Road network classification based on street-level images and its machine learning embedding features

Francisco Garrido-Valenzuela*[1], Max Lange[2], Juan C. Herrera[3,4], and Oded Cats[2]

[1]CityAI Lab, Transport and Logistics Group, Delft University of Technology, The Netherlands
[2]CityAI Lab, Transport and Planning, Delft University of Technology, The Netherlands
[3]Department of Transport Engineering and Logistics, Pontificia Universidad Católica de Chile, Chile
[4]Center for Sustainable Urban Development (CEDEUS), Pontificia Universidad Católica de Chile, Chile

*Corresponding author

Short summary

This study introduces an approach for classifying road networks using geo-tagged street-level imagery, combining representation learning and clustering techniques. The method addresses scalability issues of traditional methods by gathering images from Google Street View. The process involves matching street-level images to road sections, extracting image features through a pre-trained computer vision model, and applying clustering to categorize road sections. We apply this approach in Delft, Netherlands, classifying around 2,000 road sections with around 70 thousands images into six clusters, each representing distinct urban typologies. The clusters align with known road categories and reveal clear distinctions, such as residential, arterial, and motorway types. This method offers a scalable and adaptable solution, potentially improving urban planning, mobility studies, and automated vehicle navigation.
Keywords: Clustering, Computer vision, Embedding representation, Road classification, Street-level images.

1 Introduction

Classifying road networks is a core component of the design and planning of transport networks. Provides information on characteristics of the network structure of cities and supports a wide range of practical applications. For example, such classifications can help identify areas that require different driving behaviors for automated vehicles (Chatziioannou et al., 2024) and people (Kaptein & Claessens, 1998); regulate roads to improve potential safety concerns (Dijkstra, 2011); and plan for reduced mobility vehicles, allowing targeted transport project allocation (Gorges et al., 2018). Beyond these practical applications, road classification can influence outcomes like urban expansion and spatial development (Hu & Liang, 2024). These applications highlight the relevance of performing a meaningful road classification that is suitable for a variety of applications.
Current methods for classifying road networks include manual approaches based on rules, statistical algorithms, and automated methods, each offering distinct advantages and limitations. Several countries and cities have developed manuals with rules and algorithms for classifying roads (Tsigdinos et al., 2024). Usually, these classifications rely on individuals exploring the streets and manually collecting the attributes. On the other hand, different approaches have been developed depending on the objectives and priorities set for each city. For example, Tang & Breckon (2010); Mohammadi (2012) developed an image processing technique to incorporate automated terrain recognition into street classification; (Bosurgi et al., 2019) incorporated statistical pre-processing of road attributes to reduce subjectivity in classifications; Zhang et al. (2014) have incorporated some deep learning technique to include dynamic data like traffic flows; and Taamneh et al. (2017) include accident statistics into the classification. More recently, Leśniara & Szymański (2022) developed a method based on representation learning to derive clusters from OpenStreetMap (OSM) road infrastructure data. These diverse methods highlight the ongoing advances in road network classification, showing the need for more comprehensive and flexible strategies to meet the varied needs of modern cities.

Despite their advancements, current road classification methods face significant limitations in scalability, adaptability, integration of new characteristics, or access to comprehensive road data. Many approaches rely on a single attribute, such as terrain or road speed, to define road categories, which restricts their ability to capture the complex and multifaceted nature of road networks. Static and rule-based methods struggle to adapt to additional characteristics as the complexity of defining categories and constructing rules increases exponentially with the number of variables. Similarly, while machine learning-based techniques offer the flexibility to include a diverse range of variables, they often require high-quality datasets that thoroughly represent varied road features. Practical implementations of these methods may involve significant time and cost, mainly when applied to large or dynamic datasets. Consequently, current methods either limit the incorporation of multiple variables or lack the robust datasets needed to support more complex models, underscoring the need for innovative approaches in road network classification.

To address this gap, we propose a representation learning method combined with a clustering technique to perform road classifications based only on street-level imagery. In the field of representation learning, machine learning models are developed to transform various types of data (e.g., text or images) into vector representations known as embeddings (Mikolov, 2013). These embeddings are low-dimensional representations that preserve the essential characteristics and relationships of the original data. Using embeddings eliminates the need for manually defining clustering classes, allowing the model to autonomously group roads based on inherent visual patterns. Additionally, street-level images provide a rich source of information about streets and their surroundings, such as terrain type, urban infrastructure, land use, and vegetation. These images are also widely accessible through platforms like Mapillary, Apple Look Around, Google Street View (GSV), and social media. This approach uses the potential and availability of visual data and offers a scalable and adaptable solution for improving road network classification.

In this study, we develop and implement a method for road network classification using geo-tagged street-level images as the sole input data source. Our approach involves four key steps. First, we match street-level images to specific road sections based on the network topology of the designated area, allowing each section to be represented by a collection of images. Second, we extract features from these images using a pre-trained image embedding model. Third, we combine the image vectors assigned to each road section to generate a unique vector representation. Finally, we perform a clustering process to determine road classes through statistical analysis.

We apply this method in Delft, the Netherlands, to showcase its effectiveness. Specifically, we collect over 70 thousand street-level images from across the city using the approach described by Garrido-Valenzuela et al. (2023). These images are then used to classify around two thousands road sections in Delft. To assess the performance of our classification, we analyze various attributes within the classes, such as road type, vegetation, and surrounding infrastructure. This analysis helps validate our method's robustness and its ability to uncover meaningful distinctions between different road categories. Our findings highlight the potential of using street-level imagery for scalable and adaptable road network classification, paving the way for broader applications in urban planning and mobility studies.

2 Method

This section describes the input data and outlines our method for classifying the street network based on street-level imagery. Figure 1 provides an overview of the pipeline employed in the classification process. To illustrate the method, we include examples from its implementation in Delft, the Netherlands, though the approach is adaptable to any city with street-level images availability. Delft is a historical city in the western part of the Netherlands, covering approximately $24km^2$. Its size and varied urban landscape, featuring streets from different eras, provide diverse street styles, making it appropriate for testing our road classification approach.

Data

Our classification method utilizes street-level images (SLI) to extract visual street information and a Geographic Information System (GIS) file to represent the road network. SLIs are panoramic photographs captured at ground level, providing detailed visual and structural information about street surroundings. Several providers, including Google Street View (GSV), Mapillary (Mapillary, 2023), and Apple Look Around (Apple Inc., 2023), offer these images worldwide. For this study, we use imagery from GSV (Google, 2023), employing the image ID collection method described by

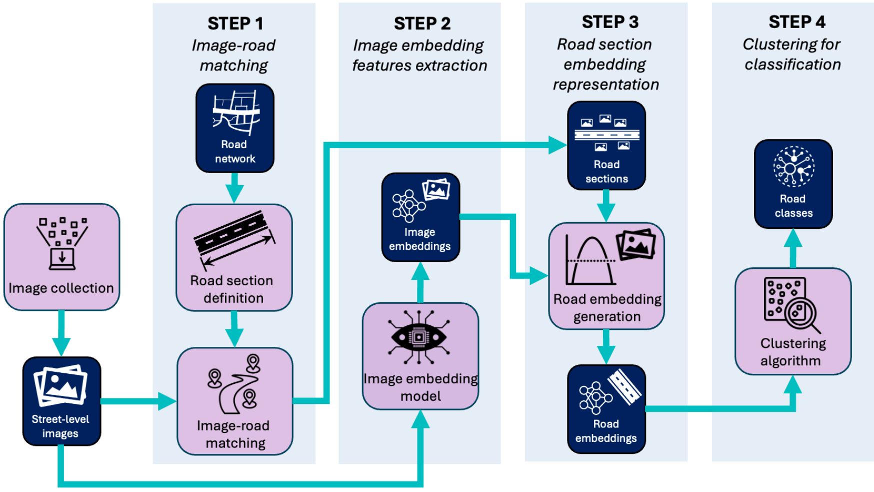

Figure 1: General pipeline of the method.

Garrido-Valenzuela et al. (2023). This approach systematically gathers geo-tagged image IDs across a study area, with each ID corresponding to a specific image and its geographical coordinates. In addition to the SLIs, the GIS road network file is available through OpenStreetMap (OSM, 2024) and it provides the coordinates needed to map and identify each street in the study area. For the city of Delft, more than 70 thousands SLIs and around $300km$ of network are gathered.

Step 1: Image-road matching

The first step involves defining and associating the road section units with the SLIs. A road section unit corresponds to a spatial line segment representing a street or a portion of it. These road sections serve as the minimal spatial units for classification in our method. The definition of a section can vary depending on the application, such as the segment between two intersections, sections based on street names, or fixed-length segments (e.g., 100 meters). Then, the SLIs are spatially associated with the road sections.

In our application for Delft, we define road sections using intersection nodes from the GIS network file. These nodes represent points where streets intersect, serving as natural boundaries for segmenting the road network. This process yields a total of $3,429$ road sections. Once the road sections are defined, we create 20-meter buffers around them to match SLIs from their surroundings. If an image corresponds to multiple sections, it is assigned to the closest one. As a result, $1,914$ road sections are matched with at least one image, and $68,178$ SLIs are used in total. Figure 2 illustrates the coverage, with road sections having at least one associated image in green and sections without images in red.

Step 2: Image embedding features extraction

The second step is designed to extract features from the road sections' SLIs. We use a pre-trained image embedding mode to achieve this, transforming the street-level images into vector representations. These embeddings capture essential visual characteristics of the images, such as textures, colors, structural patterns, and composition, while reducing the dimensionality of the data. Several pre-trained models can be used for this purpose, including convolutional neural networks (CNNs) and more recent architectures like Vision Transformers (ViTs). Once images are represented as vectors, similar images will be closer in the multidimensional vector space. The distances among images in the multidimensional space facilitate the clustering of road sections based on shared visual characteristics.

In the following, we employ ResNet152, a CNN-based model pre-trained on ImageNet (He et al., 2016). We choose this model because of its ability to extract high-quality features from images across diverse contexts, making it well-suited for analyzing the visual content of street-level imagery. For each road section, we process the associated images through the embedding model to generate a set of vectors. (i.e., one vector per image). Using ResNet152, the resulting vector for each image contains $2,048$ characteristics, preserving details that may be relevant for classification.

Figure 2: Coverage of images over the road sections. Sections in green are associated with at least one SLI. Sections in red have no images in the 20-meter surroundings.

Step 3: Road section embedding representation

The third step involves aggregating the image embeddings associated with each road section to create a unified representation. To achieve this, we compute the mean of the embedding dimensions for all images linked to a given road section. By averaging the vectors, we derive a comprehensive representation that encapsulates the overall visual content of the section, balancing the contribution of each image. This approach ensures that the resulting vector reflects the collective visual characteristics of the road section rather than relying on any single image.
Once the road section embeddings are generated, we address the curse of dimensionality issue (Crespo Márquez, 2022), which can negatively affect the performance of clustering algorithms (in step 4). High-dimensional data often leads to noise, sparse distributions, and diminished distances between points (in the embedding space), making it challenging for clustering techniques to distinguish the classes. We apply Principal Component Analysis (PCA) to reduce the dimensionality of the road section embeddings and mitigate this effect. Our PCA implementation is forced to keep at least 80% of the original variance in the data, resulting in 86 dimensions in the final road section representations.

Step 4: Clustering for classification

The final step involves classifying the road sections by applying clustering techniques to the road section embeddings. The clusters identified over the embedding features should group road sections with similar visual characteristics. We use the elbow method to determine the proper number of clusters. This method analyzes the variance explained by different numbers of clusters and identifies the point where adding more clusters yields diminishing returns, balancing accuracy and interpretability when selecting the number of clusters.
For the clustering process, we implement hierarchical clustering, which aligns well with the hierarchical nature of urban road networks. Hierarchical clustering builds a tree-like structure of nested clusters, which mirrors the way roads are often organized in cities, from motorways to smaller local streets. We also tried other clustering methods as part of an exploratory analysis of the robustness of the embedding information in the next section.

3 Results

Clustering application

Using the elbow method on hierarchical clustering, we determine that six clusters provide the optimal balance between accuracy and interpretability for classifying the road sections. Our method allows us to group the road sections into distinct categories based on the visual characteristics captured in the embedding. Figure 3 illustrates a map of Delft's road network, with color-coded roads according to their assigned cluster, highlighting the distinct road classes present in Delft.

Figure 3: Delft's road sections colored by clusters. Clusters names are set based on the visual characteristics of the images.

The model effectively differentiates some urban typologies. The city center, marked in light blue (cluster 1), is characterized by narrow streets, canals, and historic, closely narrow houses with minimal vehicle presence. Our model accurately identifies the natural boundaries of this area. Additionally, the two major highways in Delft are clearly distinguished in yellow (cluster 2). While the map reveals separation of other colored streets (red, pink, blue, and green), further analysis is necessary to explore the meaning of these classifications.

Image exploration for cluster identification

We perform a qualitative analysis to better understand the characteristics of each cluster by sampling and exploring subsets of images assigned to each group. This exploratory process complement the cluster definitions by considering visual features from the images. Figure 4 presents random images from the six clusters.
We labeled the clusters based on the visual characteristics observed in the sampled images and the spatial distribution of the road sections in Figure 3. Cluster 1 (light blue) is labeled *City Centric* due to the presence of narrow streets and historic buildings. This area aligns with the well-known historic city center of Delft. Clusters 0 and 5 (pink and red) are labeled *Residential A* and *Residential B*, respectively, characterized by residential houses, open areas, and presence of trees on the streets. Residential B encircles the city center and features streets and buildings similar in style to the historic core, while Residential A is located in the southern outskirt, with wider streets, newer buildings, and green spaces. Cluster 2 (yellow) is identified as *Motorway*, encompassing the main highways in Delft, and visible from the images showing large, high-speed roads. Cluster 3 (blue), labeled *Arterial*, features wider streets with multiple lanes, reflecting their role as major traffic arteries connecting different city areas. Lastly, Cluster 4 (green), labeled *Greenery*, includes roads surrounded by more rural landscapes and vegetation. While similar to

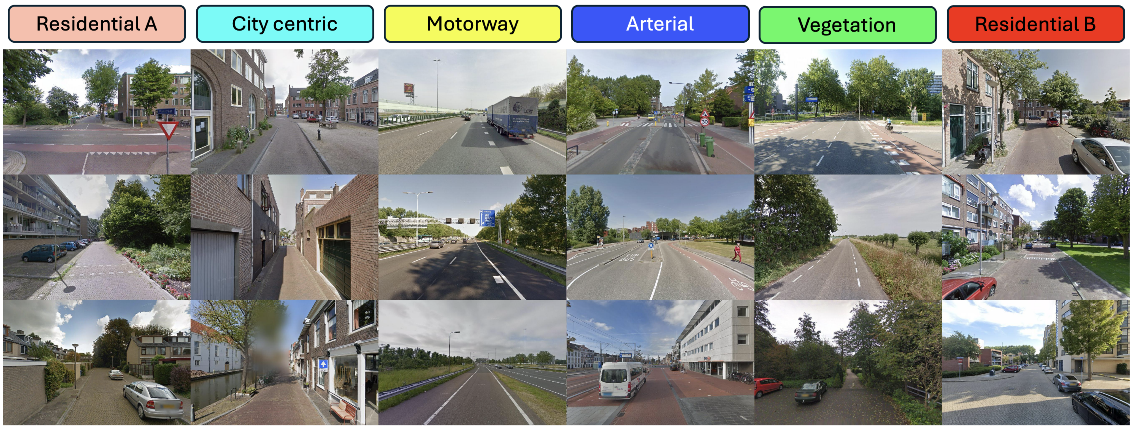

Figure 4: Randomly-sampled images from each cluster.

Arterial roads in their connectivity function, the visual cues of abundant greenery distinguish this cluster. This labels serve as a preliminary interpretation of the clusters, which can be further refined by incorporating additional data and expert knowledge.

Exploring street attributes within the classes

We explore the distribution of various street attributes within the identified clusters to evaluate the ability of our method to classify features not explicitly present in the input imagery. Specifically, we examine data related to road types, surface materials, speed limits, and the presence of vegetation in Delft. This assessment provides insights into how well the visual characteristics captured in the street-level images correlate with other significant urban features. Figure 5 illustrates the distribution of these attributes across the six clusters.

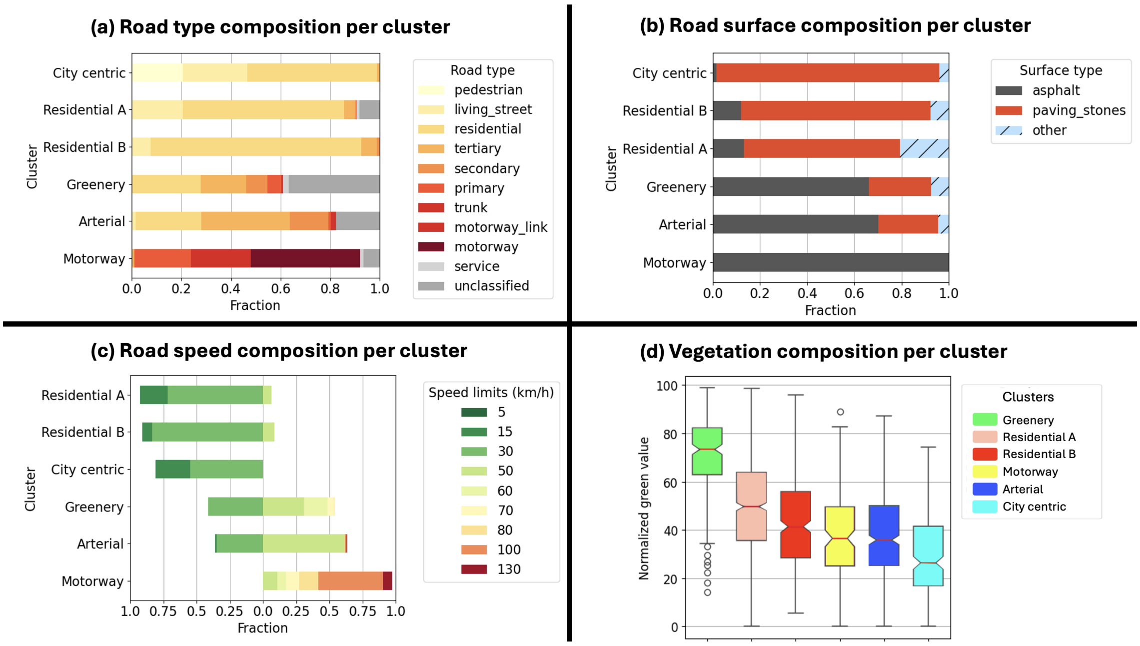

Figure 5: Distribution of road attributes across the clusters. (a) Road types , (b) Surface materials, (c) Speed limits, and (d) Greenery across the clusters.

Figure 5a illustrates the distribution of road types across the clusters, sourced from OSM data. The Motorway cluster predominantly includes trunk, motorway link, and motorway types, highlighting the main highways in Delft. The City centric cluster includes the pedestrian roads, consistent with its pedestrian-only zones. Residential A and B clusters mainly comprise residential and living street types. The Arterial cluster shows a blend of primary, secondary, and tertiary roads, indicating its function as a major traffic artery. The Greenery cluster contains more service roads, aligning with

its rural nature.
Figure 5b shows the distribution of road surface materials across the clusters, focusing on paving stones and asphalt, the dominant surface materials in Delft. City centric and residential A/B clusters primarily have paving stones, indicative of older, traditional streets and lower speed zones. Conversely, the Arterial and Motorway clusters mainly consist of asphalt roads, typical for major thoroughfares.
Figure 5c depicts the road speed limits within the clusters, as reported by OSM. The Residential clusters mostly have speed limits of $30km/h$ or less, while Arterial roads often have limits of $50km/h$ or more. The Motorway cluster mainly includes speed limits above $70km/h$. The Greenery cluster (green) shows a mix of speed limits, with a notable presence of roads allowing speeds over $70km/h$, reflecting its rural character.
Figure 5d shows the composition of greenery across the clusters. This data is obtained from the *Groenkaart* of the Dutch government, which provides a normalized value between 0 and 100 for the greenery of the surroundings. We calculate the average greenery value within a 5-meter radius around each road section and plot the distribution of these values across the clusters. The Greenery cluster (green) has a significantly higher average greenery value than the other clusters, reflecting the presence of more rural landscapes and vegetation in these areas.

4 Conclusions

In this study, we introduced an approach to classifying road networks using geo-tagged street-level imagery. By applying representation learning and clustering techniques, we developed a scalable method that effectively categorized road sections in Delft, the Netherlands, into six distinct clusters. Each cluster reflected unique urban typologies and visual characteristics, demonstrating the potential of street-level imagery for detailed road network classification. This method has promising applications in urban planning, mobility studies, and automated vehicle navigation.
Our approach addresses limitations of traditional road classification methods by incorporating diverse visual data and reducing reliance on manual input or high-quality road attribute datasets. The use of image embeddings enabled the automatic grouping of roads based on visual patterns, with clustering results aligning well with known urban structures in Delft. The method also revealed nuanced distinctions in residential and arterial road types, highlighting its robustness. Moreover, the clusters captured information beyond visual features, such as road types, surface materials, speed limits, and greenery, suggesting that our method could be enhanced by integrating additional data layers like traffic density or public transport routes to further improve accuracy and utility.
However, we also acknowledge several limitations. The method's reliance on street-level imagery means that image quality, coverage, and representativeness can affect classification outcomes. In regions with sparse, outdated, or biased imagery, the results may be less reliable. Additionally, the image embedding process may overlook important road features, such as utilities or temporal changes like construction or seasonal variations which are not present in the images. The clustering process also poses challenges in determining the optimal number of clusters and interpreting the resulting classifications. Future research could explore automated techniques for cluster validation and refinement, as well as methods to improve the interpretability of the clusters.
Future work should implement specified computer vision models, such as transformer-based models, to enhance feature extraction and classification performance. Additionally, developing specialized classification models for street categorization could improve accuracy. Testing this method in various cities with diverse urban layouts would assess its generalizability and robustness. Incorporating multi-modal data sources, such as satellite imagery or traffic flow statistics, could enrich the classification process and provide an holistic view of road networks. Finally, creating user-friendly tools for urban planners and policymakers could bridge the gap between research and practical application.
Our study demonstrates the viability of using street-level imagery for road network classification. By leveraging visual data and representation learning techniques, we offer a scalable and adaptable method for classifying road networks that can support a wide range of urban planning and mobility applications.

References

Apple Inc. (2023). *Apple maps look around.* Retrieved from https://www.apple.com/maps/ (Accessed: 2024-04-12)

Bosurgi, G., Pellegrino, O., & Sollazzo, G. (2019). Road functional classification using pattern recognition techniques. *The Baltic Journal of Road and Bridge Engineering*, *14*(3), 360–383.

Chatziioannou, I., Tsigdinos, S., Tzouras, P. G., Nikitas, A., & Bakogiannis, E. (2024). Connected and autonomous vehicles and infrastructure needs: Exploring road network changes and policy interventions. In *Deception in autonomous transport systems: Threats, impacts and mitigation policies* (pp. 65–83). Springer.

Crespo Márquez, A. (2022). The curse of dimensionality. In *Digital maintenance management: Guiding digital transformation in maintenance* (pp. 67–86). Springer.

Dijkstra, A. (2011). En route to safer roads. how road structure and road classification can affect road safety.

Garrido-Valenzuela, F., Cats, O., & van Cranenburgh, S. (2023). Where are the people? counting people in millions of street-level images to explore associations between people's urban density and urban characteristics. *Computers, Environment and Urban Systems*, *102*, 101971.

Google. (2023). *Google street view.* Retrieved from https://www.google.com/maps/streetview/ (Accessed: 2024-11-03)

Gorges, C., Öztürk, K., & Liebich, R. (2018). Road classification for two-wheeled vehicles. *Vehicle system dynamics*, *56*(8), 1289–1314.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 770–778).

Hu, Y., & Liang, C. (2024). Study on the spatial relationship between road network and the diversity of urban public facilities: the case of the central area of changsha city. *Journal of Engineering and Applied Science*, *71*(1), 156.

Kaptein, N., & Claessens, F. (1998). Effects of cognitive road classification on driving behaviour: a driving simulator study.

Leśniara, K., & Szymański, P. (2022). Highway2vec: Representing openstreetmap microregions with respect to their road network characteristics. In *Proceedings of the 5th acm sigspatial international workshop on ai for geographic knowledge discovery* (pp. 18–29).

Mapillary. (2023). *Mapillary.* Retrieved from https://www.mapillary.com/ (Accessed: 2024-05-12)

Mikolov, T. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mohammadi, M. (2012). Road classification and condition determination using hyperspectral imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *39*, 141–146.

OSM. (2024). *Planet dump retrieved from https://osm.org* . Retrieved from https://www.openstreetmap.org

Taamneh, M., Taamneh, S., & Alkheder, S. (2017). Clustering-based classification of road traffic accidents using hierarchical clustering and artificial neural networks. *International journal of injury control and safety promotion*, *24*(3), 388–395.

Tang, I., & Breckon, T. P. (2010). Automatic road environment classification. *IEEE Transactions on Intelligent Transportation Systems*, *12*(2), 476–484.

Tsigdinos, S., Salamouras, G., Chatziioannou, I., Bakogiannis, E., & Nikitas, A. (2024). A worldwide review of formal national street classification plans enhanced via an analytical hierarchy process: Street classification as a tool for more sustainable cities. *Cities*, *154*, 105371.

Zhang, J., Chen, X., Xiang, Y., Zhou, W., & Wu, J. (2014). Robust network traffic classification. *IEEE/ACM transactions on networking*, *23*(4), 1257–1270.