This is the Accepted Manuscript version of an article published by Elsevier in the journal *Computers, Environment and Urban Systems* in 2024, which is available at: https://doi.org/10.1016/j.compenvurbsys.2024.102129
Cite as: Lei, B, Liu, P, Milojevic-Dupont, N, Biljecki F (2024): Predicting building characteristics at urban scale using graph neural networks and street-level context. *Computers, Environment and Urban Systems*, 111: 102129.

Predicting building characteristics at urban scale using graph neural networks and street-level context

Binyu Lei[a], Pengyuan Liu[b], Nikola Milojevic-Dupont[c], Filip Biljecki[a,d,*]

[a]*Department of Architecture, National University of Singapore, Singapore*
[b]*Future Cities Lab Global, Singapore-ETH Centre, Singapore, Singapore*
[c]*Mercator Research Institute of Global Commons and Climate Change, Germany*
[d]*Department of Real Estate, National University of Singapore, Singapore*

Abstract

Building characteristics, such as number of storeys and type, play a key role across many domains: interpreting urban form, simulating urban microclimate or modelling building energy. However, geospatial data on the building stock is often fragmented and incomplete. Here, we propose a novel and easily adaptable method to predict building characteristics in diverse cities, which attempts to mitigate such data gaps. Our method exploits the geospatial connectivity between street-level urban objects and building characteristics by employing graph neural networks, as they can model spatial relationships and leverage them for predictions. We apply this approach in three representative cities (Boston, Melbourne, and Helsinki) that offer a variety of building features as prediction targets (storeys, types, construction period and materials) and diverse urban environments as predictors. Overall, the magnitude of errors is acceptable for a series of use cases. In the prediction of building storeys, an average of 81.83% buildings in three cities have less than one-storey prediction error. We also find that the prediction of building type achieves an average of 88.33% accuracy across three cities. Meanwhile, an average of 70.5% of buildings are correctly classified by construction period in Melbourne and Helsinki, and the building material prediction accuracy is 68% in Helsinki. The results confirm that our approach is adaptable across different urban environments because comparable performance is achieved in the other two cities. Further, we assess the impact of varying local data availability on model performance. Our findings underscore the feasibility of the method in scenarios with sparse building data (10%, 30% and 50% availability). Our graph-based approach advances research on filling in incomplete building semantics from existing datasets, and showcases the potential to enable 3D city modelling. Given the broad applicability of the approach to predicting many building characteristics, diverse downstream applications exist, such as enhancing contemporary urban studies (e.g. exploring streetscapes) and facilitating the development of 3D GIS (e.g. maintaining and updating 3D building settings).

Keywords: Building semantics, Urban digital twins, Data availability, Deep learning, Urban morphology

1. Introduction

Acquiring key characteristics of buildings is fundamental for supporting city planning, contributing to various urban research (Hudson et al., 2019; Kitchin, 2014; Nouvel et al., 2017). However, the availability of building characteristics widely varies across initiatives and regions (as illustrated in Figure 1). Biljecki et al. (2021) investigated 140 open government geospatial datasets from 28 countries and found that only half of them contained more than one building feature. Detailed characteristics, such as building height, number of storeys, type, age, and status, while crucial for many analyses, are often not available. Beyond 2D urban studies, a similar situation is observed in 3D GIS where only a few datasets offer semantic building information, while they mostly inform only the building geometry (Lei et al., 2023). In urban digital twins, where 3D city models play a role in representing the physical environment, data issues (e.g. scattered building data and varied accessibility) hinder their implementations (Kolbe and Donaubauer, 2021; Park and Guldmann, 2019; Chen et al., 2019; Hong et al., 2020; Gil, 2020; Schrotter and Hürzeler, 2020). Nevertheless, a comprehensive-enough set of building characteristics is fundamental for supporting use cases of 3D city models and urban digital twins, such as simulating energy consumption and analysing urban heat island effect (Gröger and Plümer, 2012; Harter et al., 2023; Xia et al., 2022; Katal et al., 2022; Bansal and Quan, 2022).

The methods to collect information about buildings are diverse, each with pros and cons. Two streams are found in the research community. Image-based approaches are commonly used for extracting building attributes, relying on remote sensing (Liasis and Stavrou, 2016; Wang et al., 2018; Chen et al., 2018), and street

*Corresponding author

Email addresses: binyul@u.nus.edu (Binyu Lei), pengyuan.liu@sec.ethz.ch (Pengyuan Liu), milojevic@mcc-berlin.net (Nikola Milojevic-Dupont), filip@nus.edu.sg (Filip Biljecki)

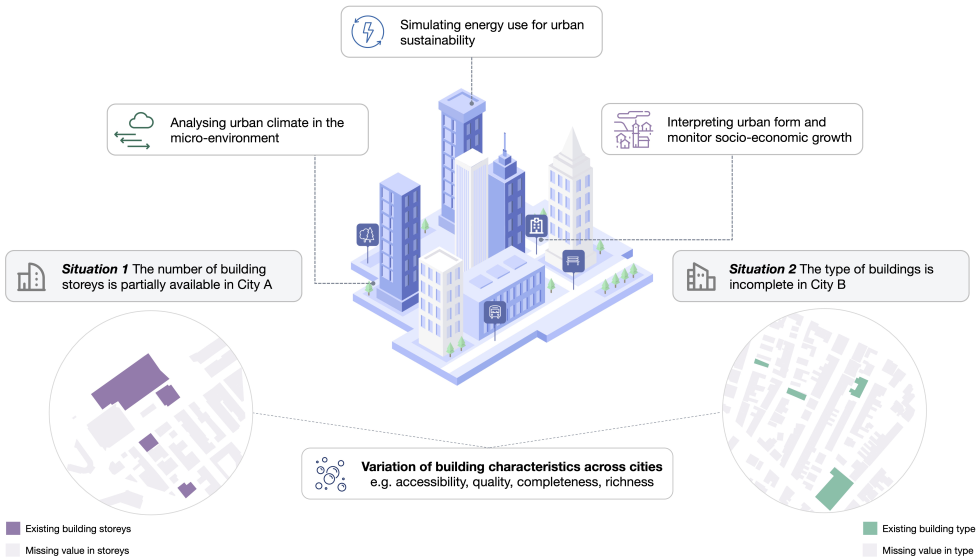


Figure 1: Geospatial building data can facilitate a diversity of use cases, but its availability varies across cities. A typical situation is described with two examples: on the left side, a building attribute (here the number of storeys) is partially available in City A, and on the right side, another attribute (the building type) is incomplete in City B. Such situations make it impossible to perform downstream analyses, such as analysing urban climate, simulating energy use and interpreting urban growth, highlighting the importance of building data completeness and richness for urban studies.

view imagery (Dimitrov and Golparvar-Fard, 2014; Sun and Gu, 2022; Ning et al., 2022; Pang and Biljecki, 2022). Other than the *direct* measurements, approaches using urban morphology to infer building characteristics with machine learning have gained attention (Biljecki et al., 2017; Milojevic-Dupont et al., 2020; Nachtigall et al., 2023). One methodological limitation of these approaches is that they typically use standard machine learning algorithms such as XGBoost and only partially encode spatial aspects through feature engineering. Further, taking into account spatial heterogeneity, machine learning methods may be failing in representing spatial characteristics locally. The lack of explicit representation may make these studies miss spatial aspects potentially rich in predictive information, such as spatial patterns and relationships.

Here, we introduce an innovation in methods for inferring building characteristics by proposing a more spatially explicit representation of urban features as predictors. In particular, we develop a generic graph-based approach to exploit the spatial relationships between buildings and street-level urban objects, such as nearby amenities and city furniture, thus predicting building characteristics.

Our approach advances the literature in three ways and does so significantly. *First*, we adopt graph neural networks (GNNs) for the first time to capture the geospatial connection between buildings and their surrounding urban entities and conduct predictions. In contrast to conventional machine learning models (e.g support vector machine, XGBoost), which merely learn urban features at attribute space (Nachtigall et al., 2023; Milojevic-Dupont et al., 2020), the graph neural network-based method can additionally learn the underlying spatial context (i.e. spatial interactions) of the urban road features; thus, enabling the method to be geospatially explicit. Further, addressing unique spatial characteristics of various urban settings, our approach is intended to make predictions locally and across cities with appropriate adjustments and customisation.

Second, we investigate new predictive features that have not been used previously in the literature – street-level urban objects. Urban objects surrounding buildings at the street level (e.g. amenities and trees) are harnessed as predictors. When designing this approach, we only exploit the spatial connectivity between buildings and urban objects from streetscapes, without considering building-related information (e.g. height and the number of storeys, which can be hard to obtain in different cities) and local regulations (e.g. population and housing price). In contrast with the previous literature, we also choose not to use geometric features from buildings (e.g. footprint area) to specifically focus on investigating the predictive power of urban objects surrounding buildings. Building footprints demonstrate the ability to infer building characteristics in current research (Biljecki et al., 2017; Milojevic-Dupont et al., 2020), but a number of developing countries or cities remain the challenge of offering building footprints (e.g. cities in China and South East Asia). However, urban objects can be available retrieved from, e.g. social media or volunteered geographical information. Further, our work is distinctive as it is the first to focus on urban objects surrounding buildings, exploiting their spatial relationships to perform predictions.

Third, the approach is generic and reproducible, applying to any kind of input data source that can be used as predictors (e.g. geotagged information from authorities or companies). In this proof-of-concept work, we use OpenStreetMap (OSM) data, an open and crowdsourced dataset available worldwide (Biljecki et al., 2023), as a source of street-level urban objects. We retrieve data on points of interest (POIs), transportation infrastructure, and vegetation objects (e.g. trees or plant covered areas) in different cities. The selection of OSM ensures that the approach can be used in a variety of cities around the world if the degree of completeness of urban objects in the city is sufficient. The method can be further developed as an end-to-end pipeline from raw data, making it easier for adoptions among a variety of use cases.

The paper is structured as follows. Section 2 introduces related work on predicting building characteristics and discusses current methodologies employed for predictions. Section 3 outlines our workflow for building the prediction model, detailing the datasets incorporated in this study. In Section 4, we adopt the proposed approach within diverse urban contexts (i.e. Boston, Melbourne and Helsinki) to derive a variety of building attributes. Further, we evaluate the impact of availability by simulating a set of data availability scenarios. The following Section 6 gathers insights from implementation together with contemplating potential avenues for future exploration and existing limitations. Section 7 concludes this work, highlighting the novelty of our graph-based approach to learning multiple building characteristics.

2. Background and related work

2.1. Geospatial data on buildings

Buildings are fundamental elements of urban landscapes. Building characteristics not only define architectural aesthetics but also offer insights into understanding and examining the multifaceted dynamics of cities and associated activities. A number of studies have highlighted an underlying connection between urban entities; for example, Wang et al. (2016) investigated how streetscape quality associated with building design (e.g. levels and usage) influenced human behaviours in the cities. When discussing the imageability of a city, Lynch (1964) underlined the importance of building information as critical urban elements that mirror the local context. Delving deeper, each building has its symbolism, encapsulating various facets — social, political, cultural, or aesthetic function. Such symbolic interpretations often intertwine with distinct physical attributes of the building, such as its size, intended use, overall shape, and façade, and subsequently influence a range of socio-economic activities, e.g. shaping urban functions (Appleyard, 1969; Chen et al., 2017; Xu et al., 2022a). The dynamic establishes a symbiotic relationship, highlighting the interaction between urban activities and constructed surroundings. Furthermore, building characteristics play a role in environmental studies, serving as foundational information for advanced semantic analyses. Amidst a growing global focus on sustainability, investigations into such topics, ranging from energy consumption to vertical farming and micro-climate, have become increasingly prevalent in contemporary discussions (Palliwal et al., 2021; Teng et al., 2021; Fathy et al., 2021; Dembski et al., 2020).

Conducting these studies requires a granular understanding of building features. For instance, in the realm of energy consumption analysis, Kumar et al. (2018) incorporated a myriad of building attributes, including roof information, surface area, and wall area, to explore heating and cooling loads, aiming to elucidate efficient strategies that could be instrumental for the energy and building sectors. Likewise, the integration of diverse building characteristics has proven essential in guiding decisions related to ventilation. Wang et al. (2019) found that varying building types have distinct requirements for installing heat recovery devices. When conducting research on urban climate, building features, such as building height (serving as a primary metric for gauging urban density) and building age (identifying architectural styles like heritage façade), are identified as pertinent factors (He et al., 2022; Zhu et al., 2022). Such information is instrumental in both analysing the phenomenon and devising mitigation strategies, but is often missing.

In 3D GIS, building characteristics act as the backbone of city models and urban digital twins, driving the quality and richness of models. For example, detailed 3D models with abundant semantic information have greater potential for a variety of applications, such as simulating disaster scenarios, facilitating advanced indoor analysis (Lei et al., 2023). While building characteristics offer a valuable lens for diverse urban studies, the accessibility and completeness of geospatial data on building levels can vary across cities. Prediction of building information from other data sources, therefore, emerges as a strategy to bridge the data gap and enable analyses that demand specific semantics.

2.2. Inferring building characteristics

In the realm of predicting building features, there has been a surge in research targeting the estimation of building heights, which can be used for interpreting vertical spatial morphology and employed to generate 3D building models when combined with building footprints (Silva et al., 2017; Tu et al., 2016; Chen et al., 2022; Usui, 2023b). One popular method involves leveraging the correlation between building morphology to enable the prediction of height information. By investigating straightforward building-related indicators, researchers can infer missing details (Biljecki and Chow, 2022). For example, Biljecki et al. (2017) utilised a set of ten predictors, categorised across three domains: (1) building attributes, encompassing information such as usage, construction year, number of storeys, and floor area; (2) geometry, such as footprint area and shape complexity; and (3) statistical insights, e.g. population density, household size, and income. In another notable study, Milojevic-Dupont et al. (2020) expanded the scope by integrating urban form, gleaning 152 metrics that span building-specific indicators, street networks, and block parameters. Such an exhaustive dataset then served as the foundation for a supervised machine learning model, contributing to predicting building heights.

Meanwhile, a contemporary approach towards height estimation is harnessing remote sensing. For example, building height can be measured from high-resolution synthetic aperture radar (SAR) images by computing the slant range and distance, necessitating a deep knowledge of photogrammetry (Sun et al., 2022b; Wang et al., 2015). Another technique capitalises on shadows, in particular, by assessing their lengths and solar elevation angles, one can deduce height attributes (Liasis and Stavrou, 2016). However, these modalities come with their own challenges concerning data accessibility and timeliness relevance (Biljecki et al., 2017). With the increasing awareness of open science, the paradigm has shifted to more publicly available data and open sources. A number of studies have attempted to derive height information by correlating SVIs with building locations, inferring distance through angular perspectives (Díaz and Arguello, 2016; Zhao et al., 2019b). Yan and Huang (2022) employed deep neural networks to derive height data by detecting architectural nuances (e.g. vanishing points, line segments and semantic segmentation) from standalone SVIs. Following suit, Xu et al. (2022b) integrated the mask region-based convolutional neural network (Mask R-CNN) to identify buildings, correlating detected buildings with ground truths to estimate height — achieving a mean height error of 0.78 meters. Related to building heights, an emergent research domain centres around predicting the number of building storeys, which can reflect architectural structure in detail and act as a proxy of building height in some cases (Goetz and Zipf, 2012; Biljecki, 2020). For example, Roy et al. (2022) explored building morphology to infer storey data, particularly for residential structures. A key motivation to predict heights is to create 3D building models from 2D building footprints, significantly broadening their usability (Biljecki et al., 2017).

In addition to the popularity of inferring height information, other building characteristics play a role in research across domains, contributing to a variety of studies. For example, building usage, as a part of shaping urban functions and forms (Hecht et al., 2015; Henn et al., 2012; Kong et al., 2024), has the potential to drive applications ranging from various scales, e.g. energy modelling at the building level, disaster risk assessment in the cities energy modelling to disaster risk assessment (Fan et al., 2014; Li et al., 2022; Xia et al., 2020; Xu et al., 2019; Kim et al., 2022). In terms of deducing building types, convolutional neural networks

(CNNs) have proven instrumental in performing such predictions in the research landscape. Recent research applies CNN-based image segmentation techniques to dissect and analyse vast heterogeneity of urban structures from high-resolution satellite imagery (Huang et al., 2020; Lu et al., 2014; Belgiu et al., 2014). Other than satellites, Wurm et al. (2015) exploited stereoscopic airborne images combined with building metrics (e.g. footprint, floor area and volume) to capture the distinct physiognomies of individual buildings. Moreover, Du et al. (2015) integrated GIS data with very high-resolution imagery and employed the random forest algorithm to categorise buildings into seven distinct clusters. Taking advantage of crowdsourced data in urban analytics (Hoffmann et al., 2023), present research leans on such public accessibility, providing a lens on inferring building usage in an open manner (Bandam et al., 2022). For instance, Atwal et al. (2022) adopted a supervised learning paradigm to predict building classifications, filling the data gap in building stocks. By leveraging OpenStreetMap data, they distinguished residential and non-residential buildings across three American cities. More recently, Kong et al. (2024) proposed a graph-based neural network (GNN) approach for classifying building functions, integrating multiple data sources (e.g. OSM data and images) and leveraging contextual information between buildings.

In parallel, the predictions of other building characteristics, such as building age and building material, are explored with growing interest in the research communities. For example, some research uses black-box-based approaches (e.g. neural networks and machine learning) to infer building age through LiDAR data (Tooke et al., 2014). In a similar vein, Google Street View images, rich in visual detail, have been used as a data source for age prediction models (Li et al., 2018). Notably, such age-prediction efforts make a potential contribution to broader urban studies, e.g. the construction year of buildings can be an instrumental factor when estimating energy consumption (Gassar and Cha, 2020). Beyond a single prediction of particular building features, current research has explored methods for inferring multiple building attributes. For example, leveraging a deep learning framework, the extraction of building façade information from street view images determines both building style and age (Sun et al., 2022a). Furthering this trend, Meng et al. (2022) designed a CNN architecture that specialises in classifying a set of building features, particularly in rural landscapes, using oblique aerial photography. In this regard, artificial intelligence techniques, such as neural networks and machine learning algorithms, have shown promise in processing vast amounts of heterogeneous data and conducting predictions of multiple building characteristics.

2.3. Urban digital twins and GNNs

Building characteristics play a role in multiple domains, supporting not only 2D urban analysis but also 3D urban realm. Urban digital twins, where 3D buildings serve as the backbone, are digital replicas of urban environments that integrate a variety of urban data and techniques, facilitating city planning and decision making by simulating and analysing dynamic urban scenarios. However, the availability and richness of building characteristics in urban digital twins vary from initiatives and cities, impeding further operation of urban digital twins. The increasing adoption of urban digital twins require detailed semantics, supporting dynamic simulations in city planning, infrastructure development, and environmental sustainability. Therefore, recognising the importance of building-level geospatial data in 3D modelling tasks, the rapid progression of AI in urban analytics provides promising tools for inferring missing information and transferring knowledge across multi-disciplines, enriching semantic information of urban digital twins (Ye et al., 2023). However, the complex and irregular spatial patterns of the urban streets and the urban objects (e.g. buildings, trees and other facilities) often pose obstacles to conventional AI models (Mai et al., 2022). Graph neural networks (GNNs), due to their capabilities of handling non-euclidean structured data (Zhu et al., 2021) and encoding the spatial interactions of studied objects in their computational process, have been increasingly recognised to hold the key in developing better AI frameworks that can offer high-quality predictions in spatial analytics (Liu and Biljecki, 2022; Yan et al., 2019). GNNs have proven to be successful in a wide range of street-level prediction tasks (Liang et al., 2023; De Sabbata and Liu, 2023), such as traffic analysis (Zhao et al., 2022, 2019a) and human-centred urban planning tasks (Liu et al., 2023b). However, most of the applications where GNNs are proven to be successful are 'horizontal' tasks. That is, those tasks focus on predictions of certain attributes that do not require 3D modelling; hence, they do not fall into the area of urban digital twins. Therefore, the capability of GNNs to predict 'vertical' information of urban objects remains underexplored. Our study will address such a limitation by developing a GNN-based framework, aiming to integrate GNNs into the development of urban digital twins, providing accurate predictions and classifications with the spatial data available at the urban scale.

3. Methodology

3.1. Workflow

As outlined in Section 1 and Section 2, our study focuses on predicting building characteristics by leveraging geospatial relationships between buildings and their surrounding street contexts (Figure 2). A primary motivation to design such a method is to exploit spatial connectivity, bridge the geospatial data gap for buildings, and pursue broad applicability across diverse urban environments. While several approaches exist for inferring building characteristics, spatial connections are not fully investigated for this purpose. As discussed in the previous section, recent techniques (e.g. random forest) fall short of capturing the geospatial connectivity between buildings and their surroundings.

In this regard, the GraphSAGE algorithm (Hamilton et al., 2017), a variant of graph neural networks, aligns well with our objectives, boasting features such as inductive learning, information aggregation, dynamic spatial representation, and scalability. The graph-based model constructs spatial networks among buildings and determines node embeddings, leveraging context-based similarities for predictions. It also allows a high degree of customisation, as a number of hyperparameters can be tuned to determine the size or the depth of graphs. Therefore, we employ the GraphSAGE algorithm in a bifurcated manner, addressing both regression and classification tasks. First, the regression task focuses on learning continuous target variables, such as building heights and storeys. Second, the classification task facilitates predicting discrete class labels, e.g. building types, roof details and construction materials. While a plethora of studies have tackled building semantics predictions, our methodology distinguishes itself by using GraphSAGE to exploit the spatial connectivity between building stocks and street settings, thus estimating multiple building characteristics.

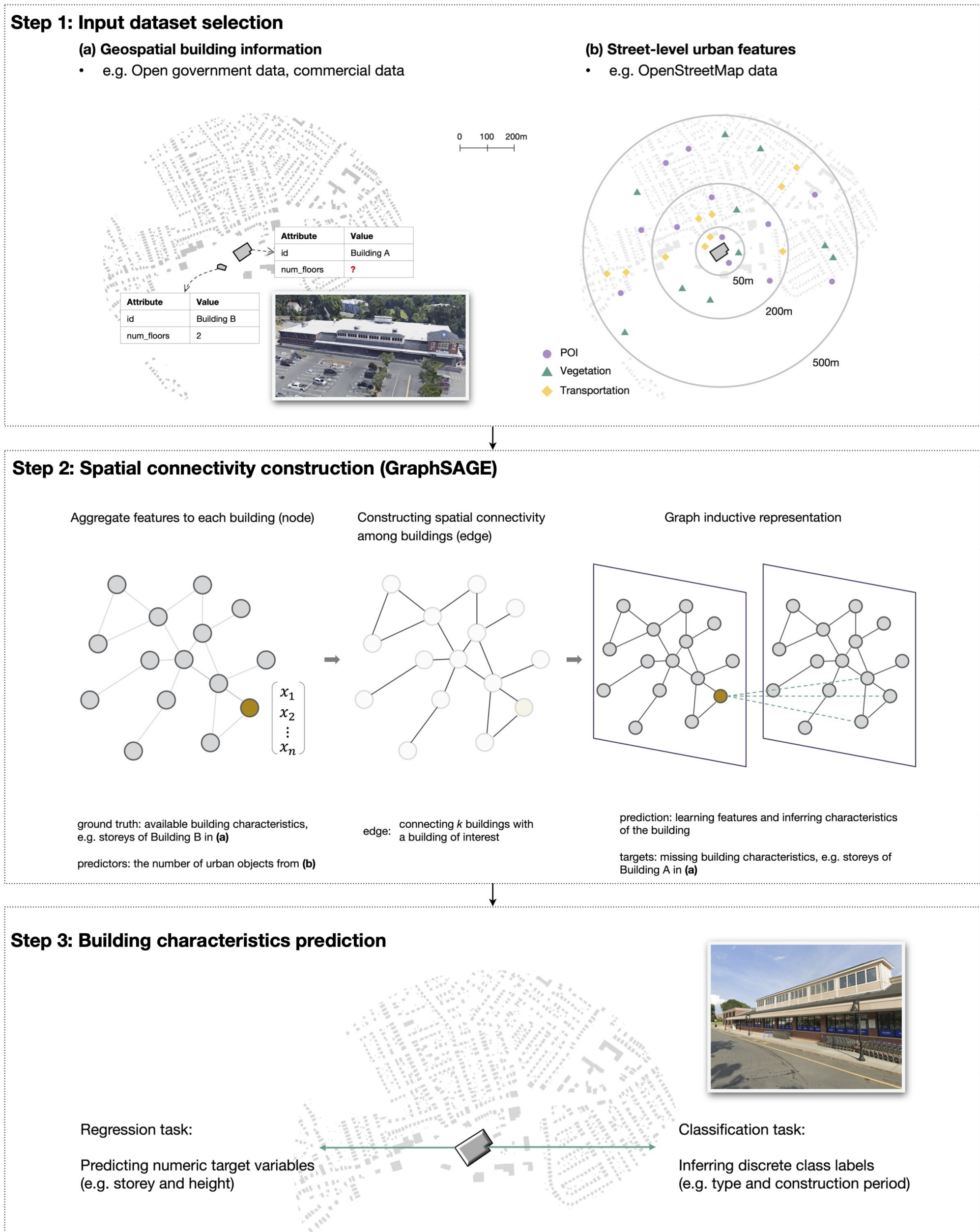


Figure 2: A three-step workflow for developing the graph-based method for predicting building characteristics, taking the prediction of building storeys as an example. Step 1 involves dataset selection, and combining geospatial building information from open government data and street-level urban objects from OpenStreetMap. Step 2 delineates the graph construction process, aggregating features, establishing spatial connectivity via edges, and forming an inductive graph representation. Step 3 consists in predicting building characteristics based on the learned graph for two possible types of prediction tasks – regression for numeric attributes like building storeys and heights, and classification for discrete attributes such as building type and age. Source of the imagery: Google Maps 3D mode and Google Street View.

3.2. Datasets

We base our dataset on OpenStreetMap (OSM), a globally accessible and editable geographic database. The dynamic platform offers rich, volunteer-contributed information on varied geographical features, which can be retrieved using the overpass-turbo API[1]. OSM data's growing completeness and reliability have led to its extensive use in urban morphology studies, as evident in recent literature (Biljecki and Chow, 2022). Such wealth of data offers detailed insights into a wide range of urban objects, such as amenities, which are crucial for understanding spatial information and inferring urban patterns (Chen et al., 2024; Krapu et al., 2023). Our selection of points of interest (POIs), transportation facilities, and vegetation objects from OSM data aims to leverage the spatial relationships between buildings and their immediate surroundings, elucidating how these urban elements interact with and influence buildings within their spatial contexts. These elements are chosen for their significant roles in characterising the urban landscape from three distinct but complementary perspectives: human activity, urban mobility, and the natural environment.

First, *POIs* represent focal points of human activity within an urban area, including such as commercial centres, educational institutions, healthcare facilities, cultural venues, and public spaces. The diversity and density of POIs around a building can signal its function, the population it serves, and its architectural style and design. For instance, a building in a neighbourhood with numerous business POIs may be more likely to serve office workers and therefore may have features catering to their mixed needs, such as apartment living units and restaurants. The selection of POIs as a predictor can acknowledge a implicit correlation between the vibrancy of human activities and the characteristics of nearby buildings. Second, the category of *transport* is a critical component of urban mobility, encompassing bus stops, railway stations, biking racks, and parking facilities. The accessibility and variety of transportation options available around a building reflect the attractiveness of a building for certain uses, such as residential and commercial purposes, while suggesting its capacity to accommodate population density as well. Buildings in areas with rich transportation resources may prioritise access and connectivity, impacting their design and layout. Including transportation facilities as predictors aim to capture the mobility patterns around buildings, reflecting how buildings integrate into a broader urban transit network. Third, *vegetation* in urban areas, from street trees to public parks and natural reserves, contributes to

[1]https://overpass-turbo.eu

environmental landscape within urban settings. The presence of vegetation can inform about the environmental considerations of a building. Further, the access to natural resources can represent the ecological sustainability of buildings, varying from building locations and usage. Selecting vegetation as a predictor, we tend to understand how green elements impact building design and functionality.

Further, we retrieve open government data, primarily focused on geospatial building characteristics, as our target data. It enables us to establish a dataset, integrating urban objects and ground truth labels. The variety of labels we use depends on the availability and detail of government building data. The data information is detailed in Figure 3.

Dataset	Feature	Input	Example	Function	Value
Open government data	Building information	Geospatial network	*k* buildings with a building of interest	Spatial connectivity	Adjacency matrix
		Available building characteristics	Heights, storeys, type, etc.	Ground truth	Numerical value or encoded categorical variables
OpenStreetMap data	Street-level urban objects	Points of interest	Restaurant, school, etc.	Predictors	Total count within 50, 200 & 500 m around a building of interest
		Transport	Bus stops, train stations, bike racks, etc.		
		Vegetation	Trees, natural reserves, etc		

Figure 3: Detailed breakdown of datasets. The table highlights the features and subgroups extracted from 'Open government data' and 'OpenStreetMap data'. Examples are provided for clarity, along with information on counting method for street-level urban objects.

3.3. GraphSAGE

GraphSAGE is a graph neural network algorithm developed for inductive node-level representation learning in graphs (Hamilton et al., 2017), denoted as $G = (V, E)$, where V represents vertices (or nodes) and E indicates edges. GraphSAGE aggregates information from a node's local neighbourhood, linking nodes and their adjacent neighbours to share and propagate information and computing embeddings to make predictions. One of GraphSAGE's prevailing functions is to handle large graphs with a large number of nodes and edges efficiently and effectively. Given the scale and heterogeneity of urban data, in particular regarding amounts of buildings associated with various street contexts, the GraphSAGE algorithm delivers strengths in scalability and flexibility, making it suitable in our method for expansive urban environments. Meanwhile, GraphSAGE presents the advantages of facilitating graphs with diverse aggregation techniques, such as mean aggregator, pooling aggregator, and LSTM aggregator. For our model, we choose the pooling method, given its efficacy in transmitting neighbours' embeddings:

$$\text{AGGREGATE}_k^{pool} = \max\left(\left\{\sigma\left(\mathbf{W}_{pool}\mathbf{h}_{ui}^k + \mathbf{b}, \forall u_i \in N_{(v)}\right)\right\}\right) \tag{1}$$

where *max* denotes the element-wise max operator and σ is a nonlinear activation function. Such aggregation architecture balances symmetric and trainable characteristics, ensuring each neighbour's vector is independently fed through a fully connected layer. By learning from graph network structures and creating node embeddings, GraphSAGE enables various downstream tasks, for example, node classification, node regression, and edge predictions, which can be further applied to infer multiple building characteristics, such as building levels and usage. In this regard, GraphSAGE captures the spatial distribution of buildings and learns from the relationship between buildings and urban objects through an explicit sense of its surroundings, therefore setting the stage for predictions of various building characteristics.

3.4. Model design

The architecture of our model, illustrated in Figure 4, comprises two main phases: graph construction and inference. In the first phase, we construct the graph structure, where each building is conceptualised as a *node*. We employ the principle of kd-trees to efficiently identify and locate street-level objects around buildings (Bentley, 1979). Each building node is then characterised by its surroundings as features, including the number of POIs, transport elements, and vegetation. In our graph, the spatial connectivity between buildings is vital and represented by *edges*. Using the k-nearest neighbours (KNNs) algorithm, we identify a set number of buildings closest to a building of interest, interlinking them within our graph to create a dense network of spatial connections. For our approach, each building node is linked to its 10 nearest buildings based on preliminary analysis comparing different neighbourhood sizes, ensuring a balance between model complexity and computational efficiency. However, the number of nearest buildings serves as a hyperparameter in our method, allowing customisation to construct graphs of varying complexity based on specific contexts. We adopt dual embedding techniques in the inference phase: node regression for numerical building characteristics and node classification for categorical attributes. Hidden layers are constructed to optimise the network's depth, ensuring efficient information propagation between layers. We split the dataset into a 70:30 ratio for training and testing, a distribution well-regarded for its effectiveness in related work. Next, we train the model on the pre-processed data, deriving node embeddings that capture the underlying network features. After certain epochs of training, we assess the model's performance using evaluation metrics such as RMSE (root mean squared error), MAE (mean absolute error), classification accuracy, F1-score, precision and recall scores (Hyndman and Koehler, 2006). Further, we employ the median relative error as a complementary metric to evaluate how the method perform in predicting building storeys. It allows us to take a deeper look at the magnitude of prediction inaccuracies, and gauge the extent of information loss in our model performance. In particular, we consider a range of acceptable errors for evaluating building storey predictions, defined as the percentage of buildings with an prediction error below 0.25 storeys, 0.5 storeys, 0.75 storeys and 1 storey. For example, a prediction error of 0.5 storeys can be translated into 1.3-1.8 metres (e.g. the average storey height is around 2.65 m in the Netherlands (Roy et al., 2022) and around 3.5-3.9 m in Japan (Usui, 2023a)). We believe that having the acceptable error range can effectively and diversely conclude the evaluation results, as well as take into account various scenarios in which the model is being used for specific cases and contexts.

3.5. Implementation

The prediction model is implemented in Python:

- Google Colab[2] environment for data processing and model training and evaluation;
- `scikit-learn` library (Pedregosa et al., 2011) for matching nearest features surrounding buildings and measuring the model performance, modelling the baseline random forest, etc.;
- `DGL` library (Paszke et al., 2019) and `Pytorch` (Paszke et al., 2019) for constructing graph network;
- `loss function` of MAE and RMSE for regression task (e.g. predicting numeric building characteristics) and cross-entropy loss for classification task (e.g. building types), for evaluating the model's accuracy.

[2]https://colab.research.google.com

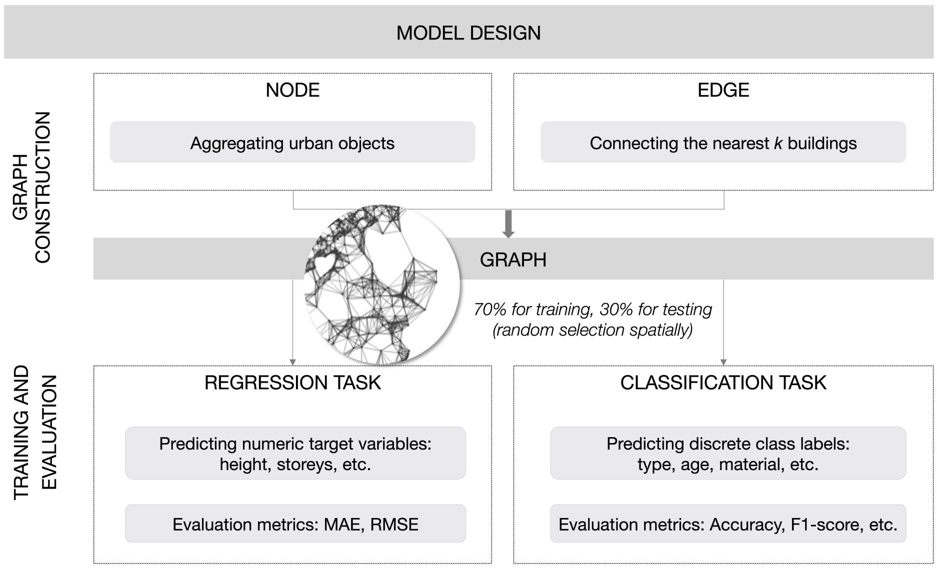


Figure 4: Our prediction pipeline: model design, detailing graph construction and evaluation. For graph construction, urban objects at street scale are aggregated to be building (node) features, while edges are formed by connecting the nearest k buildings (in this work, k=10). Post construction, the resultant graph is randomly divided into 70% for training and 30% for testing. Model evaluation comprises two main tasks: node regression, where numeric target variables like height and storeys are predicted with associated RMSE and MAE, and node classification, where discrete class labels such as building types and materials are predicted, evaluated by classification accuracy and other metrics.

3.6. Experiments

To evaluate our method and assert its wide applicability, we designate our experiments in three illustrative cities, each carefully chosen to encapsulate various dimensions of our study. The selection process is based on two inclusion criteria. First, geospatial heterogeneity is considered to ensure the worldwide applicability. Each chosen city will then serve as a representative case study, showcasing the method's adaptability and effectiveness across diverse urban contexts and geospatial variances. Second, we seek cities with varied levels of data completeness of building characteristics. While some cities offer comprehensive building attributes, others may possess only a subset of such semantic content. The diversity in data richness allows us to gauge model performance across a spectrum of data availability. Given these criteria, three cities are selected for their distinctive geospatial and building attributes: Boston, Melbourne, and Helsinki (as illustrated in Figure 5). Each city, hailing from a distinct continent, provides a unique urban backdrop, making our study both comprehensive and globally relevant.

Experiment 1. We set up our first experiment in Boston (USA) to validate our method's applicability. The open government geospatial data on buildings in Boston[3] provides key semantic information. For this experiment, we chose two building characteristics as prediction targets: building storeys and building type. Given the quantitative nature of building storeys, we employ regression task for its predictions. Meanwhile, the classification task is used to learn building types.

Experiment 2. For this trial, our approach is applied in Melbourne (Australia). Our objectives include evaluating its applicability across cities and exploring its capability to predict a broader range of building characteristics. Leveraging the comprehensive open government building data in Melbourne [4], we target building storeys, type, and additionally construction period for prediction. Besides storeys and types, introducing construction period as a prediction target provides insights into the historical context of buildings, a feature with substantial relevance to downstream applications (e.g. building energy calculation).

Experiment 3. Our method is further extended to Helsinki (Finland), allowing it to adapt to diverse contexts and geospatial variations. Following the established workflow, building data is sourced from the local government [5]. The exemplary development of 3D city model in Helsinki, especially with its rich semantic building attributes (Lei et al., 2023), provides an ideal foundation for our approach. The diverse building semantics facilitate the provision of multiple ground truth values, guiding our specific predictions of building characteristics. Hence, for this experiment, we target a range of building characteristics: storeys, type, construction period and building materials. A novel aspect of this experiment is our effort to estimate building materials, further highlighting the versatility of our method.

4. Results

We train and cross-validate the designed method on three cities. Overall, the graph-based approach performs well in predicting a variety of building characteristics across various urban environments, achieving competitive results through

[3] https://data.boston.gov/dataset/boston-buildings-inventory

[4] https://data.melbourne.vic.gov.au/explore/dataset/buildings-with-name-age-size-accessibility-and-bicycle-facilities/information/?refine.census_year=2021

[5] https://hri.fi/data/en_GB/dataset/helsingin-rakennukset

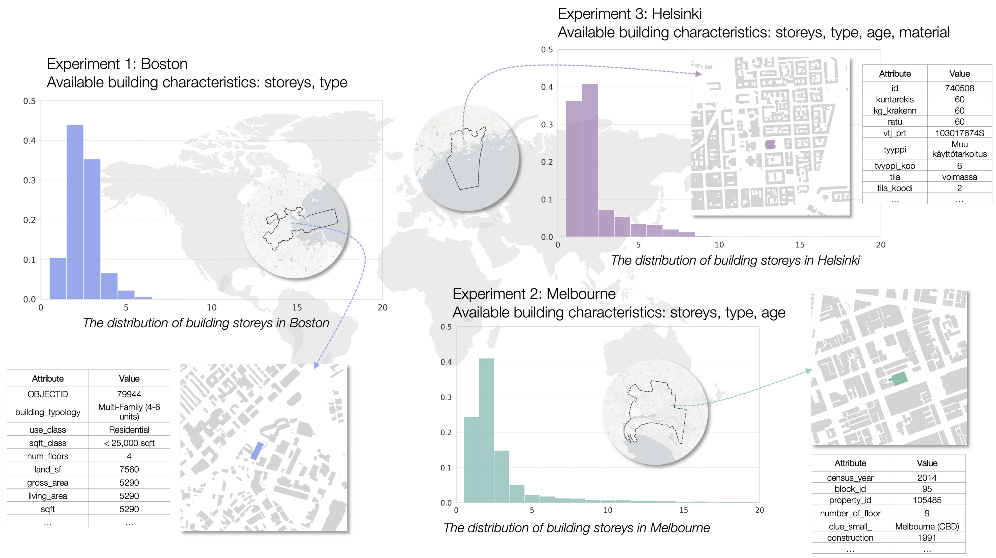


Figure 5: The three cities in focus in this study: Boston, Melbourne and Helsinki. The plots show the distribution of building storeys to indicate the diversity of settings included in the experiments. For each city, there is an example of a building showing the data content, illustrating the variation of building characteristics cross cities.

comparing with the present literature and the baseline model random forest. Across the prediction of building storeys in three cities, we find that on average 81.83% of the predictions are of acceptable (below one building storey), with best case being 91.58% in Boston and worst case being 69.92% in Melbourne. For example, Figure 6 spatial-explicitly illustrates the distribution of building storeys and type information in Boston. In this zoomed-in area, most buildings are inferred correctly in storeys and type, where the prediction error is within a ±1 storey range. Notably, our method indicates potential to adapt different scenarios, for example, a city with limited available building data.

4.1. Model comparison

With the cross-validation process, we benchmark the graph-based approach against a widely accepted method in the literature: the random forest algorithm. We adopt the same features that used in the GNN framework building the baseline, with specific parameters, (i.e. `max_depth=100`, `random_state=0`). Thus, it allows us to gauge the relative strengths and weaknesses of our method in comparison to conventional techniques. The comparative evaluation is consolidated in

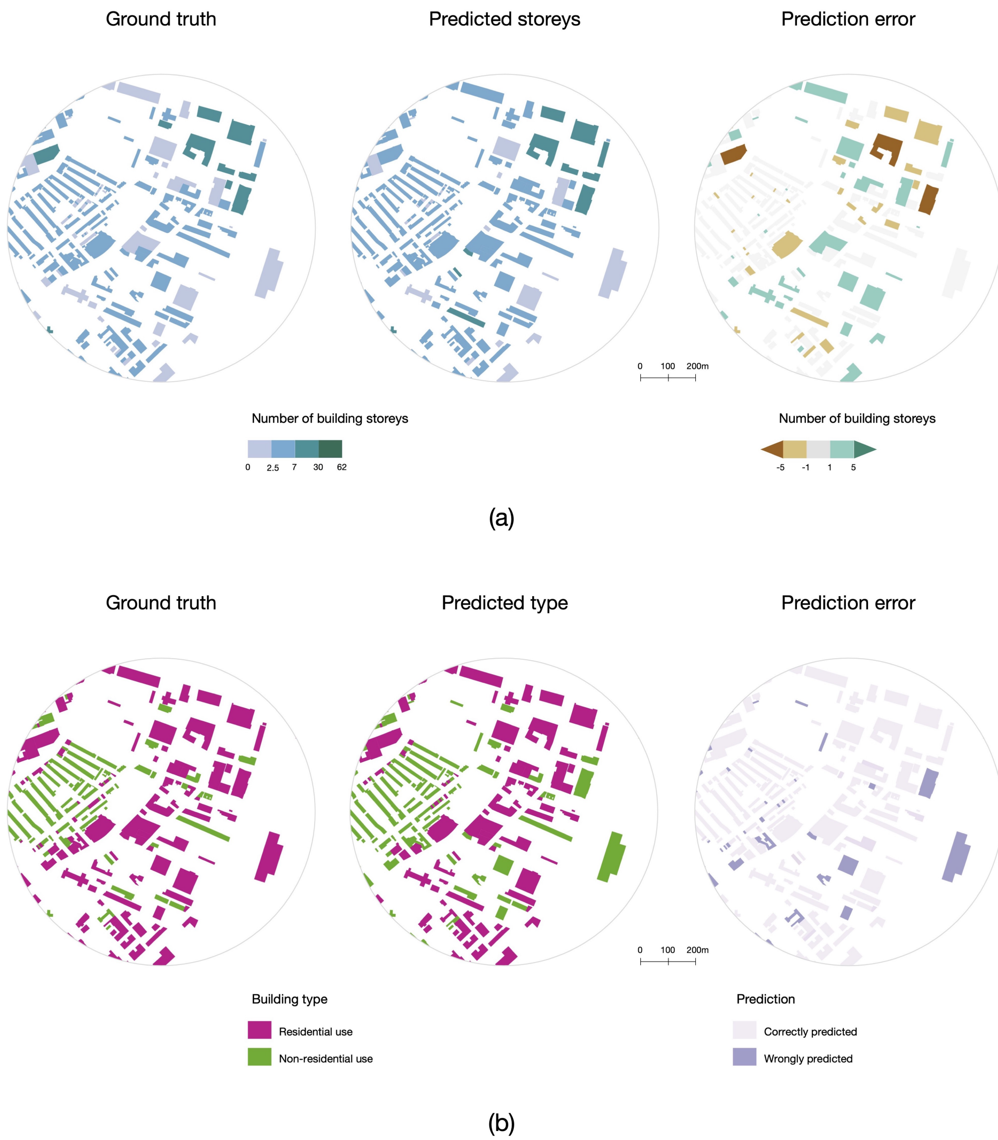


Figure 6: Geospatial map of predicted building characteristics in Boston. The top plot (a) demonstrates the actual building storeys, predicted storeys and prediction errors. The bottom one (b) depicts the actual building types, classified predictions and errors.

Table 1.

For the regression task to predict continuous target variables, the evaluation metrics are the RMSE and MAE, both of which measure the differences between predicted and actual values. In particular, such metrics are employed to assess the prediction accuracy of building storeys. Further, for the testing datasets, we compute the percentage of buildings with a prediction error range deemed acceptable for downstream studies, defined as less than 0.25, 0.5, 0.75 and 1 building storey (indicated as 'acceptable error' in Table 1). In the task of classification, our method learns to categorise buildings based on discrete class labels such as building type, construction period, and material. Performance in this task is evaluated using classification accuracy, F1-score, recall, and precision. Each metric provides insight into the model's ability to correctly classify buildings based on the aforementioned criteria.

The results confirm that our graph-based method outperforms the random forest baseline. When predicting building storeys, our model delivers an average RMSE of 1.23 storeys and an MAE of 0.76 storeys with an average median relative error of 36.5% across all three experiments. For the prediction error below one storey, our model results reveal that on average of 81.83% of buildings are of acceptable in three cities. In contrast, the baseline records an RMSE of 2.61 storeys and an MAE of 1.09 storeys, with 66.99% of buildings predicted within an error range of one storey. For building classifications, our method boasts an average accuracy of 88.33% for correctly predicted building types, 70.5% for construction period, and 68% for building materials. Outperforming the state of the art, we believe that these results present an advancement in the body of knowledge in this domain. The subsequent sections dive into more details and present the most significant results and findings.

Table 1: Comparative performance evaluation of random forest and GraphSAGE models across three experiments. Metrics include acceptable error range, median relative error, RMSE and MAE for building storeys; Accuracy (Acc), F1-score, Precision, and Recall for building type, construction period and material.

Prediction task	**Model**	**Evaluation metrics**			
Experiment 1 – Boston		RMSE	MAE	Median relative error	
Storey	Random forest	2.23	1.07	27.0%	
	GraphSAGE	1.20	0.55	20.9%	
		Acceptable error (0.25, 0.5, 0.75, 1 storey)			
	Random forest	17.0%	34.4%	50.2%	74.8%
	GraphSAGE	28.4%	45.5%	60.3%	91.6%
		Acc	F1-score	Precision	Recall
Type	Random forest	0.79	0.80	0.81	0.79
	GraphSAGE	0.90	0.85	0.81	0.89
Experiment 2 – Melbourne		RMSE	MAE	Median relative error	
Storey	Random forest	3.47	1.11	114.4%	
	GraphSAGE	1.26	0.91	55.3%	
		Acceptable error (0.25, 0.5, 0.75, 1 storey)			
	Random forest	11.5%	20.2%	27.5%	46.2%
	GraphSAGE	33.1%	50.5%	64.3%	69.9%
		Acc	F1-score	Precision	Recall
Type	Random forest	0.86	0.86	0.86	0.85
	GraphSAGE	0.90	0.86	0.87	0.87
Construction period	Random forest	0.79	0.78	0.79	0.79
	GraphSAGE	0.86	0.86	0.86	0.86
Experiment 3 – Helsinki		RMSE	MAE	Median relative error	
Storey	Random forest	2.14	1.10	34.2%	
	GraphSAGE	1.24	0.83	33.3%	
		Acceptable error (0.25, 0.5, 0.75, 1 storey)			
	Random forest	16.4%	37.1%	56.3%	80.1%
	GraphSAGE	30.6%	46.8%	60.9%	84.0%
		Acc	F1-score	Precision	Recall
Type	Random forest	0.81	0.80	0.79	0.81
	GraphSAGE	0.85	0.78	0.72	0.85
Construction period	Random forest	0.53	0.52	0.53	0.53
	GraphSAGE	0.55	0.54	0.52	0.54
Material	Random forest	0.64	0.62	0.63	0.64
	GraphSAGE	0.68	0.67	0.65	0.66

4.2. GraphSAGE outperforms random forests for storeys, but issues remain with tall buildings

The GraphSAGE model exhibits comparable performance in Boston and Helsinki, and notably outperforms the baseline model in Melbourne, when conducting the prediction of building storeys. Indeed, the RMSE values are 1.20 storeys for Boston, 1.26 storeys for Melbourne, and 1.24 storeys for Helsinki. Regardless of nuances, the results from all three experiments indicate that our method can predict the number of building storeys throughout a city with an average error margin of approximately one floor, which is sufficient for a number of use cases. In Melbourne, the RMSE is higher, likely due to unusual buildings and heterogeneous distributions of tall buildings.

To get richer insights into model performance for high- and low-rise buildings, we divide the datasets into two categories: buildings with 5 storeys or fewer and taller buildings. The model demonstrates fine precision for low-rise buildings, maintaining a prediction error margin within 1 storey, as suggested in Table 2. However, the model accuracy tends to decrease as the number of storeys increases. Across all experiments, the model consistently undervalued high-rise buildings. For instance, while the results for buildings below 5 storeys in *Experiment 1* and *Experiment 2* are within half a storey's margin of error, model performance deteriorates for taller ones, evidenced by a RMSE of 7.67 storeys and a MAE of 4.32 storeys in Boston. Interestingly, the model showcases superior accuracy in Helsinki, achieving a MAE of 1.13 storeys for high-rise buildings. Such performance aligns closely with the findings of Roy et al. (2022), where their best model recorded an MAE of 1.00 storeys for Dutch cities. Considering the disparate storey distributions across the three cities, the subset of high-storey buildings may require more training instances to improve prediction accuracy. It is inherently complex to estimate the number of storeys in higher buildings. Further, the quality and accuracy of official data on building storeys can impact the predicted results as well. For example, inaccuracies in the storey information provided by the governments as ground truth can affect the predictions.

4.3. Accuracy is very high for residential and high for non-residential buildings

The characteristic of building types is available for all three cities. Considering that our work is an illustration, we consolidated the buildings into two primary categories: residential (R) and non-residential (NR), despite variations in type classifications. We deem that the binary classification is a foundational step towards distinguishing between broadly different uses of urban space. It can be tailored to

Table 2: Model performance on storey prediction, categorised by ‘other’ buildings (with storeys less than or equal to 5) *versus* buildings and ‘tall’ buildings (with more than 5 storeys) in the three experiments.

Experiment 1 — Boston: 98.62% other buildings, 1.38% tall buildings						
	RMSE		MAE		Mean value	
	Storey <= 5	Storey >5	Storey <= 5	Storey >5	Storey <= 5	Storey >5
Random forest	0.89	7.80	0.71	4.95	2.38	9.75
GraphSAGE	0.75	7.67	0.59	4.32		
Experiment 2 — Melbourne: 87.64% other buildings, 12.36% tall buildings						
	RMSE		MAE		Mean value	
	Storey <= 5	Storey >5	Storey <= 5	Storey >5	Storey <= 5	Storey >5
Random forest	0.99	13.242	0.81	9.96	2.09	14.73
GraphSAGE	0.75	11.21	0.59	6.65		
Experiment 3 — Helsinki: 93.00% other buildings, 7.00% tall buildings						
	RMSE		MAE		Mean value	
	Storey <= 5	Storey >5	Storey <= 5	Storey >5	Storey <= 5	Storey >5
Random forest	1.39	2.31	1.05	1.30	1.91	7.00
GraphSAGE	1.09	2.22	0.81	1.13		

specific cases and contexts, e.g. with a need of inferring multiple building functions. In this study, we adopt a binary classification for its universal applications across three cities where classifying building types in different ways.

The *accuracy* is the key metric to evaluate prediction performance, defined as the fraction of buildings correctly classified based on their type. We incorporate three additional metrics to conduct a more comprehensive assessment. The *precision* quantifies the accuracy of positive predictions, while the *recall* measures the model’s ability to identify all relevant instances. The *F1-score*, derived as the harmonic mean of *precision* and *recall*, offers a consolidated measure of the model’s accuracy and recall capabilities (Goutte and Gaussier, 2005). In this context, the *precision* represents the portion of buildings accurately labelled as residential from all buildings predicted as the residential type. Conversely, the *recall* refers to the fraction of correctly predicted residential buildings out of the total number of actual residential buildings. Overall, our model discerns building types with an average accuracy of 87.9% across the three geographically disparate areas (Figure 7). In *Experiment 1*, considering the extensive dataset of local buildings, an impressive 89.5% of buildings are correctly categorised. Drawing a comparison with Atwal et al. (2022), our model obtains the highest *F1-score* of 86.0% in Melbourne, prevailing over their best model performance with an *F1-score* of 83.1%.

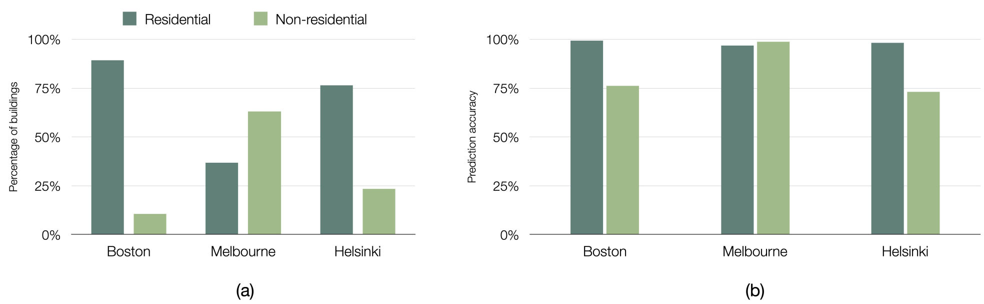


Figure 7: Comparison of predictions for building types across all three cities. (a) Data distribution for residential and non-residential buildings. (b) Prediction accuracy of two building types in each city.

4.4. *Promising accuracy for construction period and building materials, with some heterogeneity across areas*

Given the uneven distribution of construction years, we categorised building ages into five distinct classes: (0-1900], (1901-1950], (1951-1975], (1976-2000], and (2001-2023]. Such classification not only alleviates the challenges posed by unevenly distributed labels on model performance but also furnishes categories that are beneficial for downstream applications, such as estimating building energy (Aksoezen et al., 2015), in which the exact year of construction is not essential but rather the approximate era. In both *Experiment 2* and *Experiment 3*, the accuracy for predicting construction period stands at 85.7% and 55.1%, respectively. Predictions are more precise for buildings in Melbourne but occasionally misjudged the age of older buildings, as described in Figure 8. Notably, the outcomes of *Experiment 3* address that prediction performance fluctuates across cities. For instance, in Helsinki, buildings constructed post-1975 are predominantly predicted accurately, while those between the 1900s and 1950s face frequent misclassifications. The prediction of construction period may be challenged by several factors. Alongside the reliability of construction year data, the variation in sample sizes (as illustrated in Figure 8) influences prediction accuracy. Moreover, the diversity of historical buildings in terms of their locations and surroundings makes it more difficult to learn construction periods. For example, heritage buildings in Melbourne are usually mixed with modern buildings, displaying various patterns of renewal and reconstruction. Thus, we find that construction period prediction should be tailored to specific countries or cities.

In *Experiment 3*, we also venture into predicting building materials. As per official documentation, we identified five construction materials: concrete, wood, brick, steel, and others. The model accurately classifies the construction materials for 68.1% of buildings. In particular, it showcases remarkable accuracy with wood (93.8%) and concrete (87.9%) buildings. However, the accuracy decreases for brick (77.5%) and other materials (71.0%). Notably, steel-constructed buildings exhibit a prediction accuracy of only 68.8%. The challenges encountered here mirror those in the construction period prediction task, stemming from uneven sample sizes and vague material classifications, like the 'other' category.

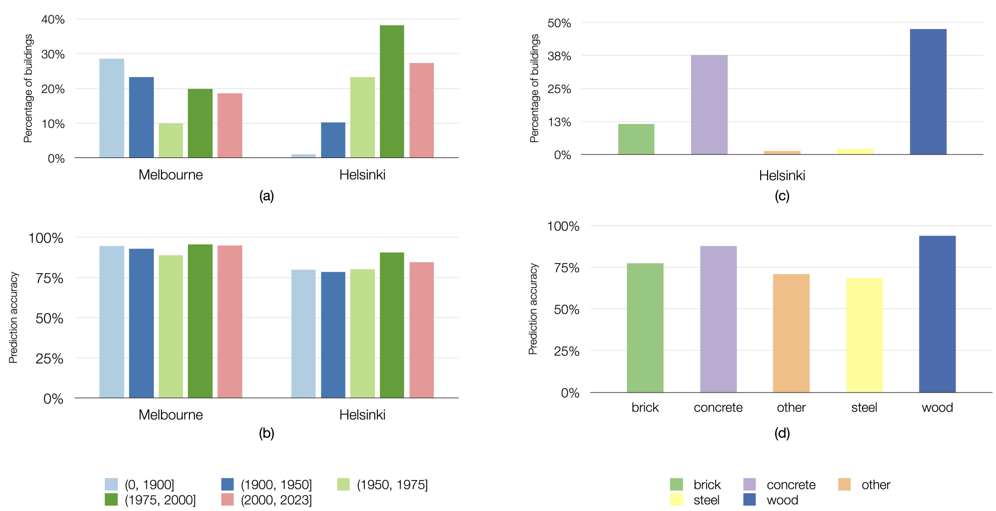


Figure 8: Comparison of predictions for construction period and building materials in Melbourne and Helsinki. (a) Distribution of buildings by construction periods for each city. (b) Prediction accuracy showcasing the classified construction period in Melbourne and Helsinki. (c) Material composition of buildings in Helsinki. (d) Prediction accuracy of each building material in Helsinki.

4.5. Limited impact on results of buffer size and objects included

To gain a deeper understanding of the importance of selected features, we conducted an ablation study in Boston to evaluate model performance by omitting certain features, allowing for an assessment of each feature's contribution. The ablation study considers two aspects: (i) how different types of street-level urban objects impact the model; (ii) how different distances of urban objects to buildings affect the model. In this regard, two groups are designed: (i) a single category of urban objects is examined at different distances from buildings; (ii) different urban objects are kept at the same distance. Table 3 summarises the results of the study.

Overall, incorporating a diverse set of urban objects enhances model performance, including only one category of urban objects, which is evident when comparing two groups. When assessing the impact of feature categories, POIs that cover multiple distances from buildings are found to prove more effective than the other two types. It may hint at the significance of services and amenities in characterising the surroundings. Meanwhile, urban objects captured within different circular buffers of 50, 200 and 500 metres around buildings play a role. Cases 4-6 describe the impact of various scales on our model performance. Indeed, urban features distributed in a broader neighbourhood (e.g. in a 500-metre buffer) are more informative than the immediate surroundings (e.g. a 50-metre buffer). The findings from the ablation study not only examine the importance of features regarding their categories and distance but also verify our initial selection of 9 predictors. While our selected features provide a solid starting point, researchers and practitioners can customise it for particular cases as well (e.g. increasing the categories of urban objects). The ablation study, however, enhances the understanding of feature significance in our method and adapts accordingly in future work. Moreover, the results imply the availability of POIs alone can aid in predicting building characteristics in a specific city, when the information regarding surrounding vegetation is missing. It indicates that future users can customise their selection of urban objects depending on the local data availability.

The ablation study attempts to interpret and explain our models by exploring the causal links between predictors and targets. However, our primary focus is on designing and evaluating a predictive framework. Therefore, we do not intend to delve deeply into model explainability in this work, despite its crucial role in interpretability (Liu et al., 2023a).

4.6. Modest training data often suffices for acceptable results

Considering the scalability and reproducibility of our method, it is important to understand its performance under varying data availability across cities. Thus, we

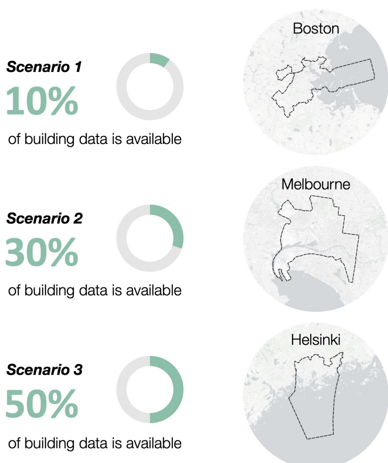
Table 3: Ablation study of the graph-based approach, evaluating two groups in the impact of urban object categories and buffer areas, using Boston as a case study.

		Impact of features selected			Impact of the buffer size			Pipeline
		Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	
POI_50		✓			✓			✓
Transport_50			✓		✓			✓
Vegetation_50				✓	✓			✓
POI_200		✓				✓		✓
Transport_200			✓			✓		✓
Vegetation_200				✓		✓		✓
POI_500		✓					✓	✓
Transport_500			✓				✓	✓
Vegetation_500				✓			✓	✓
Storey	RMSE	1.28	1.30	1.29	1.25	1.27	1.31	1.20
	MAE	0.65	0.68	0.69	0.60	0.64	0.73	0.55
Type	Accuracy	0.894	0.893	0.893	0.895	0.892	0.893	0.895

simulate three distinct scenarios: cities with 10%, 30%, and 50% data availability, relative to the total dataset. Such scenarios are designed to mirror real-world conditions where researchers and practitioners may have to work with diverse data accessibility levels. We perform three random samples, taking care to average the results over three iterations for each experiment to mitigate any anomalies caused by sampling or unbalanced label distribution. The scenarios with results are demonstrated in Figure 9. Such a case is especially important for a VGI instance such as OpenStreetMap because of its partial and varying completeness in virtually all cities around the world, e.g. from 0% to 100% completeness (Biljecki et al., 2023). Analysing the performance of our work through this lens would help understanding whether the existing data in OSM, e.g. when available for 30% of the buildings, could be used to predict the remaining values.

Overall, the results underscore its stability and applicability amidst fluctuations in data availability; however, more training data indicates better predictions. For instance, when predicting building storeys in Boston, the disparity in performance between the 10% and 50% data availability scenarios is slight, with RMSE values peaking at 1.22 and 1.27 storeys and MAE values at 0.56 and 0.71 storeys, respectively. However, for more intricate tasks like predicting construction period and materials, the accuracy diminishes as the available building data dwindles, as evidenced in *Experiment 2* and *Experiment 3*.

The inclusion of scenarios not only sheds light on the potential challenges in areas with sparse building data but also investigate the feasibility of our method for broader applications. It is particularly pertinent in areas where building information may be scarce or inaccessible. By applying the graph-based method to geographically diverse cities, we intend to provide valuable insights into its potential in heterogeneous urban contexts and across different levels of data availability.



		Scenario 1	Scenario 2	Scenario 3
Experiment 1 – Boston				
Storey	RMSE	1.27	1.25	1.22
	MAE	0.71	0.62	0.56
Type	Accuracy (%)	0.86	0.87	0.89
Experiment 2 – Melbourne				
Storey	RMSE	5.01	3.82	1.51
	MAE	2.12	1.59	1.28
Type	Accuracy (%)	0.83	0.86	0.86
Age	Accuracy (%)	0.59	0.74	0.81
Experiment 3 – Helsinki				
Storey	RMSE	1.42	1.38	1.35
	MAE	0.89	0.85	0.84
Type	Accuracy (%)	0.78	0.80	0.82
Age	Accuracy (%)	0.46	0.49	0.52
Material	Accuracy (%)	0.59	0.62	0.64

Note: each experiment was conducted three times with random sampling of training data, and the results presented are the averages over three experiments

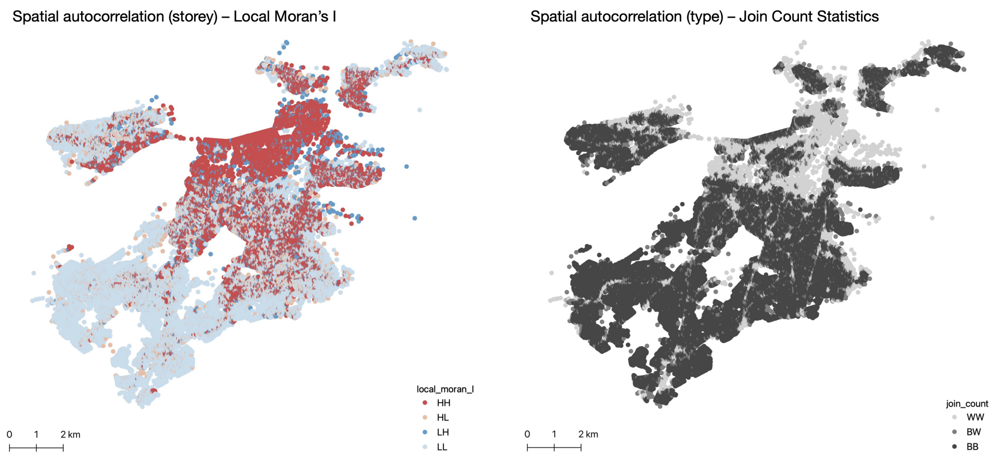
Figure 9: The impact of data availability on model performance across three cities. Each experiment was conducted three times with random sampling of training data, and the results presented are the averages over three experiments.

4.7. Spatial autocorrelation sampling enhance model robustness in predicting building storeys

To enhance the understanding of spatial heterogeneity impacts on our framework, we closely examine spatial cross-validation on a local scale. Consider the scenario where several identical buildings are present in both training and testing datasets, which could artificially inflate model accuracy due to the model's familiarity with these buildings. To address potential bias and evaluate model performance, we move beyond traditional random sampling techniques (as described in Section 4.6), implementing a spatial sampling strategy rooted in spatial autocorrelation. This approach involves spatial-based cross-validation, leveraging the concept of spatial autocorrelation, a measure that assesses the degree to which a spatial attribute correlates with itself across neighbouring locations. Spatial autocorrelation reveals how the occurrence or intensity of a phenomenon at one site mirrors that in its vicinity, manifesting either as positive (similar values clustering together) or negative (dissimilar values clustering) autocorrelation. Note that this experiment also aims to mirror the real-world process of humanitarian mapping, where mappers often begin by identifying and mapping buildings clustered in areas that exhibit shared characteristics.

Using Boston as an example, we employ two spatial analysis techniques: computing Local Moran's I for the continuous data of building storeys, and performing join counts for categorical data of building types. The findings are showcased in Figure 10. For building storeys, we apply terms HH, HL, LL, and LH to delineate distinct spatial associations identified by local Moran's I index. Here, 'HH' denotes a 'hot spot' or a clustering of similarly high values; 'LL' signifies a 'cold spot' or a clustering of similarly low values; 'HL' identifies a spatial outlier characterised by a high value amidst low-value surroundings; and 'LH' points to a spatial outlier marked by a low value amidst high-value surroundings. In parallel, for building types, 'WW' highlights clusters predominantly featuring non-Residential buildings, 'BB' denotes clusters primarily composed of residential buildings, and 'BW' signifies areas exhibiting a mixed distribution of both building types.

In our investigation of building storey predictions, we conducted three distinct experiments to assess model performance. By utilising information on building storeys from areas characterised by 'HL', 'LH', and 'HH' spatial clusters, our model aims to forecast the number of storeys in buildings situated within 'LL' clusters. The outcomes reveals significant enhancements over those presented in Table 1, underscoring the importance of incorporating geographical proximity regarding building storeys into our sampling strategy. It is an addition to acknowledging our framework performing regression tasks with notably improved results. However, when predicting building types, the classification model exhibits limited generalisation capabilities when trained exclusively on one type of data and tested on another. Notably, the model exhibits superior predictive capabilities for categorical outcomes when trained with a mix of data types, as observed in the 'BW' cluster, illustrated on the right side of the figure. This enhanced performance extends to building story prediction, where the model, trained on data from 'HL' and 'LH' clusters to predict outcomes for the 'LL' cluster, showcases exceptional accuracy. Such contrast highlights the critical role of diverse training data in enhancing the model's predictive performance across different spatial contexts.



Experiment – Boston		HH -> LL	HH -> HL	HL -> LL	HL -> LH	LH -> LL
Storey	RMSE	1.107	0.442	0.713	0.939	0.308
	MAE	0.976	0.329	0.625	0.663	0.236

Experiment – Boston		BB -> BW	BW -> WW
Type	Accuracy (%)	0.330	0.668

Figure 10: Evaluating local model performance in predicting building characteristics, by applying spatial autocorrelation in the sampling process.

5. Use cases

5.1. Computing building volume to investigate acceptable predictions for downstream modelling

Given that a set of evaluation metrics can provide more than a statistical overview of our storey predictions, we consider a downstream case to demonstrate how acceptable our prediction errors can be for specific applications and contexts. Therefore, we utilise the predicted building storeys to calculate aggregated building volume, which is crucial for facilitating, for example, building energy simulations and urban growth analysis (Teng et al., 2021; Fathy et al., 2021; Mahtta et al., 2019; Kong et al., 2012). Employing common methods found in the literature (Resch et al., 2016; Allegrini and Carmeliet, 2017), we compute the area of the 2D building footprint, then multiply by the number of storeys and the average height per storey. Taking Boston as an example, we compare the actual building volume and predicted volume by using officially documented building storeys and estimated storeys from the random forest baseline and our designed

GraphSAGE method, where we assume that one storey is 3 metres high. Further, we focus on residential buildings at a large scale, which is a topic garnering increasing attentions in recent scholar works, suggesting a profound research need. Indeed, the potential of volume information for residential buildings is not limited to enabling environmental simulations at the household level, but also supports extensive analysis related to social and economic studies (Galimshina et al., 2024). The results of comparisons are illustrated in Figure 11, representing actual volume calculation against the predicted values generated by two models. Our GraphSAGE model exhibits competitive prediction precision, as evidenced by an R^2 value of 0.76, representing that 76% of the actual volume's variance can be accounted for by the model's predictions. This high level of accuracy is further substantiated by a relatively low MAE of 1124.52 cubic metres and a RMSE of 8527.31 cubic metres, coupled with a median relative error of 19.14%. In contrast, the random forest baseline demonstrated a more modest level of accuracy with an R^2 of 0.66, indicating a substantial portion of prediction variance remained unexplained by the model. It was further corroborated by a higher MAE of 1630.50 cubic metres and a significantly greater RMSE of 10252.21 cubic metres, with a median relative error rising to 25.33%.

Such a comparison reveals that while both models possess the capability to predict building storeys in Section 4.2, the estimated storeys from our graph-based model prove superior in this instance. Further, this use case prompts a deeper investigation into downstream modelling and thus, asserts the robustness of our predictions. Such findings underscore the prediction errors of building storeys can be acceptable in a number of scenarios. However, considering various cases and contexts in which our predicted building storeys will be used for, we acknowledge that our predictions may have limitations for specific downstream use cases that require more accurate and precise height-related building information. For example, when conducting disaster management in the city, such as in case of flooding and earthquake, accurate building storeys are essential for developing plans and simulating evacuation. Therefore, our graph-based model may need improvements in future, supporting diverse downstream applications in specific contexts with more acceptable prediction errors.

5.2. Using the estimated number of storeys to generate 3D building models

The predicted building characteristics have potential to support a variety of use cases. For instance, the number of storeys can be used for generating a 3D city model, which is a common use case as it is often based on combining the

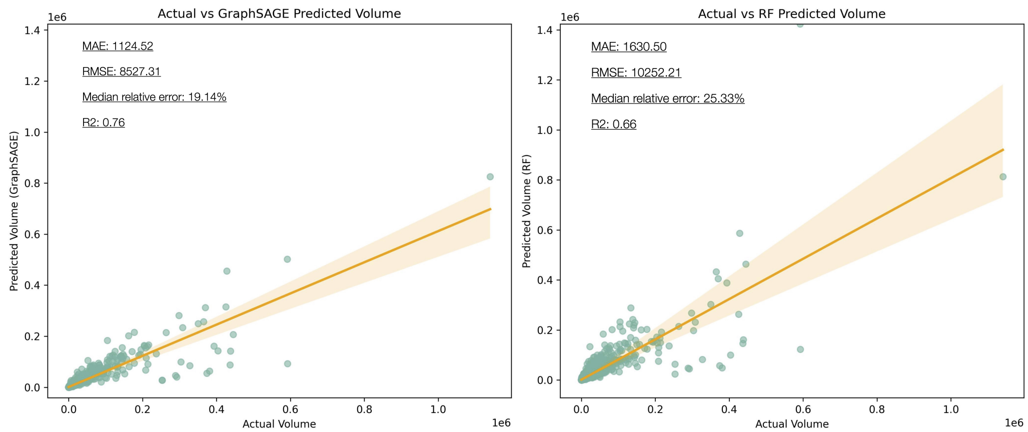


Figure 11: Comparative analysis of building volume in cubic metre, using predicted building storeys from the random forest baseline and the GraphSAGE model.

2D geometry of footprints and semantic information on heights, but latter is often missing (Ledoux et al., 2021). We take Boston as an example to demonstrate the adoption of height-related information in 3D GIS (see Figure 12). The predicted building storeys and type are processed in tabular form and associated with building footprints from OpenStreetMap. Subsequently, the dataset is converted to CityJSON LoD1 buildings (Ledoux et al., 2019) by extruding footprints according to the number of storeys, serving as a proxy for height. These block models are widely used in practice and serve a variety of purposes, from noise and wind simulations to urban planning (Pađen et al., 2022; Stoter et al., 2020; Ledoux et al., 2021; Peters et al., 2022). The city model and its building characteristics are visualised using a web viewer – ninja (Vitalis et al., 2020). This semantic city model is generated using open-source tools, and the code is publicly available on GitHub [6].

Further, given the capability of our graph-based method to infer a comprehensive set of characteristics, it can facilitate the creation of a 3D city model enriched with building semantics. With an integration of other building characteristics such as construction period and materials, the city model can be further applied for environmental simulations, e.g. building energy calculation (Teng et al., 2021; Fathy et al., 2021), serving as use cases of 3D GIS and urban digital twins. Thus, the graph-based method not only infers a variety of building characteristics but also

[6] https://github.com/binyulei/gnn-building-characteristics-prediction

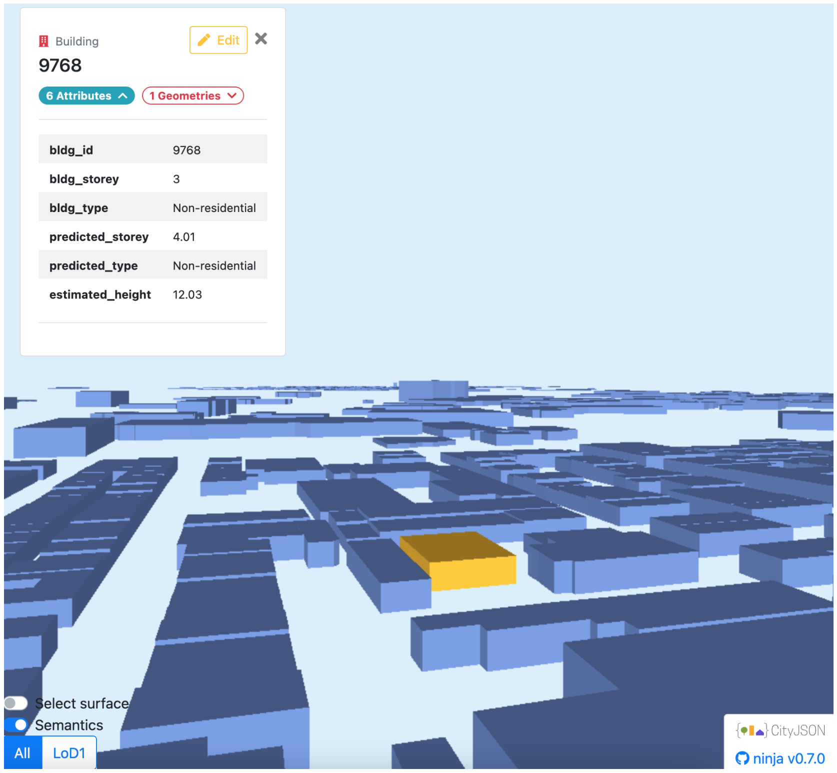


Figure 12: The generated semantic 3D building models for Boston, using predicted building storeys and type information. Our method can be used to enhance existing building datasets and increase their usability in downstream analyses.

substantiates the applicability in 3D urban studies.

6. Discussion

6.1. GNNs can model spatial links between building attributes and surrounding urban objects

Our methodology, built upon the GraphSAGE and entrenched in the geospatial distribution of street-level contexts, brings forth an innovative approach to inferring building characteristics. It explores the use of a spatially explicit GeoAI method in tackling 'vertical' urban prediction tasks, which complement traditional 3D modelling of urban environment. Given the special nature of spatial data, the First Law of Geography – '*everything is related to everything else, but near things are more related than distant things*' – suggests that valued information can be derived from spatial patterns and spatial interactions (Anselin, 1989; Tobler, 1979; Goodchild, 1987; Gould, 1981).

While many studies are indeed adopting random forest algorithm as a useful method to conduct spatial analysis, but we find such a method is rather difficult to fully exploit spatial characteristics through a detailed comparison. Indeed, the random forest baseline has limitations to explicitly incorporate spatial relationships and connectivity due to its inherent design and operational mechanics. By comprising multiple decision trees to make predictions based on the values of individual features, this design principle determines that random forest treats data points as isolated entities without considering their spatial context or the potential influence of neighbouring data points.

Spatial relationships, which denote how data points are related or influenced by their physical locations or proximity to one another, require a modelling approach that can account for such dependencies explicitly. The graph-based method allows for modelling spatial characteristics of urban objects, encoding such particularity of spatial information. Depending on the spatial connectivity between street-level urban objects and building stocks, which are often considered the fabric of the urban environment, our method successfully serves multiple urban tasks in terms of predicting building characteristics, such as building storey estimation and building type classification. These results echo the existing findings that urban fabrics are often the direct reflections of cities' images (Hull IV et al., 1994). Therefore, the results of this study show that our graph-based approach offers a well-suited and effective new approach to fill the data gap in geospatial buildings (e.g. data omission and inconsistency) by leveraging spatial networks within the context of urban morphology.

By testing our approach in three cities from three continents, we can demonstrate across a variety of contexts that our method is superior in terms of predictive performance for predicting multiple building characteristics, compared to a standard machine learning approach that does not represent space explicitly, random forests. Our model yields the most acceptable predictions in building storeys and types (performing well in all three cities, with an average of 81.83% and 88.83% respectively), compared to construction period (70.5%) and materials (68%) as discussed in Section 4.1. Moreover, examining the impact of building data availability on our method in Section 4.6 and Section 4.7, suggests that prediction errors are within an acceptable range across availability scenarios, although less building data leads to slightly worse prediction accuracy. The employment of different sampling methods allows for a confident conclusion that our graph-based framework represents predictive performance across modest training data randomly and spatially.

6.2. Limitations

The limitations of our work are discussed from two perspectives. First, the quality of OpenStreetMap data regarding inaccuracies and incompleteness may influence method performance in some cases, for example, a lack of sufficient predictors in a city of interest. The method works well in our experiments; however, variations in OSM data completeness and accuracy can indeed affect the adoption of our framework in such settings. Therefore, using OSM data to represent street-level contexts may be not enough for predicting building characteristics in specific areas. That is, some cities will be facing challenges to collecting reliable OSM data in terms of insufficient volunteered mapping activities and human-generated errors. Hence, integrating multiple types of publicly accessible data will tackle such cases when adopting our method in specific urban contexts in future. For example, street view imagery can be considered as a supplementary source to capture urban information around building stocks, as well as leveraging localised datasets when tailored our framework in cities where the reliability of OSM data is a challenge. Indeed, we believe that multiple data integration will enhance the framework for further reproducibility.

Second, the uneven distribution of target labels has an impact on prediction accuracy. For example, when gauging building storeys, our graph-based approach underpins the estimation of storeys; however, it needs more training instances to learn high-rise buildings. For example, the spatial visualisation of predicted results in Boston (Figure 6) explicitly indicates that our model tends to underestimate higher buildings with an error above 5 storeys. Meanwhile, given the larger number of residential buildings in Boston, non-residential ones are more likely to be inferred as residential buildings. Therefore, a sample size for cases that are hard to predict correctly is required to trim prediction errors, making it more digestible for the method.

7. Conclusion

Given the significance of building characteristics in the geospatial research landscape, their availability and richness are the backbone to facilitate downstream studies. However, in practice, there are often data gaps, in both authoritative (government) and crowdsourced (e.g. OpenStreetMap) datasets. In this study, we introduce a new method to predict building characteristics. We exploit the spatial connectivity between building stocks and street-level urban objects from an urban morphology perspective. Adopting Graph Neural Networks (GNNs), the GraphSAGE algorithm is used to learn morphology features and enable inductive representation in spatial connections. Departing from the state of the art, which is largely focused on predicting a single building characteristic, we design the model for handling two tasks – regression task for predicting continuous target variables (e.g. building storeys) and classification task for inferring discrete class labels (e.g. type, construction period and materials). We train and evaluate the model in three geographical areas – Boston, Helsinki and Melbourne, which have a subset of building characteristics or more comprehensive building information. Multiple cases are included to mirror the reality and validate the applicability, that is, of 10%, 30% and 50% building data available in these three cities. The evaluation metrics of the experimental studies indicate that our model performs well in various scenarios and cross-city generalisation. Meanwhile, through a comprehensive comparison to the random forest, our model achieves competitive performance in predicting multiple building characteristics; however, training graphs may require more computational resources. Nevertheless, given the initial motivation of inferring building characteristics in a spatially explicit perspective, we deem GNNs to deliver the potential to represent spatial connectivity between buildings and the surroundings. In this case, our method indicates that there remains room to investigate urban morphology information to learn building semantics, in particular from a geospatial dimension. The method, considered as a means to tackle data challenges, facilitates the completeness of building data and enables downstream work, such as generating 3D city models and enriching the properties of urban digital twins. Further, researchers and practitioners can adopt this method in customised cases considering its generic features, for example, integrating additional input datasets and tuning hyperparameters.

Declaration of Competing Interest

The authors report there are no competing interests to declare.

Data availability

The code of this work is available on GitHub: https://github.com/binyulei/gnn-building-characteristics-prediction.

Acknowledgements

We thank the members of the NUS Urban Analytics Lab for the discussions. We thank Dr Tianhong Zhao (Shenzhen Technology University) for insightful discussion in model architecture. We thank Florian Nachtigall (Technical University of Berlin) for his constructive comments and discussions on the early version of the paper. Map data is copyrighted by OpenStreetMap contributors and is available at https://www.openstreetmap.org. This research is part of the project Multi-scale Digital Twins for the Urban Environment: From Heartbeats to Cities, which is supported by the Singapore Ministry of Education Academic Research Fund Tier 1. This research is part of the project Large-scale 3D Geospatial Data for Urban Analytics, which is supported by the National University of Singapore under the Start Up Grant R-295-000-171-133.

References

Aksoezen, M., Daniel, M., Hassler, U., and Kohler, N. (2015). Building age as an indicator for energy consumption. *Energy and Buildings*, 87:74–86.

Allegrini, J. and Carmeliet, J. (2017). Coupled cfd and building energy simulations for studying the impacts of building height topology and buoyancy on local urban microclimates. *Urban Climate*, 21:278–305.

Anselin, L. (1989). What is special about spatial data? alternative perspectives on spatial data analysis (89-4).

Appleyard, D. (1969). Why buildings are known: A predictive tool for architects and planners. *Environment and Behavior*, 1(2):131–156.

Atwal, K. S., Anderson, T., Pfoser, D., and Züfle, A. (2022). Predicting building types using openstreetmap. *Scientific Reports*, 12(1):1–13.

Bandam, A., Busari, E., Syranidou, C., Linssen, J., and Stolten, D. (2022). Classification of building types in germany: A data-driven modeling approach. *Data*, 7(4):45.

Bansal, P. and Quan, S. J. (2022). Relationships between building characteristics, urban form and building energy use in different local climate zone contexts: An empirical study in seoul. *Energy and Buildings*, 272:112335.

Belgiu, M., Tomljenovic, I., Lampoltshammer, T. J., Blaschke, T., and Höfle, B. (2014). Ontology-based classification of building types detected from airborne laser scanning data. *Remote Sensing*, 6(2):1347–1366.

Bentley, J. L. (1979). Multidimensional binary search trees in database applications. *IEEE Transactions on Software Engineering*, (4):333–340.

Biljecki, F. (2020). Exploration of open data in Southeast Asia to generate 3D building models. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, VI-4/W1-2020:37–44.

Biljecki, F., Chew, L. Z. X., Milojevic-Dupont, N., and Creutzig, F. (2021). Open government geospatial data on buildings for planning sustainable and resilient cities.

Biljecki, F. and Chow, Y. S. (2022). Global building morphology indicators. *Computers, Environment and Urban Systems*, 95:101809.

Biljecki, F., Chow, Y. S., and Lee, K. (2023). Quality of crowdsourced geospatial building information: A global assessment of OpenStreetMap attributes. *Building and Environment*, 237:110295.

Biljecki, F., Ledoux, H., and Stoter, J. (2017). Generating 3d city models without elevation data. *Computers, Environment and Urban Systems*, 64:1–18.

Chen, F.-C., Subedi, A., Jahanshahi, M. R., Johnson, D. R., and Delp, E. J. (2022). Deep learning–based building attribute estimation from google street view images for flood risk assessment using feature fusion and task relation encoding. *Journal of Computing in Civil Engineering*, 36(6):04022031.

Chen, J., Yang, L., Qin, C., Yang, Y., Peng, L., and Ge, X. (2024). Heterogeneous graph traffic prediction considering spatial information around roads. *International Journal of Applied Earth Observation and Geoinformation*, 128:103709.

Chen, K., Lu, W., Xue, F., Tang, P., and Li, L. H. (2018). Automatic building information model reconstruction in high-density urban areas: Augmenting multi-source data with architectural knowledge. *Automation in Construction*, 93:22–34.

Chen, Y., Hong, T., Luo, X., and Hooper, B. (2019). Development of city buildings dataset for urban building energy modeling. *Energy and Buildings*, 183:252–265.

Chen, Y., Liu, X., Li, X., Liu, X., Yao, Y., Hu, G., Xu, X., and Pei, F. (2017). Delineating urban functional areas with building-level social media data: A dynamic time warping (dtw) distance based k-medoids method. *Landscape and Urban Planning*, 160:48–60.

De Sabbata, S. and Liu, P. (2023). A graph neural network framework for spatial geodemographic classification. *International Journal of Geographical Information Science*, pages 1–23.

Dembski, F., Wössner, U., Letzgus, M., Ruddat, M., and Yamu, C. (2020). Urban digital twins for smart cities and citizens: The case study of herrenberg, germany. *Sustainability*, 12(6):2307.

Díaz, E. and Arguello, H. (2016). An algorithm to estimate building heights from google street-view imagery using single view metrology across a representational state transfer system. In *Dimensional Optical Metrology and Inspection for Practical Applications V*, volume 9868, pages 58–65. SPIE.

Dimitrov, A. and Golparvar-Fard, M. (2014). Vision-based material recognition for automated monitoring of construction progress and generating building information modeling from unordered site image collections. *Advanced Engineering Informatics*, 28(1):37–49.

Du, S., Zhang, F., and Zhang, X. (2015). Semantic classification of urban buildings combining vhr image and gis data: An improved random forest approach. *ISPRS journal of photogrammetry and remote sensing*, 105:107–119.

Fan, H., Zipf, A., and Fu, Q. (2014). Estimation of building types on openstreetmap based on urban morphology analysis. In *Connecting a digital Europe through location and place*, pages 19–35. Springer.

Fathy, Y., Jaber, M., and Nadeem, Z. (2021). Digital twin-driven decision making and planning for energy consumption. *Journal of Sensor and Actuator Networks*, 10(2):37.

Galimshina, A., Moustapha, M., Hollberg, A., Lasvaux, S., Sudret, B., and Habert, G. (2024). Strategies for robust renovation of residential buildings in switzerland. *Nature Communications*, 15(1):2227.

Gassar, A. A. A. and Cha, S. H. (2020). Energy prediction techniques for large-scale buildings towards a sustainable built environment: A review. *Energy and Buildings*, 224:110238.

Gil, J. (2020). City information modelling: A conceptual framework for research and practice in digital urban planning. *Built Environment*, 46(4):501–527.

Goetz, M. and Zipf, A. (2012). Openstreetmap in 3d–detailed insights on the current situation in germany. In *Proceedings of the AGILE 2012 International Conference on Geographic Information Science, Avignon, France*, volume 2427, page 2427.

Goodchild, M. F. (1987). A spatial analytical perspective on geographical information systems. *International journal of geographical information system*, 1(4):327–334.

Gould, P. (1981). Letting the data speak for themselves. *Annals of the Association of American Geographers*, 71(2):166–176.

Goutte, C. and Gaussier, E. (2005). A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In *Advances in Information Retrieval: 27th European Conference on IR Research, ECIR 2005, Santiago de Compostela, Spain, March 21-23, 2005. Proceedings 27*, pages 345–359. Springer.

Gröger, G. and Plümer, L. (2012). Citygml–interoperable semantic 3d city models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 71:12–33.

Hamilton, W., Ying, Z., and Leskovec, J. (2017). Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30.

Harter, H., Willenborg, B., Lang, W., and Kolbe, T. H. (2023). Life cycle assessment of building energy systems on neighbourhood level based on semantic 3d city models. *Journal of Cleaner Production*, 407:137164.

He, Y., Yuan, C., Ren, C., Wang, W., Shi, Y., and Ng, E. (2022). Urban ventilation assessment with improved vertical wind profile in high-density cities–investigations in nighttime extreme heat. *Building and Environment*, 216:109018.

Hecht, R., Meinel, G., and Buchroithner, M. (2015). Automatic identification of building types based on topographic databases–a comparison of different data sources. *International Journal of Cartography*, 1(1):18–31.

Henn, A., Römer, C., Gröger, G., and Plümer, L. (2012). Automatic classification of building types in 3d city models: Using svms for semantic enrichment of low resolution building data. *GeoInformatica*, 16:281–306.

Hoffmann, E. J., Abdulahhad, K., and Zhu, X. X. (2023). Using social media images for building function classification. *Cities*, 133:104107.

Hong, T., Chen, Y., Luo, X., Luo, N., and Lee, S. H. (2020). Ten questions on urban building energy modeling. *Building and Environment*, 168:106508.

Huang, X., Li, S., Li, J., Jia, X., Li, J., Zhu, X. X., and Benediktsson, J. A. (2020). A multispectral and multiangle 3-d convolutional neural network for the classification of zy-3 satellite images over urban areas. *IEEE Transactions on Geoscience and Remote Sensing*, 59(12):10266–10285.

Hudson, P., Dennett, A., Russell, T., and Smith, D. (2019). Colouring london–a crowdsourcing platform for geospatial data related to london's building stock. In *Proceedings of the 27th Annual Gis Research UK Conference, Newcastle University, Newcastle, UK*, pages 23–26.

Hull IV, R. B., Lam, M., and Vigo, G. (1994). Place identity: symbols of self in the urban fabric. *Landscape and urban planning*, 28(2-3):109–120.

Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688.

Katal, A., Mortezazadeh, M., Wang, L. L., and Yu, H. (2022). Urban building energy and microclimate modeling–from 3d city generation to dynamic simulations. *Energy*, 251:123817.

Kim, J., Hatzis, J. J., Klockow, K., and Campbell, P. A. (2022). Building classification using random forest to develop a geodatabase for probabilistic hazard information. *Natural Hazards Review*, 23(3):04022014.

Kitchin, R. (2014). The data revolution: Big data, open data, data infrastructures and their consequences. *The Data Revolution*, pages 1–240.

Kolbe, T. H. and Donaubauer, A. (2021). Semantic 3d city modeling and bim. *Urban informatics*, pages 609–636.

Kong, B., Ai, T., Zou, X., Yan, X., and Yang, M. (2024). A graph-based neural network approach to integrate multi-source data for urban building function classification. *Computers, Environment and Urban Systems*, 110:102094.

Kong, F., Yin, H., Nakagoshi, N., and James, P. (2012). Simulating urban growth processes incorporating a potential model with spatial metrics. *Ecological Indicators*, 20:82–91.

Krapu, C., Hayes, N., Stewart, R., Kurte, K., Rose, A., Sorokine, A., and Urban, M. (2023). A bayesian model for multivariate discrete data using spatial and expert information with application to inferring building attributes. *Spatial Statistics*, 55:100745.

Kumar, S., Pal, S. K., and Singh, R. P. (2018). A novel method based on extreme learning machine to predict heating and cooling load through design and structural attributes. *Energy and Buildings*, 176:275–286.

Ledoux, H., Arroyo Ohori, K., Kumar, K., Dukai, B., Labetski, A., and Vitalis, S. (2019). Cityjson: A compact and easy-to-use encoding of the citygml data model. *Open Geospatial Data, Software and Standards*, 4(1):1–12.

Ledoux, H., Biljecki, F., Dukai, B., Kumar, K., Peters, R., Stoter, J., and Commandeur, T. (2021). 3dfier: automatic reconstruction of 3D city models. *Journal of Open Source Software*, 6(57):2866.

Lei, B., Stouffs, R., and Biljecki, F. (2023). Assessing and benchmarking 3d city models. *International Journal of Geographical Information Science*, 37(4):788–809.

Li, J., Huang, X., Tu, L., Zhang, T., and Wang, L. (2022). A review of building detection from very high resolution optical remote sensing images. *GIScience & Remote Sensing*, 59(1):1199–1225.

Li, Y., Chen, Y., Rajabifard, A., Khoshelham, K., and Aleksandrov, M. (2018). Estimating building age from google street view images using deep learning (short paper). In *10th international conference on geographic information science (GIScience 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Liang, X., Zhao, T., and Biljecki, F. (2023). Revealing spatio-temporal evolution of urban visual environments with street view imagery. *Landscape and Urban Planning*, 237:104802.

Liasis, G. and Stavrou, S. (2016). Satellite images analysis for shadow detection and building height estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 119:437–450.

Liu, P. and Biljecki, F. (2022). A review of spatially-explicit geoai applications in urban geography. *International Journal of Applied Earth Observation and Geoinformation*, 112:102936.

Liu, P., Yan, Z., and Biljecki, F. (2023a). Explainable spatially explicit geospatial artificial intelligence in urban analytics. *Environment and Planning B: Urban Analytics and City Science*.

Liu, P., Zhao, T., Luo, J., Lei, B., Frei, M., Miller, C., and Biljecki, F. (2023b). Towards human-centric digital twins: Leveraging computer vision and graph models to predict outdoor comfort. *Sustainable Cities and Society*, 93:104480.

Lu, Z., Im, J., Rhee, J., and Hodgson, M. (2014). Building type classification using spatial and landscape attributes derived from lidar remote sensing data. *Landscape and Urban Planning*, 130:134–148.

Lynch, K. (1964). *The image of the city*. MIT press.

Mahtta, R., Mahendra, A., and Seto, K. C. (2019). Building up or spreading out? typologies of urban growth across 478 cities of 1 million+. *Environmental Research Letters*, 14(12):124077.

Mai, G., Janowicz, K., Hu, Y., Gao, S., Yan, B., Zhu, R., Cai, L., and Lao, N. (2022). A review of location encoding for geoai: methods and applications. *International Journal of Geographical Information Science*, 36(4):639–673.

Meng, C., Song, Y., Ji, J., Jia, Z., Zhou, Z., Gao, P., and Liu, S. (2022). Automatic classification of rural building characteristics using deep learning methods on oblique photography. In *Building Simulation*, pages 1–14. Springer.

Milojevic-Dupont, N., Hans, N., Kaack, L. H., Zumwald, M., Andrieux, F., de Barros Soares, D., Lohrey, S., Pichler, P.-P., and Creutzig, F. (2020). Learning from urban form to predict building heights. *Plos one*, 15(12):e0242010.

Nachtigall, F., Milojevic-Dupont, N., Wagner, F., and Creutzig, F. (2023). Predicting building age from urban form at large scale. *Computers, Environment and Urban Systems*, 105:102010.

Ning, H., Li, Z., Ye, X., Wang, S., Wang, W., and Huang, X. (2022). Exploring the vertical dimension of street view image based on deep learning: a case study on lowest floor elevation estimation. *International Journal of Geographical Information Science*, 36(7):1317–1342.

Nouvel, R., Zirak, M., Coors, V., and Eicker, U. (2017). The influence of data quality on urban heating demand modeling using 3d city models. *Computers, Environment and Urban Systems*, 64:68–80.

Pađen, I., García-Sánchez, C., and Ledoux, H. (2022). Towards automatic reconstruction of 3D city models tailored for urban flow simulations. *Frontiers in Built Environment*, 8(899332).

Palliwal, A., Song, S., Tan, H. T. W., and Biljecki, F. (2021). 3d city models for urban farming site identification in buildings. *Computers, Environment and Urban Systems*, 86:101584.

Pang, H. E. and Biljecki, F. (2022). 3D building reconstruction from single street view images using deep learning. *International Journal of Applied Earth Observation and Geoinformation*, 112:102859.

Park, Y. and Guldmann, J.-M. (2019). Creating 3d city models with building footprints and lidar point cloud classification: A machine learning approach. *Computers, environment and urban systems*, 75:76–89.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830.

Peters, R., Dukai, B., Vitalis, S., van Liempt, J., and Stoter, J. (2022). Automated 3d reconstruction of lod2 and lod1 models for all 10 million buildings of the netherlands. *Photogrammetric Engineering & Remote Sensing*, 88(3):165–170.

Resch, E., Bohne, R. A., Kvamsdal, T., and Lohne, J. (2016). Impact of urban density and building height on energy use in cities. *Energy Procedia*, 96:800–814.

Roy, E., Pronk, M., Agugiaro, G., and Ledoux, H. (2022). Inferring the number of floors for residential buildings. *International Journal of Geographical Information Science*, pages 1–25.

Schrotter, G. and Hürzeler, C. (2020). The digital twin of the city of zurich for urban planning. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 88(1):99–112.

Silva, M., Oliveira, V., and Leal, V. (2017). Urban form and energy demand: A review of energy-relevant urban attributes. *Journal of Planning Literature*, 32(4):346–365.

Stoter, J., Peters, R., Commandeur, T., Dukai, B., Kumar, K., and Ledoux, H. (2020). Automated reconstruction of 3D input data for noise simulation. *Computers, Environment and Urban Systems*, 80:101424.

Sun, M., Zhang, F., Duarte, F., and Ratti, C. (2022a). Understanding architecture age and style through deep learning. *Cities*, 128:103787.

Sun, Y. and Gu, Z. (2022). Using computer vision to recognize construction material: A trustworthy dataset perspective. *Resources, Conservation and Recycling*, 183:106362.

Sun, Y., Mou, L., Wang, Y., Montazeri, S., and Zhu, X. X. (2022b). Large-scale building height retrieval from single sar imagery based on bounding box regression networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 184:79–95.

Teng, S. Y., Touš, M., Leong, W. D., How, B. S., Lam, H. L., and Máša, V. (2021). Recent advances on industrial data-driven energy savings: Digital twins and infrastructures. *Renewable and Sustainable Energy Reviews*, 135:110208.

Tobler, W. R. (1979). Cellular geography. *Philosophy in geography*, pages 379–386.

Tooke, T. R., Coops, N. C., and Webster, J. (2014). Predicting building ages from lidar data with random forests for building energy modeling. *Energy and Buildings*, 68:603–610.

Tu, J., Sui, H., Feng, W., and Song, Z. (2016). Automatic building damage detection method using high-resolution remote sensing images and 3d gis model. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 3(8).

Usui, H. (2023a). Comparison of precise and approximated building height: Estimation from number of building storeys and spatial variations in the tokyo metropolitan region. *Environment and Planning B: Urban Analytics and City Science*, 50(2):487–499.

Usui, H. (2023b). Relative spatial variability in building heights and its spatial association: Application for the spatial clustering of harmonious and inharmonious building heights in tokyo. *Environment and Planning B: Urban Analytics and City Science*, page 23998083231204691.

Vitalis, S., Labetski, A., Boersma, F., Dahle, F., Li, X., Arroyo Ohori, K., Ledoux, H., and Stoter, J. (2020). Cityjson+ web= ninja. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 6:167–173.

Wang, L., Ma, G., Zhou, F., Liu, Y., and Tian, T. (2019). Multicriteria decision-making approach for selecting ventilation heat recovery devices based on the attributes of buildings and the preferences of decision makers. *Sustainable Cities and Society*, 51:101753.

Wang, W., Xu, Y., Ng, E., and Raasch, S. (2018). Evaluation of satellite-derived building height extraction by cfd simulations: a case study of neighborhood-scale ventilation in hong kong. *Landscape and Urban Planning*, 170:90–102.

Wang, Y., Chau, C. K., Ng, W., and Leung, T. (2016). A review on the effects of physical built environment attributes on enhancing walking and cycling activity levels within residential neighborhoods. *Cities*, 50:1–15.

Wang, Z., Jiang, L., Lin, L., and Yu, W. (2015). Building height estimation from high resolution sar imagery via model-based geometrical structure prediction. *Progress In Electromagnetics Research M*, 41:11–24.

Wurm, M., Schmitt, A., and Taubenböck, H. (2015). Building types' classification using shape-based features and linear discriminant functions. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(5):1901–1912.

Xia, B., Li, X., Shi, H., Chen, S., and Chen, J. (2020). Style classification and prediction of residential buildings based on machine learning. *Journal of Asian Architecture and Building Engineering*, 19(6):714–730.

Xia, H., Liu, Z., Efremochkina, M., Liu, X., and Lin, C. (2022). Study on city digital twin technologies for sustainable smart city design: A review and bibliometric analysis of geographic information system and building information modeling integration. *Sustainable Cities and Society*, 84:104009.

Xu, G., Su, J., Xia, C., Li, X., and Xiao, R. (2022a). Spatial mismatches between nighttime light intensity and building morphology in shanghai, china. *Sustainable Cities and Society*, 81:103851.

Xu, X., Wang, W., Hong, T., and Chen, J. (2019). Incorporating machine learning with building network analysis to predict multi-building energy use. *Energy and Buildings*, 186:80–97.

Xu, Z., Zhang, F., Wu, Y., Yang, Y., and Wu, Y. (2022b). Building height calculation for an urban area based on street view images and deep learning. *Computer-Aided Civil and Infrastructure Engineering*.

Yan, X., Ai, T., Yang, M., and Yin, H. (2019). A graph convolutional neural network for classification of building patterns using spatial vector data. *ISPRS journal of photogrammetry and remote sensing*, 150:259–273.

Yan, Y. and Huang, B. (2022). Estimation of building height using a single street view image via deep neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 192:83–98.

Ye, X., Du, J., Han, Y., Newman, G., Retchless, D., Zou, L., Ham, Y., and Cai, Z. (2023). Developing human-centered urban digital twins for community infrastructure resilience: A research agenda. *Journal of Planning Literature*, 38(2):187–199.

Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M., and Li, H. (2019a). T-gcn: A temporal graph convolutional network for traffic prediction. *IEEE transactions on intelligent transportation systems*, 21(9):3848–3858.

Zhao, T., Huang, Z., Tu, W., He, B., Cao, R., Cao, J., and Li, M. (2022). Coupling graph deep learning and spatial-temporal influence of built environment for short-term bus travel demand prediction. *Computers, Environment and Urban Systems*, 94:101776.

Zhao, Y., Qi, J., and Zhang, R. (2019b). Cbhe: Corner-based building height estimation for complex street scene images. In *The World Wide Web Conference*, pages 2436–2447.

Zhu, D., Liu, Y., Yao, X., and Fischer, M. M. (2021). Spatial regression graph convolutional neural networks: A deep learning paradigm for spatial multivariate distributions. *GeoInformatica*, pages 1–32.

Zhu, W., Zhang, L., Mei, S.-J., and Yuan, C. (2022). Developing urban heat mitigation strategies for a historic area using a high-fidelity parametric numerical simulation: A case study in singapore. *Buildings*, 12(9):1311.