41
vegetation	3,874	0.63
car	33,267	5.45

Table 1. Distribution of semantic classes in ALS point clouds across three representative urban scenes.

MMS Classes scene 1	n. of point	%
ground	1,709,823	33.51
building	13,562,519	24.63
vegetation	17,992,463	32.66
car	533,222	0.97
poles	1,678,869	3.05
MMS Classes scene 2		
ground	8,675,896	29.80
building	8,172,502	28.07
vegetation	10,026,363	34.43
car	518,107	1.78
poles	565,333	1.94

https://doi.org/10.5194/isprs-archives-XLIX-B1-2026-565-2026

MMS Classes scene 3		
ground	19,610,678	46.74
building	13,947,422	33.24
vegetation	2,980,056	7.09
car	527,328	1.26
poles	4,331,181	10.32

Table 2. Distribution of semantic classes in MMS point clouds across three representative urban scenes.

3.3 Point Cloud supervised classification

To automatically classify the LiDAR point clouds, a supervised machine learning workflow based on multi-scale geometric feature extraction and a Random Forest classifier was implemented, following established approaches for geometry-based urban point cloud classification (Weinmann et al., 2015). The method was designed to operate on both MMS and ALS data, enabling the classification of heterogeneous urban point clouds characterized by different acquisition geometries, densities, and viewpoints.

The training data consisted of manually labelled scenes, where each point was assigned to a predefined semantic class, including ground, building, poles, vehicles, and unclassified elements. These annotated datasets were used to train the classification model and define the feature space.

For each point, a set of geometric features was computed from its local spatial neighborhood using a multi-scale approach. Neighborhoods were defined through spherical queries at different radii within a KD-tree spatial structure. For each neighborhood, the covariance matrix was computed and decomposed into eigenvalues, which were then used to derive geometric descriptors such as linearity, planarity, and scattering. These descriptors characterize the local spatial organization of points, enabling the discrimination between linear structures, planar surfaces, and volumetric distributions.

Additional attributes were incorporated to improve classification performance. These include the absolute elevation (Z coordinate), local height variation, and LiDAR return intensity. Such features are particularly effective in distinguishing between ground and roof surfaces, identifying irregular structures such as vegetation, and detecting vertically developed elements like poles.

The same workflow was applied to ALS data, with minor adaptations to account for its lower point density and top-down acquisition geometry. In this case, geometric features were computed using larger neighborhood radii in order to capture meaningful spatial structures. The classification focused on a reduced set of classes (ground, building, vegetation, vehicles, and unclassified).

The experiments highlighted the strong influence of class distribution across different scenes. Significant variations were observed, particularly for vegetation, which ranged from less than 1% to over 30% depending on the tile. This variability directly impacted classification performance, especially for minority classes. Additional experiments conducted by swapping training and testing scenes demonstrated that the model is capable of accurately learning all classes, but its performance is highly dependent on their representation in the test data. This confirms the presence of dataset shift and emphasizes the importance of spatial representativeness in supervised point cloud classification.

Classification performance was evaluated using standard metrics, including precision, recall, and F1-score for each class, along with overall accuracy. Finally, the trained models were applied to previously unseen point clouds to generate automatically labelled datasets. The predicted semantic labels were written back into the LiDAR files, producing semantically enriched point clouds that can be directly used for further analysis and integration into infrastructure monitoring workflows and DT environments.

3.4 Validation of semantic classification

The enhanced classification framework, integrating multi-scale feature extraction and additional geometric descriptors, leads to a substantial improvement in performance on MMS data. The model achieves an overall accuracy of 66% and a macro-averaged F1-score of 0.62. These metrics capture complementary aspects of performance: overall accuracy reflects the proportion of correctly classified points across all classes and is influenced by class frequency, whereas the macro-averaged F1-score assigns equal weight to each class, making it more sensitive to performance on underrepresented categories such as vehicles and poles. Together, these results indicate that the proposed feature set effectively captures the geometric complexity of urban environments while maintaining balanced performance across classes.

High performance is observed for well-defined geometric classes such as buildings (F1-score of 0.80), vegetation (0.76), and poles (0.75), demonstrating the capability of the multi-scale approach to characterize both planar and vertically developed structures. Vehicle detection also shows strong reliability, with very high precision (0.97) and an F1-score of 0.77, indicating that small but geometrically distinct objects can be effectively identified in high-density MMS data. Ground surfaces are classified with high recall (0.97), although some confusion with adjacent classes persists, particularly with low vegetation. This reflects a known limitation of geometry-based approaches, where classes with similar local structure remain difficult to separate without additional radiometric or contextual information.

The application of the same methodology to ALS data highlights both its robustness and its limitations when dealing with different acquisition modalities. The ALS-based classification achieves an overall accuracy of up to 77% and a weighted F1-score of 0.77 when evaluated on an independent test tile. Ground surfaces are consistently well identified (F1-score of 0.82), confirming the suitability of airborne LiDAR for terrain-related tasks, while building classification shows stable performance (F1-score between 0.65 and 0.69), despite the reduced representation of vertical structures due to the top-down acquisition geometry.

To assess the impact of dataset composition, two training/testing configurations were considered. In the Scenario 1 , Scenes 1 and 3 were used for training, while Scene 2 was used for testing. In this case, where vegetation is well represented in the test data (approximately 19%), the model achieves a very high F1-score of 0.94 for this class. However, vehicle detection performance is significantly lower (F1-score of 0.28), due to the limited presence of cars in the test set (less than 1%).

In contrast, in the alternative configuration Scenario 2 (Scenes 1 and 2 used for training, Scene 3 for testing), the class distribution is effectively reversed. Here, vehicle detection improves substantially (F1-score of 0.60), while vegetation performance drops dramatically (F1-score close to 0.01) due to its extremely limited representation in the test data. This clearly demonstrates that the model is capable of learning both vegetation and vehicle features, and that performance degradation is primarily driven by class imbalance and distribution mismatch rather than intrinsic model limitations.

These results highlight the presence of dataset shift and confirm that classification performance in LiDAR data is strongly

dependent on the spatial distribution and relative abundance of semantic classes.
Overall, the comparison between MMS and ALS results emphasizes the complementary nature of the two data sources. MMS data enables detailed classification of small and vertically structured objects, while ALS provides more robust large-scale classification, particularly for ground and building classes. The proposed methodology demonstrates good adaptability across both domains, while also underlining the importance of representative training data for achieving reliable generalization.
Detailed class-level performance metrics for the two ALS training/testing configurations are reported in Tables 3 and 4, while the corresponding MMS classification results are summarized in Table 5.
These findings support the use of the proposed workflow as a scalable and transferable solution for semantic enrichment of urban point clouds, with direct applications in infrastructure monitoring and DT environments.

ALS class scenario 1	**Precision**	**Recall**	**F1-score**
Ground	0.79	0.85	0.82
Building	0.79	0.61	0.69
vegetation	0.91	0.96	0.94
car	0.17	0.77	0.28
Accuracy			0.77
Macro avg	0.57	0.68	0.58
Weighted avg	0.79	0.77	0.77

Table 3. Classification performance metrics (precision, recall, F1-score) for ALS point clouds in Scenario 1.

ALS class scenario 2	**Precision**	**Recall**	**F1-score**
Ground	0.93	0.74	0.83
Building	0.60	0.72	0.65
vegetation	0.01	0.01	0.01
car	0.48	0.79	0.60
Accuracy			0.72
Macro avg	0.44	0.52	0.46
Weighted avg	0.79	0.72	0.74

Table 4. Classification performance metrics (precision, recall, F1-score) for ALS point clouds in Scenario 2.

MMS class	**Precision**	**Recall**	**F1-score**
Ground	0.46	0.97	0.62
Building	0.72	0.90	0.80
vegetation	0.85	0.69	0.76
car	0.97	0.64	0.77
poles	0.76	0.75	0.75
Accuracy			0.62
Macro avg	0.62	0.66	0.62
Weighted avg	0.62	0.66	0.62

Table 5. Classification performance metrics (precision, recall, F1-score) for MMS point clouds.

3.5 Damage assessment dataset

Beyond semantic scene classification, the proposed dataset provides a multi-layered resource designed to support both infrastructure monitoring and DT applications. It integrates high-density MMS point clouds with ALS data, combining detailed local geometry with broader contextual information, and enabling a unified representation of pavement condition through the joint use of radiometric and geometric features.
Building upon previous work by the authors (Scolamiero et al, 2025), which demonstrated the effectiveness of multi-sensor data integration for pavement condition assessment, the proposed dataset formalizes this approach into a structured and reusable ground-truth resource.
The identification and classification of pavement distresses were performed through a combination of radiometric analysis of 360° imagery and geometric analysis of MMS-derived point clouds. Image-based inspection supported the detection of surface-level anomalies such as cracks and material degradation, while point cloud processing enabled the extraction of depth-related features for volumetric defects, including rutting and depressions. Detected anomalies were subsequently vectorized and classified according to standardized distress categories, ensuring consistency and reproducibility of the annotation process. This hybrid approach enables a more robust and unambiguous characterization of pavement distresses by leveraging complementary data sources.
Each detected defect is described through geometric descriptors, including spatial footprint and depth, and semantic attributes specifying damage type and severity (figure 6). The defects are stored in georeferenced vector formats compatible with both GIS and BIM environments, resulting in a dataset comprising more than 7,000 annotated anomalies, including cracks, potholes, repairs, rutting, bumps, corrugation, and depressions.
The dataset is spatially aligned with MMS point clouds, ALS data, and BIM entities representing road infrastructure components, enabling a direct association between observed surface anomalies and the corresponding assets. In this way, pavement condition information is not treated as an external annotation layer, but as an integral component of a structured infrastructure model. This integration supports multiple downstream applications, including the development of AI-based defect detection models, automated computation of pavement condition indices, and spatially explicit maintenance prioritization strategies. While automated detection models have not yet been quantitatively evaluated in terms of precision, recall, or F1-score, the dataset provides a robust foundation for future AI-driven validation and infrastructure monitoring within a DT framework.

Figure 6. Visualization of the damage dataset. From left to right: LiDAR point cloud, panoramic image, and attribute table of the damage vectors. Red vectors represent the identified defects, while blue markers indicate the positions of the panoramic images utilized in the analysis.

3.6 BIM entities and DT linkage

A key contribution of the proposed dataset lies in the integration of survey-derived data with structured BIM entities, enabling a coherent linkage between reality-based measurements and semantic infrastructure modeling within a DT environment. Following semantic classification of point clouds, selected urban elements are abstracted into BIM objects representing real-world infrastructure components. Unlike point clouds, which primarily encode geometric information, BIM entities encapsulate both geometry and semantics, including material properties, functional roles, condition states, and potential maintenance actions, thus supporting a higher level of abstraction suitable for infrastructure management.

The BIM model includes pavement surfaces reconstructed from MMS data, urban street elements such as poles and lighting structures, road markings, and pavement defect geometries derived from georeferenced vector data. The pole class has been further structured into specific subclasses, including lighting systems, traffic signals, and vertical road signs, extending previous classification approaches toward infrastructure-oriented BIM representation. Centroids associated with these elements were derived from segmented point cloud clusters (figure 7), following the methodology proposed in previous work by the authors (Scolamiero et al., 2026). These centroids were then used as reference points for the placement and georeferencing of parametric BIM objects (figure 8), ensuring both geometric consistency and semantic alignment between survey data and infrastructure representation.

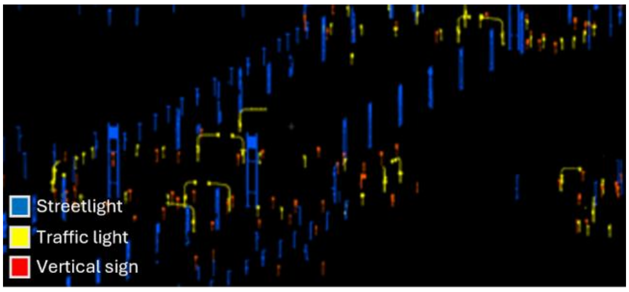

Figure 7. The image shows the further subdivision of the vertical poles class into 3 main sub-classes: blue – streetlight, yellow – traffic light, red – vertical sign.

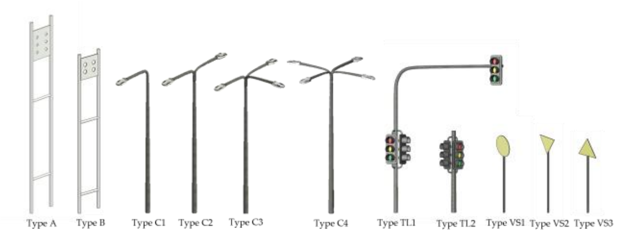

Figure 8.Abacus of vertical element families. Types A, B, C1, C2, C3, and C4 correspond to streetlight elements; types TL1 and TL2 correspond to traffic light elements; and types VS1, VS2, and VS3 correspond to vertical sign elements. Specifically, VS1 represents mandatory and prohibitory road signs, VS2 corresponds to right-of-way signs, and VS3 refers to warning signs.

The explicit linkage between BIM entities and more than 7,000 pavement defects transforms the model from a purely geometric representation into a condition-aware and semantically enriched infrastructure system.

The BIM structure reflects the dual nature of infrastructure modeling, consisting of a geometric layer derived from survey data and an informational layer composed of semantic attributes, condition indicators, and metadata. From a DT perspective, this integration represents a fundamental step toward data-driven infrastructure management, enabling asset-level querying, condition-based analysis, and the potential for automated updates driven by future survey campaigns. Interoperability is ensured through the adoption of a consistent spatial reference system across all data sources, standardized semantic classifications compatible with IFC-based modeling, a clear separation between raw survey data and structured asset representations, and comprehensive metadata documentation. Although the current implementation represents an initial stage toward a fully operational DT, these foundations enable future developments such as real-time updates, bidirectional synchronization, and increased automation of DT workflows.

This structured integration enables a direct linkage between observed data, semantic classification, and infrastructure representation, which is a key requirement for scalable and operational DT systems.

4. Results and Discussion

This section discusses the results achieved across the different components of the Turin Urban Road Dataset, interpreting their significance in relation to the objectives stated in the introduction and to comparable initiatives in the literature. The aim is to evaluate the maturity and relevance of the proposed framework, identify its current limitations, and position it within the broader context of urban DT research.

The classification experiments conducted on both MMS and ALS data confirm the suitability of the multi-scale Random Forest workflow for urban point cloud analysis, while also revealing clear dependencies on training data composition. On MMS data, overall accuracy reached 66% with a macro-averaged F1-score of 0.62, with particularly strong results for buildings (F1 = 0.80), vegetation (0.76), and poles (0.75). These figures are broadly consistent with comparable geometry-based classification approaches reported in the literature for dense urban environments, where class overlap and occlusion remain persistent challenges. On ALS data, overall accuracy reached up to 77%, with ground (F1 = 0.82) and vegetation performing strongly in favourable conditions, while vehicle detection degraded sharply when cars were underrepresented in the test set (F1 = 0.28 in Scenario 1 vs. 0.60 in Scenario 2). This sensitivity to class distribution confirms the presence of dataset shift and underlines a known limitation of supervised approaches applied to spatially heterogeneous environments. Rather than representing a weakness specific to the proposed workflow, this behaviour reflects a general challenge for the community and reinforces the need for diverse, geographically representative training data — a gap the Turin Urban Road Dataset is explicitly designed to help address as it expands.

Compared to existing resources, the proposed dataset occupies a distinct position. General-purpose urban benchmarks such as SemanticKITTI, Paris-Lille-3D, and Toronto-3D provide large-scale annotated data for autonomous driving and scene understanding, but do not incorporate pavement condition information or explicit links to infrastructure models. More recent initiatives such as TUM2TWIN and SUD extend toward richer multi-modal integration and DT-oriented representation, yet remain primarily focused on general urban modelling rather than road-specific infrastructure monitoring. Camhighways, conversely, addresses road infrastructure in a structured and semantically explicit manner but is oriented toward highway environments, where acquisition geometry, object density, and management requirements differ substantially from dense urban contexts. The Turin Urban Road Dataset addresses a specific combination of requirements — high-density MMS acquisition, airborne integration, pavement defect annotation, and BIM linkage — that is not simultaneously satisfied by any of the above resources.

The BIM integration layer represents the most distinctive contribution of the current framework. While point cloud datasets typically encode geometric information without reference to asset identity or condition, the proposed approach establishes an explicit linkage between survey-derived geometry, semantic classification, pavement defect records, and parametric BIM objects. This connection enables asset-level querying and condition-aware analysis that goes beyond what is achievable from annotated point clouds alone. From a DT perspective, it constitutes a foundational step toward operational infrastructure management: maintenance teams can, in principle, query specific assets, retrieve their associated condition indicators, and prioritise interventions based on spatially explicit evidence. Although automated workflows for real-time updating and bidirectional synchronisation remain a target for future development, the data architecture established here is explicitly designed to support such extensions.

It is important to contextualise the current state of the dataset accurately. The Turin Urban Road Dataset is an actively developing resource: the annotated subset covers three representative urban scenes totalling approximately 200 million labelled points, and the defect catalogue currently includes over 7,000 georeferenced anomalies. While these figures are substantial and sufficient to support initial experimentation in semantic segmentation and defect detection, the dataset does not yet constitute a fully validated benchmark in the traditional sense — that is, a fixed, publicly released resource with standardised evaluation splits and reported baseline model performance. The present contribution formalises the dataset's structure, integration strategy, and semantic organisation, establishing the methodological foundation upon which future benchmark formalisation will build. This framing is consistent with established practice in the community, where foundational dataset papers such as Toronto-3D and SUD precede full benchmark maturation and serve a different but equally important purpose: enabling the community to understand, adopt, and extend a new data infrastructure.

Confusion matrices (figure 9) illustrate both the complementarity of MMS and ALS acquisitions and the scene-dependent nature of classification performance. ALS data excel in terrain and vegetation characterisation owing to their top-down geometry, while MMS data better resolve vertically structured elements such as poles and building facades. These patterns suggest that multi-sensor fusion strategies — exploiting the geometric specificity of each modality — represent a natural direction for improving classification robustness, particularly for classes such as low vegetation or parked vehicles that remain challenging across both acquisition types.

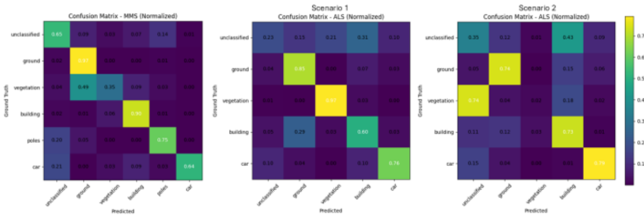

Figure 9. Normalized confusion matrices for ALS (Scenario 1 and Scenario 2) and MMS datasets. The results highlight the influence of acquisition geometry on classification performance. ALS shows strong performance in vegetation detection due to its top-down perspective, while MMS better captures structured urban elements such as buildings and poles.

5. Conclusion and future work

This research presents the Turin Urban Road Dataset as a replicable and scalable framework for urban road infrastructure management within a DT perspective. By integrating MMS and ALS point clouds, high-resolution imagery, georeferenced pavement defects, and BIM-based semantic entities, the dataset provides a structured and semantically rich representation of urban roads that bridges raw survey data with actionable infrastructure models.

The dataset uniquely combines high-density 3D data with detailed pavement condition annotations and BIM-linked assets, supporting semantic segmentation, defect detection, and infrastructure monitoring within a unified, interoperable framework. By documenting acquisition parameters, semantic definitions, and processing workflows, the dataset adheres to DT principles of geometric accuracy, semantic coherence, and traceable metadata.

Future work will focus on extending the dataset to larger urban areas and enhancing automated annotation strategies, integrating machine learning pipelines within human-in-the-loop frameworks to reduce manual effort while maintaining high semantic accuracy. Further developments include refining semantic stratification, such as differentiating vegetation layers and road infrastructure elements by function. Advances in AI-driven 3D point cloud segmentation and multi-sensor data fusion are expected to improve the discrimination of complex urban classes, increasing the dataset's applicability for DT workflows.

Ultimately, this dataset establishes a robust foundation for research and operational applications, enabling hybrid use of MMS and ALS data, seamless BIM integration, and condition-aware analysis of urban roads. By providing a high-quality, multi-layered and semantically structured resource, it supports the development of scalable, interoperable, and intelligent infrastructure management systems capable of bridging geospatial data acquisition and real-world decision-making.

Data Availability Statement

The Turin Urban Road Dataset described in this paper is currently under active development and is not yet publicly released as a standalone repository. Given the sensitivity of certain data components and the ongoing nature of the Turin Digital Twin initiative, full open access will be evaluated upon dataset consolidation. In the interim, the dataset, or representative subsets thereof, may be made available to researchers upon reasonable request to the authors, subject to appropriate data sharing agreements and institutional approval.

Requests should specify the intended research application and the data components required.

References

Batty, M. (2018). Digital twins. *Environment and planning B: Urban analytics and City Science*, *45*(5), 817-820.

Behley, J., Garbade, M., Milioto, A., Quenzel, J., Behnke, S., Stachniss, C., & Gall, J. (2019). Semantickitti: A dataset for semantic scene understanding of lidar sequences. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 9297-9307).

Boje, C., Guerriero, A., Kubicki, S., & Rezgui, Y. (2020). Towards a semantic Construction Digital Twin: Directions for future research. *Automation in construction*, *114*, 103179.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., ... & Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3213-3223).

d'Avigneau, A. M., Potseluyko, L., Anvo, N. R., Taha, H. M., Reja, V. K., Davletshina, D., ... & Brilakis, I. (2025).

Camhighways: The cambridge highways dataset. *Advanced Engineering Informatics*, *64*, 103036.

Fuller, A., Fan, Z., Day, C., & Barlow, C. (2020). Digital twin: Enabling technologies, challenges and open research. *IEEE access*, *8*, 108952-108971.

Gonzalez-Collazo, S. M., Balado, J., Garrido, I., Grandio, J., Rashdi, R., Tsiranidou, E., ... & Lorenzo, H. (2024). Santiago urban dataset SUD: Combination of Handheld and Mobile Laser Scanning point clouds. *Expert Systems with Applications*, *238*, 121842.

Hackel, T., Savinov, N., Ladicky, L., Wegner, J. D., Schindler, K., & Pollefeys, M. (2017). Semantic3d. net: A new large-scale point cloud classification benchmark. *arXiv preprint arXiv:1704.03847*.

La Riccia, L., Scolamiero, V., Yadav, Y., & Eusebio, A. (2024). Turin digital twins: Initiatives and challenges. *Trends Earth Obs*, *16*, 16-20.

Roynard, X., Deschaud, J. E., & Goulette, F. (2018). Paris-Lille-3D: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification. *The International Journal of Robotics Research*, *37*(6), 545-557.

Scolamiero, V., Boccardo, P., & La Riccia, L. (2025). Mobile Mapping System for Urban Infrastructure Monitoring: Digital Twin Implementation in Road Asset Management. *Land*, *14*(3), 597.

Scolamiero, V., & Boccardo, P. (2026). A BIM-Based Digital Twin Framework for Urban Roads: Integrating MMS and Municipal Geospatial Data for AI-Ready Urban Infrastructure Management. *Sensors*, *26*(3), 947.

Tan, W., Qin, N., Ma, L., Li, Y., Du, J., Cai, G., ... & Li, J. (2020). Toronto-3D: A large-scale mobile LiDAR dataset for semantic segmentation of urban roadways. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 202-203).

Weil, C., Bibri, S. E., Longchamp, R., Golay, F., & Alahi, A. (2023). Urban digital twin challenges: A systematic review and perspectives for sustainable smart cities. *Sustainable Cities and Society*, *99*, 104862.

Weinmann, M., Jutzi, B., Hinz, S., & Mallet, C. (2015). Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing*, *105*, 286-304.

Wysocki, O., Schwab, B., Biswanath, M. K., Greza, M., Zhang, Q., Zhu, J., ... & Jutzi, B. (2026). TUM2TWIN: Introducing the large-scale multimodal urban digital twin benchmark dataset. *ISPRS Journal of Photogrammetry and Remote Sensing*, *232*, 810-830.