



UST-GNN: A unified spatial–topological graph neural network framework for urban analytics demonstrated through a case study on urban health prediction[☆]

Minwei Zhao^{a,c,1}, Sanja Šćepanović^{b,d}, Stephen Law^{c,*,*}, Ivica Obadić^{f,g}, Cai Wu^a, Daniele Quercia^{b,e}

^a The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China

^b Nokia Bell Labs, Cambridge, United Kingdom

^c University College London, London, United Kingdom

^d University of Oxford, Oxford, United Kingdom

^e Politecnico di Torino, Turin, Italy

^f Technical University of Munich, Munich, Germany

^g Munich Center for Machine Learning, Munich, Germany

ARTICLE INFO

Keywords:

Urban health
Graph neural networks
Health prescription modelling
Spatial data science
Environmental inequality
Urban analytics
Geographic information science

ABSTRACT

Understanding how social, demographic, environmental, and spatial factors jointly shape urban outcomes is essential for sustainable urban development and evidence-based policy. Traditional statistical approaches often struggle to capture complex non-linear relationships, while many machine learning methods overlook the joint roles of spatial autocorrelation and network topology in urban systems. Recent advances in GeoAI have addressed these challenges only partially, often treating spatial effects, graph structure, evaluation, and interpretability separately. We present **UST-GNN**, a unified spatial–topological graph neural network framework that integrates neighbourhood connectivity, heterogeneous urban features, and positional/location embeddings into a single representation. Using the MedSAT dataset, which contains over 150 environmental and socio-demographic variables and six prescription outcomes across 4835 neighbourhoods in Greater London, UST-GNN outperforms strong statistical, geographically enhanced, and graph Machine Learning baselines, improving out-of-sample R^2 by 8.4–13.2% under strict spatial cross-validation. We further introduce a lightweight principal-component module to interpret learned node embeddings geographically and relate them to policy-relevant covariates. The resulting analyses recover established patterns, offer new perspectives on debated associations, and reveal novel predictors warranting further causal investigation. Together, these findings demonstrate the value of graph-based spatial machine learning for urban health analytics, environmental inequality assessment, and evidence-based urban policy. Beyond predictive gains, UST-GNN provides a unified GeoAI analytical pipeline that can be embedded into urban digital twin workflows for scenario testing, monitoring, and data-informed decision-making for healthier, more sustainable cities.

1. Introduction

Urban outcomes, including health, are shaped by the complex interplay of socio-demographic, environmental, and infrastructural factors, making their systematic study essential for sustainable urban development and informed public policy. Recent global trends—such as rapid urbanisation (Vlahov & Galea, 2002), the increasing frequency

and severity of extreme weather events (Bell et al., 2018), and pandemics such as COVID-19 (Capolongo et al., 2020) have heightened the urgency of understanding how urban characteristics influenced population health and contributed to existing inequalities (Cole et al., 2021; Fiscella & Williams, 2004). As cities worldwide strove to meet

[☆] This article is part of a Special issue entitled: ‘Digital Twins and AI for Cities’ published in Computers, Environment and Urban Systems.

* Correspondence to: UCL Geography, North West Wing, London WC1E 6BT, United Kingdom.

E-mail address: stephen.law@ucl.ac.uk (S. Law).

¹ Currently affiliated with the Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China.

the United Nations Sustainable Development Goals (SDGs), particularly those related to good health and wellbeing (SDG 3) and sustainable cities and communities (SDG 11), policymakers and planners increasingly required systematic, spatially explicit, and evidence-based insights to guide equitable urban planning and targeted interventions (Batty, 2018; Corburn, 2004; Obradovich et al., 2018).

Urban and public health studies have traditionally employed statistical methodologies (Alegría et al., 2018; Cousens et al., 2011; Hayat et al., 2017) and, more recently, machine learning (ML) approaches (dos Santos et al., 2019; Maitra et al., 2025; Šćepanović et al., 2023b) to explore relationships between the urban environment and health outcomes. Recent advances in Artificial Intelligence (AI) and Digital Twin technologies are increasingly transforming urban analytics by coupling data-driven models with dynamic, real-time city representations. This convergence enables continuous simulation and feedback between physical and virtual urban systems, supporting proactive responses to health, mobility, and environmental challenges (Birks et al., 2020; Malleson et al., 2024). Integrating GeoAI models into digital-twin frameworks offers new opportunities for evidence-based planning and scenario evaluation towards healthy and sustainable cities.

Despite these advances, important methodological limitations remain. First, conventional statistical models often assumed linear relationships for interpretability, limiting their ability to capture complex non-linear interactions among environmental exposures, socio-economic conditions, and other urban factors (Kihal-Talantikite et al., 2013; Lalloué et al., 2013; Pineo et al., 2018; Song et al., 2004). Second, many machine learning approaches overlooked the simultaneous spatial dependencies and network topology inherent in urban systems, failing to account for neighbourhood severance and structural linkages such as transport or social networks that may propagate risks (Arcaya et al., 2016; Sarkar & Webster, 2017). These omissions can reduce predictive accuracy and hinder understanding of spatial disparities (Ijeh et al., 2024). Third, although location encoder has improved the representation of spatial context in machine learning by embedding geographic information directly into feature space (Klemmer et al., 2025; Mai et al., 2020), such methods remain rarely integrated with architectures that jointly model spatial proximity and topological connectivity in urban health analytics. Graph Neural Networks (GNNs) offer a promising avenue for this challenge because they are well suited to graph-structured data and can capture complex spatial-topological relationships, with demonstrated success in domains including social networks, molecular chemistry, recommender systems, and, more recently, urban informatics tasks such as traffic prediction and urban function classification (Gao et al., 2024; Li et al., 2023; Liu & Liu, 2024; Sun et al., 2022; Wu et al., 2022; Zhang et al., 2023). However, their application to urban health remains limited, partly because their black-box nature constrains uptake among public health researchers and policymakers (Kjellstrom, 2016; Liu et al., 2024). This is especially important as urban and public health research increasingly calls for machine learning and AI methods with strong predictive performance (Gilardi et al., 2021; Kong et al., 2023; Mhasawade et al., 2021). Existing post hoc explainability methods, such as GNNExplainer and gradient-based attribution, can identify influential nodes, edges, and features, but they typically provide instance-specific explanations rather than global, policy-relevant concepts for urban analytics.

In summary, recent GeoAI advances have largely addressed these geospatial challenges in isolation, modelling spatial effects through location encoders, topological dependencies through graph-based learning, spatial leakage through spatial cross-validation, and interpretability through post hoc explanation frameworks such as GNNExplainer. Few studies have unified or systematically ablated these components within a single spatially explicit GeoAI pipeline designed to generate health-policy-relevant insights. To address this gap, we propose **UST-GNN**, a unified spatially and topologically explicit graph neural network framework for urban analytics, demonstrated here through

a case study on health outcomes. UST-GNN combines position/location encoders, neighbourhood connectivity, and diverse urban features within a unified graph representation to model interactions among population, environmental, and spatial variables. Using a spatial graph constructed from neighbourhood-level data, it integrates heterogeneous features, including air quality, greenness, climate, and population measures, together with positional and locational embeddings, to predict neighbourhood-level prescription outcomes within an ecological study design. UST-GNN also introduces a lightweight interpretability module that summarises the principal components of the learned representation, identifies spatial patterns, and links these components back to policy-relevant input features. We show that UST-GNN outperforms state-of-the-art baselines under strict spatial cross-validation while generating interpretable insights that align with, challenge, and extend current urban health knowledge. Our contributions are threefold:

1. We developed **UST-GNN**, a unified spatial-topological graph neural network pipeline that explicitly captured geographic dependencies and neighbourhood topology for urban analytics. Through systematic spatial ablation analyses, we rigorously evaluated alternative graph construction strategies (e.g., K -hop adjacency, K -nearest neighbours), diverse positional and locational encoding techniques (e.g., Laplacian, random-walk positional encodings, and SatCLIP locational encodings (Klemmer et al., 2025)), multiple GNN architectures (GCN, GIN, GraphSAGE, GATv2) and multiple spatial cross validation strategies. This established a spatially-explicit methodological pipeline for integrating graph machine learning into urban analytics research, including but not limited to health applications (Section 4.1).
2. We applied UST-GNN to the MedSAT dataset (Šćepanović et al., 2023b), a recently released urban health resource compiling over 150 high-resolution environmental and socio-demographic indicators, together with six health outcome measures (per-capita prescription rates for depression, anxiety, diabetes, hypertension, asthma, and opioid use) across 4835 neighbourhoods in Greater London. UST-GNN achieved an average error reduction exceeding 10% compared to competitive baselines, outperforming gradient-boosted trees, Spatial Lag Models, Geographically Weighted Regression, and geo-informed machine learning methods (Grekousis, 2025; Maitra et al., 2025; Šćepanović et al., 2023b). These results confirmed UST-GNN's ability to capture complex non-linear effects and spatial-topological effects that were often overlooked by conventional spatial statistical models and standard ML approaches (Section 4.2).
3. Using antidepressant prescription rates as a representative case study, we developed a novel lightweight module that integrates principal component analysis (PCA) with spatial and socio-economic correlation analysis. The module compresses the learned graph representations into two intrinsic components, maps them geographically to reveal spatial structure, and relates each component back to health-policy-relevant input features. Antidepressants were selected because they have been extensively studied in prior work, enabling comparison with the existing literature. Unlike commonly used interpretability techniques, such as feature permutation, gradient based attribution, or GNNExplainer (see Appendix B), which primarily operate at the input or node level, our approach interprets the dominant component and its global associations with policy-relevant input features, which are often overlooked by local perturbation or feature-wise methods. By relating the intrinsic components back to the original input features, our framework recovered socio-economic and demographic relationships consistent with findings from more interpretable models in health research (e.g., lower prescriptions among individuals aged 20–39, commuters, and those in professional employment), provided new perspectives on debated associations (e.g., the paradoxical

relationship between increased urban greenery and higher antidepressant use), and identified novel spatial predictors (e.g., a negative correlation between canopy evaporation and depression prescriptions) that may warrant further causal investigation (Section 4.3). Discovering such health-policy-relevant insights from a complex GeoAI model represents a key contribution of this study, demonstrating how advanced spatial deep learning architectures can be linked to actionable evidence for urban health decision-making.

2. Literature review

2.1. Urban socio-demographic and environmental correlates of outcomes

A substantial body of research has examined how urban socio-demographic and environmental characteristics are linked to a range of outcomes, particularly public health and related inequalities. In the health domain, Kihal-Talantikite et al. (2013) demonstrated that low-income populations residing in environmentally disadvantaged neighbourhoods faced compounded risks, especially in contexts of pronounced social inequities. Similarly, access to urban green and blue spaces has been consistently associated with improved health outcomes; for example, Zhou et al. (2023) found that increased availability of green spaces and waterways significantly mitigated hypertension risks. Other studies identified strong associations between urban environmental conditions and chronic health outcomes (Hunter et al., 2023), including depression and diabetes (Mazumdar et al., 2021; Van Draanen et al., 2020; White et al., 2021). Beyond health, socio-demographic and environmental factors have also been linked to variations in mobility, environmental quality, and social well-being (Gao et al., 2022; Yue et al., 2024). Collectively, these findings underscored the critical role that environmental exposures and socio-economic contexts played in shaping urban disparities. However, many existing studies tended to adopt a relatively narrow analytical scope, often investigating single factors in isolation rather than addressing the broader, multifaceted urban context. As a result, the complex interactions and potential non-linear effects among multiple urban characteristics remained insufficiently explored, limiting comprehensive understanding of both urban health and other spatially distributed outcomes.

2.2. Machine learning applications in urban analytics and health research

The advent of extensive, high-dimensional urban datasets, including those derived from remote sensing, wearable technology, crowdsourced platforms, and administrative records, has accelerated the adoption of machine learning (ML) techniques in urban analytics, with public health emerging as a key application domain (Ohanyan et al., 2022). Traditional statistical models, while valuable, are often constrained by assumptions of linearity and independence among variables, which limits their ability to capture the complex, non-linear interactions common in urban health contexts (Pineo et al., 2018; Rothenberg et al., 2014; Shane & Graedel, 2000; Song et al., 2004). For instance, Laloué et al. (2013) illustrated that linear models struggled to account for intricate urban health patterns arising from interdependent variables. In contrast, contemporary ML methods including random forests, gradient boosting, and neural networks offered enhanced flexibility, automatically capturing complex interactions and non-linearities inherent to diverse urban datasets (Jordan & Mitchell, 2015). Recent studies have successfully employed such algorithms to predict diverse health outcomes, including depression incidence, diabetes prevalence, and hypertension risks, based on multifaceted urban data sources (Bhakta & Sau, 2016; Su et al., 2021; Thotad et al., 2023).

Nevertheless, conventional ML approaches typically neglected the geographic context essential for accurately modelling community-level phenomena (Arcaya et al., 2016; Sarkar & Webster, 2017). Standard

ML treated spatial units, such as neighbourhoods, as independent entities, overlooking spatial heterogeneity and dependencies. Factors such as the influence of adjacent areas, the presence of spatial barriers (e.g., rivers or highways), and spatially structured socio-economic gradients could significantly impact health and other urban outcomes, yet were frequently ignored (Comber et al., 2011; Shrestha et al., 2016). While spatial statistical methods, including spatial lag models and geographically weighted regression (GWR), explicitly accounted for spatial interactions, they often exhibited limitations in scalability and in capturing non-linear spatial relationships in high-dimensional contexts (Cheng et al., 2021; Zhao & Pu, 2025). Moreover, the limited interpretability of many ML models remained an ongoing challenge. In urban health applications, understanding why particular regions exhibited elevated risks was as critical as achieving accurate predictions, in order to inform targeted and effective interventions (Caruana et al., 2015; Rudin, 2019). The lack of transparency inherent in many “black box” models thus constrained their acceptance and broader use in both public health and other urban policy domains (Wirtz et al., 2021).

2.3. GeoAI and graph neural networks in urban analytics and health research

Recognising the significance of spatial context in urban studies, the emerging field of Geographic Artificial Intelligence (GeoAI) has integrated geographical information science (GIScience) with advanced machine learning and AI methodologies (Klemmer et al., 2025; Mai et al., 2020; Scheider & Richter, 2023; VoPham et al., 2018; Wang et al., 2025). GeoAI approaches have shown considerable potential in modelling location-based factors and spatial relationships, with applications spanning public health, environmental monitoring, and mobility analysis. In the health domain, research has underscored how urban spatial heterogeneity — including variations in built and social environments — contributed to disparities in health outcomes (Tung et al., 2017). The COVID-19 pandemic further highlighted the critical role of spatial context, prompting greater attention to neighbourhood characteristics and mobility patterns in analysing disease transmission and health impacts (Krenz et al., 2023; Li et al., 2023).

Within the GeoAI landscape, Graph Neural Networks (GNNs) have emerged as a particularly promising approach, capable of naturally capturing network topology and spatial relationships. In such frameworks, urban areas can be explicitly represented as interconnected nodes, with edges encoding spatial proximity, infrastructure connectivity, or socio-economic interactions (Klemmer et al., 2023; Li et al., 2021; Wu et al., 2021). GNN-based methods have been successfully applied to complex phenomena such as human mobility modelling and infectious disease dynamics, both of which involve pronounced network effects (Choi et al., 2020). Helbich (2018) highlighted the value of integrating diverse spatial datasets (e.g., environmental, infrastructural, socio-economic) to explain disparities in urban outcomes, a task well suited to graph-based approaches given their ability to model heterogeneous, relational data structures (De Sabbata & Liu, 2023; Mai et al., 2020).

Despite this promise, the use of GNNs in urban health research remained limited, in large part due to challenges of interpretability. The complexity of GNN architectures often obscured intuitive understanding, posing a barrier for domain experts who require model reasoning to align with established theoretical and empirical knowledge. Existing applications of GNNs to health contexts, such as those by Lu and Uddin (2021) and Fritz et al. (2022), have largely emphasised predictive accuracy or epidemiological modelling, rather than describing the underlying socio-environmental drivers of health disparities. As a result, the potential of GNNs to combine predictive performance with domain-relevant interpretability in urban health analytics has remained underexplored. Recent post hoc explainability techniques, including GNNExplainer and gradient-based attribution, seek to isolate the minimal set of nodes, edges, and features that most influence a model's

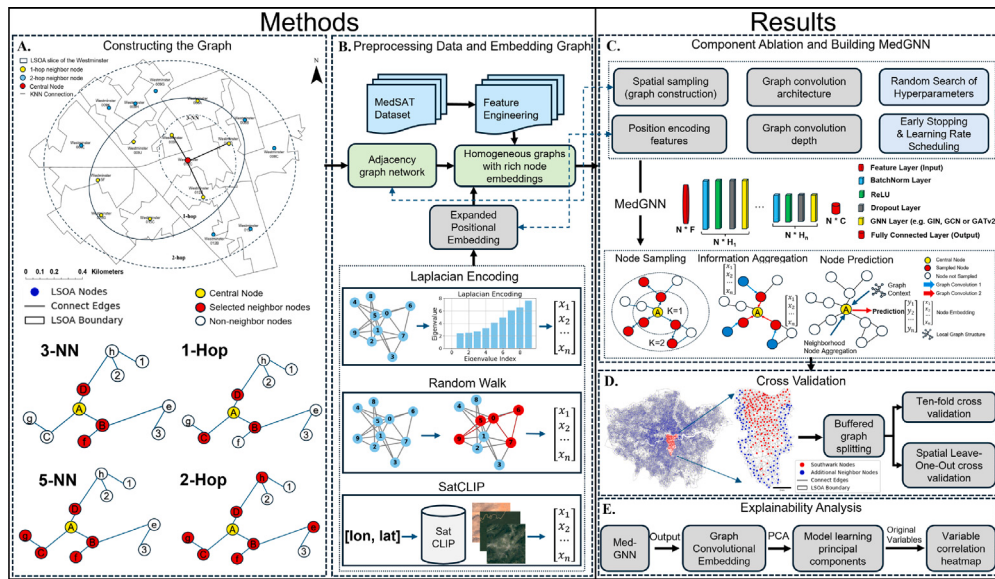


Fig. 1. UST-GNN methodology overview (Steps A–E).

prediction. However, these approaches largely explain the black box after training via saliency masks that are instance-specific rather than interpreting global policy-relevant concepts for health urban analytics.

To address this gap, we proposed **UST-GNN**, a unified spatial and topologically explicit GNN framework designed for urban analytics, demonstrated here through a case study on health outcomes (per capita medical prescription distribution). UST-GNN combines robust predictive capabilities with interpretable spatial insights, enabling the identification of neighbourhood-level patterns and feature associations relevant in health policies. This approach aligns with core GeoAI objectives, providing a pathway to jointly address the intertwined challenges of modelling non-linear relationships, capturing spatial dependencies, and delivering interpretable outputs to support sustainable urban research and policy holistically. Moreover, such graph-based GeoAI approaches can form the analytical backbone of emerging urban digital twins, integrated virtual representations of cities, that enable continuous monitoring, simulation, and policy testing (Malleson et al., 2024). Embedding models like UST-GNN could facilitate dynamic assessment of urban health and sustainability scenarios, supporting real-time, data-informed decision-making for healthier and more sustainable cities.

3. Methodology

Our methodology comprised five steps (Fig. 1): (A) spatial graph construction, (B) data integration and positional/location encoding, (C) GNN architecture development and ablation, (D) model evaluation, and (E) interpretability analysis. We describe each step below.

3.1. Study area

Our study focused on Greater London in the United Kingdom, comprising 32 London boroughs and the City of London (Fig. 2). The analysis was conducted at the Lower Layer Super Output Area (LSOA) level, a small-area statistical geography in England, with each LSOA typically containing 1000–3000 residents and covering 1–4 km². Fig. 2 provides a geographic overview of the study area, the spatial distribution of total per-capita prescription counts (TPC), and example socio-demographic and environmental predictors from the MedSAT dataset, highlighting substantial spatial variation across Greater London.

3.2. Step A: Constructing the graph

To cast the prediction of prescription outcomes in geographical neighbourhoods as a node-level learning task, we represented the study area as a graph $G = (V, E)$. Each LSOA corresponded to a node $v_i \in V$, positioned at its geographic centroid. An edge $e_{ij} \in E$ connected two nodes if the corresponding LSOAs are considered spatial neighbours. The adjacency relationships are encoded in a binary spatial proximity matrix $A \in \mathbb{R}^{n \times n}$, where $A_{ij} = 1$ if LSOA i and j are adjacent (and 0 otherwise). Each node v_i was associated with a feature vector $\mathbf{x}_i$, comprising its socio-environmental attributes, and a target value y_i , indicating the per capita prescription outcome of interest.

As shown in Part A of Fig. 1, multiple definitions of spatial adjacency, such as k -hop contiguity (Chen et al., 2020) and k -nearest neighbours (Feng et al., 2022), were evaluated to determine the most informative graph topology. We addressed this selection process in Section 3.4, where an ablation study empirically evaluated the impact of adjacency construction methods.

3.3. Step B: Preprocessing data and embedding graph

Having established the spatial graph structure in Step A, the next step was to associate each node with a set of attributes representing the socio-environmental context of its corresponding geographic unit. This involved selecting, processing, and embedding multi-source urban indicators into the graph structure, ensuring that the feature representation was both comprehensive and robust. While our case study focused on neighbourhood-level health outcomes, the procedure described here is generalisable to other urban analytics tasks involving spatially structured data.

3.3.1. MedSAT dataset

We utilised the MedSAT dataset (Šćepanović et al., 2023b), which integrates detailed socio-demographic and environmental indicators at the LSOA level. They are the smallest geographic units in England for which social and health indicators can be publicly released due to privacy constraints. MedSAT compiled data from multiple authoritative sources, including the Office for National Statistics (ONS) (2025), the National Health Service (NHS) (2025), and remote sensing platforms such as Copernicus and Google Earth Engine (GEE) (2025), ensuring consistency and accuracy across variables.

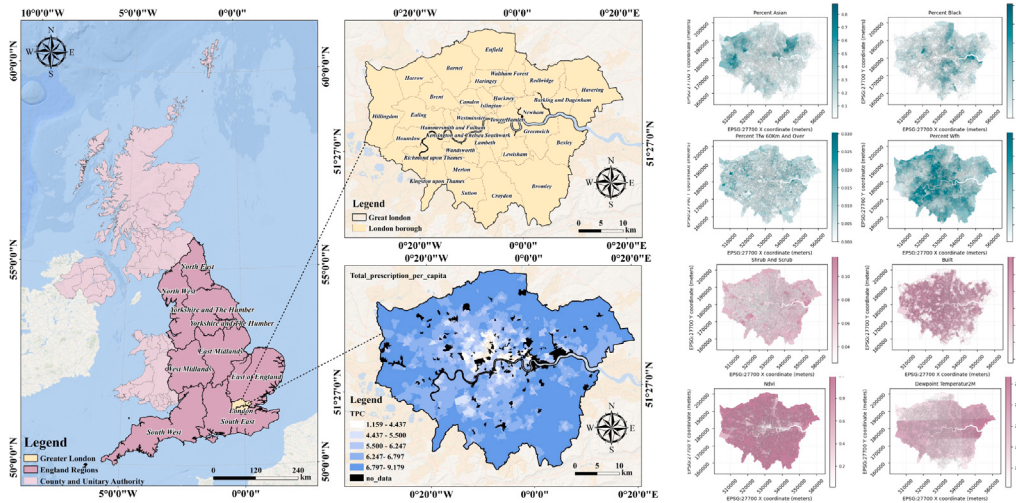


Fig. 2. Study area (Greater London) and example spatial variables. Left: location within England. Centre: borough boundaries, LSOAs, and total per-capita prescription counts (TPC). Right: example socio-demographic and environmental predictors from the MedSAT dataset.

The dataset covered 111 socio-demographic indicators — such as gender ratio, age distribution, and measures of deprivation — and 55 environmental variables, including air quality metrics (e.g., NO_2 , $\text{PM}_{2.5}$), greenness, land cover, and climate characteristics. All variables were aligned with the 2020 statistical and geographic boundaries of Greater London. In addition, MedSAT provided six per-capita medical prescription outcomes (anxiety, depression, diabetes, hypertension, asthma, and opioid use) derived from NHS administrative records. Although these outcomes are health-specific, the underlying multi-modal indicator set is structurally similar to datasets used in other domains such as environmental monitoring, mobility analysis, and socio-economic modelling, enabling broader applicability of the proposed framework.

3.3.2. Feature selection and preprocessing

Given the dataset's high dimensionality and potential multicollinearity, we performed systematic feature selection to enhance model stability, interpretability, and the attribution clarity of subsequent GNN analyses. The MedSAT dataset initially contained 166 variables (111 socio-demographic and 55 environmental indicators).

Our feature selection process proceeded in two stages. First, we conducted pairwise correlation analysis to identify and remove highly redundant features ($|r| > 0.90$), retaining within each correlated set the most representative variable informed by domain knowledge (Apter et al., 1999; Chen & Kan, 2008; Kjellstrom et al., 2007; Lantz et al., 1998; Marmot, 2005; Nutsford et al., 2013). Second, we applied additional dimensionality reduction techniques to further streamline the feature set while preserving theoretically essential predictors commonly used as key confounders in public health and socio-spatial research (Bor et al., 2014; Compton & Shim, 2015; Harvey et al., 2017; Hossain et al., 2020; Hynie, 2018).

This process resulted in a refined feature set of 67 variables: 13 environmental indicators and 54 socio-demographic variables. The selected features encompass key demographic characteristics (age distribution, gender, ethnicity), socioeconomic indicators (occupation, housing, deprivation), and environmental exposures (air quality, greenness, land cover). Table 1 lists the key control and environmental variables retained across all configurations. Detailed information about the complete feature selection methodology and variable statistics is provided in the supplementary materials (Appendix A).

After filtering, all retained variables were normalised and embedded as node attributes $\mathbf{x}_i$ for each LSOA in the graph.

3.3.3. Positional and locational encoding

To supplement the socio-demographic and environmental attributes derived from MedSAT, we enriched each node with additional locational (spatial) and positional (topological) encodings. This design enabled the model to better capture both relative topological positions within the graph and absolute spatial context in the urban environment. Specifically, we incorporated three distinct encoding strategies — Laplacian positional encoding (Dwivedi et al., 2021), random walk encoding (Schaub et al., 2020), and SatCLIP locational encoding (Klemmer et al., 2025) — to enhance spatial expressiveness. The locational encoder is implemented as a modular component that maps geographic context to a fixed-dimensional embedding. Alternative locational encoders, including other compatible geospatial foundation models, can therefore be substituted without requiring changes to the downstream UST-GNN pipeline. These were systematically evaluated in our ablation study (see Fig. 1, Part C), and implementation details are as follows:

Laplacian positional encoding. Laplacian encoding computes node-level position features from the eigenvectors of the graph Laplacian matrix. This spectral technique captures the connectivity and smoothness of local graph neighbourhoods, assigning similar embeddings to structurally proximate nodes (Dwivedi et al., 2021). Let the graph Laplacian be defined as

$$\mathbf{L} = \mathbf{D} - \mathbf{A},$$

$\mathbf{A}$ is the adjacency matrix and $\mathbf{D}$ is the degree matrix. We then extracted the first three non-trivial eigenvectors of $\mathbf{L}$ to represent each node's relative position within the graph topology:

$$\mathbf{L}\mathbf{u}_k = \lambda_k \mathbf{u}_k, \quad \mathbf{p}_i = [u_1(i), u_2(i), u_3(i)].$$

where λ_k and $\mathbf{u}_k$ are the k th eigenvalue and eigenvector of $\mathbf{L}$, respectively, and $\mathbf{p}_i$ is the 3-dimensional Laplacian positional encoding of node i , obtained from the entries of the first three non-trivial eigenvectors evaluated at node i .

Random walk positional encoding. Random walk encoding represents each node by its probability distribution over a one-step traversal to its 1-hop neighbours (Dwivedi et al., 2021; Yeh et al., 2023). This approach reflects localised relational structures by modelling how information could propagate across directly connected areas. Given the moderate size and shallow depth of our spatial graph, a single-step walk was sufficient to capture local adjacency without overcomplicating the feature space:

$$P(v_i, t) = \frac{1}{Z} \sum_{j \in \mathcal{N}(v_i)} \frac{A_{ij}}{d_i} P(v_j, t-1)$$

Table 1

Key control and environmental variables retained across all model configurations.

Variable name	Description
Demographics (Age)	Population proportion across age groups (e.g., Aged 10–14, 15–19, 20–24, ..., 85+); age structure is a known determinant of chronic disease prevalence and healthcare utilisation (Lantz et al., 1998; Marmot, 2005).
Sex and ethnicity	Male share, Mixed-race share, White population share; demographic composition influences disease risk profiles and access to care (Apter et al., 1999; Marmot, 2005).
Socioeconomic status	Proportion in professional occupations, population density; socio-economic indicators are established predictors of health inequalities (Bixby et al., 2015; Lantz et al., 1998; Marmot, 2005).
Environment	NDVI (greenness), NO ₂ concentration, water body, tree, grassland, and bare land coverage ratios; environmental exposures such as air quality and access to green space have demonstrated impacts on respiratory, cardiovascular, and mental health (Chen & Kan, 2008; Fecht et al., 2015; Kjellstrom et al., 2007; Twohig-Bennett & Jones, 2018).

where $P(v_i, t)$ is the probability distribution of node v_i at time t , d_i is the degree of v_i , and Z is a normalisation factor.

Geo-foundation-model-based locational encoding. SatCLIP (Klemmer et al., 2025) provides absolute spatial descriptors by combining Sentinel-2 satellite imagery and geospatial coordinates into a unified 256-dimensional embedding, generated by a pretrained earth-observation (EO) contrastive learning GeoAI model. These embeddings encode physical and environmental characteristics from satellite images of each LSOA, such as vegetation, surface materials, and land use. To reduce dimensionality and ensure comparability with other node features, we applied PCA and retained the top principal components as the final locational encoding vector. To assess the modularity of this component, we also considered Google's AlphaEarth (Brown et al., 2025) as an alternative geo-foundation-model *GeoFM* locational encoder in the ablation analysis.

By appending these position and location embeddings to each node's attribute vector $\mathbf{x}_i$, UST-GNN was equipped to learn from both relative topological patterns and absolute geographic contexts. This dual perspective is particularly valuable in urban health applications, where both network position (e.g., connectedness to high-risk areas) and absolute location (e.g., proximity to green space or pollution sources) can link to health outcomes. Section 3.4.2 evaluates the individual and combined contributions of these encoding methods, including different position and *GeoFM* locational encoders, through comparative experiments.

3.4. Step C: Building UST-GNN via component ablation

With the spatial graph structure and enriched node attributes prepared in Steps A and B, we next designed a graph neural network architecture capable of leveraging both socio-environmental attribute and spatial-topological information to predict neighbourhood-level outcomes. Our aim was to identify an architecture that could effectively capture the socio-environmental and spatial patterns most relevant to the case study on health outcomes, while remaining robust to alternative data contexts. To this end, we conducted an extensive component-wise evaluation, systematically varying graph definitions, feature encodings, and GNN architectures to determine the combination yielding the best predictive performance (Fig. 1, Part C).

3.4.1. UST-GNN model architecture

UST-GNN is a graph neural network that predicts the outcome for each LSOA node using socio-environmental attributes together with the spatial-topological relationships encoded in the graph. In the case study, the target outcome was the per-capita prescription rate. The

model takes as input the MedSAT-derived socio-demographic and environmental features, augmented with positional and locational encodings, and propagates them through multiple graph neural network layers. At each layer, a node aggregates information from its neighbours (as defined by the adjacency A) and combines it with its current representation, enabling iterative propagation of relevant signals across the urban network.

Through these stacked message-passing operations, node embeddings progressively integrate both local context and broader neighbourhood influences. After L such layers, the resulting node embeddings encapsulate multi-scale structural patterns relevant to the prediction task. A final fully connected layer maps each node embedding to the target value, producing predictions for all LSOAs (Figure 1E). Formally, the model learns a function $f : \mathbf{x}_i \mapsto \hat{y}_i$ that minimises the mean squared error (MSE) between predicted and observed values:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2, \quad (1)$$

where N is the number of nodes, y_i is the ground-truth outcome for node i , and $\hat{y}_i = f(\mathbf{x}_i)$ is the model prediction.

3.4.2. Spatial component ablation study

To determine the most effective design that considers spatial effects, we conducted a comprehensive sequential ablation study assessing the influence of key model components. All experiments were performed on the training set using 3-fold internal cross-validation to ensure stable validation performance. In addition, within each training fold, the data were further divided into a 70:20:10 ratio, where 10% of the validation subset was reserved as a held-out early-stopping set to prevent overfitting and ensure convergence stability (Kedzie et al., 2018; Meyes et al., 2019). The specific data splitting strategy for cross-validation, including how spatial adjacency relationships were preserved in train-validation partitions, is detailed later in Step D: *Spatial Cross-Validation and Baseline Comparison*. Hyperparameters were tuned via 10 rounds of random search within a commonly used GNN hyperparameter range (Adnan et al., 2022; Zhou et al., 2022), covering learning rate, hidden dimension, dropout rate, number of layers, and aggregation functions.

We systematically tested multiple options across three core components of the model. First, different definitions of the spatial adjacency graph were assessed (as described in Section 3.2) to capture meaningful spatial dependencies among neighbourhoods. Second, the contribution of positional and locational encodings was evaluated by testing the inclusion of no encoding, single encodings, and combined encodings (detailed in Section 3.3.3). Third, we compared four widely used GNN architectures under different layer configurations: Graph Convolutional Network (GCN) (Kipf & Welling, 2016), Graph Isomorphism

Table 2
GNN hyperparameter search space for random search.

Hyperparameter	Search range
Learning rate	$\{1 \times 10^{-4}, 5 \times 10^{-4}, 1 \times 10^{-3}, 5 \times 10^{-3}\}$
Hidden dimension	$\{16, 32, 64, 128, 256\}$
Dropout rate	$\{0.0, 0.2, 0.3\}$
Aggregation function	$\{\text{mean, sum, max}\}$
Activation function	$\{\text{ReLU, ELU, LeakyReLU}\}$
Weight decay	$\{0, 1 \times 10^{-5}, 5 \times 10^{-5}, 1 \times 10^{-4}, 5 \times 10^{-4}\}$
Batch size	$\{32, 64, 128\}$
Optimiser	$\{\text{Adam, AdamW}\}$

Network (GIN) (Xu et al., 2018), GraphSAGE (Hamilton et al., 2017), and Graph Attention Network v2 (GATv2) (Bhatti et al., 2023; Brody et al., 2021; Zhang et al., 2019). These architectures span a range of message-passing strategies, including linear aggregation (GCN), MLP-based transformations (GIN), neighbourhood sampling (GraphSAGE), and attention-based mechanisms (GATv2).

For each configuration, we recorded validation performance using mean squared error (MSE) and coefficient of determination (R^2) to identify the most effective combination of components and hyperparameter settings. Table 2 summarises the hyperparameter ranges used in our random search, based on commonly adopted configurations in prior work (Adnan et al., 2022; Hamilton et al., 2017; Kipf & Welling, 2016; Velicković et al., 2018; Xu et al., 2018; Zhou et al., 2022).

3.5. Step D: Spatial cross-validation and baseline comparison

To rigorously evaluate UST-GNN's performance, we compared its predictive accuracy against several baseline models under consistent spatial cross-validation settings. Following recent urban health modelling studies (Maitra et al., 2025; Šćepanović et al., 2023b), we selected a diverse set of baselines covering statistical, machine learning, geographically-enhanced, and graph-based approaches. The statistical baselines included the Spatial Lag Model (SLM) (Anselin, 2009) and Geographically Weighted Regression (GWR) (Fotheringham et al., 2009), which explicitly incorporated spatial effects within global and local linear frameworks, respectively. The machine learning baselines consisted of Random Forest and LightGBM (Ke et al., 2017), representing non-spatial, non-linear ensemble learners. To further examine the benefit of spatial feature augmentation, we implemented geographically-enhanced variants of tree-based models: Geo-LightGBM and Geo-XGBOOST using two strategies: (i) directly providing geographic coordinates as additional input features, and (ii) incorporating a spatial weight matrix to encode neighbourhood structure (Li, 2022; Nikparvar & Thill, 2021).

Finally, representative GNN baselines (GCN, GraphSAGE, and GIN) were also evaluated in the ablation study for comparison with UST-GNN's spatial-topological architecture. For a fair comparison, all baseline models were trained and evaluated using the same 3-fold data splits and identical feature sets as UST-GNN. In practice, the statistical models (SLM and GWR) were fit on the training portion of each fold and applied directly to the test fold, whereas the machine learning and geographically-enhanced models were tuned within the training folds using internal validation splits to ensure optimal hyperparameter selection and stable performance.

Splitting strategies and spatial considerations. Random node-level splits can severely fragment the spatial graph, often producing isolated test nodes or very small disconnected components. Spatial continuity is therefore critical for meaningful cross-validation. Fig. 3 compares a naïve random split (Panel A) with our spatially continuous, buffer-based strategy (Panel B). In the main evaluation setting, this strategy uses a fixed 2-hop buffer, while in the topology ablation study the exclusion region is scaled with the tested adjacency definition to ensure

a fairer comparison across graph constructions with different receptive fields.

In our approach, subgraphs for cross-validation were randomly selected but constrained to contain at least 30 nodes, ensuring that each test region had sufficient spatial extent and internal connectivity. We further adopted a community-based graph partitioning scheme (using Louvain or METIS clustering, depending on graph size (Shao et al., 2015)) to generate spatially coherent folds, thereby preserving spatial coherence and graph connectivity across folds.

To mitigate spatial information leakage, we introduced a k -hop buffer around the test subgraph. In the main evaluation setting, we used a fixed 2-hop buffer, excluding this region from both training and testing so that no node used for prediction had seen direct or second-order neighbours during training. For the topology ablation study, however, we further adapted this protocol so that the exclusion region scales with the tested adjacency definition: for an n -hop graph, the held-out centre nodes are separated from training visibility by an $n+2$ -hop invisible neighbourhood. This topology-aware exclusion rule is intended to reduce boundary-effect asymmetry when comparing graph constructions with different effective receptive fields. Conceptually, this corresponds to an inductive evaluation setting: model parameters are learned from one part of the city (training subgraphs) and tested on spatially disjoint regions, rather than reusing the same adjacency structure as in a transductive setup. This design better approximates real-world deployment scenarios, such as predicting health outcomes in unseen districts while preserving the contextual dependencies necessary for spatial reasoning.

3.5.1. Ten-fold spatial cross-validation

In the first strategy, we applied standard k -fold cross-validation with $k = 10$ (Fushiki, 2011), incorporating the buffered splitting procedure described above. In each fold, the test set and its immediate 2-hop neighbours were withheld from training, minimising border effects and preserving spatial structure. Performance was averaged over the 10 folds, providing a robust estimate of out-of-sample accuracy across the city while utilising 90% of the data in each training round.

3.5.2. Leave-one-borough-out spatial cross-validation

The second strategy was a geographically structured leave-one-out cross-validation (Wong, 2015). In contrast to ten-fold CV, which partitions the data into multiple smaller folds, this approach held out an entire London borough during each evaluation. Six diverse boroughs: Camden, Barking and Dagenham, Barnet, Southwark, Croydon, and Ealing, these were selected to represent various geographic locations and socio-demographic profiles (Raco et al., 2014). The splitting procedure followed the same 2-hop buffered approach as in ten-fold CV to maintain spatial continuity.

This strategy served as a stringent test of spatial generalisability, revealing how well UST-GNN performed in unseen regions. Both cross-validation setups were evaluated using root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). Hyperparameters were fixed across all folds, based on prior tuning through random search on each of the specific training sets.

3.6. Step E: Interpretability analysis

To complement UST-GNN's predictive capacity with meaningful interpretation, we introduced a light-weight principal component interpretation framework that globally interpret the input features through geographical mapping and correlational analysis as shown in Fig. 4. The approach first applies PCA (Cohen et al., 2013; Greenacre et al., 2022) to the final-layer node embeddings to extract two intrinsic components that summarises the latent spatial structures and attributes learned by the model.

The top two principal components (explaining over 80% of variance) were then mapped back to neighbourhoods, enabling interpretation of embeddings as coherent spatial patterns rather than abstract

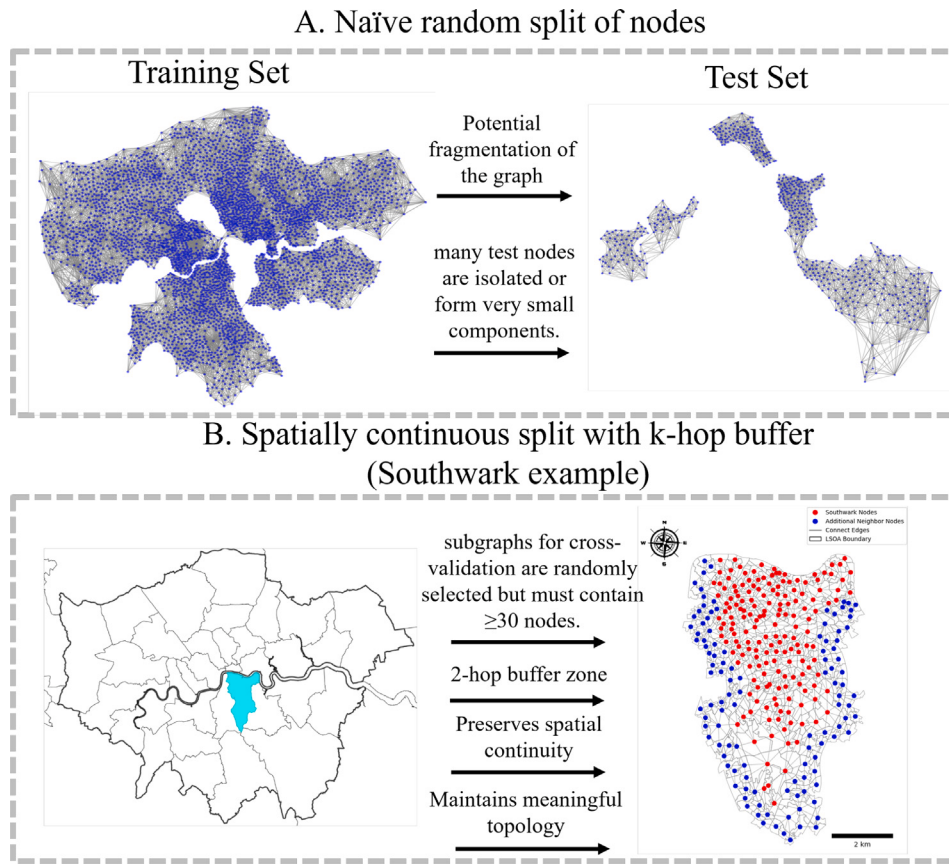


Fig. 3. Comparison of splitting strategies for spatial graph cross-validation: (A) naïve random split; (B) spatially continuous split with a k -hop buffer.

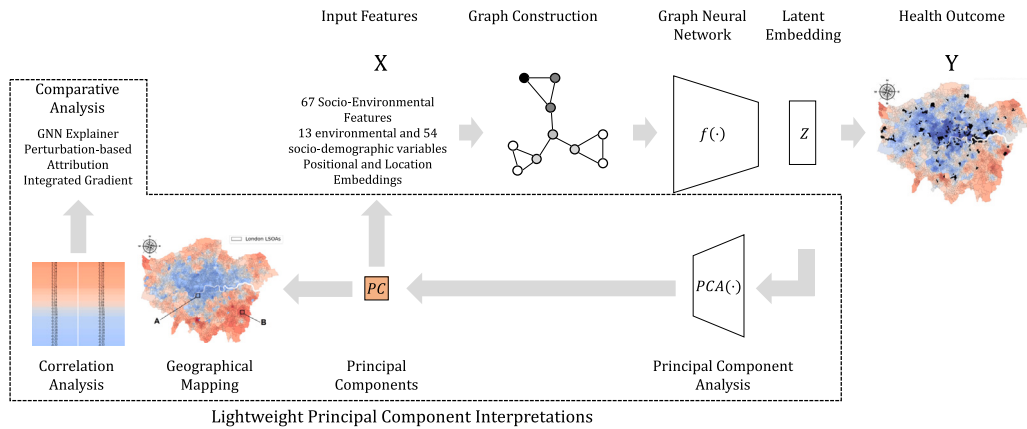


Fig. 4. Principal-component-based interpretability framework and comparison with existing explainability methods.

features. The two principal components were then correlated with the original socio-demographic and environmental variables, thereby anchoring the intrinsic representation space back to real-world policy-relevant urban characteristics and revealing how the model internally organises outcome-relevant information. This approach provides a simplified association between the learned latent space and its policy-relevant input variables.

Unlike input-level explainers such as GNNExplainer or gradient-based methods, this global approach offers a holistic and distilled view of how message passing encodes spatial and socio-environmental dependencies. PCA provides a light-weight and interpretable means of identifying the dominant and intrinsic spatial hierarchies captured during learning, without relying on model-specific assumptions or task labels.

To further contextualise this design, we established a comparative protocol (Appendix B) against established explanation methods, Permutation-based attribution, Integrated Gradients, and GNNExplainer, under identical data splits and hyperparameters. We evaluated (i) consistency of variable importance, (ii) ability to produce spatially continuous gradients, and (iii) interpretability across local and global levels. This positions our PCA-based framework within a unified methodological benchmark for spatial model interpretation.

4. Results

In this section, we evaluate the UST-GNN framework through a series of experiments designed to: (1) identify optimal model components via systematic ablation (Section 4.1), (2) assess predictive

performance in a health-related case study using the MEDSAT dataset (Section 4.2), and (3) illustrate interpretability insights from a domain-specific analysis of antidepressant prescriptions (Section 4.3). While the case study focuses on health outcomes, the experimental design and evaluation procedures are general to other urban analytics tasks involving spatially structured data.

4.1. Component ablation

To determine the optimal design of the UST-GNN framework, we performed a sequential ablation study covering four core components: message passing architecture, spatial adjacency definition, network depth, and location/positional encoding strategy.

Fig. 5 illustrates the spatial adjacency construction strategies considered in the ablation study. KNN graphs generate relatively sparse and distance-based connections, whereas h -hop contiguity graphs produce progressively denser and more interconnected neighbourhood networks, directly affecting information flow, spatial scale, and computational cost in graph learning.

Table 3 summarises the sequential ablation procedure, in which the best-performing component at each stage is retained and fixed for the next stage. The selected backbone inherited by the later stages is 2-head GATv2 with 1-HOP adjacency and 2 layers; the subsequent position-and-location-encoding blocks further refine this backbone, including comparisons across alternative encoders.

Among the adjacency configurations, the 1-HOP contiguity graph achieved the strongest cross-validated performance ($R^2 = 0.860 \pm 0.008$), outperforming both KNN alternatives and denser multi-hop graphs. In particular, extending the graph to 2-HOP and 3-HOP reduced performance to 0.821 ± 0.027 and 0.799 ± 0.011 , respectively. This suggests that, for neighbourhood-level prescription prediction in Greater London, immediate contiguity captures the most relevant local spatial context, whereas denser topologies may introduce less informative meso-scale structure and increase the risk of over-smoothing. We therefore retain 1-HOP adjacency as the default spatial representation for UST-GNN.

The Graph Attention Network v2 (GATv2) emerged as the most effective architecture. Two graph convolutional layers provided the best balance between expressiveness and stability, with three layers remaining competitive but slightly less robust, and the additional head ablation showed that two attention heads performed best overall. For position encoding, the combination of SatCLIP and Random Walk yielded the highest performance, suggesting that both absolute geographic positioning and local structural context contribute to stronger spatial representations. The additional AlphaEarth comparison further shows that different *GeoFM*-based representation enhancements provide broadly similar improvements within the same UST-GNN pipeline, indicating that the framework can benefit from multiple interchangeable locational encoding modules. Together, these design choices define the final UST-GNN configuration used in the subsequent experiments.

4.2. Model performance

Addressing our second contribution, we evaluated UST-GNN in a 10-fold as well as leave-one-out cross-validation.

4.2.1. Ten-fold spatial cross-validation

We evaluated UST-GNN against a comprehensive suite of baseline models, including statistical methods (Spatial Lag Model, SLM; Geographically Weighted Regression, GWR), machine learning approaches (Random Forest, LightGBM), geographically-enhanced models (Geo-XGBOOST and Geo-LightGBM incorporating either coordinate or spatial-weight features), and graph neural network baselines (GCN, GraphSAGE) that share the same adjacency structure as UST-GNN. All models were trained and tested under an identical 10-fold cross-validation protocol to ensure comparability.

Table 4 summarises predictive performance across all six prescription outcomes, while Fig. 6 provides a detailed view of fold-wise performance and prediction stability for the representative depression task. UST-GNN consistently outperformed all baselines across diabetes, hypertension, asthma, depression, anxiety, and opioid prescriptions. Best and second-best results are highlighted in bold and underline, respectively. While geographically enhanced models achieved notable gains over traditional machine learning baselines, confirming the value of incorporating spatial context, UST-GNN further improved predictive accuracy by explicitly modelling graph-based spatial dependencies and nonlinear socio-environmental interactions. Across outcomes, it improves upon the strongest baseline by 8.4%–13.2% in out-of-sample R^2 . Fig. 6 further shows that these gains are accompanied by accurate and stable out-of-fold predictions with limited variation across spatial folds.

To further illustrate predictive performance, we visualised observed versus predicted per-capita prescription rates across three representative health outcomes: depression (mental health), diabetes (chronic disease), and opioids (pain management) in Fig. 7. For each outcome, the UST-GNN configuration was individually fine-tuned. Across outcomes, the model successfully reproduced major spatial trends — such as high depression prescription rates in South East London, elevated diabetes prescriptions in eastern boroughs, and opioid hotspots in outer boroughs — while capturing low-prescription clusters in Central London. Residual maps indicate generally small errors, with localised discrepancies in parts of Central and East London, suggesting potential effects of unobserved socio-environmental factors or limitations in current graph connectivity definitions.

4.2.2. Leave-one-out spatial cross-validation (LOOSCV)

To further test spatial generalisability under a more stringent setting, We conducted LOOSCV using six Boroughs in London—Camden, Barking, Barnet, Southwark, Croydon, and Ealing, which were selected for their diverse geographic, socio-demographic, and environmental characteristics (Kearns & Ron, 1997; Ng, 1997). Table 5 summarised the results across six health outcomes.

Compared with the 10-fold cross-validation, UST-GNN experienced a slight performance decline under LOOSCV for most outcomes, with R^2 scores generally reduced by less than 0.1. A more pronounced drop was observed for diabetes in Camden and Ealing, which likely reflects greater generalisation difficulty under borough-level hold-out. Plausible explanations include distribution shift between held-out and training boroughs, variation in feature–outcome relationships across boroughs, and higher intra-borough heterogeneity. Similarly, asthma prediction in Camden, Southwark, and Ealing showed reduced R^2 , suggesting that neighbourhood-level patterns in these boroughs were more heterogeneous or less well represented by the current graph attributes.

Despite these expected reductions, UST-GNN still maintained an average R^2 above 0.80 across all outcomes and boroughs. Notably, prediction in Croydon consistently achieved the highest scores, indicating that certain boroughs may have more regular socio-environmental patterns that the model can capture more effectively under spatial hold-out.

Overall, even under this more stringent validation regime, where entire unseen Borough subgraphs were held out, UST-GNN continued to outperform traditional baselines (SLM, GWR, LightGBM) evaluated under the less challenging 10-fold setting (Table 5). This suggests that UST-GNN remains robust under heterogeneous spatial configurations, although borough-specific generalisation difficulty is clearly non-uniform.

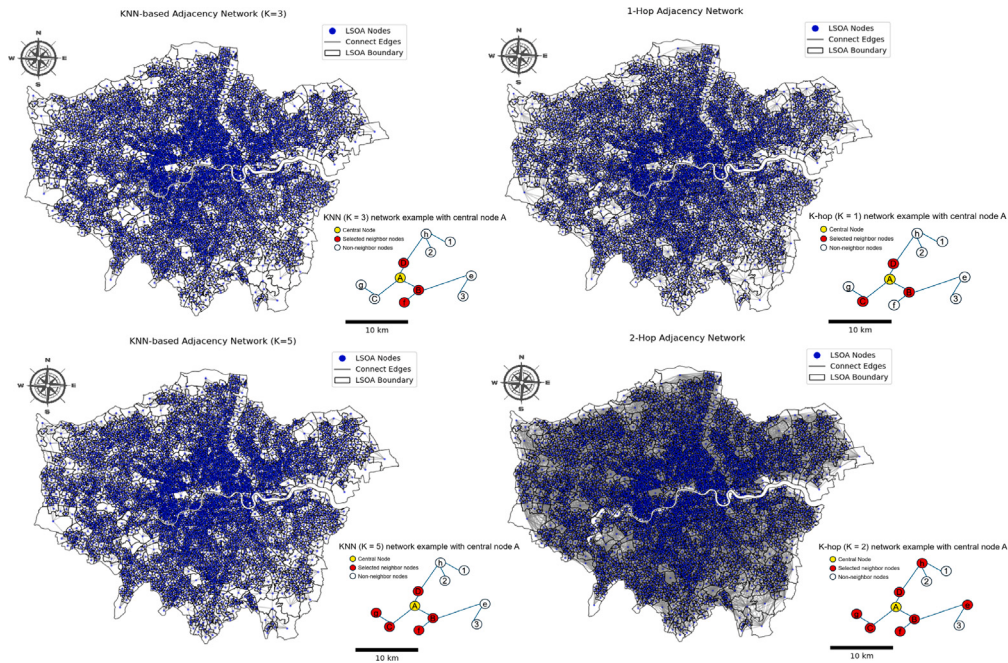


Fig. 5. Comparison of spatial adjacency construction strategies for London LSOAs: KNN graphs ($k = 3, 5$) and h -hop graphs ($h = 1, 2$).

Table 3

Component-wise ablation results for depression prediction. The ablation study was conducted sequentially starting from a preset backbone configuration (GCN; 1-hop adjacency; 2 layers; 2 attention heads; no positional/location embeddings). At each step, one component was varied across candidate settings, previously selected components were fixed to their selected values, and all remaining untested components retained their preset settings. The final selected model is reported in the last row.

Tested Pluggable Module	Setting / Candidate	3-fold R^2
Architecture	GCN	0.805 ± 0.013
	GIN	0.811 ± 0.006
	GraphSAGE	0.828 ± 0.009
	GAT	0.829 ± 0.013
	GATv2*	0.860 ± 0.008
	Best architecture updated to GATv2.	
Adjacency	3-NN	0.837 ± 0.009
	5-NN	0.854 ± 0.006
	1-hop*	0.860 ± 0.008
	2-hop	0.821 ± 0.027
	3-hop	0.799 ± 0.011
	Best adjacency = 1-hop.	
Convolution Depth	1 layer	0.767 ± 0.014
	2 layers*	0.860 ± 0.008
	3 layers	0.858 ± 0.011
	4 layers	0.833 ± 0.009
	Best depth = 2 layers.	
Attention Head	1 head	0.840 ± 0.011
	2 heads*	0.860 ± 0.008
	4 heads	0.853 ± 0.004
	8 heads	0.853 ± 0.013
	Best number of attention heads = 2.	
Position/Location	None	0.860 ± 0.008
	Random Walk (1-D)*	0.861 ± 0.008
	Laplace (3-D)	0.854 ± 0.011
	SatCLIP (2-D)	0.855 ± 0.008
	AlphaEarth (2-D)	0.861 ± 0.008
	AlphaEarth + Laplace (2+3-D)	0.857 ± 0.018
	AlphaEarth + Random Walk (2+2-D)	0.863 ± 0.011
	SatCLIP + Laplace (2+3-D)	0.859 ± 0.006
	SatCLIP + Random Walk (2+2-D)*	0.868 ± 0.008
	Best positional/location encoding updated to SatCLIP + Random Walk.	
Final model	UST-GNN = GATv2; 1-hop; 2 layers; 2 heads; SatCLIP + Random Walk	0.868 ± 0.008

4.3. Interpretability analysis

Aligned with our third contribution, we applied our proposed Principal Component Interpretation framework to the learned node embeddings from the depression prescription task. As shown in Fig. 8, the

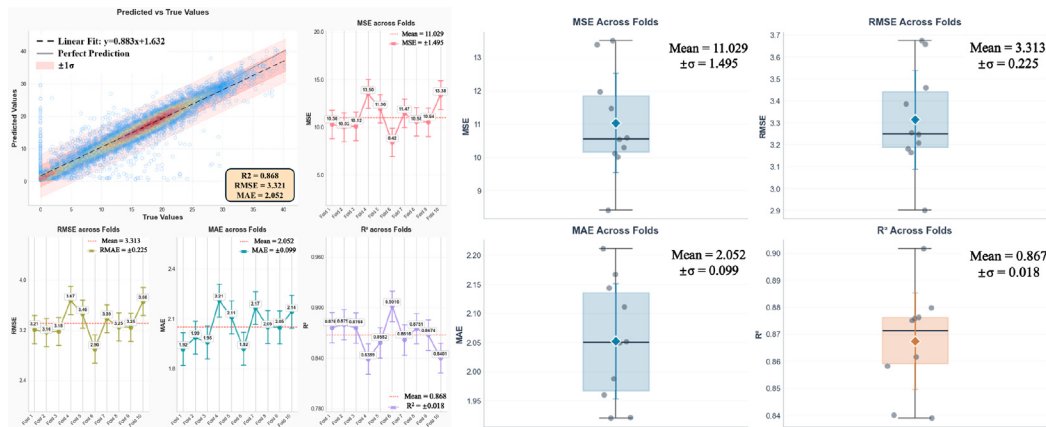
first two components (PCA1 and PCA2) explained 61.26% and 23.83% of the total variance, respectively, capturing the dominant spatial and socio-environmental structures embedded by UST-GNN.

The spatial maps of PCA1 and PCA2 revealed distinct but complementary patterns. PCA1 expressed a pronounced urban-rural gradient,

Table 4

Ten-fold cross-validation results across six prescription outcomes. Best results are in bold, second-best are underlined.

Model	Diabetes	Hypertension	Asthma	Depression	Anxiety	Opioids
Statistical Baselines						
SLM	0.711	0.725	0.684	0.714	0.678	0.677
GWR	0.726	0.752	0.710	0.734	0.705	0.691
Machine Learning Baselines						
Random Forest	0.637 ± 0.021	0.671 ± 0.016	0.621 ± 0.021	0.633 ± 0.023	0.599 ± 0.023	0.555 ± 0.022
LightGBM	0.632 ± 0.029	0.649 ± 0.017	0.589 ± 0.024	0.628 ± 0.020	0.586 ± 0.021	0.535 ± 0.032
Geographically-Enhanced Models						
Geo-XGBOOST(Coordinates)	0.691 ± 0.023	0.725 ± 0.013	0.678 ± 0.016	0.699 ± 0.019	0.662 ± 0.017	0.6546 ± 0.021
Geo-LightGBM (Coordinates)	0.684 ± 0.022	0.720 ± 0.016	0.670 ± 0.021	0.689 ± 0.017	0.655 ± 0.021	0.627 ± 0.020
Geo-XGBOOST(Spatial weight matrix)	0.781 ± 0.023	0.801 ± 0.017	0.764 ± 0.022	0.767 ± 0.022	0.762 ± 0.021	0.751 ± 0.015
Geo-LightGBM (Spatial weight matrix)	0.772 ± 0.026	0.790 ± 0.019	0.754 ± 0.023	<u>0.779 ± 0.015</u>	0.750 ± 0.020	0.735 ± 0.013
Graph Neural Network Baselines						
Evaluated (GCN, GraphSAGE, GIN) in ablation; see Table 3						
Proposed Method						
UST-GNN (Ours)	0.864 ± 0.024	0.869 ± 0.020	0.849 ± 0.018	0.868 ± 0.018	0.853 ± 0.021	0.850 ± 0.025
Relative gain over the best baseline (R _S)	10.6%	8.4%	11.1%	11.4%	11.9%	13.2%

**Fig. 6.** Predictive performance and cross-validation stability of UST-GNN for depression prescriptions: predicted vs. observed values, fold-wise metrics, and metric distributions across folds.**Table 5**Performance of UST-GNN under Leave-One-Out Spatial Cross-Validation (LOOSCV) across six London boroughs and six prescription outcomes. Best R^2 values per medical outcome are shown in bold.

Outcome	Metric	Camden	Barking	Barnet	Southwark	Croydon	Ealing
Diabetes	RMSE	3.93	7.03	6.83	5.61	4.60	10.43
	MAE	2.80	5.28	4.84	4.11	3.34	6.80
	R^2	0.78	0.88	0.83	0.80	0.90	0.79
Hypertension	RMSE	6.70	12.34	9.88	8.51	7.12	11.64
	MAE	4.55	7.86	6.81	5.69	5.17	7.31
	R^2	0.83	0.81	0.87	0.84	0.93	0.73
Asthma	RMSE	1.14	1.55	1.41	1.49	1.25	1.62
	MAE	0.87	1.03	0.93	1.13	0.89	1.11
	R^2	0.76	0.82	0.84	0.76	0.88	0.74
Depression	RMSE	1.98	2.53	2.54	2.63	2.82	2.68
	MAE	1.37	1.89	1.65	1.90	1.74	1.84
	R^2	0.89	0.90	0.88	0.87	0.88	0.77
Anxiety	RMSE	3.81	3.70	4.27	3.83	3.49	4.49
	MAE	2.39	2.70	2.87	2.83	2.32	2.82
	R^2	0.81	0.89	0.84	0.87	0.90	0.75
Opioids	RMSE	1.44	2.28	1.62	1.99	1.98	1.63
	MAE	0.94	1.50	1.09	1.37	1.45	1.13
	R^2	0.83	0.84	0.85	0.77	0.89	0.76

with low values in central boroughs (e.g., Region A) and higher values in outer boroughs (e.g., Region B), aligning with known socio-demographic centrality. PCA2, in contrast, reflected intra-urban heterogeneity (e.g., Regions C and D), capturing finer distinctions among neighbourhoods that traditional spatial regression models often fail to represent.

The correlation heatmap between embedding components and MedSAT features (Fig. 8) indicated that PCA1 was negatively correlated with population density (−0.54), air pollution (−0.79), and income levels (−0.23), and positively correlated with housing deprivation (0.53). Together, these associations suggest that PCA1 captures a latent gradient related to urban inequality and environmental stress. PCA2 was

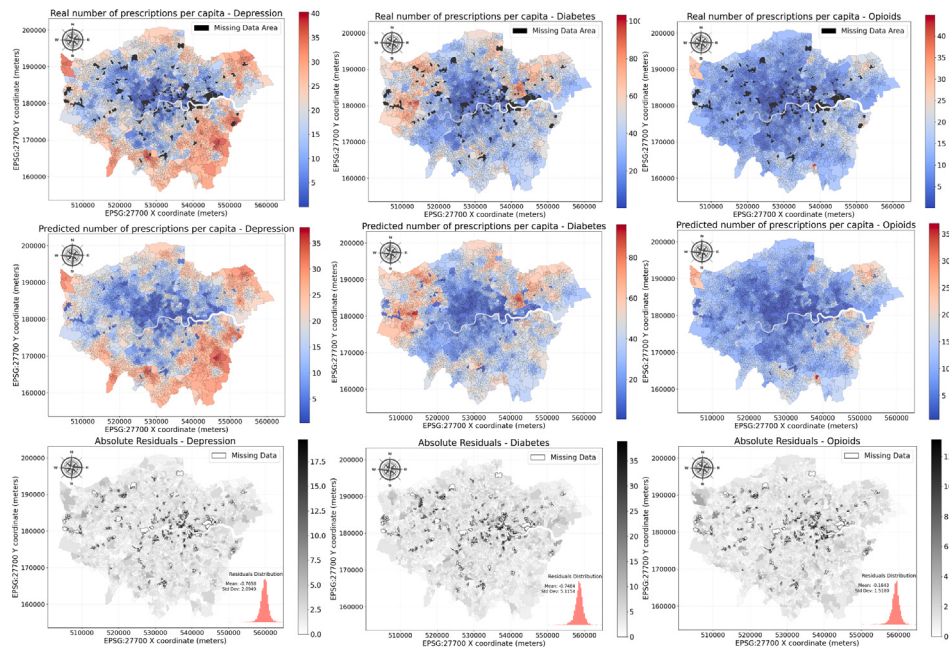


Fig. 7. Spatial distribution of real values, predicted values, and absolute residuals for three representative prescription categories in London: (left) Depression (mental health), (middle) Diabetes (chronic disease), and (right) Opioids (pain management). Rows correspond to: (top) real number of prescriptions per capita, (middle) predicted number of prescriptions per capita by the UST-GNN framework, and (bottom) absolute residuals between predictions and real values. Blue-to-red gradients indicate relative prescription intensity, while greyscale in the bottom row shows residual magnitude (darker = larger error). Black polygons mark LSOAs with missing target values; histograms in the residual maps show the residual distribution with mean and standard deviation.

associated with cultural and demographic variation, showing positive correlations with the proportion of White residents (0.34) and Christian beliefs (0.24), and negative correlations with Asian residents (−0.32) and Muslim beliefs (−0.22). It further captured built-environment attributes, such as housing quality and commuting modes, indicative of contextual differences in access and exposure.

Compared with established interpretability techniques such as Permutation-based attribution, Integrated Gradients, and GNNExplainer (see [Appendix B](#)), our PCA-embedding approach yielded more consistent and spatially continuous interpretations. It effectively summarised global structure–function relationships, offering a stable, transferable understanding of how the model internalises spatial context—thereby advancing interpretability in spatial graph learning beyond purely local attribution methods. Together, these findings demonstrate that UST-GNN embeddings encode interpretable, spatially coherent dimensions linking socio-environmental structure to health outcomes.

5. Discussion

5.1. UST-GNN: Model insights and general implications

The proposed UST-GNN framework consistently outperformed all baseline models across six distinct prescription outcomes, demonstrating its ability to capture spatial dependencies, topological structures, and feature heterogeneity in complex urban systems. By combining expressive graph-based aggregation with complementary positional and location encodings, UST-GNN generated neighbourhood representations that were both more accurate and more spatially coherent than those produced by traditional spatial regression or tree-based machine learning approaches. Consistent with the view that “if a model approximates the data-generating process well, its interpretation should yield better insights into that process” ([Molnar et al., 2020](#)), our results showed that UST-GNN not only improved predictive accuracy but also uncovered interpretable latent structures that aligned with known spatial inequalities.

A key contribution of this study lies in the design of a transferable feature engineering pipeline for high-dimensional urban datasets. Existing studies often mitigate multicollinearity through simple VIF filtering or single-metric ranking ([Cheng et al., 2022](#)), yet such approaches overlook the compounded dependencies among socio-demographic and environmental variables ([Trushna et al., 2021](#); [Wu et al., 2023](#)). Our comprehensive multi-dimensional feature selection framework ([Appendix A](#)) integrates four complementary evaluation metrics, mutual information, F-statistics, random forest importance, and Lasso regularisation, within an adaptive cross-validation procedure and correlation-based redundancy control ($|r| > 0.8$). The approach balances theoretical relevance with statistical robustness by mandatorily retaining core control variables (e.g., demographics, deprivation, environmental exposure) while optimising the remaining feature set in a data-driven manner. The resulting subset achieves near-saturated explanatory capacity with substantially reduced dimensionality, providing a stable, interpretable, and generalisable representation of urban socio-environmental systems. This integrated pipeline is broadly applicable to other spatial prediction and health geography tasks, supporting reproducible, domain-grounded feature learning across diverse urban contexts.

Beyond performance gains, the systematic spatial component ablation (covering graph architecture, spatial adjacency definitions, neighbourhood depth, and positional encoding) reinforced that spatially explicit GNN effectiveness depends as much on how spatial relationships are encoded as on the neural architecture itself. While automated methods such as Neural Architecture Search (NAS) ([Zhou et al., 2022](#)) may further refine configurations, their computational demands often limit applicability in high-resolution urban datasets. Our sequential ablation procedure offers a more tractable yet still rigorous pathway for spatial model optimisation.

More broadly, our findings highlight a methodological point for GeoAI-based urban analytics: spatial effects, network topology, leakage-aware evaluation, and interpretability are often treated as separate modelling concerns, even though they jointly shape the reliability and usefulness of urban prediction systems. In policy-facing settings

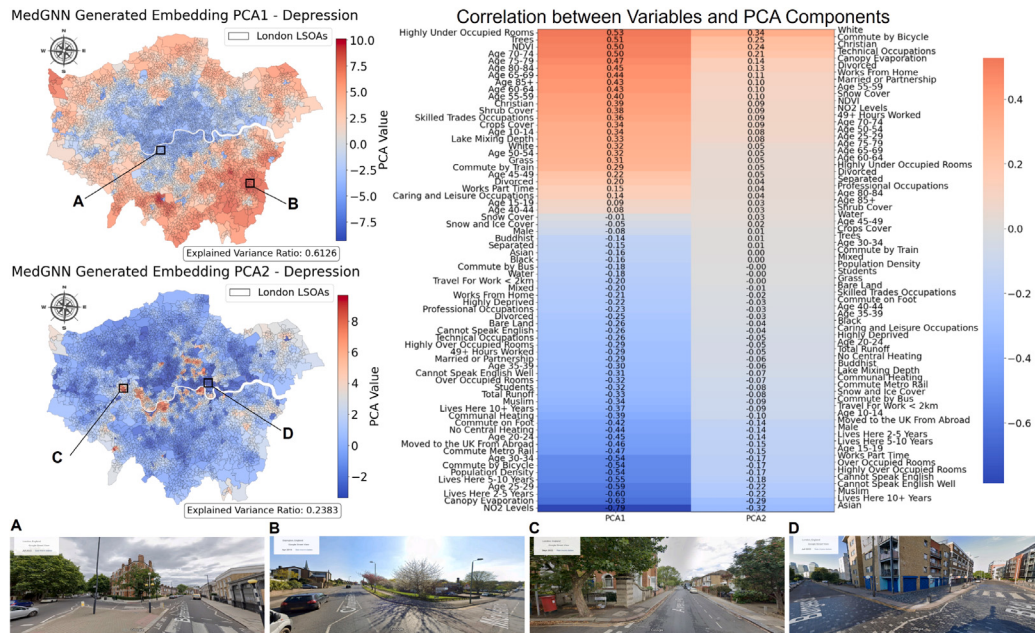


Fig. 8. Principal Component Interpretation of learned embeddings and their explanatory links to antidepressant spatial patterns. (Top left) Spatial distributions of PCA1 and PCA2 across London LSOAs. (Top right) Correlation heatmap between embedding components and original socio-environmental features. (Bottom A–D) Representative street-view samples from annotated neighbourhoods.

such as urban health, predictive performance alone is rarely sufficient; models must also generate explanations that are spatially meaningful and interpretable at a level relevant to intervention and planning. By integrating these elements within a single spatially explicit unified framework, UST-GNN moves beyond prediction towards insight generation, showing that graph-based urban models can support both robust forecasting and policy-relevant interpretation. In this sense, the framework also provides a useful foundation for future urban digital twin workflows, where heterogeneous urban data, scenario testing, monitoring, and transparent analytical outputs must be combined to support evidence-based decision-making.

Finally, although evaluated here in the context of health outcomes, UST-GNN is outcome-agnostic and readily transferable to other neighbourhood-level prediction tasks, ranging from housing prices and energy demand to accessibility, environmental risk, and socio-economic deprivation. This flexibility underscores the potential of the framework as a reusable GeoAI pipeline for diverse urban analytics applications and for broader evidence-based planning in sustainable and equitable cities.

5.2. Interpretation through PCA-embedding method

To interpret the deep representation learned by UST-GNN, we applied PCA to the final-layer node embeddings (Fig. 8). The first two components (PCA1 and PCA2) explained 61.3% and 23.8% of the total variance, respectively, capturing the dominant spatial and structural gradients encoded by the network. PCA1 exhibited a pronounced centre-periphery gradient consistent with established urban form and demographic structures in Greater London (Aiello et al., 2019; Hatch et al., 2011), with high scores linked to denser, younger, and less green areas, and low scores characterising suburban and greener neighbourhoods. PCA2 captured finer-scale socio-cultural variations, identifying clusters of inner-London communities with elevated antidepressant prescription rates and strong associations with ethnicity and religion.

Regression of the observed antidepressant prescription rates on PCA1 and PCA2 yielded $R^2 > 0.80$, indicating that these components

collectively explained the majority of spatial variation captured by UST-GNN. Together, PCA1 and PCA2 revealed complementary concepts: the first represented macro-level structural-environmental gradients, while the second seemed to exhibit localised socio-cultural contrasts underlying heterogeneous mental health outcomes.

Unlike conventional attribution-based explainers such as SHAP, Integrated Gradients, or GNNExplainer (see Appendix B), our Principal Component Interpretation analysis does not attempt to assign importance to individual input features, nor does it seek to disentangled neurons like mechanistic interpretability. Instead, it focuses on identifying and interpreting the most informative and intrinsic component of the model's learned representation. This inductive, correlation-oriented approach trades off granular instance-level precision for global interpretability and computational efficiency, an advantage in high-dimensional, spatially dependent systems where strict feature-level causality is often ill-posed. While acknowledging limitations with analysing these entangled, concept-level components, this method reveals spatial gradients linking *raw features*, *learned embeddings*, and *target outcomes*, a level of structural insight that most local attribution methods cannot provide.

In this sense, the Principal Component Interpretation framework serves as a scalable and domain-grounded interpretability strategy: it transforms complex GNN representations into policy-relevant axes of socio-environmental variation, facilitating cross-model comparison and rapid exploration of structure-function relationships in urban systems. This approach, though deliberately simplified, provides a transferable and analytically transparent pathway to understanding how deep spatial models generalise across diverse geographic contexts.

5.2.1. Alignment with existing literature

Our results affirm prior findings regarding age, ethnicity, and housing conditions in relation to mental health outcomes. Antidepressant prescription rates were highest among elderly populations (Cahoon, 2012) and adolescents aged 10–19 (Armitage, 2021; Lalji et al., 2021), and lower among adults aged 20–39. This younger adult group tends to exhibit health-protective behaviours such as active commuting (Berrie

et al., 2024; Soini et al., 2024), professional employment (Remes et al., 2021), and remote work—an arrangement shown to buffer mental distress, especially during the COVID-19 pandemic (Bonacini et al., 2021).

Notably, PCA1 showed strong correlation with under-occupied housing (correlation coefficient: 0.53), a potential proxy for social isolation and loneliness (Erzen & Çikrikci, 2018). Both PCA1 and PCA2 also exhibited positive associations with the proportion of White residents, consistent with prior evidence suggesting higher antidepressant usage in this demographic (Sclar et al., 2008).

Conversely, lower prescription rates were observed in areas with high proportions of Muslim residents, recent immigrants, and individuals with limited English proficiency. These trends may reflect not only cultural stigma or differing health beliefs (Assari, 2017; Loewenthal et al., 2002), but also structural barriers to care. For instance, language limitations have been identified as a barrier to receiving appropriate depression treatment (Lesser et al., 2008). This suggests that observed disparities may arise from inequities in access rather than differences in need. Figs. 8 A–D illustrate representative neighbourhood typologies from these areas.

5.2.2. Adding to ongoing debates

Our findings contribute to several ongoing debates within the urban health literature by challenging commonly accepted associations and highlighting context-specific dynamics.

First, we observe a positive correlation between green land cover (e.g., crops, shrubland) and antidepressant prescriptions. While most research links exposure to green spaces with lower depression incidence (Browning et al., 2019), some recent studies have reported similar counterintuitive trends (Astell-Burt et al., 2022; Hyam, 2020). These discrepancies may reflect heterogeneity in green space accessibility, usage patterns, or the socio-demographic composition of green areas. For example, peripheral neighbourhoods rich in vegetation may coincide with older or more isolated populations, partially explaining the observed increase in prescriptions.

Second, working more than 49 h per week (exceeding the UK Working Time Directive²) is negatively correlated with antidepressant prescriptions ($r = -0.29$ with PCA1). This pattern differs from prior literature associating long working hours with elevated depression risk (Melchior et al., 2007). One possible interpretation is that this variable may proxy for correlated factors such as stable employment and income, both of which are often associated with better mental health outcomes. This aligns with newer perspectives suggesting that certain types of work-related stress may be positively associated with purpose and psychological resilience (Šćepanović et al., 2023a).

Third, we find a negative correlation between NO₂ levels and antidepressant use, a result that diverges from the prevailing consensus linking air pollution to adverse mental health outcomes (Lim et al., 2012). Similar trends were observed in other geospatial studies (Fleury et al., 2021) and within the MedSAT dataset itself (Maitra et al., 2025; Šćepanović et al., 2023b). A possible explanation lies in urban policy interventions: London's Ultra Low Emission Zone (ULEZ), introduced in stages since 2019,³ may have spatially decoupled pollution levels from mental health burdens by redistributing high-emission zones away from vulnerable populations.

5.2.3. Insights warranting future research

UST-GNN's pipeline, coupled with the comprehensive MedSAT dataset, enables a comprehensive exploration of urban-environment-health interrelations beyond the scope of conventional public health analyses. This modelling approach surfaced several previously under-explored associations that merit further investigation.

Among them, we find a positive correlation between antidepressant prescriptions and both train commuting rates and employment in caring and leisure occupations. While public transport and caregiving roles are typically viewed as health-supportive due to increased accessibility and social engagement, their observed links with higher prescription rates may reflect context-dependent stressors such as emotional labour, shift work, or pandemic-related occupational burnout. These findings suggest that presumed protective urban or occupational features may carry hidden mental health burdens depending on exposure intensity and socio-environmental context.

Furthermore, environmental hydrological features, specifically-total surface runoff and canopy evaporation—show consistent negative correlations with antidepressant prescriptions. Total runoff, which quantifies the volume of precipitation-induced surface water flow, may serve as a proxy for effective stormwater infrastructure and urban green design, often characteristic of walkable, eco-resilient neighbourhoods. Similarly, higher canopy evaporation, primarily driven by extensive tree cover, contributes to urban cooling and microclimatic regulation, mitigating thermal stress and promoting psychological comfort. These environmental buffers may form indirect yet impactful pathways for mental well-being in dense urban settings.

Together, these associations highlight the ability of graph-based spatial machine learning to reveal hidden structural and environmental drivers of urban health. Advancing this work requires transdisciplinary collaboration between urban science, causal inference, and public health to connect model-derived hypotheses with individual-level data, qualitative interviews and targeted interventions. Integrating such insights into policy can help shape more equitable and responsive urban health environments.

6. Conclusion

This study demonstrates the effectiveness of a unified spatio-topological graph neural network (UST-GNN) for neighbourhood-level health outcome prediction from rich socio-demographic and environmental datasets. In a London case study on health prescriptions, UST-GNN consistently outperformed statistical, machine learning and geographically enriched baselines by jointly modelling nonlinear spatial dependencies, network topology, with heterogeneous features in strict spatial cross validation settings. Principal component interpretation analysis further demonstrated that the model not only achieved high predictive accuracy but also produced interpretable and meaningful links to real-world urban patterns that validated established drivers, contributed to ongoing debates, and revealed novel associations warranting further causal investigation.

While demonstrated here in an urban health context, UST-GNN is inherently outcome-agnostic and can be readily applied to a wide range of spatial prediction tasks, such as modelling energy demand, housing prices, or environmental risk indices, offering a reproducible GeoAI pipeline for urban analytics, spatial epidemiology, and equitable city planning. Beyond these applications, the proposed framework can serve as an analytical pipeline within emerging *urban digital twin* systems, that continuously integrate real-time data for monitoring, simulation, and policy experimentation (Malleson et al., 2024). Embedding UST-GNN into digital-twin workflows (e.g. agent-based-models) would enable data-driven equitable urban planning to foster healthier and more sustainable cities.

CRedit authorship contribution statement

Minwei Zhao: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sanja Šćepanović:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration,

² UK Government guidance on maximum weekly working hours.

³ Transport for London Ultra Low Emission Zone policy.

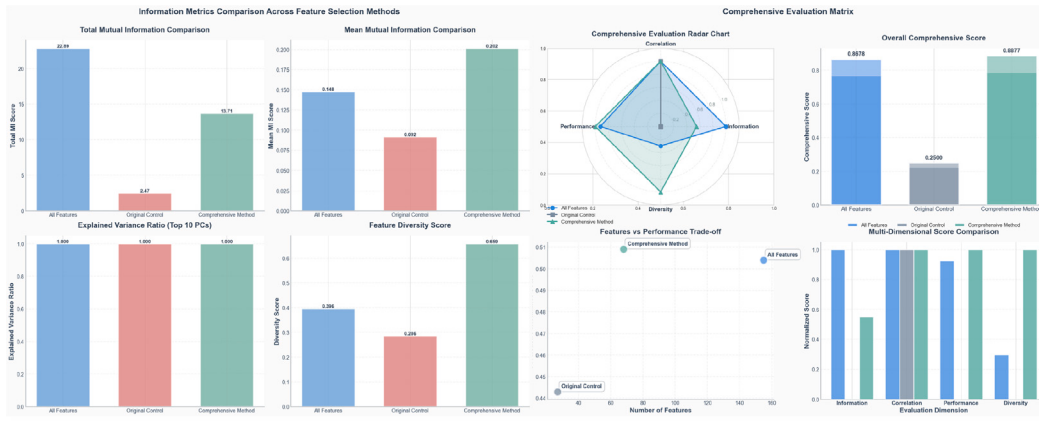


Fig. A.9. Comprehensive evaluation of feature selection strategies. Panels summarise: (top-left) total mutual information (MI) of retained features; (top-middle) mean MI; (top-right) radar plot comparing four normalised dimensions (information, correlation control, performance, diversity); (top-far-right) overall composite score. Bottom row: (left) explained variance ratio (EVR) of the top 10 principal components; (middle-left) diversity score measuring category coverage; (middle-right) trade-off between number of features and performance; (right) per-dimension normalised scores. The “Comprehensive Method” balances information content and diversity while maintaining high performance with a compact feature set, improving over the naive “All Features” and the minimal “Original Control” set.

Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Stephen Law:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Ivica Obadić:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Formal analysis. **Cai Wu:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Investigation, Formal analysis. **Daniele Quercia:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Investigation, Formal analysis, Conceptualization.

AI Use Statement

The authors used ChatGPT and Google Gemini to assist with language editing and improving readability. All output was reviewed, edited, and verified by the authors, who take full responsibility for the final content of the manuscript

Acknowledgments

This work was supported by the UK Research and Innovation (UKRI) Economic and Social Research Council (ESRC) [grant number ES/L011840/1], the German Federal Ministry of Research, Technology and Space [grant number 50EE2201C], and Smart Data Research UK, a UKRI investment [grant number UKRI4017].

Appendix A. Feature selection methodology

This appendix provides detailed information about the feature selection process used to reduce the dimensionality of the MedSAT dataset from 155 variables to 67 variables.

Fig. A.9 compares three strategies evaluated in this study: using all variables (*All Features*), a minimal confounder set (*Original Control*), and our multi-criteria pipeline (*Comprehensive Method*). The latter maximises information (total/mean MI) under explicit correlation control and achieves near-saturated EVR with markedly fewer variables, yielding higher composite scores and a more diverse, policy-relevant feature portfolio. This selection underpins the stability of UST-GNN and the clarity of subsequent interpretability analyses.

A.1. Comprehensive multi-dimensional feature selection strategy

Our feature selection strategy prioritised theoretical relevance and causal inference robustness while leveraging data-driven insights to optimise predictive performance. We selected 67 variables from 155 available features through a systematic multi-dimensional evaluation framework that integrates domain knowledge with statistical learning approaches, ensuring comprehensive coverage of socio-demographic, environmental, and behavioural determinants of health outcomes.

The selection process comprised four key stages:

(i) Multi-dimensional feature scoring: We computed feature importance scores using four complementary evaluation metrics: mutual information (MI) regression, F-statistics from univariate linear regression, random forest feature importances, and Lasso regularisation coefficients. Each metric captured different aspects of feature–outcome relationships: MI quantified non-linear associations, F-statistics assessed linear dependencies, random forest importances captured interactive effects, and Lasso coefficients reflected sparsity-regularised contributions. Individual scores were normalised to [0, 1] and combined using weighted averaging (weights: 0.3 for MI, 0.2 for F-statistics, 0.3 for random forest, and 0.2 for Lasso), producing a unified *combined score* that balanced multiple statistical perspectives.

(ii) Adaptive feature number determination: Rather than fixing a predefined feature count, we applied cross-validation (5-fold) over candidate feature set sizes ($k \in [30, 80]$) to identify the optimal dimensionality that maximised predictive performance while maintaining model stability. For each candidate k , we selected the top- k features ranked by combined scores and evaluated their collective performance using Random Forest regression with cross-validation R^2 as the selection criterion.

(iii) Mandatory control variable inclusion: Core control variables commonly used as key confounders in public health and broader socio-spatial research, including age distributions, gender, ethnicity, occupation types, population density, and key environmental exposures (NDVI, NO_2 , land cover types)—were automatically retained regardless of their combined scores. This ensured that theoretically essential predictors remained represented across all model configurations. Because a small set of theory-driven control variables was mandatorily retained, some residual collinearity may remain in the final feature set. This reflects a deliberate trade-off between strict redundancy removal and the preservation of substantively important controls.

(iv) Redundancy elimination: To address multicollinearity while preserving feature diversity, we computed pairwise correlation matrices

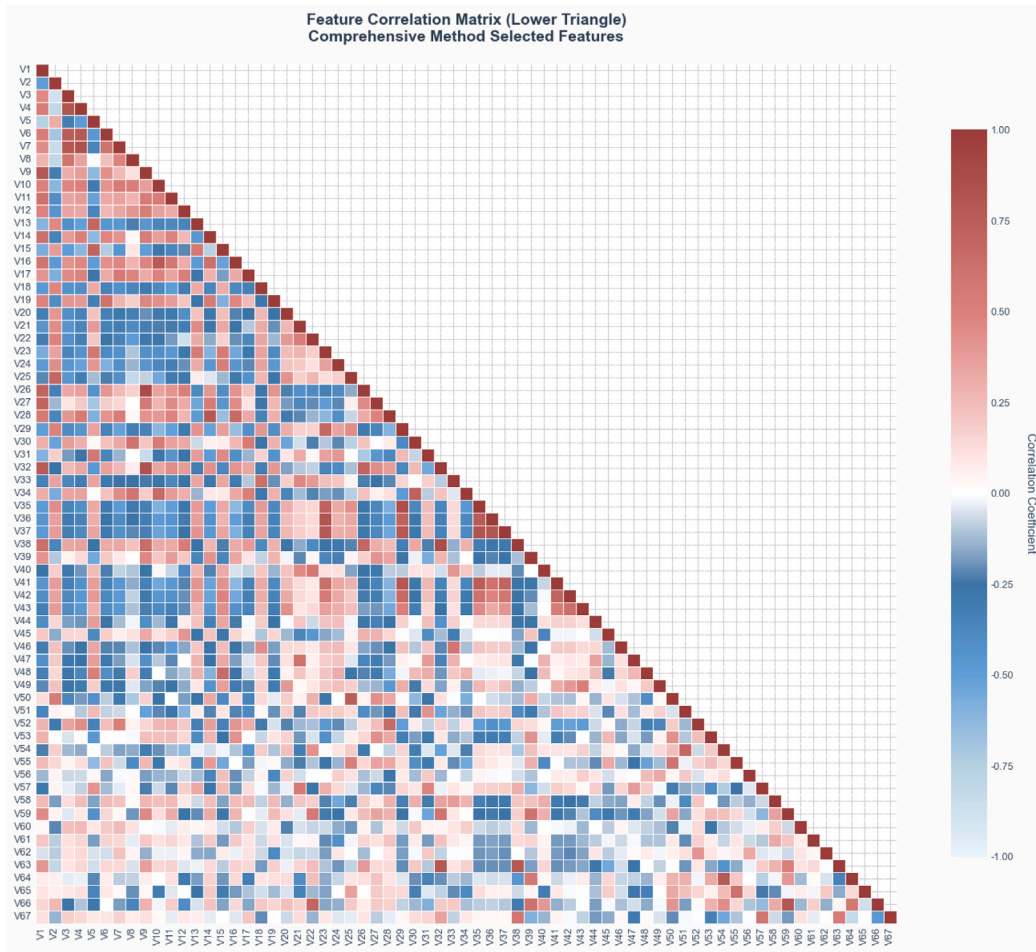


Fig. A.10. Pairwise correlation matrix (lower triangle) of the 67 variables retained by the comprehensive feature selection method. Each cell represents the Pearson correlation coefficient between two features (red: positive; blue: negative). The pattern illustrates that strong pairwise dependencies ($|r| > 0.8$) were effectively minimised through the redundancy elimination step, resulting in a feature set that maintains informational diversity across socio-demographic, environmental, and behavioural dimensions.

for selected features and identified highly correlated pairs ($|r| > 0.8$). For each pair, we retained the feature with the higher combined score (or prioritised control variables when one feature was a mandatory control), removing redundant predictors without substantial information loss.

This multi-stage approach balances dimensionality reduction with the retention of sufficient feature diversity to capture complex dependencies, thereby supporting both model stability and interpretable embedding–feature associations. The final feature set comprised 67 variables (54 socio-demographic and 13 environmental), representing an optimal balance between comprehensiveness and parsimony. Fig. A.10 visualises the pairwise correlation structure among the 67 variables selected by our comprehensive multi-dimensional feature selection framework. The lower-triangular correlation map demonstrates that the retained variables exhibit a balanced mix of moderate associations without strong linear redundancy. Clusters of moderate correlations (notably within age-related or occupation-related variables) indicate internally coherent thematic groups, yet the overall absence of dense high-correlation blocks ($|r| > 0.8$).

A.2. Selected variables summary

Table A.6 presents the complete list of 67 selected variables, organised by category. The selection encompasses key demographic characteristics (age distribution, gender, ethnicity), socioeconomic indicators

(occupation, housing, deprivation), and environmental exposures (air quality, greenness, land cover).

Appendix B. Interpretability

See Figs. B.11 and B.12.

B.1. Motivation and evaluation criteria

Interpreting spatial graph models requires balancing methodological faithfulness and practical communicability. In the context of urban analytics, meaningful explanations should (i) remain faithful to the trained model, (ii) exhibit stability under random seeds and data perturbations, (iii) incorporate spatial topology rather than treating nodes independently, and (iv) provide outputs interpretable to domain experts such as planners or epidemiologists. Guided by these principles, we evaluate and compare four complementary interpretability families — embedding PCA (ours), gradient-based saliency, permutation importance, and GNNExplainer — each representing a different trade-off between global generality and local fidelity.

B.2. PCA-embedding (ours)

Our PCA-embedding interpretation departs from traditional attribution frameworks by focusing on the structure of the learned representation rather than direct input–output gradients. Applying PCA

Table A.6
Complete list of 67 selected variables organised by category.

Category	Selected variables
Demographics	Age groups (10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+), Male proportion
Ethnicity	Mixed race, White, Asian, Black populations
Socioeconomic	Professional occupations, Population density, Household deprivation
Environment	NO ₂ , NDVI, Water bodies, Trees, Grass, Bare land
Housing	Central heating availability, Communal heating, Occupancy ratings
Transport	Commute modes (foot, metro rail, bus, bicycle, train), Distance to work
Employment	Working hours, Part-time work, Work from home, Student population
Residential	Length of residence, Marital status, English proficiency
Religion	Christian, Buddhist, Muslim populations
Climate	Lake depth, Snow cover, Surface runoff, Snow/ice, Crops, Canopy evaporation, Shrub cover

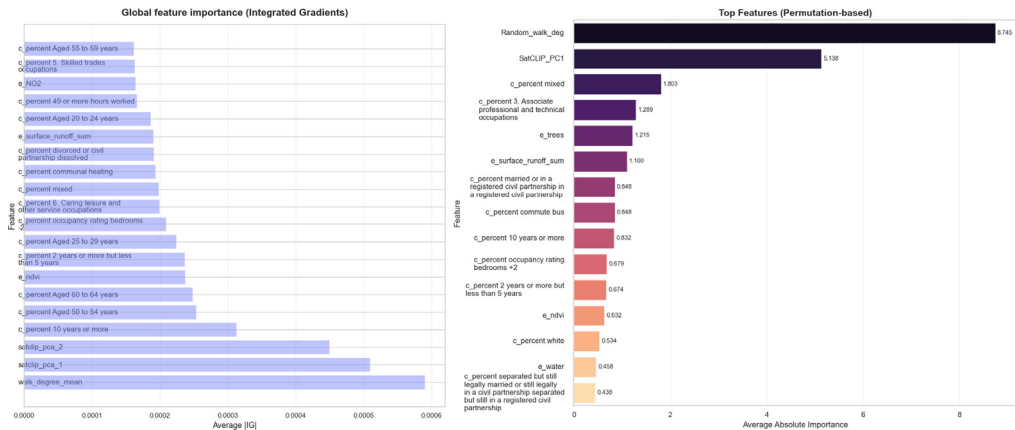


Fig. B.11. Permutation-based global feature importance. Bars show performance degradation after shuffling each feature, summarising city-wide contribution. Labels are wrapped for readability.

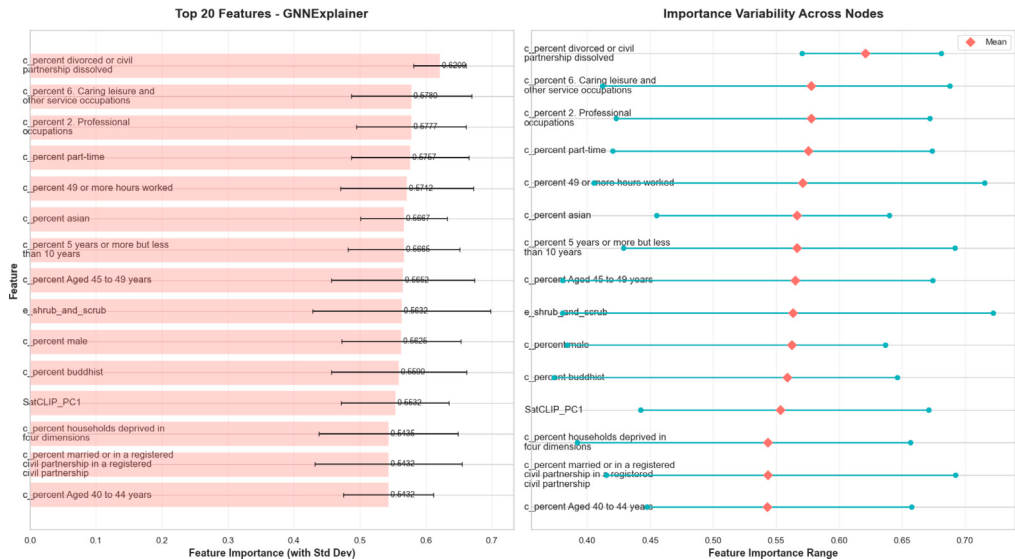


Fig. B.12. GNNExplainer-based summaries over five representative nodes (10th–90th percentiles). Left: bar chart with uncertainty; Right: range of importance across nodes. Stochastic optimisation and sparse sampling can yield local deviations from global rankings.

to the final-layer node embeddings enables us to identify dominant axes of variation that capture the macro-scale organisation of urban socio-environmental patterns. The approach is architecture-agnostic, computationally efficient, and stable across folds, producing interpretable spatial gradients that link *raw covariates*, *learned embeddings*, and *observed outcomes* within a single analytic space.

Although PCA does not assign additive feature importances and may blend correlated influences, its inductive and correlation-oriented nature makes it particularly suited to complex urban systems where strict causal decomposition is rarely identifiable. By summarising high-dimensional latent spaces into continuous spatial surfaces, the PCA method transforms model internal representations into city-scale narratives—revealing coherent gradients and clusters that conventional attribution methods fail to convey.

B.3. Gradient-based and integrated gradients methods

Gradient-based saliency and Integrated Gradients (IG) provide node-level sensitivity analyses that are directly faithful to model derivatives. They excel at highlighting how small input perturbations influence predictions and are theoretically grounded (IG satisfies axioms such as completeness). However, in graph contexts these gradients entangle self-node and neighbour effects, complicating their aggregation into global or spatially coherent interpretations. In our experiments, IG maps yielded consistent but spatially fragmented patterns, confirming their strength in local attribution but limited communicability at city scale.

B.4. Permutation-based importance

Permutation-based importance evaluates how shuffling each feature affects model performance, offering a model-agnostic and intuitively “what-if” assessment of feature relevance. The method produced stable global rankings aligned with known socio-environmental correlates (e.g., population density, greenery, occupation type), corroborating PCA-derived associations. Yet, because it ignores graph topology and can inflate correlated variables’ importance, permutation analysis complements but cannot substitute embedding-level structure discovery.

B.5. GNNExplainer

GNNExplainer directly optimises a sparse subgraph and feature mask that most strongly influence a node’s prediction. This yields rich local explanations incorporating topology, suitable for case-level insights or motif discovery. Nevertheless, its stochastic optimisation introduces variability across runs, and aggregating multiple local explanations into a consistent global interpretation is computationally intensive. As shown in Fig. B.12, feature rankings fluctuate across nodes, illustrating the challenge of scaling local explainers to city-wide contexts.

B.6. Cross-method comparison and discussion

Figs. B.11 and B.12 summarise the three feature-level explanation methods, while Fig. 8 (main text) presents the embedding-level PCA interpretation. Across methods, key socio-demographic and environmental variables—such as age structure, occupation type, commuting mode, greenness, and ethnic composition—consistently emerge as explanatory factors, affirming the substantive coherence of the learned spatial patterns.

In essence, PCA-embedding interpretation reframes interpretability from an attribution problem to a structural one—seeking to understand how the model organises information across space rather than merely how it responds to feature perturbations. This paradigm aligns more naturally with the goals of urban science, where understanding spatial heterogeneity and systemic dependencies often outweighs pinpointing individual variable causality.

Summary. PCA offers a transparent, stable, and computationally light interpretability pathway, revealing global structure–function relationships that complement the local precision of gradient and GNNExplainer methods. Together, these complementary techniques form a triangulated interpretability framework: PCA provides macro-scale coherence, gradients ensure model faithfulness, permutation tests assess robustness, and GNNExplainer yields local topology awareness. Their convergence substantiates that UST-GNN captures genuine, spatially embedded mechanisms rather than artefacts of data correlation or model leakage.

Takeaway. Our PCA-embedding method excels at revealing fold-stable, spatially coherent gradients and relating them to socio-environmental variables, offering a compact, communicable narrative. Gradient/IG and permutation importance provide faithful, feature-level corroboration; GNNExplainer offers case-level, topology-aware insight.

Appendix C. Robustness and reliability tests

C.1. Fast permutation test (OOB-based)

We conducted a fast permutation test using out-of-fold (OOF) predictions to assess whether the observed performance could arise by chance. Holding the OOF predictions fixed and randomly permuting the ground-truth labels 1000 times produced a null distribution of R^2 centred at -0.8819 with standard deviation 0.0093 . The observed OOF R^2 was 0.926 , yielding a permutation p -value of < 0.001 (resolution-limited by 1000 permutations). The complete separation between the observed value and the null confirms that the model captures genuine structure rather than accidental fit.

C.2. Lightweight learning curve

To verify generalisation with reduced runtime, we computed a lightweight learning curve using five training sizes (20%–100%), capped epochs, and a fixed test fold. Test R^2 increased monotonically with training size (approx. $0.39 \rightarrow 0.83$), while train R^2 decreased towards the test curve, indicating reduced variance and healthy capacity control. Together with the permutation test, these results support the reliability and robustness of UST-GNN.

Data availability

I have shared a figshare link to my data/code in the response to reviewer documents as part of the reviewing process.

References

- Adnan, M., Alarood, A. A. S., Uddin, M. I., & ur Rehman, I. (2022). Utilizing grid search cross-validation with adaptive boosting for augmenting performance of machine learning models. *PeerJ Computer Science*, 8, Article e803.
- Aiello, L. M., Schifanella, R., Quercia, D., & Del Prete, L. (2019). Large-scale and high-resolution analysis of food purchases and health outcomes. *EPJ Data Science*, 8(1), 1–22.
- Alegría, M., NeMoyer, A., Falgàs Bagué, I., Wang, Y., & Alvarez, K. (2018). Social determinants of mental health: where we are and where we need to go. *Current Psychiatry Reports*, 20, 1–13.
- Anselin, L. (2009). Spatial regression. *The SAGE Handbook of Spatial Analysis*, 1, 255–276.
- Apter, A. J., Reisine, S. T., Affleck, G., Barrows, E., & ZuWallack, R. L. (1999). The influence of demographic and socioeconomic factors on health-related quality of life in asthma. *Journal of Allergy and Clinical Immunology*, 103(1), 72–78.
- Arcaya, M. C., Tucker-Seeley, R. D., Kim, R., Schnake-Mahl, A., So, M., & Subramanian, S. V. (2016). Research on neighborhood effects on health in the United States: a systematic review of study characteristics. *Social Science & Medicine*, 168, 16–29.
- Armitage, R. (2021). Antidepressants, primary care, and adult mental health services in England during COVID-19. *The Lancet Psychiatry*, 8(2), Article e3.
- Assari, S. (2017). Social determinants of depression: The intersections of race, gender, and socioeconomic status. *Brain Sciences*, 7(12), 156.

- Astell-Burt, T., Navakatikyan, M., Eckermann, S., Hackett, M., & Feng, X. (2022). Is urban green space associated with lower mental healthcare expenditure? *Social Science & Medicine*, 292, Article 114503.
- Batty, M. (2018). Artificial intelligence and smart cities.
- Bell, J. E., Brown, C. L., Conlon, K., Herring, S., Kunkel, K. E., Lawrimore, J., & Uejio, C. (2018). Changes in extreme events and the potential impacts on human health. *Journal of the Air & Waste Management Association*, 68(4), 265–287.
- Berrie, L., Feng, Z., Rice, D., Clemens, T., Williamson, L., & Dibben, C. (2024). Does cycle commuting reduce the risk of mental ill-health? An instrumental variable analysis using distance to nearest cycle path. *International Journal of Epidemiology*, 53(1), dyad153.
- Bhakta, I., & Sau, A. (2016). Prediction of depression among senior citizens using machine learning classifiers. *International Journal of Computer Applications*, 144(7), 11–16.
- Bhatti, U. A., Tang, H., Wu, G., Marjan, S., & Hussain, A. (2023). Deep learning with graph convolutional networks: An overview and latest applications in computational intelligence. *International Journal of Intelligent Systems*, 2023(1), Article 8342104.
- Birks, D., Heppenstall, A., & Maleson, N. (2020). Towards the development of societal twins. 325, In *Frontiers in artificial intelligence and applications* (pp. 2883–2884). Leeds.
- Bixby, H., Hodgson, S., Fortunato, L., Hansell, A., & Fecht, D. (2015). Associations between green space and health in english cities: an ecological, cross-sectional study. *PLoS One*, 10(3), Article e0119495.
- Bonacini, L., Gallo, G., & Scicchitano, S. (2021). Working from home and income inequality: risks of a 'new normal' with COVID-19. *Journal of Population Economics*, 34(1), 303–360.
- Bor, W., Dean, A. J., Najman, J., & Hayatbakhsh, R. (2014). Are child and adolescent mental health problems increasing in the 21st century? A systematic review. *Australian & New Zealand Journal of Psychiatry*, 48(7), 606–616.
- Brody, S., Alon, U., & Yahav, E. (2021). How attentive are graph attention networks?. arXiv preprint arXiv:2105.14491.
- Brown, C. F., Kazmierski, M. R., Pasquarella, V. J., Rucklidge, W. J., Samskova, M., Zhang, C., Shelhamer, E., Lahera, E., Wiles, O., Ilyushchenko, S., et al. (2025). Alphaearth foundations: An embedding field model for accurate and efficient global mapping from sparse label data. arXiv preprint arXiv:2507.22291.
- Browning, M. H., Lee, K., & Wolf, K. L. (2019). Tree cover shows an inverse relationship with depressive symptoms in elderly residents living in US nursing homes. *Urban Forestry & Urban Greening*, 41, 23–32.
- Cahoon, C. G. (2012). Depression in older adults. *AJN the American Journal of Nursing*, 112(11), 22–30.
- Capolongo, S., Rebecchi, A., Buffoli, M., Appolloni, L., Signorelli, C., Fara, G. M., & D'Alessandro, D. (2020). COVID-19 and cities: From urban health strategies to the pandemic challenge. a decalogue of public health opportunities. *Acta Bio Medica: Atenei Parmensis*, 91(2), 13.
- Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1721–1730).
- Chen, B., & Kan, H. (2008). Air pollution and population health: a global challenge. *Environmental Health and Preventive Medicine*, 13(2), 94–101.
- Chen, Y., Wu, L., & Zaki, M. (2020). Iterative deep graph learning for graph neural networks: Better and robust node embeddings. vol. 33, In *Advances in neural information processing systems* (pp. 19314–19326).
- Cheng, T., Haworth, J., Anbaroglu, B., Tanaksaranond, G., & Wang, J. (2021). Spatiotemporal data mining. *Handbook of Regional Science*, 1691–1709.
- Cheng, J., Sun, J., Yao, K., Xu, M., & Cao, Y. (2022). A variable selection method based on mutual information and variance inflation factor. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 268, Article 120652.
- Choi, E., Xu, Z., Li, Y., Dusenberry, M., Flores, G., Xue, E., & Dai, A. (2020). Learning the graphical structure of electronic health records with graph convolutional transformer. vol. 34, In *Proceedings of the AAAI conference on artificial intelligence* (01), (pp. 606–613). AAAI Press.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.
- Cole, H. V., Mehdiapanah, R., Gullón, P., & Triguero-Mas, M. (2021). Breaking down and building up: gentrification, its drivers, and urban health inequality. *Current Environmental Health Reports*, 8, 157–166.
- Comber, A. J., Brunson, C., & Radburn, R. (2011). A spatial analysis of variations in health access: linking geography, socio-economic status and access perceptions. *International Journal of Health Geographics*, 10(1), 44.
- Compton, M. T., & Shim, R. S. (2015). The social determinants of mental health. *Focus*, 13(4), 419–425.
- Corburn, J. (2004). Confronting the challenges in reconnecting urban planning and public health. *American Journal of Public Health*, 94(4), 541–546.
- Cousens, S., Hargreaves, J., Bonell, C., Armstrong, B., Thomas, J., Kirkwood, B. R., & Hayes, R. (2011). Alternatives to randomisation in the evaluation of public-health interventions: statistical analysis and causal inference. *Journal of Epidemiology & Community Health*, 65(7), 576–581.
- De Sabbata, S., & Liu, P. (2023). A graph neural network framework for spatial geodemographic classification. *International Journal of Geographical Information Science*, 37(12), 2464–2486.
- dos Santos, B. S., Steiner, M. T. A., Fenerich, A. T., & Lima, R. H. P. (2019). Data mining and machine learning techniques applied to public health problems: A bibliometric analysis from 2009 to 2018. *Computers & Industrial Engineering*, 138, Article 106120.
- Dwivedi, V. P., Luu, A. T., Laurent, T., Bengio, Y., & Bresson, X. (2021). Graph neural networks with learnable structural and positional representations. arXiv preprint arXiv:2110.07875.
- Erzen, E., & Çikrikci, Ö. (2018). The effect of loneliness on depression: A meta-analysis. *International Journal of Social Psychiatry*, 64(5), 427–435.
- Fecht, D., Fischer, P., Fortunato, L., Hoek, G., De Hoogh, K., Marra, M., Kruize, H., Vienneau, D., Beelen, R., & Hansell, A. (2015). Associations between air pollution and socioeconomic characteristics, ethnicity and age profile of neighbourhoods in England and the Netherlands. *Environmental Pollution*, 198, 201–210.
- Feng, J., Chen, Y., Li, F., Sarkar, A., & Zhang, M. (2022). How powerful are k-hop message passing graph neural networks. vol. 35, In *Advances in neural information processing systems* (pp. 4776–4790).
- Fiscella, K., & Williams, D. R. (2004). Health disparities based on socioeconomic inequities: implications for urban health care. *Academic Medicine*, 79(12), 1139–1147.
- Fleury, V., Hims, R., Joost, S., Nicastro, N., Bereau, M., Guessous, I., & Burkhard, P. R. (2021). Geospatial analysis of individual-based parkinson's disease data supports a link with air pollution: A case-control study. *Parkinsonism & Related Disorders*, 83, 41–48.
- Fotheringham, A. S., Brunson, C., & Charlton, M. E. (2009). Geographically weighted regression. *The Sage Handbook of Spatial Analysis*, 1, 243–254.
- Fritz, C., Dorigatti, E., & Rügamer, D. (2022). Combining graph neural networks and spatio-temporal disease models to improve the prediction of weekly COVID-19 cases in Germany. *Scientific Reports*, 12(1), 3930.
- Fushiki, T. (2011). Estimation of prediction error by using K-fold cross-validation. *Statistics and Computing*, 21, 137–146.
- Gao, G., Hu, J., Wang, Y., & Wang, G. (2022). Regional inequalities and influencing factors of residents' health in China: analysis from the perspective of opening-up. *International Journal of Environmental Research and Public Health*, 19(19), 12069.
- Gao, X., Jiang, X., Haworth, J., Zhuang, D., Wang, S., Chen, H., & Law, S. (2024). Uncertainty-aware probabilistic graph neural networks for road-level traffic crash prediction. *Accident Analysis and Prevention*, 208, Article 107801.
- Gilardi, F., Gessler, T., Kubli, M., & Müller, S. (2021). Social media and policy responses to the COVID-19 pandemic in Switzerland. *Swiss Political Science Review*, 27(2), 243–256.
- Google Earth Engine (GEE) (2025). Google earth engine: A cloud-based platform for earth science data analysis. (Accessed 12 June 2025), <https://earthengine.google.com>.
- Greenacre, M., Groenen, P. J., Hastie, T., d'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, 2(1), 100.
- Grekousis, G. (2025). Geographical-xgboost: a new ensemble model for spatially local regression based on gradient-boosted trees. *Journal of Geographical Systems*, 1–27.
- Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems*, 30.
- Harvey, S. B., Modini, M., Joyce, S., Milligan-Saville, J. S., Tan, L., Mykletun, A., Bryant, R. A., Christensen, H., & Mitchell, P. B. (2017). Can work make you mentally ill? A systematic meta-review of work-related risk factors for common mental health problems. *Occupational and Environmental Medicine*, 74(4), 301–310.
- Hatch, S. L., Frissa, S., Verdecchia, M., Stewart, R., Fear, N. T., Reichenberg, A., Morgan, C., Kankulu, B., Clark, J., Gazard, B., et al. (2011). Identifying socio-demographic and socioeconomic determinants of health inequalities in a diverse London community: the south east London community health (SELCoH) study. *BMC Public Health*, 11, 1–17.
- Hayat, M. J., Powell, A., Johnson, T., & Cadwell, B. L. (2017). Statistical methods used in the public health literature and implications for training of public health professionals. *PLoS One*, 12(6), Article e0179032.
- Helbich, M. (2018). Toward dynamic urban environmental exposure assessments in mental health research. *Environmental Research*, 161, 129–135.
- Hossain, M. M., Tasnim, S., Sultana, A., Faizah, F., Mazumder, H., Zou, L., McKyer, E. L. J., Ahmed, H. U., & Ma, P. (2020). Epidemiology of mental health problems in COVID-19: a review. *F1000Research*, 9.
- Hunter, R. F., Nieuwenhuijsen, M., Fabian, C., Murphy, N., O'Hara, K., Rappe, E., Sallis, J. F., Lambert, E. V., Duenas, O. L. S., Sugiyama, T., et al. (2023). Advancing urban green and blue space contributions to public health. *The Lancet Public Health*, 8(9), e735–e742.
- Hyam, R. (2020). Greenness, mortality and mental health prescription rates in urban Scotland: a population level, observational study. *Research Ideas and Outcomes*, 6, Article e53542.
- Hynie, M. (2018). The social determinants of refugee mental health in the post-migration context: A critical review. *The Canadian Journal of Psychiatry*, 63(5), 297–303.
- Ijeh, S., Okolo, C. A., Arowoogun, J. O., Adeniyi, A. O., & Omotayo, O. (2024). Predictive modeling for disease outbreaks: a review of data sources and accuracy. *International Medical Science Research Journal*, 4(4), 406–419.

- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30.
- Kearns, M., & Ron, D. (1997). Algorithmic stability and sanity-check bounds for leave-one-out cross-validation. In *Proceedings of the tenth annual conference on computational learning theory (COLT 1997)* (pp. 152–162). ACM.
- Kedzie, C., McKeown, K., & Daume III, H. (2018). Content selection in deep learning models of summarization. arXiv preprint arXiv:1810.12343.
- Kihal-Talantikite, W., Padilla, C. M., Lalloué, B., Gelormini, M., Zmirou-Navier, D., & Deguen, S. (2013). Green space, social inequalities and neonatal mortality in France. *BMC Pregnancy and Childbirth*, 13, 1–9.
- Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.
- Kjellstrom, T. (2016). Impact of climate conditions on occupational health and related economic losses: a new feature of global and urban health in the context of climate change. *Asia Pacific Journal of Public Health*, 28(2 suppl), 28S–37S.
- Kjellstrom, T., Friel, S., Dixon, J., Corvalan, C., Rehfuess, E., Campbell-Lendrum, D., Gore, F., & Bartram, J. (2007). Urban environmental health hazards and health equity. *Journal of Urban Health*, 84(Suppl 1), 86–97.
- Klemmer, K., Rolf, E., Robinson, C., Mackey, L., & Rußwurm, M. (2025). Satclip: Global, general-purpose location embeddings with satellite imagery. 39, In *Proceedings of the AAAI Conference on Artificial Intelligence* (4), (pp. 4347–4355).
- Klemmer, K., Safir, N. S., & Neill, D. B. (2023). Positional encoder graph neural networks for geographic data. In *International conference on artificial intelligence and statistics* (pp. 1379–1389). PMLR.
- Kong, J. D., Akpudo, U. E., Effoduh, J. O., & Bragazzi, N. L. (2023). Leveraging responsible, explainable, and local artificial intelligence solutions for clinical public health in the global south. 11, In *Healthcare* (4), (p. 457). MDPI.
- Krenz, K., Dhanani, A., McEachan, R. R., Sohal, K., Wright, J., & Vaughan, L. (2023). Linking the urban environment and health: an innovative methodology for measuring individual-level environmental exposures. *International Journal of Environmental Research and Public Health*, 20(3), 1953.
- Lalji, H. M., McGrogan, A., & Bailey, S. J. (2021). An analysis of antidepressant prescribing trends in England 2015–2019. *Journal of Affective Disorders Reports*, 6, Article 100205.
- Lalloué, B., Monnez, J. M., Padilla, C., Kihal, W., Le Meur, N., Zmirou-Navier, D., & Deguen, S. (2013). A statistical procedure to create a neighborhood socioeconomic index for health inequalities analysis. *International Journal for Equity in Health*, 12, 1–11.
- Lantz, P. M., House, J. S., Lepkowski, J. M., Williams, D. R., Mero, R. P., & Chen, J. (1998). Socioeconomic factors, health behaviors, and mortality: results from a nationally representative prospective study of US adults. *Jama*, 279(21), 1703–1708.
- Lesser, I., Zisook, S., Flores, D., Sciolla, A., Wisniewski, S., Cook, I., Epstein, M., Rosales, A., Gonzalez, C., Trivedi, M., et al. (2008). Depression outcomes of spanish- and english-speaking hispanic outpatients in STAR* D. *Psychiatric Services*, 59(11), 1273–1284.
- Li, Z. (2022). Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost. *Computers, Environment and Urban Systems*, 96, Article 101845.
- Li, M., Gao, S., Lu, F., Liu, K., Zhang, H., & Tu, W. (2021). Prediction of human activity intensity using the interactions in physical and social spaces through graph convolutional networks. *International Journal of Geographical Information Science*, 35(12), 2489–2516.
- Li, X., Sun, L., Ling, M., & Peng, Y. (2023). A survey of graph neural network based recommendation in social networks. *Neurocomputing*, 549, Article 126441.
- Lim, Y. H., Kim, H., Kim, J. H., Bae, S., Park, H. Y., & Hong, Y. C. (2012). Air pollution and symptoms of depression in elderly adults. *Environmental Health Perspectives*, 120(7), 1023–1028.
- Liu, Z., & Liu, C. (2024). The association between urban density and multiple health risks based on interpretable machine learning: A study of American urban communities. *Cities*, 153, Article 105170.
- Liu, P., Zhang, Y., & Biljecki, F. (2024). Explainable spatially explicit geospatial artificial intelligence in urban analytics. *Environment and Planning B: Urban Analytics and City Science*, 51(5), 1104–1123.
- Loewenthal, K. M., MacLeod, A. K., & Cinnirella, M. (2002). Are women more religious than men? Gender differences in religious activity among different religious groups in the UK. *Personality and Individual Differences*, 32(1), 133–139.
- Lu, H., & Uddin, S. (2021). A weighted patient network-based framework for predicting chronic diseases using graph neural networks. *Scientific Reports*, 11(1), 22607.
- Mai, G., Janowicz, K., Yan, B., Zhu, R., Cai, L., & Lao, N. (2020). Multi-scale representation learning for spatial feature distributions using grid cells. arXiv preprint arXiv:2003.00824.
- Maitra, I., Lin, R., Chen, E., Donnelly, J., Šćepanović, S., & Rudin, C. (2025). How your location relates to health: Variable importance and interpretable machine learning for environmental and sociodemographic data. arXiv preprint arXiv:2501.02111.
- Malleson, N., Franklin, R., Arribas-Bel, D., Cheng, T., & Birkin, M. (2024). Digital twins on trial: Can they actually solve wicked societal problems and change the world for better?
- Marmot, M. (2005). Social determinants of health inequalities. *The Lancet*, 365(9464), 1099–1104.
- Mazumdar, S., Chong, S., Astell-Burt, T., Feng, X., Morgan, G., & Jalaludin, B. (2021). Which green space metric best predicts a lowered odds of type 2 diabetes? *International Journal of Environmental Research and Public Health*, 18(8), 4088.
- Melchior, M., Caspi, A., Milne, B. J., Danese, A., Poulton, R., & Moffitt, T. E. (2007). Work stress precipitates depression and anxiety in young, working women and men. *Psychological Medicine*, 37(8), 1119–1129.
- Meyes, R., Lu, M., de Puiseau, C. W., & Meisen, T. (2019). Ablation studies in artificial neural networks. arXiv preprint arXiv:1901.08644.
- Mhasawade, V., Zhao, Y., & Chunara, R. (2021). Machine learning and algorithmic fairness in public and population health. *Nature Machine Intelligence*, 3(8), 659–666.
- Molnar, C., König, G., Herbringer, J., Freisleben, T., Dandl, S., Scholbeck, C. A., Casalicchio, G., Grosse-Wentrup, M., & Bischl, B. (2020). Pitfalls to avoid when interpreting machine learning models.
- National Health Service (NHS) (2025). NHS health data. (Accessed 12 June 2025), <https://www.nhs.uk>.
- Ng, A. Y. (1997). Preventing "overfitting" of cross-validation data. In *Proceedings of the 14th international conference on machine learning (ICML 1997)* (pp. 245–253). Morgan Kaufmann.
- Nikparvar, B., & Thill, J.-C. (2021). Machine learning of spatial data. *ISPRS International Journal of Geo-Information*, 10(9), 600.
- Nutsford, D., Pearson, A. L., & Kingham, S. (2013). An ecological study investigating the association between access to urban green space and mental health. *Public Health*, 127(11), 1005–1011.
- Obradovich, N., Migliorini, R., Paulus, M. P., & Rahwan, I. (2018). Empirical evidence of mental health risks posed by climate change. *Proceedings of the National Academy of Sciences*, 115(43), 10953–10958.
- Office for National Statistics (ONS) (2025). National statistics. (Accessed 12 June 2025), <https://www.ons.gov.uk>.
- Ohanyan, H., Portengen, L., Kaplani, O., Huss, A., Hoek, G., Beulens, J. W., & Vermeulen, R. (2022). Associations between the urban exposome and type 2 diabetes: Results from penalised regression by least absolute shrinkage and selection operator and random forest models. *Environment International*, 170, Article 107592.
- Pineo, H., Glonti, K., Rutter, H., Zimmermann, N., Wilkinson, P., & Davies, M. (2018). Urban health indicator tools of the physical environment: a systematic review. *Journal of Urban Health*, 95, 613–646.
- Raco, M., Kesten, J., & Colomb, C. (2014). *Urban policies on diversity in London, United Kingdom*. Bartlett School of Planning, University College London, London.
- Remes, O., Mendes, J. F., & Templeton, P. (2021). Biological, psychological, and social determinants of depression: a review of recent literature. *Brain Sciences*, 11(12), 1633.
- Rothenberg, R., Weaver, S. R., Dai, D., Stauber, C., Prasad, A., & Kano, M. (2014). A flexible urban health index for small area disparities. *Journal of Urban Health*, 91(5), 823–835.
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215.
- Sarkar, C., & Webster, C. (2017). Urban environments and human health: Current trends and future directions. *Current Opinion in Environmental Sustainability*, 25, 33–44.
- Šćepanović, S., Constantinides, M., Quercia, D., & Kim, S. (2023a). Quantifying the impact of positive stress on companies from online employee reviews. *Scientific Reports*, 13(1), 1603.
- Šćepanović, S., Obadić, I., Joglekar, S., Giustarini, L., Nattero, C., Quercia, D., & Zhu, X. (2023b). MedSat: A public health dataset for England featuring medical prescriptions and satellite imagery. *Advances in Neural Information Processing Systems*, 36, 77810–77851.
- Schaub, M. T., Benson, A. R., Horn, P., Lippner, G., & Jadbabaie, A. (2020). Random walks on simplicial complexes and the normalized hodge 1-Laplacian. *SIAM Review*, 62(2), 353–391.
- Scheider, S., & Richter, K. F. (2023). Pragmatic geoai: Geographic information as externalized practice. *KI-Künstliche Intelligenz*, 37(1), 17–31.
- Sclar, D. A., Robison, L. M., & Skaer, T. L. (2008). Ethnicity/race and the diagnosis of depression and use of antidepressants by adults in the United States. *International Clinical Psychopharmacology*, 23(2), 106–109.
- Shane, A. M., & Graedel, T. E. (2000). Urban environmental sustainability metrics: a provisional set. *Journal of Environmental Planning and Management*, 43(5), 643–663.
- Shao, J., Han, Z., Yang, Q., & Zhou, T. (2015). Community detection based on distance dynamics. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1075–1084).
- Shrestha, R., Flacke, J., Martinez, J., & Van Maarseveen, M. (2016). Environmental health related socio-spatial inequalities: identifying "hotspots" of environmental burdens and social vulnerability. *International Journal of Environmental Research and Public Health*, 13(7), 691.
- Soini, E., Rosenström, T., Määttänen, I., & Jokela, M. (2024). Physical activity and specific symptoms of depression: A pooled analysis of six cohort studies. *Journal of Affective Disorders*, 348, 44–53.
- Song, X., Mitnitski, A., Cox, J., & Rockwood, K. (2004). Comparison of machine learning techniques with classical statistical models in predicting health outcomes. In *MEDINFO 2004* (pp. 736–740). IOS Press.

- Su, D., Zhang, X., He, K., & Chen, Y. (2021). Use of machine learning approach to predict depression in the elderly in China: a longitudinal study. *Journal of Affective Disorders*, 282, 289–298.
- Sun, R., Dai, H., & Yu, A. W. (2022). Does GNN pretraining help molecular representation? vol. 35, In *Advances in neural information processing systems* (pp. 12096–12109).
- Thotad, P. N., Bharamagoudar, G. R., & Anami, B. S. (2023). Diabetes disease detection and classification on Indian demographic and health survey data using machine learning methods. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 17(1), Article 102690.
- Trushna, T., Dhiman, V., Raj, D., & Tiwari, R. R. (2021). Effects of ambient air pollution on psychological stress and anxiety disorder: a systematic review and meta-analysis of epidemiological evidence. *Reviews on Environmental Health*, 36(4), 501–521.
- Tung, E. L., Cagney, K. A., Peek, M. E., & Chin, M. H. (2017). Spatial context and health inequity: reconfiguring race, place, and poverty. *Journal of Urban Health*, 94, 757–763.
- Twohig-Bennett, C., & Jones, A. (2018). The health benefits of the great outdoors: A systematic review and meta-analysis of greenspace exposure and health outcomes. *Environmental Research*, 166, 628–637.
- Van Draanen, J., Tsang, C., Mitra, S., Karamouzian, M., & Richardson, L. (2020). Socioeconomic marginalization and opioid-related overdose: a systematic review. *Drug and Alcohol Dependence*, 214, Article 108127.
- Velivcković, P., Fedus, W., Hamilton, W. L., Liò, P., Bengio, Y., & Hjelm, R. D. (2018). Deep graph infomax. arXiv preprint arXiv:1809.10341.
- Vlahov, D., & Galea, S. (2002). Urbanization, urbanicity, and health. *Journal of Urban Health*, 79, S1–S12.
- VoPham, T., Hart, J. E., Laden, F., & Chiang, Y. Y. (2018). Emerging trends in geospatial artificial intelligence (geoai): potential applications for environmental epidemiology. *Environmental Health*, 17, 1–6.
- Wang, X., Cheng, T., Law, S., Zeng, Z., Yin, L., & Liu, J. (2025). Multi-modal contrastive learning of urban space representations from POI data. *Computers, Environment and Urban Systems*, 120, Article 102299.
- White, M. P., Elliott, L. R., Grellier, J., Economou, T., Bell, S., Bratman, G. N., & Fleming, L. E. (2021). Associations between green/blue spaces and mental health across 18 countries. *Scientific Reports*, 11(1), 8903.
- Wirtz, B. W., Langer, P. F., & Fenner, C. (2021). Artificial intelligence in the public sector-a research agenda. *International Journal of Public Administration*, 44(13), 1103–1128.
- Wong, T. T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*, 48(9), 2839–2846.
- Wu, S., Sun, F., Zhang, W., Xie, X., & Cui, B. (2022). Graph neural networks in recommender systems: a survey. *ACM Computing Surveys*, 55(5), 1–37.
- Wu, C., Zhao, M., & Ye, Y. (2023). Measuring urban nighttime vitality and its relationship with urban spatial structure: A data-driven approach. *Environment and Planning B: Urban Analytics and City Science*, 50(1), 130–145.
- Wu, Y., Zhuang, D., Labbe, A., & Sun, L. (2021). Inductive graph neural networks for spatiotemporal kriging. vol. 35, In *Proceedings of the AAAI conference on artificial intelligence* (5), (pp. 4478–4485). AAAI Press.
- Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2018). How powerful are graph neural networks?. arXiv preprint arXiv:1810.00826.
- Yeh, P. K., Chen, H. W., & Chen, M. S. (2023). Random walk conformer: learning graph representation from long and short range. vol. 37, In *Proceedings of the AAAI conference on artificial intelligence* (9), (pp. 10936–10944). AAAI Press.
- Yue, H., He, C., Huang, Q., Zhang, D., Shi, P., Moallemi, E. A., Xu, F., Yang, Y., Qi, X., Ma, Q., et al. (2024). Substantially reducing global PM2. 5-related deaths under SDG3. 9 requires better air pollution control and healthcare. *Nature Communications*, 15(1), 2729.
- Zhang, Y., Liu, P., & Biljecki, F. (2023). Knowledge and topology: A two layer spatially dependent graph neural networks to identify urban functions with time-series street view image. *ISPRS Journal of Photogrammetry and Remote Sensing*, 198, 153–168.
- Zhang, S., Tong, H., Xu, J., & Maciejewski, R. (2019). Graph convolutional networks: a comprehensive review. *Computational Social Networks*, 6(1), 1–23.
- Zhao, J., & Pu, Y. (2025). Characterization and estimation of heterogeneous spatial autocorrelation in spatial autoregressive models. *PloS One*, 20(7), Article e0327316.
- Zhou, K., Huang, X., Song, Q., Chen, R., & Hu, X. (2022). Auto-gnn: Neural architecture search of graph neural networks. *Frontiers in Big Data*, 5, Article 1029307.
- Zhou, W., Wang, Q., Li, R., Kadier, A., Wang, W., Zhou, F., & Ling, L. (2023). Combined effects of heatwaves and air pollution, green space and blue space on the incidence of hypertension: a national cohort study. *Science of the Total Environment*, 867, Article 161560.