Mukara: a deep learning alternative to the four-step travel demand model with a case study on interurban highway traffic prediction in the UK

Yue Li[1*], Shujuan Chen[1], Ying Jin[1],

1 Martin Centre for Architectural and Urban Studies, University of Cambridge, Cambridge, UK

* yl901@cam.ac.uk

Abstract

Accurate traffic volume prediction is essential for managing congestion, enhancing road safety, mitigating environmental impacts, and supporting long-term transportation planning. The traditional four-step travel demand model (FSM) is a well-established framework, but it depends on static survey data and simplified assumptions that may not fully capture complex travel behaviour. Deep learning offers the potential to improve prediction accuracy by uncovering non-linear relationships and leveraging diverse, high-resolution data sources. However, existing deep learning models are mainly time-series models that rely heavily on historical traffic data, making them ineffective for predictions in locations beyond the existing sensor network. This study proposes "Mukara", a novel deep learning framework developed to predict weekday daily traffic volumes for all highway trunk road segments in England using only external features. The model is trained on eight years of data from England and Wales, incorporating inputs such as population, employment, land use, road networks, and points of interest. Mukara demonstrates moderate predictive performance that outperforms existing studies using FSM. The study highlights key limitations and outlines promising directions for future research. These include validating the feasibility of replacing individual steps of the FSM using synthetic data and exploring improved model architectures tailored to this domain.

Introduction

Road traffic prediction plays a pivotal role in addressing critical challenges such as pollution reduction, carbon emissions mitigation, and improving road safety and efficiency [1–4]. In the United Kingdom, road transport is the predominant mode of travel, accounting for 86% of all passenger kilometres in 2022 [5]. This trend is consistent with many OECD countries, where road transport similarly dominates passenger travel [6]. Simultaneously, vehicle ownership is rapidly increasing in the Global South, with projections indicating that by 2030, 56% of the world's vehicles will be owned by non-OECD countries, compared to 24% in 2002 [7]. A robust traffic prediction system can help travellers plan routes effectively, assist traffic operators in informed decision-making, and enhance overall traffic management efficiency [8].

Despite advancements in traffic prediction, research has predominantly focused on urban traffic, leaving interurban traffic networks relatively underexplored [9]. Interurban settings, however, present a unique opportunity for testing novel traffic prediction models due to the abundance of high-quality data and the relatively lower complexity of traffic patterns compared to urban areas. Urban traffic is often influenced by localized factors such as pedestrian activity, public transit systems, and highly variable demand patterns, making it noisier and more challenging to model. In contrast, interurban traffic data typically exhibits more stable and predictable patterns, making it ideal for evaluating the feasibility of innovative approaches like Mukara.

As shown in Table 1, traffic prediction has traditionally relied on two main approaches: the four-step travel demand model (FSM) and deep learning-based models. The FSM has long served as a foundational framework in transportation planning, consisting of four key stages: trip generation, trip distribution, modal split, and assignment [10]. These stages estimate travel demand based on socioeconomic factors and allocate it across different modes and network links using gravity models, discrete choice models, and network assignment algorithms. Despite its structured design and utility for long-term planning, the FSM often suffers from static assumptions, simplified network representations, and an inability to capture dynamic, real-time traffic variations. Moreover, the model's reliance on assumptions such as perfect user knowledge and homogeneous behaviour can lead to inaccuracies and limit its transferability. A case study in Istanbul found huge discrepancies between daily traffic predicted by FSM and the actual observed daily traffic volume [11].

Table 1. A comparison between the four-step travel demand model and existing deep learning models.

	The four-step model	**Deep learning models**
Model	A rule-based model based on planning assumptions	Empirical model extrapolating complex patterns from historical data, very accurate and detailed
Inputs	Population, household size, employment, land use characteristics, travel costs…	Recent historical sensor readings from the traffic network
Outputs	Trips generated, distribution among the means of transport, traffic flows for next 1, 5, 20 years	Immediate or short-term traffic prediction of traffic in the next 5, 15, 60 minutes (nowcasting)
Limitations	Low accuracy and resolution in its predictions due to reliance on simplified assumptions and rules	Black box, rely heavily on historical input, cannot predict locations with no sensors, low scalability and transferability, no insights into traffic determinants

On the other hand, deep learning models have emerged as powerful alternatives due to their ability to model complex spatial-temporal dependencies in traffic data [9, 12]. Architectures such as CNNs [13], RNNs [14], LSTMs [15], and GRUs [16] have significantly improved the modelling of temporal trends, while recent advancements like Transformers [17] and Graph Neural Networks (GNNs) [18] further enable spatial reasoning within graph-structured networks. State-of-the-art models such as ST-ResNet [19], DCRNN [20], ConvLSTM [21], PFNet [22], and STGAT [23] incorporate these techniques to produce highly accurate short-term predictions. However, most of these models treat traffic prediction as a time-series forecasting task based on historical sensor readings. As such, their outputs often reflect patterns seen in the past rather than insights into the determinants of traffic dynamics. This is illustrated in recent work such as ST-MetaNet [24, 25], where the predictions strongly mirror input trends. While some models incorporate external features (e.g., point of interests (POIs), road types, and event data) [19, 26], these are typically used as auxiliary inputs rather than primary determinants

This research attempts to combine the strengths of the FSM and deep learning approaches. The proposed model, Mukara, is designed to capture intricate spatial patterns and deliver accurate predictions while maintaining interpretability by focusing on the determinants of traffic, such as population, employment, land use, road networks, and POIs. Inspired by models like Deep Gravity [27], which uses external features to predict origin-destination (O-D) flows, Mukara goes a step further by using external determinants exclusively to predict traffic volumes, thereby replacing all four steps of the FSM. We name this model "Mukara", derived from the Japanese term meaning "from nothing", reflecting its objective of predicting traffic without relying on historical sensor readings.

Materials and methods

Overview

We now introduce the workflow of this study. First, we construct a highway trunk road network, where the graph structure encodes connectivity information relevant to trip distribution across O-D pairs. Each road segment is enriched with attributes such as road type, number of lanes, and functional classification, which correspond to the generalized travel cost inputs typically used in the assignment step of FSM. Next, we integrate rasterized geographic datasets including population density, employment statistics, land use areas, the number of various types of POIs, and aggregated measures of local road infrastructure. These inputs serve to approximate key elements of trip generation and modal choice, capturing the spatial and socioeconomic determinants of travel demand that are traditionally modelled using regression and discrete choice models within FSM. Ground truth traffic volume data is aligned with corresponding road segments to enable supervised learning.

Mukara is trained using sensor-labelled road segments in the training set and evaluated on geographically distinct test segments to assess spatial generalizability. By structuring inputs around the core components of FSM and learning end-to-end mappings to observed traffic volumes, Mukara should be able to learn the function of all four steps in a unified, data-driven, and interpretable framework. Fig 1 illustrates the complete methodological pipeline. All code, data, and additional materials are publicly available at `https://github.com/yueli901/mukara`.

Fig 1. Overview of the Mukara workflow.

Data

Highway network

To study interurban traffic dynamics, a highway network graph covering the entire region of England was constructed based on the National Highways Strategic Road Network [28]. This graph serves as the backbone for information propagation in Mukara. The network consists of 181 nodes ($N_v = 181$), representing highway trunk road junctions, typically located near major cities, and 498 edges ($N_e = 498$), corresponding to 249 highway trunk road segments in both directions. Each edge is assigned a sensor from the National Highway Traffic Information System (TRIS) [29], resulting in a total of 498 sensors. Nodes and edges were manually selected to capture traffic flow along trunk roads connecting major cities and towns, while avoiding bypasses and highway exits. Preference was given to sensors located closer to the midpoint of each road segment and with lower rates of missing data. The segmentation of the highway network, along with the selection of nodes, edges, and sensors, is illustrated in Figure 2.

Fig 2. Highway trunk road network map of England, showing the nodes (junctions) and edges (road segments) used in the Mukara model.

Features on highway trunk road segments in both directions were collected using the Google Routes API. The extracted edge features include driving duration, driving distance, straight-line distance, average driving speed, and detour factor. Together, these edge features, $\mathcal{E} = \{e_{ij}\}$, are structured into a tensor with dimensions $N_e \times D_e = 498 \times 5$, where N_e is the total number of edges and D_e is the number of features. All edge features are normalized before being input into the model. Figure 3 provides a visual summary of the distributions of these edge features.

Fig 3. Distributions of the five edge features used in the Mukara model: driving distance, driving duration, straight-line distance, average driving speed, and detour factor. These metrics were normalized for input into the model.

Population and employment

Population data was sourced from the Population Estimates - Small Area dataset [30], provided by the Office for National Statistics (ONS) through the NOMIS service. This dataset provides annual population estimates for England and Wales at the Lower Layer Super Output Area (LSOA) level, stratified by age group and sex. Employment data was obtained from the Business Register and Employment Survey (BRES) [31], also provided by ONS through NOMIS. This dataset includes employment counts, covering full-time, part-time, and self-employed workers across all industries within England and Wales.

For both datasets, data from the years 2015 to 2022 were selected. The LSOA-based data was rasterized into a 1 km x 1 km grid using LSOA boundaries provided by ONS [32, 33]. This process resulted in a grid tensor $\mathcal{M}_{pe}$ with dimensions $N_t \times D_h \times D_w \times D_{pe} = 8 \times 653 \times 573 \times 94$, where N_t is the number of years, D_h and D_w are the height and width of the grid, and D_{pe} represents the number of grid channels for population and employment.

To account for differences in travel behaviour across demographic groups, separate channels were created for population and employment strata. Specifically:

In each training task, a subset of these channels was selected to analyse the model's performance under different levels of stratification of the input data. Figure 4 visualizes the aggregated population and employment data in 2022 as heat maps.

Fig 4. Heat maps of aggregated population and employment across England and Wales in year 2022. Higher intensity indicates areas with larger population and employment density, based on LSOA-level data rasterized to a 1 km x 1 km grid.

Land use, road network, and POI

Land use, road network, and POI data were sourced from OpenStreetMap (OSM) and downloaded via Geofabrik's free download server [34]. The data were extracted from a historical snapshot of the England subregions and Wales .osm.pbf files with the

Table 2. Breakdown of raster input channels for population and employment features.

Feature Type	Subcategory	Number of Channels
Population	Total population (aggregated across all demographics)	1
	By sex (male and female)	2
	By age group (five predefined age bands)	5
	By age and sex (age groups further stratified by sex)	10
Employment	Total employment (aggregated across all sectors and types)	1
	By work type (full-time, part-time, self-employed)	3
	By sector (18 industry sectors)	18
	By work type and sector (combinations of the above categories)	54

timestamp 20230101. This static snapshot was chosen for all years from 2015 to 2022 to ensure consistency and completeness, as land use and POI data are relatively stable over time.

The tags used to extract the data are summarized in Table 3. These include land use classifications (e.g., residential, industrial), road network hierarchies from high-level motorways to low-level residential roads, and a diverse set of POI categories such as transport, food, health, education, and retail facilities. The selection of these tags was based on the Deep Gravity Model [27] and their availability in the OSM database.

Table 3. The categories, OpenStreetMap (OSM) keys, and associated values used to extract thematic raster layers for input features.

Category	Channel	OSM Key	OSM Values
Land Use	Residential	`landuse`	`residential`
	Commercial	`landuse`	`commercial`
	Industrial	`landuse`	`industrial`
	Retail	`landuse`	`retail`
Road Network	High Level	`highway`	`motorway`, `trunk`, `primary`
	Medium Level	`highway`	`secondary`, `tertiary`
	Low Level	`highway`	`residential`, `unclassified`
POI	Transport	`amenity` `railway`	`bus_station`, `parking` `station`, `stop`, `tram_stop`
	Food	`amenity`	`bar`, `cafe`, `restaurant`
	Health	`amenity`	`clinic`, `hospital`, `pharmacy`
	Education	`amenity`	`school`, `college`, `kindergarten`, `university`
	Retail	`shop`	`supermarket`, `department_store`, `mall`

For each grid cell, the following metrics were calculated: total area of each land use type, total length of roads for each hierarchical level, and total number of POIs for each category. These metrics were then aggregated into a grid tensor, $\mathcal{M}_{lp}$, with dimensions $D_h \times D_w \times D_{lp} = 653 \times 573 \times 12$, where D_h and D_w represent the height and width of the grid, and D_{lp} is the number of grid channels corresponding to the land use and POI categories. Figure 5 visualizes these features as heat maps. To construct the final grid features $\mathcal{M}$, the tensor $\mathcal{M}_{lp}$ was broadcast on the time axis, and then concatenated with the population and employment tensor $\mathcal{M}_{pe}$ along the channel dimension: $\mathcal{M} = \mathcal{M}_{pe} \| \mathcal{M}_{lp}$

Fig 5. Heat maps showing the distribution of 12 features including land use, road network, and POI across England and Wales. The features were aggregated to a 1 km x 1 km grid.

Traffic volume

The ground truth traffic volume data is sourced from the Traffic Information System (TRIS), managed by National Highways [35]. TRIS provides comprehensive data on traffic speed and volume, collected in 15-minute intervals using loop sensors. Across England, 19,364 sensors are integrated into this network, which has been operational since 2014. The data used in this study were accessed and downloaded through the API provided by TRIS [29].

To align the traffic volume data with the input features, we used 8 years of traffic volume records from January 1st, 2015, to December 31st, 2022, for each of the 498 sensors included in the established highway network. The mean weekday daily traffic volume was calculated for each year and each sensor to generate the ground truth tensor $\mathcal{Y} = \{y_{t,e}\}$, with dimensions $N_t \times N_e = 8 \times 498$. The aggregation to mean weekday daily traffic volumes was done to reduce the impact of noise and missing data, allowing the model to focus on capturing general traffic patterns.

Figure 6 shows a histogram of the average traffic volumes across the 498 sensors over the 8-year period, illustrating the variability in traffic levels. Figure 7 provides a spatial visualization of these volumes, averaged over both directions.

Fig 6. Histogram of average weekday daily traffic volumes across the 498 sensors, calculated over 8 years (2015–2023).

Fig 7. Spatial visualization of average weekday daily traffic volumes across the 498 sensors, averaged over 8 years (2015–2023) and aggregated for both directions of the same highway segment.

Mukara model

The Mukara model is designed to predict traffic volume $\mathcal{Y}$ by leveraging the highway network structure $\mathcal{G}$, edge features $\mathcal{E}$, and one year of grid features $\mathcal{M}_t$. Formally, this task is defined as:

$$\hat{\mathcal{Y}}_t = \text{Mukara}(\mathcal{G}, \mathcal{E}, \mathcal{M}_t) \tag{1}$$

To efficiently pass information from inputs to outputs, the Mukara model is composed of two main building blocks: a **CNN block** for processing spatial grid features, and a **GAT block** for capturing topological and relational information from the network. An overview of these blocks is provided in Figure 8 and Figure 9.

The CNN block is responsible for processing the grid features $\mathcal{M}_t$ to generate node-specific embeddings by extracting and encoding information from a fixed-size Region of Interest (ROI) centered around each node. For each of the N_v nodes v_i in the

Fig 8. A visual representation of the CNN block in the Mukara model. The block processes grid features ($\mathcal{M}_t$) by extracting regions of interest (ROIs) around each node, applying convolutional and pooling layers, and generating node embeddings ($\mathbf{h}_{v_i}^0$) for subsequent graph attention processing.

highway network, an ROI is extracted from $\mathcal{M}_t$. The extracted ROIs are three-dimensional tensors, with the first two dimensions representing spatial extent (height and width) and the third dimension representing the number of feature channels in $\mathcal{M}_t$.

Each ROI is passed through a series of convolutional layers with depth L_{CNN}, followed by ReLU activations and max-pooling operations. The final convolutional output is flattened into a one-dimensional vector, producing the initial node embedding $\mathbf{h}_{v_i}^0$. The entire CNN transformation can be expressed as:

$$\mathbf{h}_{v_i}^0 = \text{Flatten}\Big(\big(\text{Pooling} \circ \text{ReLU} \circ \text{Conv}\big)^{L_{\text{CNN}}}(\text{ROI}_{v_i})\Big) \quad \text{for} \quad \text{ROI}_{v_i} \in \mathcal{M}_t \tag{2}$$

This initial node embedding captures the spatial and socioeconomic context surrounding each node, which is related to the trip generation process, and serves as the starting point for graph-based reasoning.

Following the CNN block, the GAT block refines the node embeddings by integrating information from neighbouring nodes and the attributes of the connecting edges. Each edge embedding $\mathbf{h}_{e_{ij}}$ is generated via a Multi-Layer Perceptron (MLP) applied to the raw edge features $\mathcal{E}$:

$$\mathbf{h}_{e_{ij}} = \text{MLP}_{e_{ij} \to h_{e_{ij}}}(\mathcal{E}) \tag{3}$$

The GAT block operates over multiple layers, each performing an attention-based message passing step. At each layer l, the node embeddings are updated by attending to their neighbours and the edges that connect them with each of their neighbours based on the network structure:

$$\mathbf{h}_{v_i}^l = \text{GAT}^l(\mathbf{h}_{v_i}^{l-1}, \mathbf{h}_{e_{ij}}, \mathcal{G}) \quad \text{for} \quad l \in 1, 2, \ldots, L_{\text{GAT}} \tag{4}$$

Fig 9. An overview of the GAT block in the Mukara model. Initial node embeddings ($\mathbf{h}_{v_i}^0$) are refined through multiple GAT layers, incorporating edge embeddings ($\mathbf{h}_{e_{ij}}$). The final embeddings are concatenated and passed through an MLP to predict traffic volumes for edges. Solid and dashed lines represent training and test edges, respectively.

The attention mechanism computes unnormalized scores β_{ij} that quantify the importance of node v_j to node v_i, based on their embeddings and the edge between them:

$$\beta_{ij} = \text{LeakyReLU}\left(\mathbf{a}^T \left[\mathbf{W}_v \mathbf{h}_{v_i}^{l-1} \| \mathbf{W}_v \mathbf{h}_{v_j}^{l-1} \| \mathbf{W}_e \mathbf{h}_{e_{ij}}\right]\right) \tag{5}$$

Here, $\mathbf{a}$, $\mathbf{W}_v$, and $\mathbf{W}_e$ are learnable parameters and $\|$ denotes vector concatenation. To integrate edge features explicitly, the edge embeddings $\mathbf{h}_{e_{ij}}$ are included in the attention mechanism. This ensures that the connectivity and attributes of the highway network, which are related to travel cost, are directly utilized, making the GAT block particularly effective for traffic prediction.

Within each layer, the attention coefficients α_{ij} are then obtained by normalizing the scores across the neighbourhood $\mathcal{N}(i)$ of node v_i:

$$\alpha_{ij} = \frac{\exp(\beta_{ij})}{\sum_{m\in\mathcal{N}(i)} \exp(\beta_{im})} \tag{6}$$

Using these attention coefficients, each node updates its embedding by aggregating the transformed embeddings of its neighbours:

$$\mathbf{h}^{l}_{v_i,k} = \sigma\left(\sum_{j\in\mathcal{N}(i)} \alpha_{ij}\mathbf{W}_v\mathbf{h}^{l-1}_{v_j}\right) \tag{7}$$

where σ denotes a non-linear activation function, such as ReLU, and k indexes different attention heads.

Outputs from all K attention heads are concatenated and processed through another MLP to produce the final embedding $\mathbf{h}^{l}_{v_i}$ at each layer:

$$\mathbf{h}^{l}_{v_i} = \mathrm{MLP}^{l}_{K\to 1}\left(\mathrm{Concat}\left(\mathbf{h}^{l}_{v_i,1}, \mathbf{h}^{l}_{v_i,2}, \ldots, \mathbf{h}^{l}_{v_i,K}\right)\right) \tag{8}$$

The GAT block sequentially updates node embeddings by calculating attention scores, normalizing these scores, and aggregating neighbour information across L_{GAT} layers. By the end of the process, the node embeddings encapsulate both local contexts and wider network information.

Finally, traffic volume predictions for each edge are obtained by concatenating the final embeddings of the origin node ($\mathbf{h}^{L}_{v_i}$), destination node ($\mathbf{h}^{L}_{v_j}$), and the edge embedding ($\mathbf{h}_{e_{ij}}$), and feeding the result through another MLP:

$$\hat{y}_{e_{ij}} = \mathrm{MLP}_{h_{ij}\to\hat{y}_{e_{ij}}}(\mathbf{h}^{L}_{v_i}\|\mathbf{h}^{L}_{v_j}\|\mathbf{h}_{e_{ij}}) \tag{9}$$

This structured design enables Mukara to effectively capture spatial, relational, and feature-based dependencies to accurately predict traffic volumes across a highway network.

Model training and experimental settings

Loss function and evaluation metrics

The Mukara model is trained using the mean of the GEH (MGEH) statistic, a metric widely used to evaluate the goodness-of-fit of traffic models [36]. The GEH statistic accounts for both the absolute difference and the percentage difference between the modeled and observed flows, making it particularly suitable for traffic volume prediction tasks. Unlike the commonly used Mean Squared Error (MSE), GEH emphasizes proportionality, allowing errors to be evaluated relative to the magnitude of the observed volumes. A recent study has shown that the GEH loss function is consistent and outperform MAE and MSE in most cases [37].

The GEH statistic for an individual edge e_{ij} is defined as:

$$\mathrm{GEH}_{e_{ij}} = \sqrt{2\cdot\frac{(y_{e_{ij}} - \hat{y}_{e_{ij}})^2}{y_{e_{ij}} + \hat{y}_{e_{ij}}}} \tag{10}$$

where $y_{e_{ij}}$ is the observed traffic volume and $\hat{y}_{e_{ij}}$ is the predicted traffic volume for edge e_{ij}.

For evaluation, in addition to MGEH, the Mean Absolute Error (MAE) is also calculated as a supplementary metric. MAE and MSE are defined as follows:

$$\mathrm{MAE} = \frac{1}{N_e}\sum_{e_{ij}\in\mathcal{E}} \left|y_{e_{ij}} - \hat{y}_{e_{ij}}\right| \tag{11}$$

$$\text{MSE} = \frac{1}{N_e} \sum_{e_{ij} \in \mathcal{E}} \left(y_{e_{ij}} - \hat{y}_{e_{ij}} \right)^2 \tag{12}$$

The MGEH loss function used for training is the mean of the GEH statistics across all edges:

$$\mathcal{L}_{\text{train}} = \text{MGEH} = \frac{1}{N_e} \sum_{e_{ij} \in \mathcal{E}} \text{GEH}_{e_{ij}} \tag{13}$$

Training algorithm

Before the training phase, a spatial train-test split is performed with a ratio of 4:1. Ground truth values for sensors in the test set are used exclusively for evaluation throughout the entire training process. This setup aligns with the study's objective of assessing the feasibility of predicting traffic volumes for unmeasured highway segments.

The training process involves iteratively selecting one year of grid features from the training data, performing a forward pass to make predictions, calculating the loss to measure prediction errors, and conducting a backward pass to compute gradients. The parameters are updated after each batch, and this cycle is repeated for a predefined number of epochs. While using the entire year of training samples as a batch is quite luxurious, it ensures that the model learns from all sensors simultaneously. Experiments showed that this approach achieves lower loss compared to splitting the samples into smaller batches.

Algorithm 1 Training algorithm for the Mukara model

Input: Highway network graph $\mathcal{G}$, edge features $\mathcal{E} = \{e_{ij}\}$, grid features $\{\mathcal{M}_t\}_{t=1}^{N_t}$.
Output: Predicted traffic volumes $\hat{\mathcal{Y}}_t$.

```
Split sensors into training and testing sets.
Initialize model parameters θ.
for epoch = 1 to N_epochs do
    for each year t in training data do
        Extract grid features M_t from M.
        Extract ground truth traffic volumes Y_t for all sensors.
        Perform a forward pass through the Mukara model:
```

$$\hat{\mathcal{Y}}_t = \text{Mukara}(\mathcal{G}, \mathcal{E}, \mathcal{M}_t).$$

```
        Compute the training loss:
```

$$\mathcal{L}_{\text{train}} = \frac{1}{N_e} \sum_{e_{ij} \in \mathcal{E}} \sqrt{2 \cdot \frac{(y_{e_{ij}} - \hat{y}_{e_{ij}})^2}{y_{e_{ij}} + \hat{y}_{e_{ij}}}}.$$

```
        Compute gradients: ∇_θ L_train.
        Update parameters using gradient descent with learning rate α:
```

$$\theta \leftarrow \theta - \alpha \nabla_\theta \mathcal{L}_{\text{train}}.$$

```
    end for
end for
Return: Trained Mukara model and predicted traffic volumes Ŷ_t.
```

Experimental settings

This section states the default settings. For grid features, we use aggregated population, aggregated employment, all land use, road network, and POI features, resulting in a total of 14 channels for the grid tensor.

The default model hyperparameters are as follows: The ROI size is set to 25, corresponding to approximately 25 km, which covers typical spatial extents of small to medium-sized UK cities and aligns with observed urban activity ranges such as commuting distances and economic catchment areas. This choice ensures that the model captures sufficient spatial context without introducing excessive noise from distant, unrelated regions. The CNN block consists of $L_{\text{CNN}} = 3$ layers with channel sizes of 16, 32, and 64 for each layer. The kernel size is set to 3, strides are set to 1, and max pooling is applied with a pool size of 2 and strides of 2, effectively reducing spatial dimensions while preserving relevant feature patterns. The output dense layer of the CNN block, which also serves as the node embedding size in the GAT block, is set to 16 to balance representational capacity and computational efficiency. The GAT block is composed of $L_{\text{GAT}} = 5$ layers, each employing 3 attention heads to capture diverse relational patterns among neighbouring nodes and edges. All MLPs used in the model have a hidden size of 16 with ReLU as the activation function and an output size of 16. Each batch corresponds to the dataset of one year, and therefore 8 batches in one epoch.

Training is performed using the Adam optimizer with a learning rate of 0.001 and gradient clipping at 5 to ensure stability. The experiments were conducted on a system running Windows 11 OS, Python 3.9.18, TensorFlow 2.10.1, and DGL 1.1.2+cu118, utilizing a single NVIDIA RTX 4060Ti GPU.

Results

Ablation study and tuning

In the first experiment, we conducted a grid search to identify better settings for the Mukara model. The tuned hyperparameters included the number of channels in each CNN layer, the ROI size, the depth of the GAT block L_{GAT}, the number of attention heads, and the dimensions of the node embeddings. Each model was trained for a maximum of 50 epochs, and the lowest MGEH and MAE losses were recorded.

The learning curve for the default model is shown in Figure 10. The curve demonstrates that the model learns effectively, with the lowest loss occurring around the 27th epoch. After this point, the model begins to overfit, as indicated by a gradual increase in validation loss.

Fig 10. Learning curve of the default Mukara model, showing MGEH and MAE losses over training epochs.

The results of the hyperparameter tuning are presented in Figure 11. The optimal settings were found to be CNN channels of [16, 32, 64], an ROI size of 21x21 km, 4 attention heads, a GAT depth (L_{GAT}) of 5, and a node embedding size of 16. Based on these findings, the default model was updated to include 4 attention heads while retaining the other hyperparameter settings.

Fig 11. Loss values (MGEH and MAE) in the test set for models with different hyperparameter configurations.

Several observations can be drawn from these experiments. First, the simplest model, which relies solely on edge features for prediction and does not use grid features or node embeddings, results in high loss. This finding emphasizes that geographic and contextual information captured in the node embeddings is essential, as edge embeddings alone are insufficient for accurate predictions. A slightly more complex model that incorporates OD node embeddings in addition to edge features, but excludes GAT layers ($L_{\mathrm{GAT}} = 0$), also yields high loss. This demonstrates that adding OD embeddings for the edge alone is not sufficient for effective predictions. Models with a GAT depth of 5 or 6 achieved the lowest losses, suggesting that incorporating information from nodes up to 5 or 6 degrees away significantly enhances the model's predictive capability. However, increasing the depth beyond this point led to overfitting.

The inclusion of multiple attention heads also improved performance, highlighting the benefit of passing multiple "channels" of information through the network. This effect is analogous to increasing the number of feature maps in CNNs, enhancing the model's ability to capture diverse patterns and relationships.

Finally, the dimensions of the CNN channels and node embeddings were most effective when balanced. Channels and embeddings that were too small resulted in underfitting, as the model failed to capture sufficient information. Conversely, excessively large dimensions led to overfitting, where the model struggled to generalize due to capturing irrelevant or noisy features.

Performance evaluation

We evaluate the Mukara model using the best configuration obtained from the hyperparameter tuning, which sets the number of attention heads to 4. The model is retrained to achieve the lowest MGEH on the validation set. This trained model is then used to predict traffic volumes for all 8 years across the 498 sensors. The results are presented in Figures 12 and 13.

Fig 12. Prediction performance of the Mukara model. (Left) Scatter plot comparing predicted traffic volumes with ground truth values for all sensor-year points, with GEH boundaries for reference. (Upper right) Histogram of mean GEH for each sensor, averaged over 8 years, for training and test sets. (Lower right) Bar plots of MGEH and MAE for sensors grouped by traffic volume quartiles.

Fig 13. Error maps showing signed MGEH values for northbound and southbound traffic. Positive values (red) indicate overestimation, while negative values (blue) indicate underestimation. The maps reveal localized errors, particularly around areas such as Manchester, but no clear geographical trends overall.

Figure 12 illustrates the model's prediction performance. The left panel is a scatter plot comparing predicted traffic volumes with ground truth values for all sensor-year points, with GEH boundaries provided for reference. The upper right panel shows a histogram of the mean GEH for each sensor, averaged over 8 years, for both training and test sets. The lower right panel displays bar plots of MGEH and MAE grouped by traffic volume quartiles. The Mukara model achieves a test MGEH of 50.98 and a test MAE of 9041 vehicles per day. Based on the literature, a GEH below 16 is considered a good match for daily traffic, while values between 16 and 32 are considered acceptable [36, 38]. In the test set, 18% of sensors have a mean GEH below 16, 49% are below 32, and the remaining sensors show higher discrepancies. For MAE, given an average daily traffic volume of 33,734.9 vehicles with a standard deviation of 23,736.4 vehicles, a test MAE of 9041 reflects moderate predictive accuracy. Statistically, the model achieves an R^2 of 0.576 on the test set, meaning 57.6% of the total variance in traffic volume is explained by our model. These results indicate that the Mukara model achieves reasonable performance, particularly given its reliance on external determinants without historical traffic data.

However, significant variations in loss values are observed among sensors grouped by traffic volume quartiles. As shown in Figure 12, both lower and higher traffic volume sensors exhibit larger loss values in both MGEH and MAE metrics. This discrepancy suggests that the model struggles with extreme traffic scenarios, either very low or very high volumes, compared to medium-range volumes.

Figure 13 presents error maps with signed MGEH values, computed as the mean GEH for each sensor over 8 years. Positive values (red) indicate overestimation, while negative values (blue) indicate underestimation. Separate maps are shown for northbound and southbound traffic. The error maps align with the findings from Figure 12, highlighting discrepancies for sensors with extreme traffic volumes. No distinct geographical patterns are observed, although localized higher errors are noticeable around specific areas such as Manchester. This suggests potential regional factors or sensor-specific anomalies affecting model performance in those areas.

Feature importance

In this section, we explore the relative importance of various input features in the Mukara model. First, we analyse how different levels of stratification in population and employment affect the model's performance. As detailed in Section **??**, level 1 stratification includes 7 channels for population (2 for sex and 5 for age) and 21 channels for employment (3 for work type and 18 for sector). Level 2 stratification expands to 10 channels for population and 54 channels for employment.

The results are illustrated in Figure 14. When population is the sole grid feature, increasing the level of stratification does not significantly reduce the loss. However, for employment, the introduction of stratified channels leads to a marked decrease in loss, particularly for level 2 stratification. Furthermore, when both population and employment are included, the model achieves its lowest loss values with higher stratification levels, surpassing the performance of either feature alone. This indicates that stratification allows the model to capture nuanced patterns in the grid features and leverage interactions between demographic and employment strata, such as age, sex, part-time/full-time employment, and sectors.

Fig 14. Test set loss values for different population and employment stratification levels. Increased stratification improves model performance for employment and combined features.

Next, we conduct a feature ablation study to evaluate the importance of each feature set. The full model, which uses all features, serves as the baseline. Six additional models are tested, each omitting one of the following features: population, employment, land use, road network, POI, and edge features. The percentage change in MGEH and MAE loss values is calculated relative to the baseline, revealing the importance of each feature. Figure 15 presents the radar plots summarizing these changes across sensors grouped by overall performance and traffic volume tertiles (low, medium, and high levels).

The results show that the removal of any feature generally increases the loss, highlighting their contribution to the model. Notably, land use emerges as the most critical feature, with its removal leading to the largest loss increase across all tertiles. Interestingly, removing employment results in a slight decrease in loss, suggesting possible redundancy or correlation with other features. For sensors with low and medium traffic volumes, employment, land use, POI, and edge features are particularly

Fig 15. Radar plots showing the percentage change in MGEH (red) and MAE (blue) when individual feature sets are removed. The analysis is presented for overall performance and traffic volume tertiles (low, medium, high). Negative changes indicate a reduction in loss, suggesting possible overfitting or redundancy.

important, whereas high-traffic sensors exhibit less sensitivity to these features. In fact, for high-volume sensors, the loss reduction upon feature removal suggests potential overfitting or misleading patterns in the training data that fail to generalize to the test set.

These findings underscore the importance of carefully selecting and incorporating features in the Mukara model, as well as the need to account for variations in their relevance across different traffic volume levels. The results also highlight the value of stratifying features to improve the model's ability to capture complex interactions in the data.

Discussion

This study proposes an innovative methodological shift in traffic volume prediction by attempting to replace all four steps in the FSM using deep learning, with simpler inputs that can be easily obtained from official statistics and the OSM database. There are important implications of this proposed framework. By using deep learning to replace rule-based calculations, Mukara is capable of learning more complex patterns between traffic determinants and traffic volume in an end-to-end manner. Mukara also fully maintains the transferability, interpretability, and scalability that the FSM offers due to its input structure. It can predict traffic volumes even in regions lacking sensor infrastructure, making it particularly valuable for data-sparse areas such as the Global South and enabling applications across diverse geographic contexts without the need for extensive prior data collection. For urban planning and transportation policy, Mukara can also provide actionable insights for infrastructure development, zoning regulations, and sustainable urban growth. This framework also challenges the conventional reliance on historical sensor readings among existing deep learning-based models, opening new possibilities for understanding the determinants of traffic patterns and thus shedding light on these "black box" models.

Recent real-world cases further highlight the need for such approaches. For example, the opening of the Shenzhen-Zhongshan Link alleviated congestion on the Humen and Nansha Bridges but caused unexpected severe congestion within Shenzhen itself due to a surge of vehicles entering the city road network [39]. Such events demonstrate that traffic forecasting models reliant purely on historical data struggle to predict changes arising from new infrastructure developments. Mukara, by relying on external features rather than historical traffic volumes, is better positioned to handle such rapidly evolving scenarios.

This study provides a successful implementation of the proposed framework using the UK as a case study. The prediction performance is modest given the complexity of the problem and data constraints. Nevertheless, the results outperform existing studies that use the FSM for traffic volume prediction. In the traditional FSM evaluation conducted on the Istanbul case study [11], the FSM exhibited very high error rates, with a best-case %RMSE of approximately 100.92%, implying that the model's average prediction error was about as large as the observed traffic volumes themselves. In contrast, Mukara achieves a test MAE of 9041 vehicles per day against an average daily traffic volume of 33,734.9 vehicles, corresponding to a relative error of roughly 26.8%, and a test R^2 of 0.576, explaining over half of the variance. Furthermore, while the FSM struggled with adjustments to parameters and feedback mechanisms, Mukara successfully leverages external features in an end-to-end deep learning framework to attain moderate GEH scores, with 49% of sensors achieving mean GEH values below 32, considered acceptable. These comparisons highlight the strength of the Mukara framework in providing more accurate and transferable traffic volume predictions compared to traditional FSM approaches. Another recent work has also proposed using GNNs as surrogates for strategic transport planning models, aiming to replicate FSM outputs via a hierarchical classification-regression framework [40]. Their approach first classifies transport demand into usage buckets and then performs fine-grained regression within each bucket, achieving a MAE of more than 20% relative to traffic volumes on synthetic networks. Our study addresses real-world highway traffic volume without synthetic simplifications. Despite the greater noise and heterogeneity inherent in real-world data, Mukara achieves a relative MAE of 26.8%, demonstrating competitive accuracy in a more challenging setting.

Despite these improvements, there are still considerable gaps between the results in this study and those achieved by deep learning models developed for time-series traffic prediction problems. In existing studies, it is common to achieve a mean absolute percentage error (MAPE) below 20% [9]. However, those studies typically focus on settings with high temporal resolution, where strong correlations exist between consecutive time points. Nevertheless, these differences highlight that there remains substantial room for improvement in developing deep learning models capable of accurately predicting traffic volumes based solely on external features.

This study also points to several limitations that should be addressed in future research. The current model demonstrates only moderate prediction ability, which can be explained by several factors. One limitation is the approach of manually selecting a single sensor to represent the traffic volume of an entire highway trunk road segment. This restricts the number of sensors available for training and overlooks spatial variations within the same segment. As a result, predicting traffic for sensors outside the current trunk road traffic network is challenging under the current settings. Moreover, the temporal resolution of the model is relatively low, as it focuses on mean weekday traffic volumes aggregated annually. Hourly traffic predictions were not explored in this study due to the significant noise in the data.

In future studies, several strategies could be employed to improve prediction accuracy. To better understand the bottlenecks in increasing prediction performance, it is necessary to identify which steps of the FSM the deep learning components struggle to replicate. For example, DeepGravity has demonstrated accurate prediction of O-D distributions using similar traffic determinant inputs [27], suggesting that the challenges may lie more in modelling modal split and trip assignment steps. A consistent and systematic evaluation using synthetic data is desirable to test the capability of deep learning models in replacing each FSM step individually, or combinations of two or more steps. Furthermore, the architecture of the deep learning model should be redesigned to better suit the problem. While CNNs are capable of aggregating spatial features, they are most adept at refining visual features; thus, more complex models such as Vision Transformers might better suit this application. GNNs could also be replaced by tailored architectures, as GNNs are generally better suited for information propagation, whereas trip assignment behaviour requires embeddings that interact in more interpretable ways. Other potential improvements include dynamically generating spatial input data tailored to each sensor to enhance spatial granularity. Incorporating temporal variations, such as weekly and daily periodicity, and dynamic features such as weather, holidays, events, and road construction, could significantly improve temporal resolution. Additionally, integrating static features related to public transportation —

such as availability, schedules, and costs of trains and air planes — could provide further contextual insights. These enhancements, along with refinements to the Mukara model architecture, could expand its applicability and improve its predictive accuracy.

Conclusion

This study proposes Mukara, an innovative deep learning framework that seeks to predict traffic volumes across the entire highway trunk road network in England based solely on external features. By leveraging data of external traffic determinants and deep learning architecture, Mukara aims to address key limitations of traditional FSM-based approaches and existing deep learning models. Our results demonstrate that Mukara achieves moderate predictive accuracy, outperforming traditional FSM models evaluated under similar conditions, while maintaining full transferability and interpretability.

This research highlights the potential for deep learning models to extend beyond time-series forecasting and offer flexible, scalable solutions to transportation planning challenges, particularly in data-sparse contexts. However, the current implementation also reveals several challenges, such as limitations in spatial and temporal resolution, and the need for better model architectures more specifically suited to replicating complex travel behaviour processes.

Future research should focus on systematically validating the feasibility of replacing individual FSM steps using synthetic data and refining model designs to better capture human mobility dynamics. Architectural innovations, such as adopting more specialized spatial components and graph models, as well as enriching spatial and temporal feature sets, represent promising directions. Overall, this study provides a foundation for further exploration into replacing traditional travel demand modelling steps with end-to-end deep learning frameworks, aiming toward more adaptive and equitable transportation systems.

Acknowledgements

The author would like to sincerely thank Dr. Damon Wischik for his valuable discussions, insightful analysis of the challenges encountered, and guidance on future directions. The author is also grateful to Dr. Ramit Debnath for his thoughtful input throughout the research process. Finally, the author acknowledges the Department of Architecture for providing the research platform and support that made this work possible.

References

1. Wang J, Li H, Wang Y, Yang H. A novel assessment and forecasting system for traffic accident economic loss caused by air pollution. Environmental Science and Pollution Research. 2021;28:49042 – 49062. doi:10.1007/s11356-021-13595-1.

2. Haryanto EG, Latifa EA. Risk Analysis of Toll Road Accident Using Fmea and Ahp Methods. 2020;1:22–28. doi:10.32722/ARCEE.V1I02.2676.

3. Geldiyev H, Nepesov R, Amanov A, Geldiyev J, Churiyev M. Mathematic modeling for dynamics of transport flows. 2020; p. 261–274. doi:10.1007/978-3-658-27110-7$_1$8.

4. Afrin T, Yodo N. A Survey of Road Traffic Congestion Measures towards a Sustainable and Resilient Transportation System. Sustainability. 2020;doi:10.3390/su12114660.

5. UK Department for Transport. Transport Statistics Great Britain 2022: Domestic Travel; 2023. Available from: https://www.gov.uk/government/statistics/transport-statistics-great-britain-2023/transport-statistics-great-britain-2022-domestic-travel.

6. Tikoudis I, Papu Carrone A, Mba Mebiame R, Lamhauge N, Hassett K, Bystrom O. Household transport choices: New empirical evidence and policy implications for sustainable behaviour. Paris: OECD Publishing; 2024. 246. Available from: https://doi.org/10.1787/0e8469ed-en.

7. Dargay J, Gately D, Sommer M. Vehicle Ownership and Income Growth, Worldwide: 1960-2030. The Energy Journal. 2007;28(4):143–170. doi:10.5547/ISSN0195-6574-EJ-Vol28-No4-7.

8. Boukerche A, Tao Y, Sun P. Artificial intelligence-based vehicular traffic flow prediction methods for supporting intelligent transportation systems. Comput Networks. 2020;182:107484. doi:10.1016/j.comnet.2020.107484.

9. Medina-Salgado B, Sánchez-DelaCruz E, Pozos-Parra P, Sierra JE. Urban traffic flow prediction techniques: A review. Sustainable Computing: Informatics and Systems. 2022;35:100739. doi:https://doi.org/10.1016/j.suscom.2022.100739.

10. Ortúzar JdD, Willumsen LG. Modelling Transport. 5th ed. Chichester, UK: Wiley; 2011. Available from: https://www.wiley.com/en-us/Modelling+Transport%2C+5th+Edition-p-9781119242813.

11. Gachanja JN. Towards Integrated Land Use and Transport Modelling: Evaluating Accuracy of the Four Step Transport Model- the Case of Istanbul, Turkey; 2010. Available from: http://essay.utwente.nl/90760/.

12. Jiang W, Luo J. Graph neural network for traffic forecasting: A survey. Expert Systems with Applications. 2022;207:117921. doi:10.1016/j.eswa.2022.117921.

13. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 1998;86(11):2278–2324. doi:10.1109/5.726791.

14. Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Computation. 1997;9(8):1735–1780. doi:10.1162/neco.1997.9.8.1735.

15. Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks. 2005;18(5):602–610. doi:https://doi.org/10.1016/j.neunet.2005.06.042.

16. Cho K, van Merrienboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al.. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation; 2014.

17. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is All you Need. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. Advances in Neural Information Processing Systems. vol. 30. Curran Associates, Inc.; 2017.Available from: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

18. Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G. The Graph Neural Network Model. IEEE Transactions on Neural Networks. 2009;20(1):61–80. doi:10.1109/TNN.2008.2005605.

19. Zhang J, Zheng Y, Qi D. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction. Proceedings of the AAAI Conference on Artificial Intelligence. 2017;31(1). doi:10.1609/aaai.v31i1.10735.

20. Li Y, Yu R, Shahabi C, Liu Y. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting; 2018.

21. Yao H, Wu F, Ke J, Tang X, Jia Y, Lu S, et al. Deep multi-view spatial-temporal network for taxi demand prediction. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence. AAAI'18/IAAI'18/EAAI'18. AAAI Press; 2018.

22. Wang C, Zuo K, Zhang S, Lei H, Hu P, Shen Z, et al. PFNet: Large-Scale Traffic Forecasting With Progressive Spatio-Temporal Fusion. IEEE Transactions on Intelligent Transportation Systems. 2023;24(12):14580–14597. doi:10.1109/TITS.2023.3296697.

23. Kong X, Xing W, Wei X, Bao P, Zhang J, Lu W. STGAT: Spatial-Temporal Graph Attention Networks for Traffic Flow Forecasting. IEEE Access. 2020;8:134363–134372. doi:10.1109/ACCESS.2020.3011186.

24. Pan Z, Liang Y, Wang W, Yu Y, Zheng Y, Zhang J. Urban Traffic Prediction from Spatio-Temporal Data Using Deep Meta Learning. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019;.

25. Pan Z, Zhang W, Liang Y, Zhang W, Yu Y, Zhang J, et al. Spatio-Temporal Meta Learning for Urban Traffic Prediction. IEEE Transactions on Knowledge and Data Engineering. 2022;34(3):1462–1476. doi:10.1109/TKDE.2020.2995855.

26. Wang S, Lv Y, Peng Y, Piao X, Zhang Y. Metro Traffic Flow Prediction via Knowledge Graph and Spatiotemporal Graph Neural Network. Journal of Advanced Transportation. 2022;2022(1):2348375. doi:https://doi.org/10.1155/2022/2348375.

27. Simini F, Barlacchi G, Luca M, Pappalardo L. A Deep Gravity Model for Mobility Flows Generation. Nature Communications. 2021;12(1):6576. doi:10.1038/s41467-021-26752-4.

28. National Highways. National Highways Strategic Road Network; 2023. `https://nationalhighways.co.uk/our-roads/roads-we-manage/`.

29. National Highways. Traffic Information System (TRIS); 2024. `https://webtris.highwaysengland.co.uk/api/swagger/ui/index`.

30. Office for National Statistics. Population Estimates - Small Area (2021 Based); 2011-2022. `https://www.nomisweb.co.uk/sources/pest`.

31. Office for National Statistics. Business Register and Employment Survey (BRES); 2015-2022. `https://www.nomisweb.co.uk/sources/bres`.

32. Office for National Statistics. Lower Layer Super Output Areas (December 2011) Boundaries EW BFC (V3); 2011. Accessed via REST URL of ArcGIS for INSPIRE Feature Download Service: https://dservices1.arcgis.com/ESMARspQHYMw9BZ9/arcgis/services/Lower_layer_Super_Output_Areas_Dec_2011_Boundaries_Full_Clipped_BFC_EW_V3/WFSServer?service=wfs&request=getcapabilities.

33. Office for National Statistics. Lower Layer Super Output Areas (December 2021) Boundaries EW BFC (V10); 2024. Accessed via REST URL of Feature Access Service: https://services1.arcgis.com/ESMARspQHYMw9BZ9/arcgis/rest/services/Lower_layer_Super_Output_Areas_December_2021_Boundaries_EW_BFC_V10/FeatureServer.

34. Geofabrik GmbH and OpenStreetMap Contributors. OpenStreetMap Data Extracts: United Kindom; 2023. https://download.geofabrik.de/europe/united-kingdom.html#.

35. National Highways. National Highways: Managing England's Strategic Road Network; 2024. https://nationalhighways.co.uk/.

36. Feldman O. The GEH Measure and Quality of the Highway Assignment Models. London, UK: Association for European Transport and Contributors; 2012.Available from: https://aetransport.org/public/downloads/V7AGa/5664-5218a2370407f.pdf.

37. Esugo M, Haas O, Lu Q. Hybrid Deep-Learning Approach with Geoffrey E. Havers-Based Loss Function and Evaluation Metric for Multilocation Traffic-Flow Forecasting. Transportation Research Record. 0;0(0):03611981241274645. doi:10.1177/03611981241274645.

38. Friedrich M, Pestel E, Schiller C, Simon R. Scalable GEH: A Quality Measure for Comparing Observed and Modeled Single Values in a Travel Demand Model Validation. Transportation Research Record. 2019;2673(4):722–732. doi:10.1177/0361198119838849.

39. Post SCM. Long Traffic Jams Greet Hong Kong Tourists Trying Out New Shenzhen-Zhongshan Link. South China Morning Post. 2024;.

40. Makarov N, Narayanan S, Antoniou C. Development of a graph neural network surrogate for travel demand modelling; 2025. Available from: https://arxiv.org/abs/2408.07726.