Imperial College London

Post-hoc Explanations for AI in Transport Planning

Date: 14 Dec 2023

Xuehao Zhai

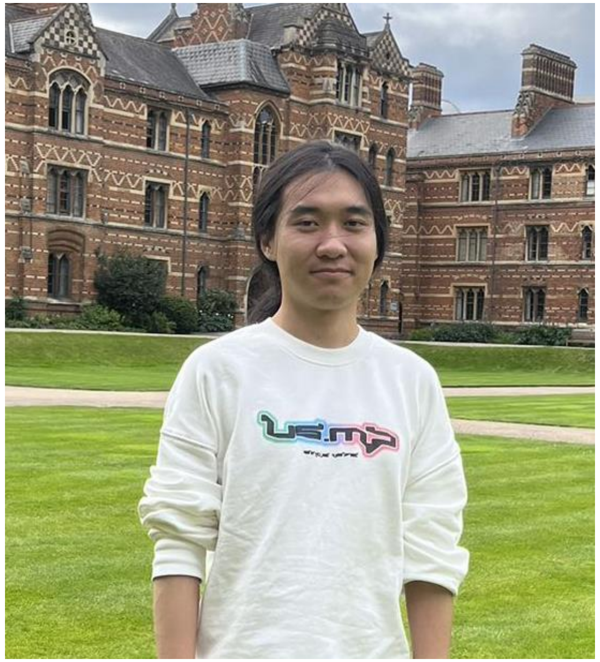
x.zhai20@imperial.ac.uk

Profile

I am a final year PhD student in Centre for Transport Studies in department of civil and environmental engineering at Imperial College London.

My research interests span the interdisciplinary study of XAI within urban planning and transport policy. Specifically, I focus on applying and explaining the graph deep learning in land-use modelling and choice behaviour analysis.

I am currently seeking postdoctoral opportunities in XAI, as well as career opportunities in the industry related to the XAI field.



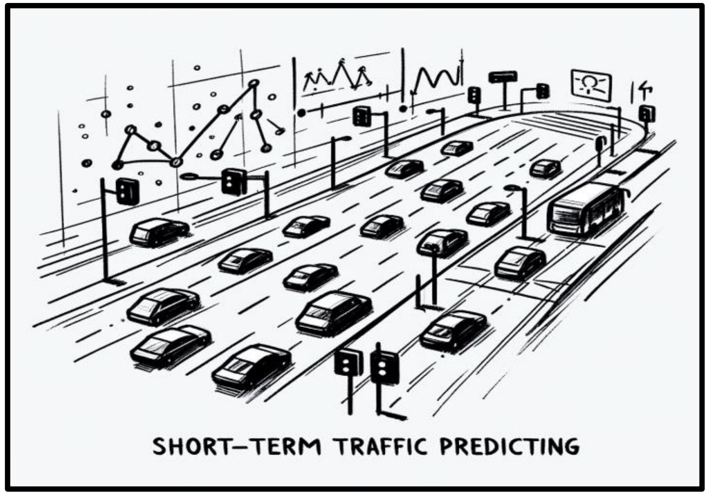
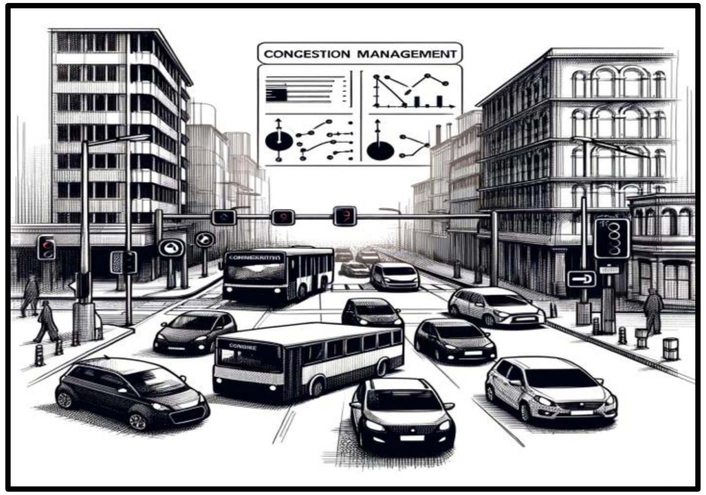
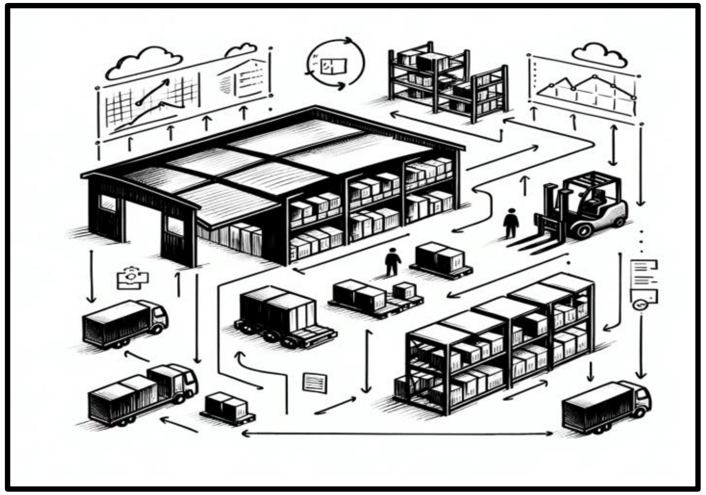
Xuehao Zhai
Email:
x.zhai20@imperial.ac.uk

Website:
https://www.imperial.ac.uk/people/x.zhai20

Outlines

- AI in transport
- Why XAI in transport?
- Research challenges and objectives
- Case 1: Land-use identification
- Case 2: Traffic anomaly detection

AI in Transport: Data

Prediction	Pattern Recognition	Optimisation
		

Active data

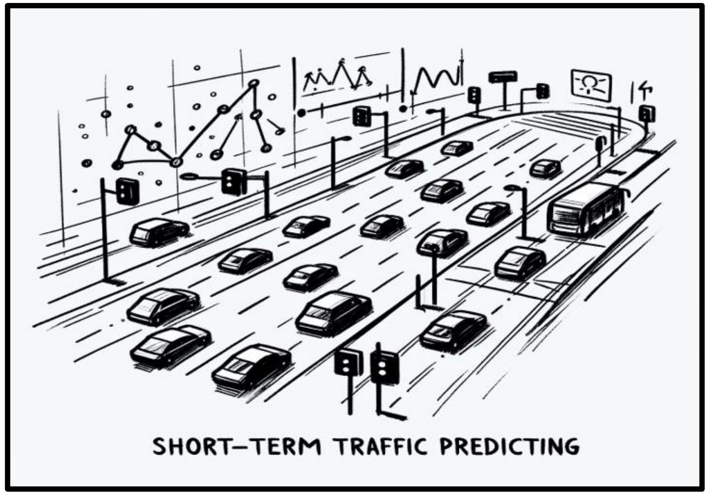
- Surveys data
  (route preferences, driving habits, …)
- Feedback and reviews
  (travel diaries or logs, interactive app inputs, …)

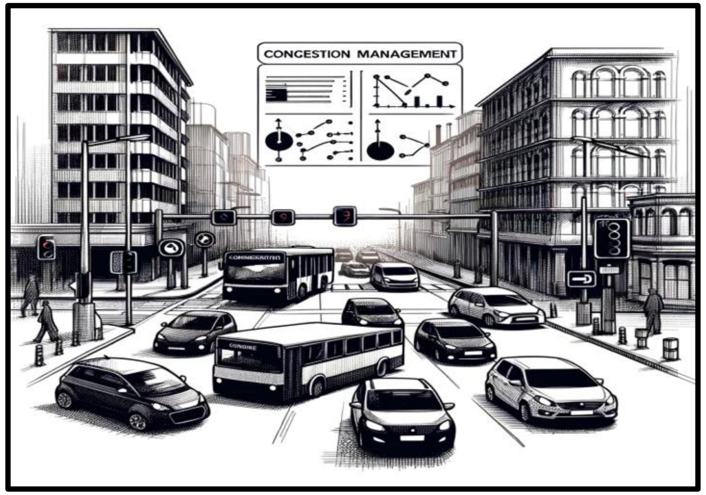
Passive data

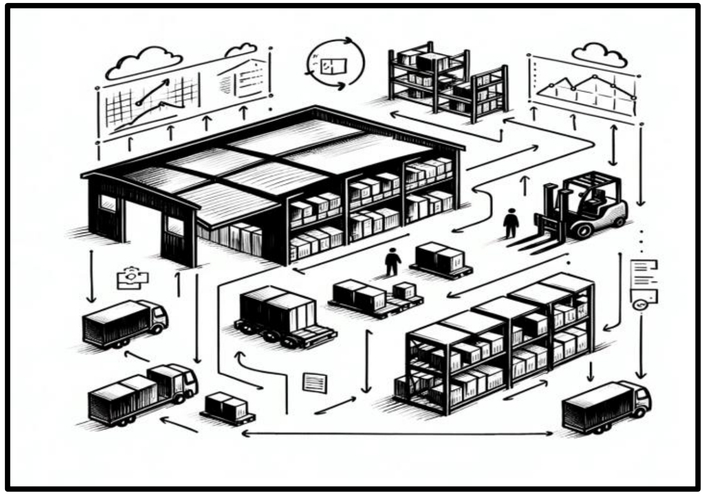
- Time series
  (traffic flow, demand, accidents, …)
- Spatial data
  (facilities location, land use,…)
- Image data
  (monitoring camera, drone, satellite, …)

AI in Transport: Models

Prediction	Pattern Recognition	Optimisation







- Basic MLs
  (SVMs, Random Forests, …)
- Temporal models
  (LSTM, Transformer, ...)
- Spatial models
  (CNNs, GNNs, …)
- CV models
  (CNNs, Diffusion, …)
- Hybrid models
  (ST-GCNs, …)
- Reinforcement learning

Why XAI in Transport?

Transportation is a high-stake domain; the successful application largely rely on trust from different stakeholders.

Why XAI in Transport?

Transportation is a high-stake domain; the successful application largely rely on trust from different stakeholders.

Who?	**Operators**	**Owners**	**Policy-makers**
Where?	Transport companies	Government	Planning departments
Why?	Reliability	Transparency Accountability	Causality
Description	• Transport services involve the movement of people and goods, making safety prior.	• It is essential to trace back the reasons behind when mistakes happen.	• How inputs X impact the outputs Y? • If modifications are made on X, what outcome could follow?

Why XAI in Transport?

Transportation is a high-stake domain; the successful application rely on users trust.

How to build Trust?

- Reliability: reliable performance, especially for critical moments.
- Accountability: tractable errors.
- Transparency: understand the working-process of a model.
- Causality: understand what will happen when change been made on inputs.
- …

Why XAI in Transport?

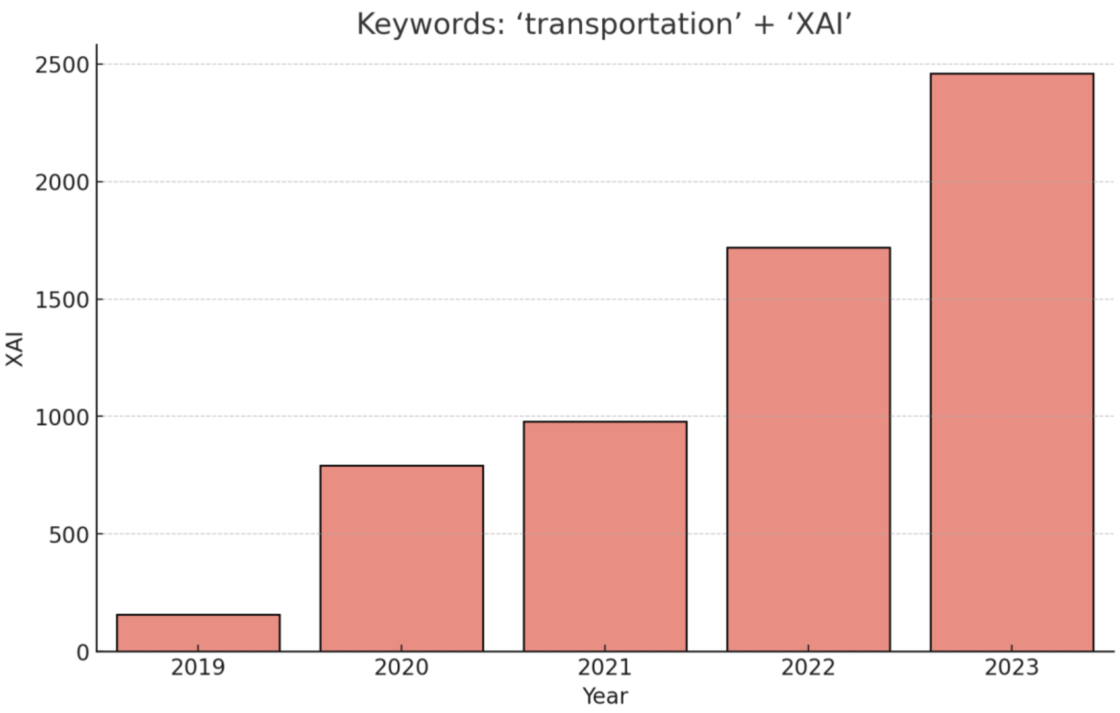
Transportation is a high-stake domain; the successful application rely on users trust.

How to build Trust?

- Reliability: reliable performance, especially for critical moments.
- Accountability: tractable errors.
- Transparency: understand the working-process of a model.
- Causality: understand what will happen when change been made on inputs.
- …

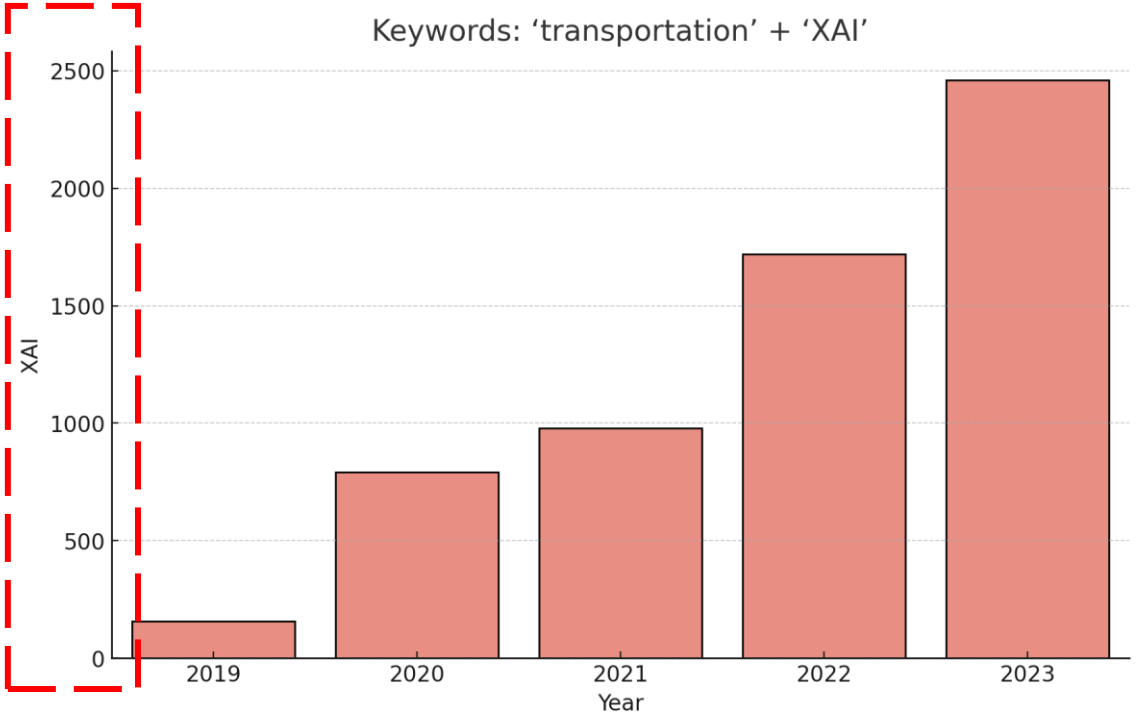
Explainability!!

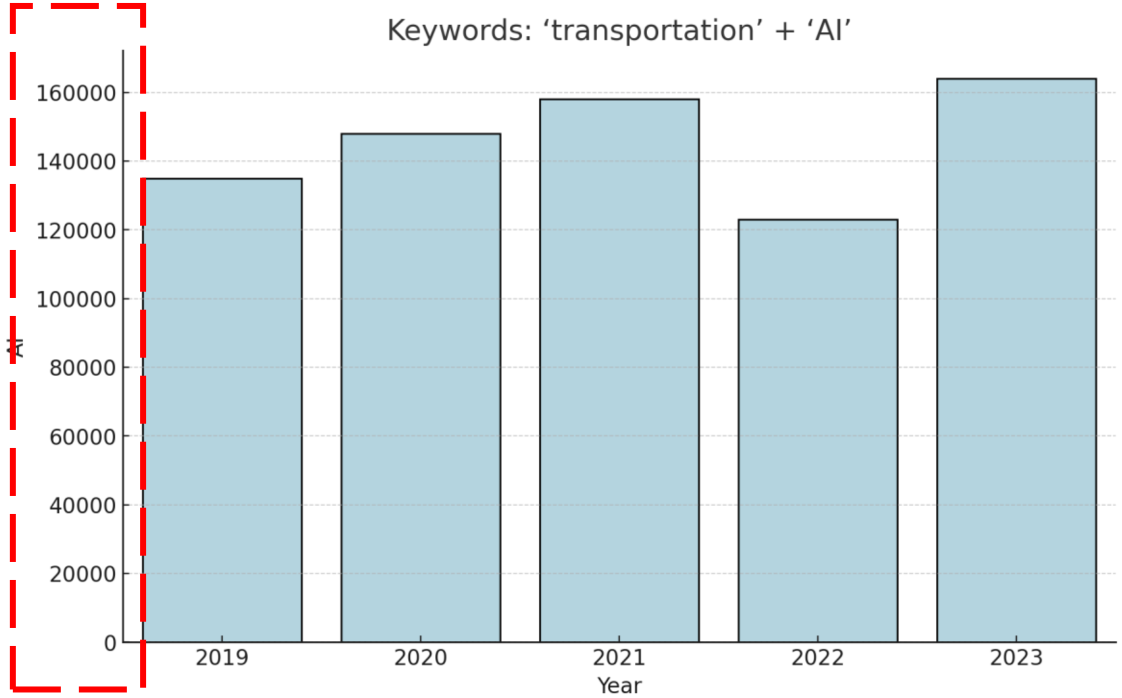
Why XAI in Transport?



Notes: statistic from google scholar

Why XAI in Transport?





Notes: statistic from google scholar

Challenge 1: Lack of focus on complex models

Highly cited papers in AI transport
Model (number of citations):

Temporal: Transformer (240), LSTM (1363), …

Spatial: GraphSage (157), …

Hybrid: STGCN(3010), T-GCN (1673), ASTGCN (1597)

Highly cited papers using XAI methods in AI transport
Model (number of citations):

Post-hoc: XGBoost + SHAP (496), MLs + LIME (34), …

Ante-hoc: GNN + Attention (888), …

Challenge 1: Lack of focus on complex models

Highly cited papers in AI transport
Model (number of citations):

Temporal: Transformer (240), LSTM (1363), …

Spatial: GraphSage (157), …

Hybrid: STGCN(3010), T-GCN (1673), ASTGCN (1597)

Highly cited papers using XAI methods in AI transport
Model (number of citations):

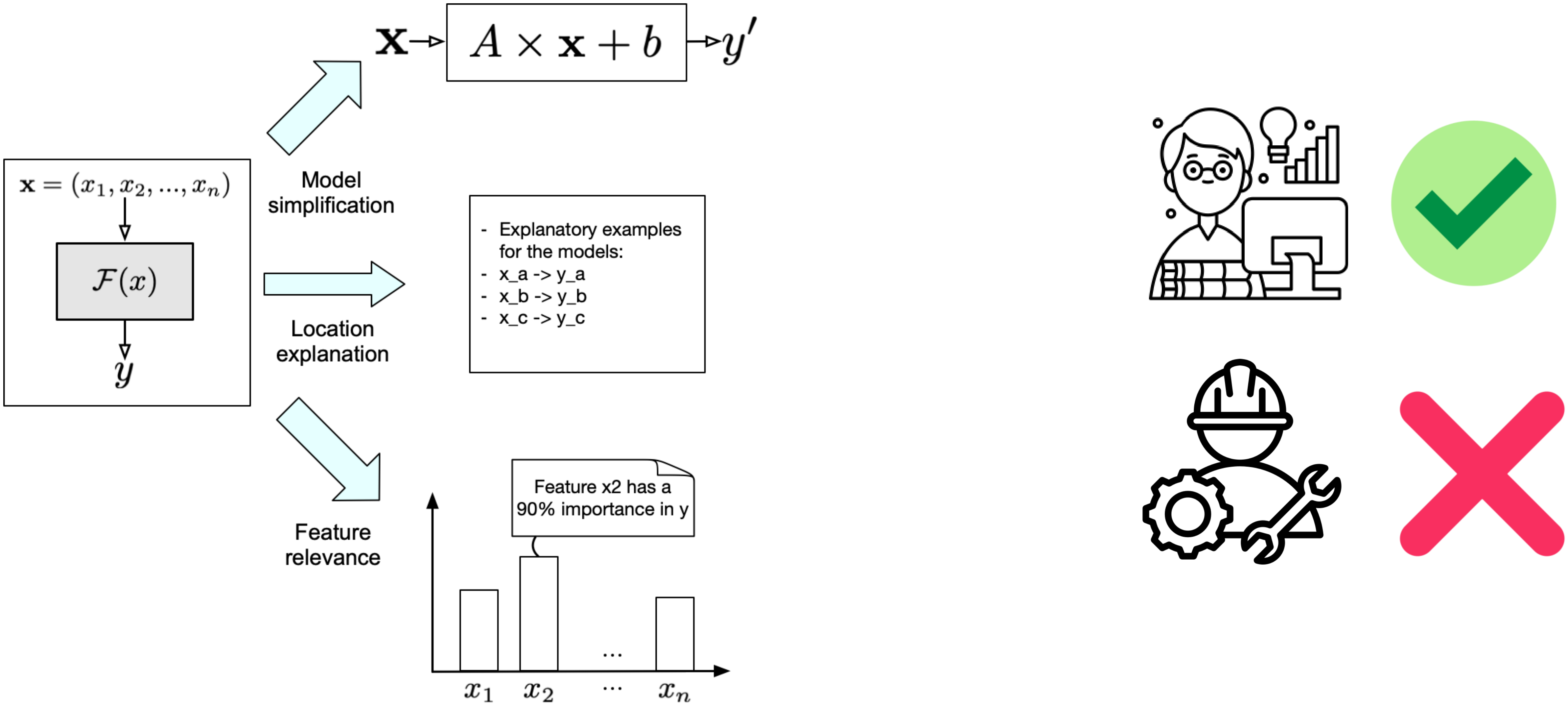
Post-hoc: XGBoost + SHAP (496), MLs + LIME (34), …

Ante-hoc: GNN + Attention (888), …

- Lack of a standardised, quantifiable measure for comparing among different models and scenarios.
- Few XAI applications focus on complex deep learning models.

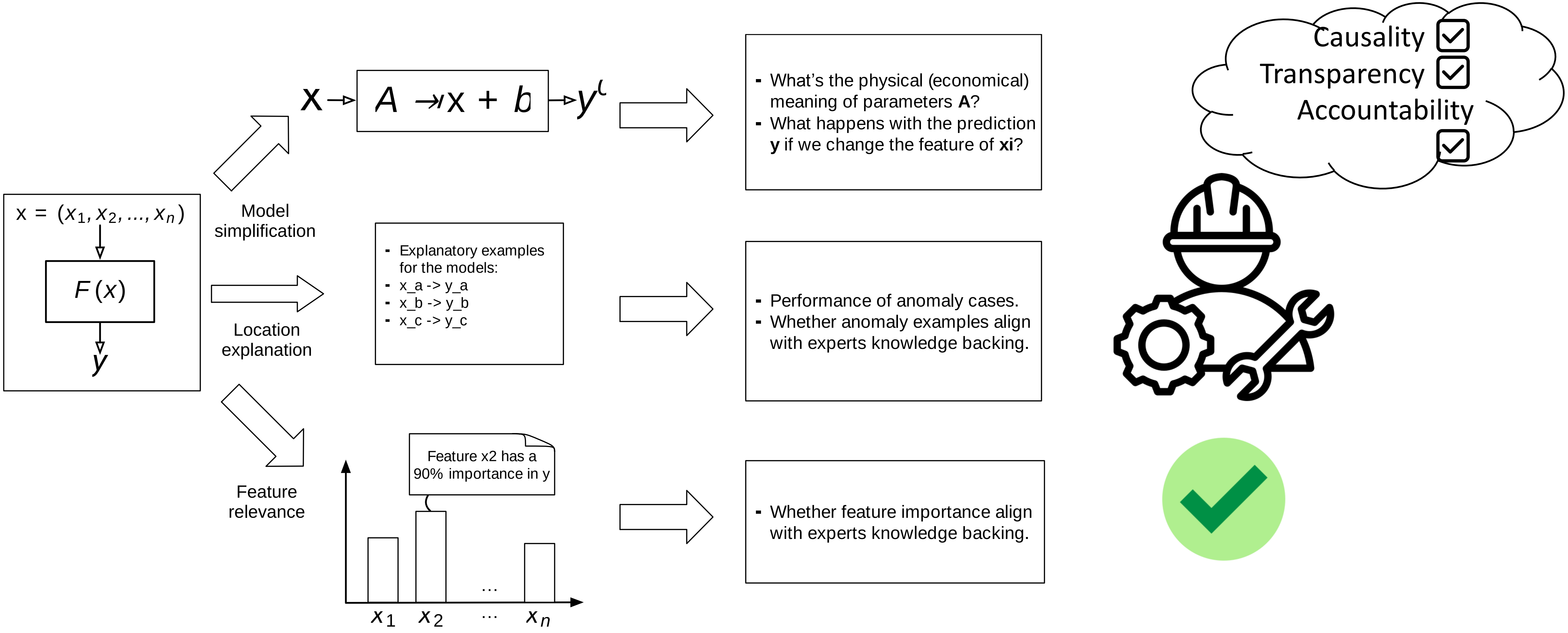
Challenge 2: Tailor XAI for stakeholders in transport

Differences in 'XAI' Perception: Computer Scientists vs. Transport Experts



Challenge 2: Tailor XAI for stakeholders in transport

Differences in 'XAI' Perception: Computer Scientists vs. Transport Experts



Research Objectives

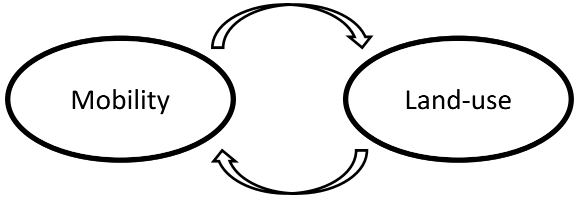
Objective A
A **Quantifiable and Model-agnostic** Framework for Deep Learning Models

Objective B
Post-hoc Explanations: Methodologies and Applications

Objective C
Ante-hoc explainable modelling: Methodologies and Applications

A1 Building of a Quantifiable Definition of explainability

A2 Proofing of land-use can be a reference for measuring transport-related model explainability



B1 Attribution-based explanation for the **causality** of models.

B2 Hidden-based explanation for **reliability and transparency**

B3 Example-based explanation for the model's **accountability**

C1 Self-learning

C2 Graph-deep-learning

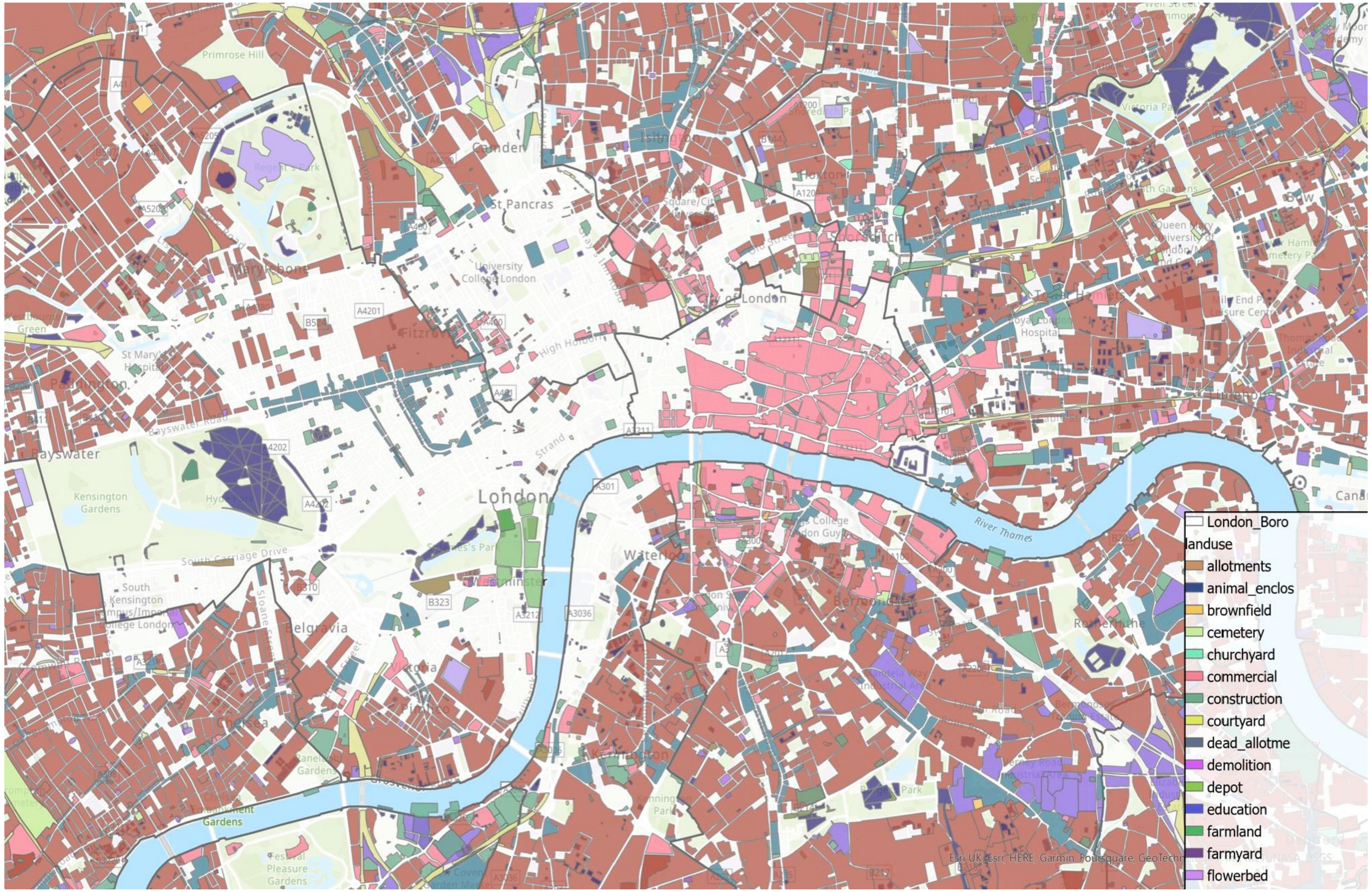
C3 Attention

Case One : Land-use Identification Using Transport Data and XAI Methods

Zhai, X.*, Guo, F., and Sivakumar, A., (2024). Identification of urban land-use from mobility data based on Graph Neural Networks and post-hoc explanations: A case in London. *Computers, Environment and Urban Systems*, Under review.

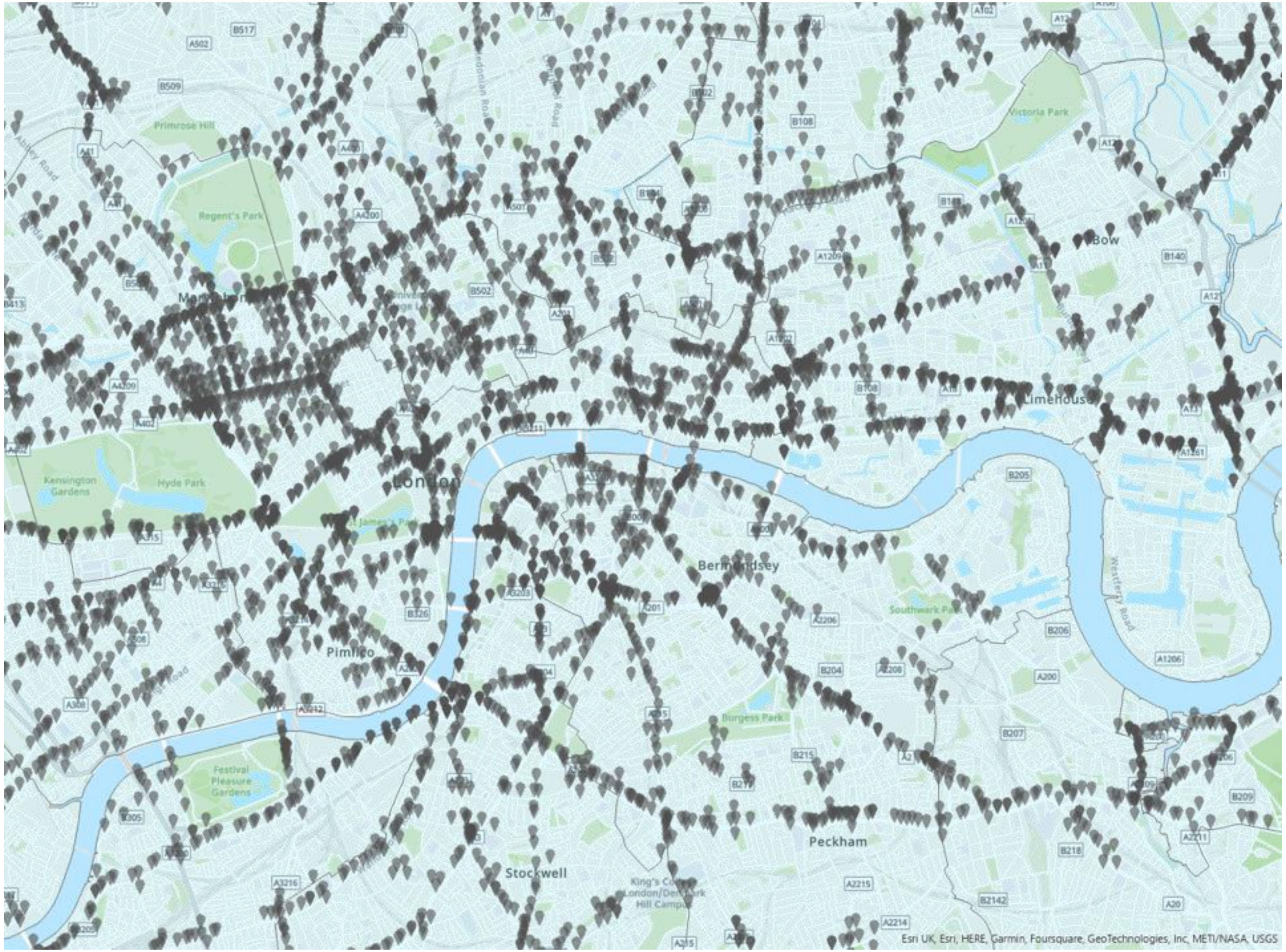
What is Urban Land Use?

Land use is used to describe the function of human-made spaces in which people live, work, and recreate on a day-to-day basis.



'Land use' in the central London collected from Open Street Map

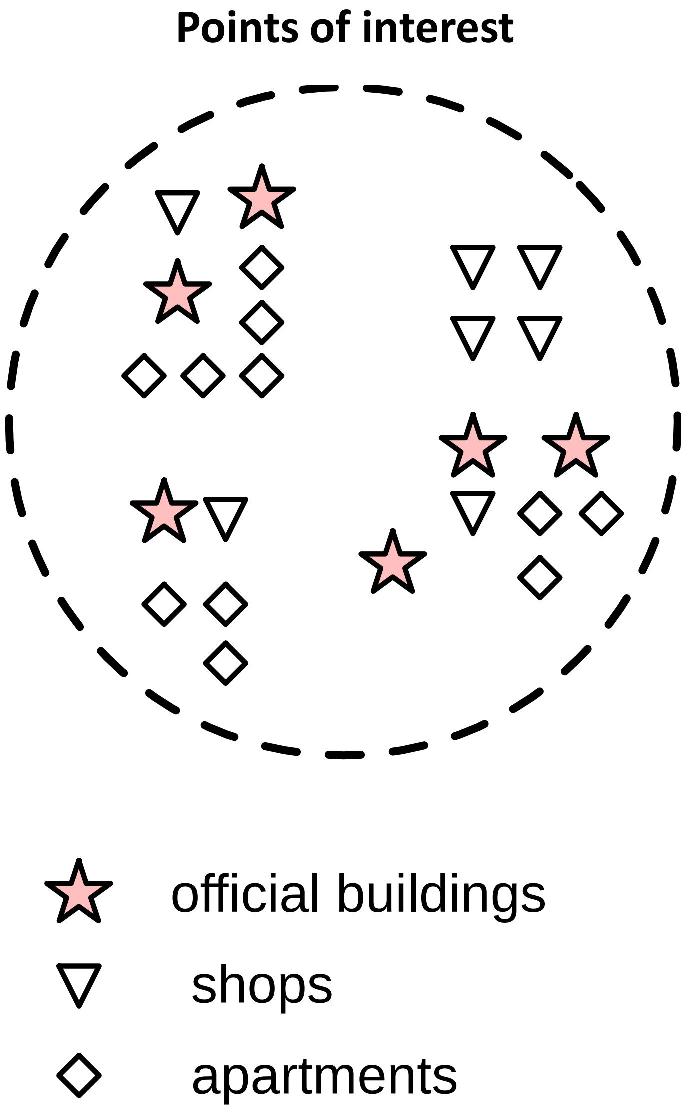
Inputs: Ridership of Transport Service

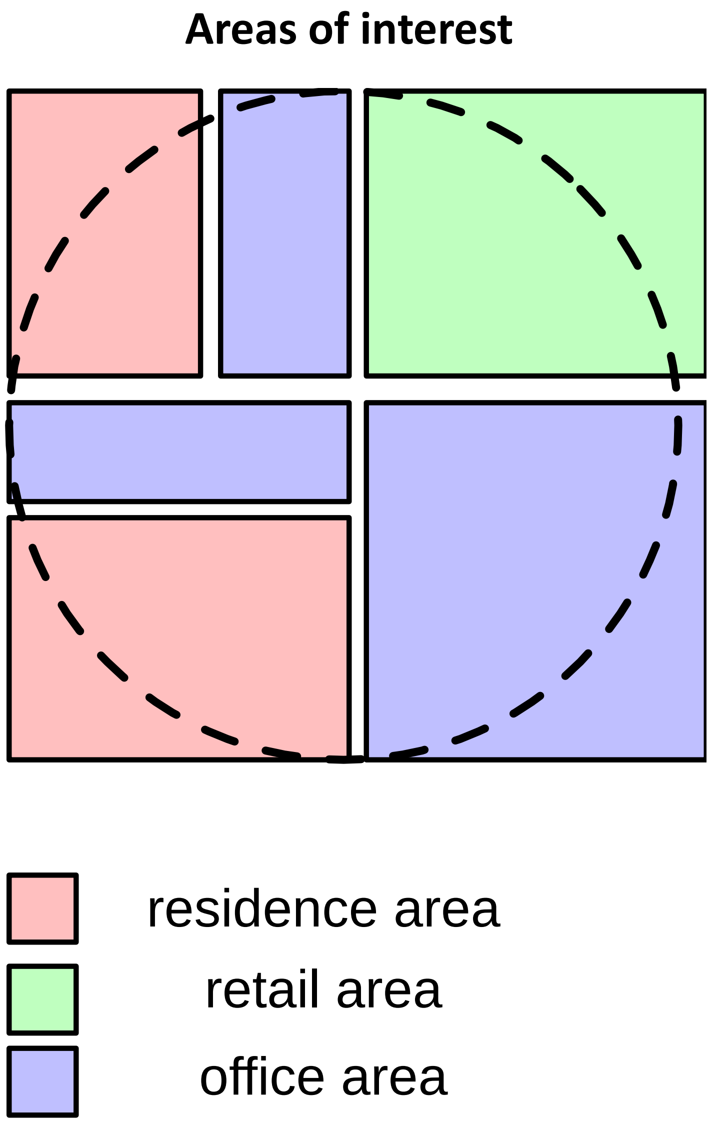


Around 15000 bus stations and railway stations in the Great London area

Ridership of 3 months (Sep - Nov) in 2017

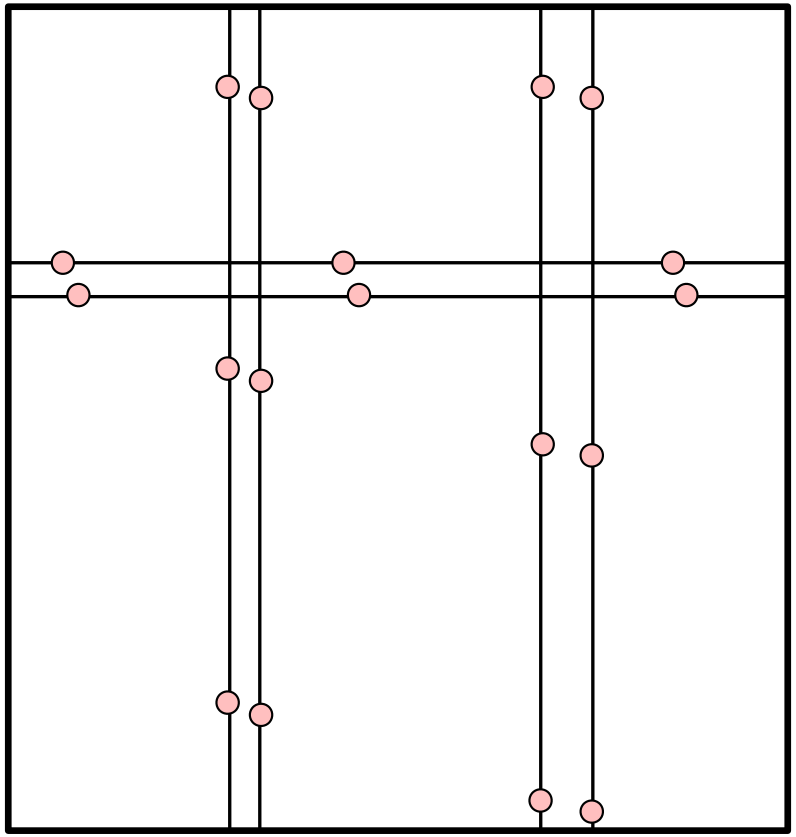
Labels: How is Urban Land Use Quantified?



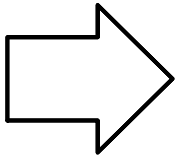


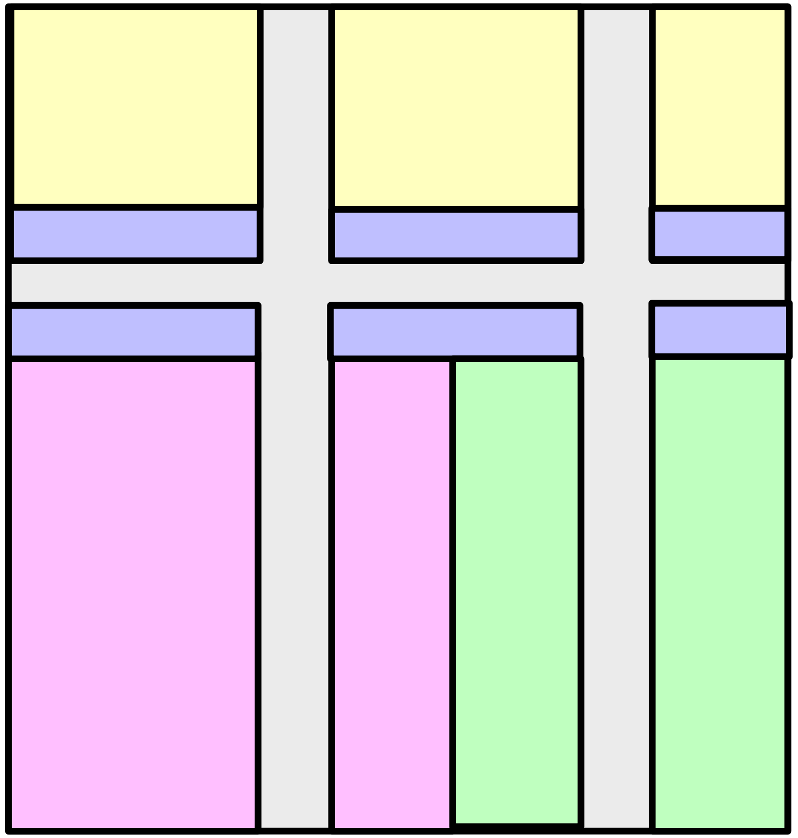
Land Use Identification at Nodes Level

The aim of Land use Identification is used to predict (regress) the land use intensity and types of a region given the transport data in the surround area.



Input data (graph of historical traffic volume)

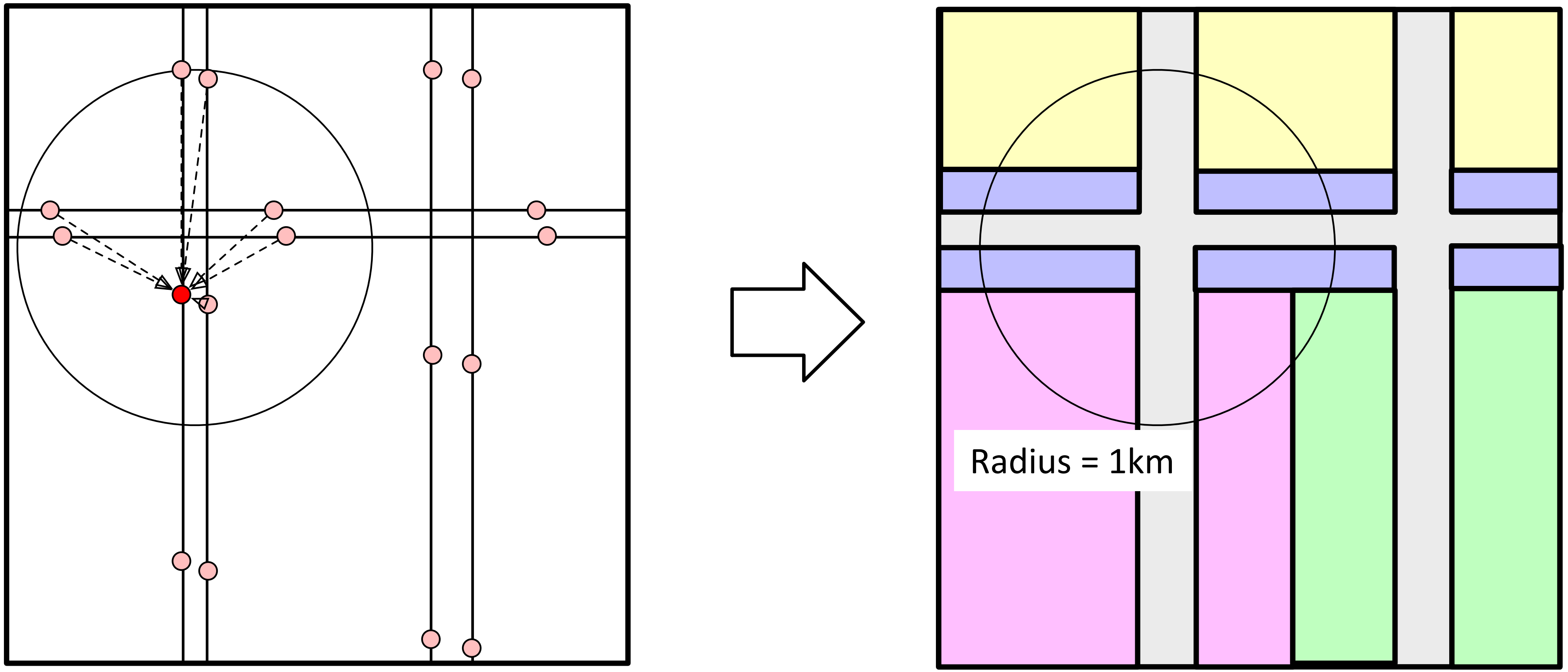




Labels: types and intensity of different land use.

Land Use Identification at Nodes Level

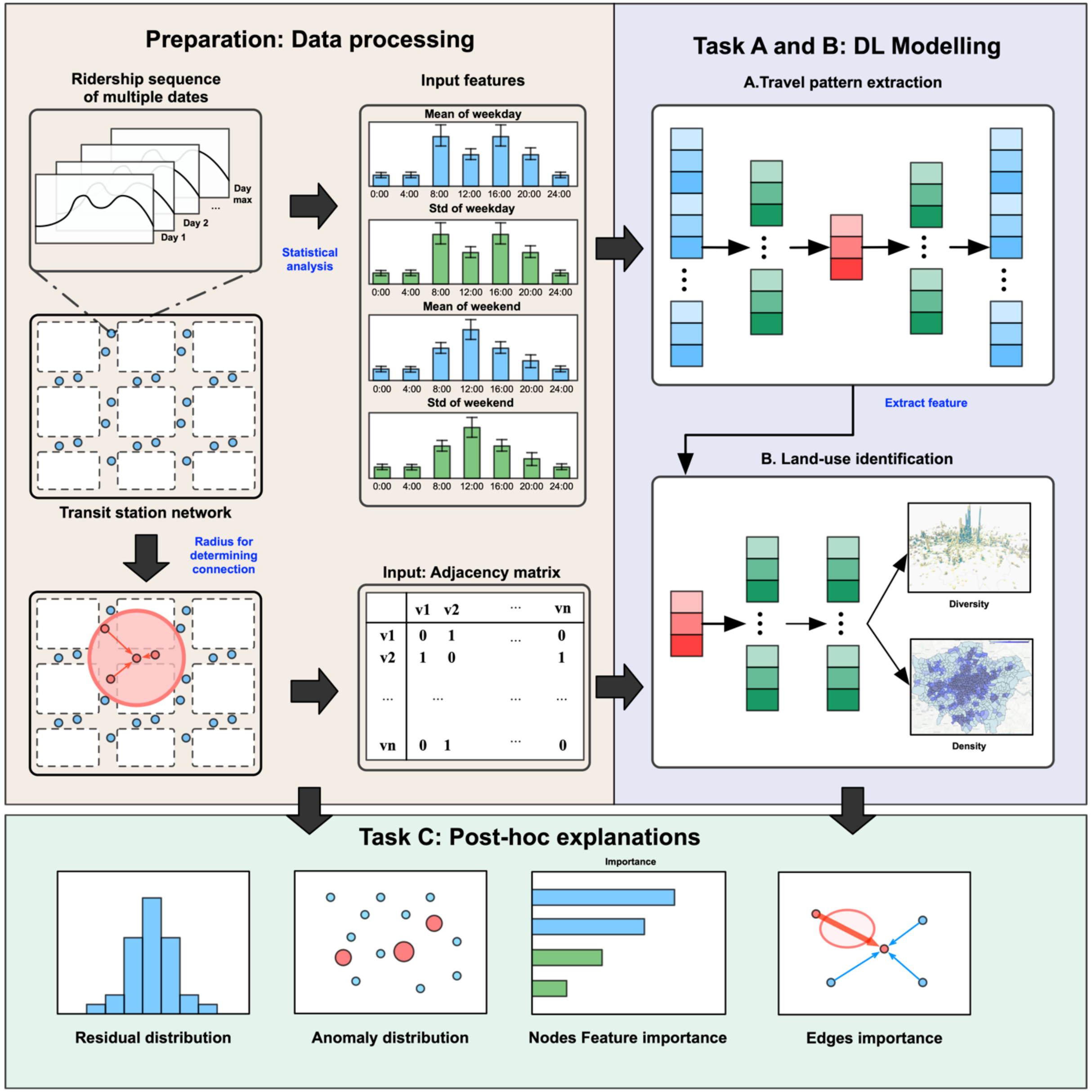
The aim of Land use Identification is used to predict (regress) the land use intensity and types of a region given the transport data in the surround area.



Input data (a graph of ridership data)

Labels: land use intensity of different types

The flow chart of modelling

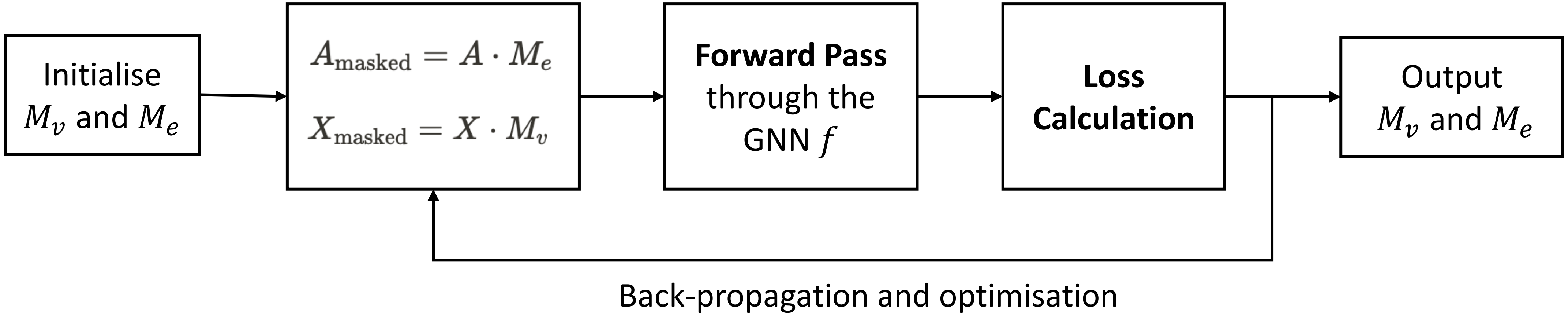


Calculating the ‘ importance’ for Graph models

- **GNN-explainers**

Given a well-trained GNN, the goal of the GNN Explainer is to learn a set of **masks** for nodes M_v and edges M_e that highlight the most important structures in the GNN prediction.

$$\min_{M_v, M_e} \underbrace{\mathcal{L}(f(X,A),y)}_{\text{prediction loss}} + \underbrace{\lambda_1 \|M_v\|_1 + \lambda_2 \|M_e\|_1}_{\text{Regularisation L1}} + \underbrace{\lambda_3 TV(M_v)}_{\text{total variation of the node masks}}$$



Ying, Z., Bourgeois, D., You, J., Zitnik, M. and Leskovec, J., 2019. Gnnexplainer: Generating explanations for graph neural networks. *Advances in neural information processing systems*, *32*.

Objectives of post-hoc explanations

Build post-hoc explanation to achieve:

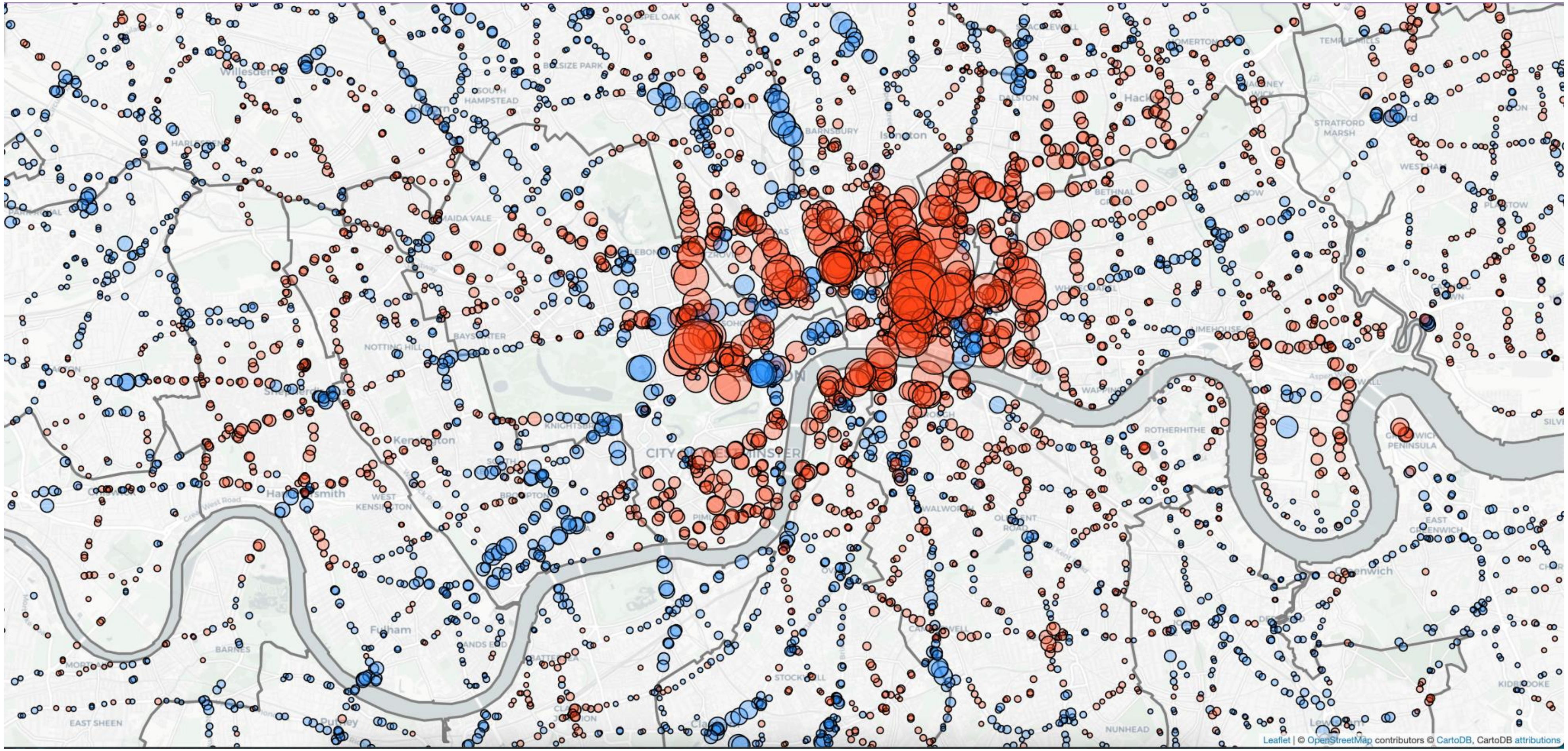
- **C1 Reliability:** trace the anomalies and the reason behind the anomalies
- **C2 Transparency:** explore the working process in an intuitive way.
- **C3 Causality:** importance (contribution to the outcome) of elements in the Graph models

C1 Anomalies distributions

- **Residual spatial distribution**

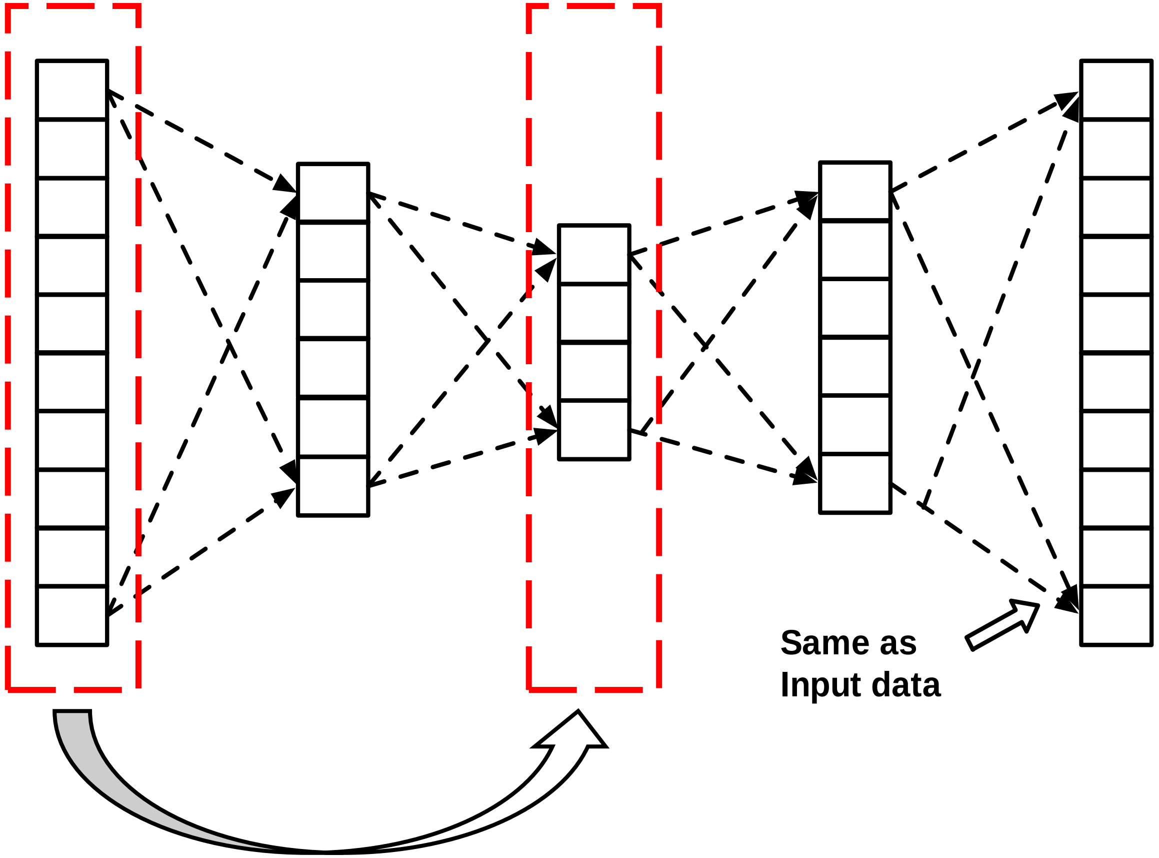
Actual > prediction

Actual < prediction

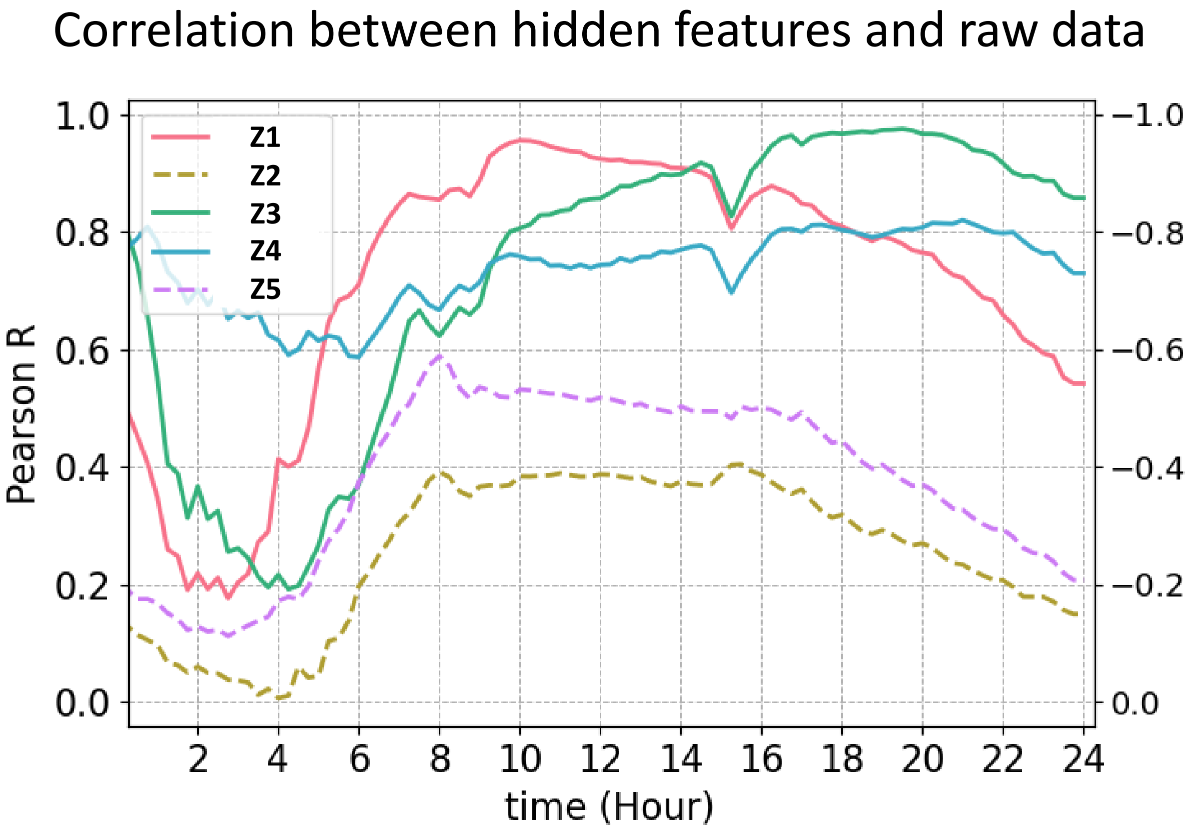


Office

C2 Intuitive function of hidden layer

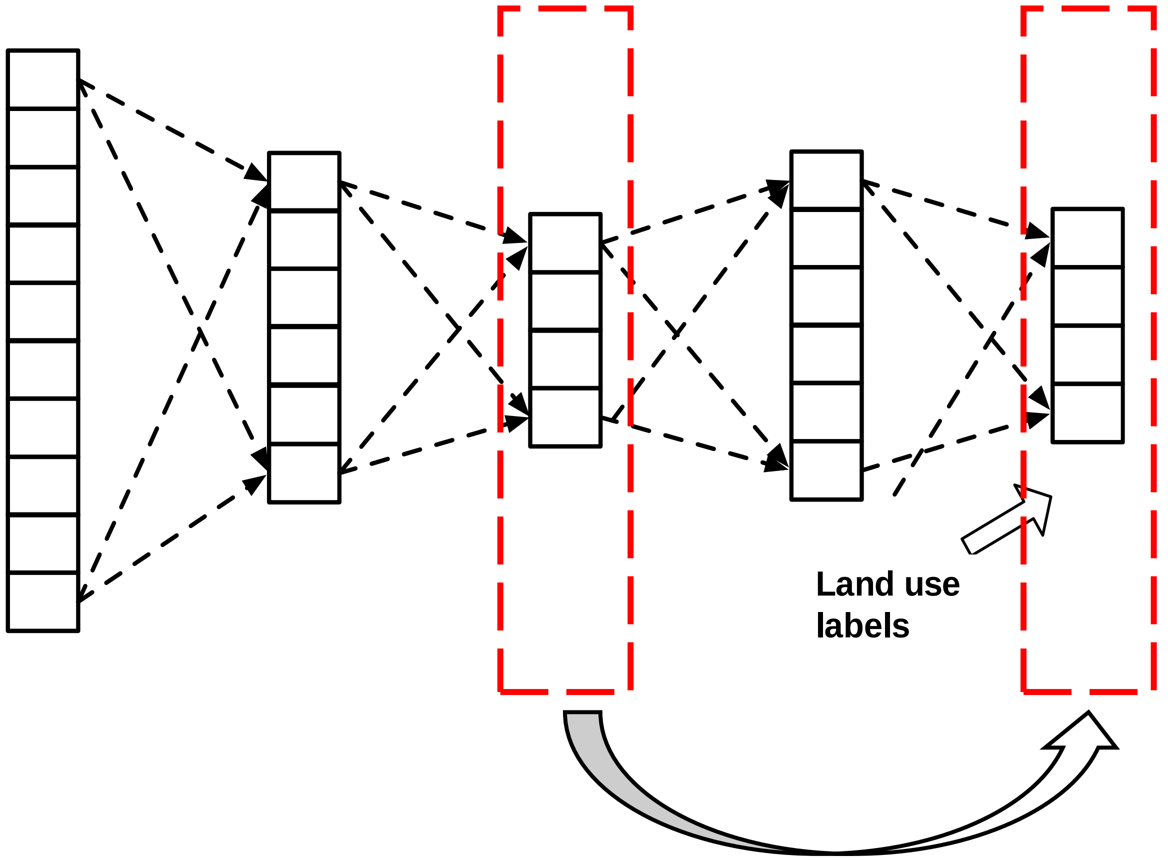


How the compressed features represent the raw data?



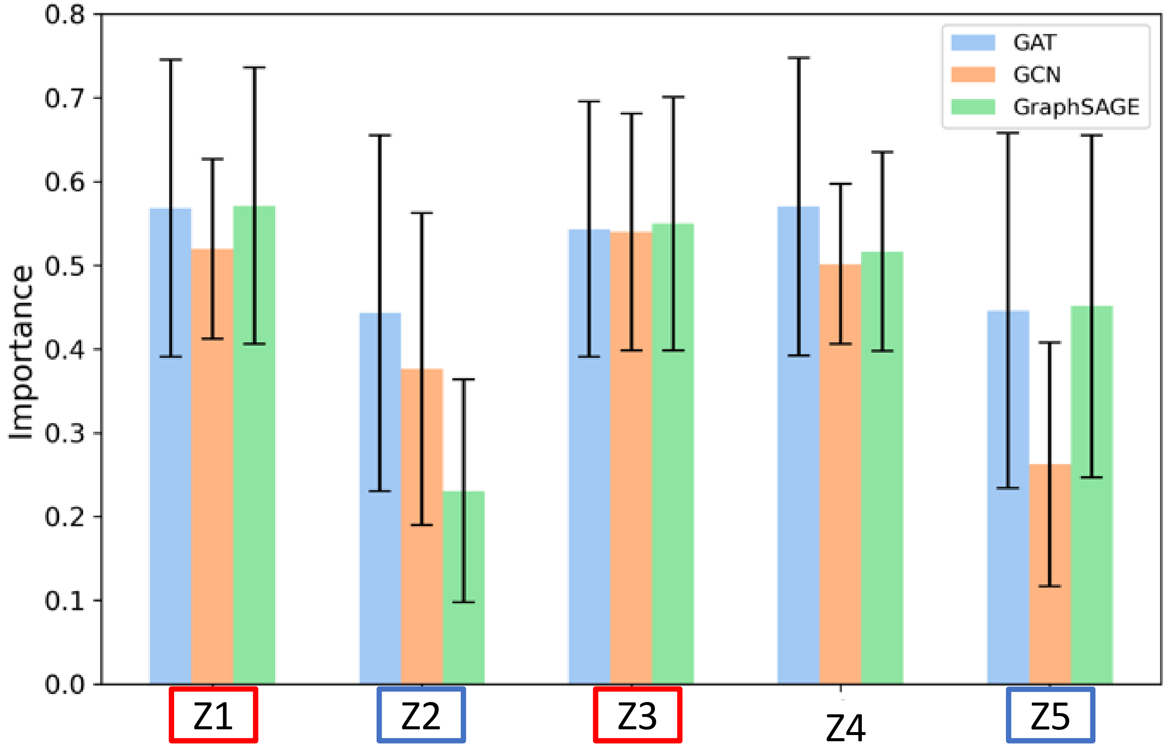
Z1 and Z3 represent the peak hours;
Z2 and Z5 redundant features

C2 Intuitive function of hidden layer



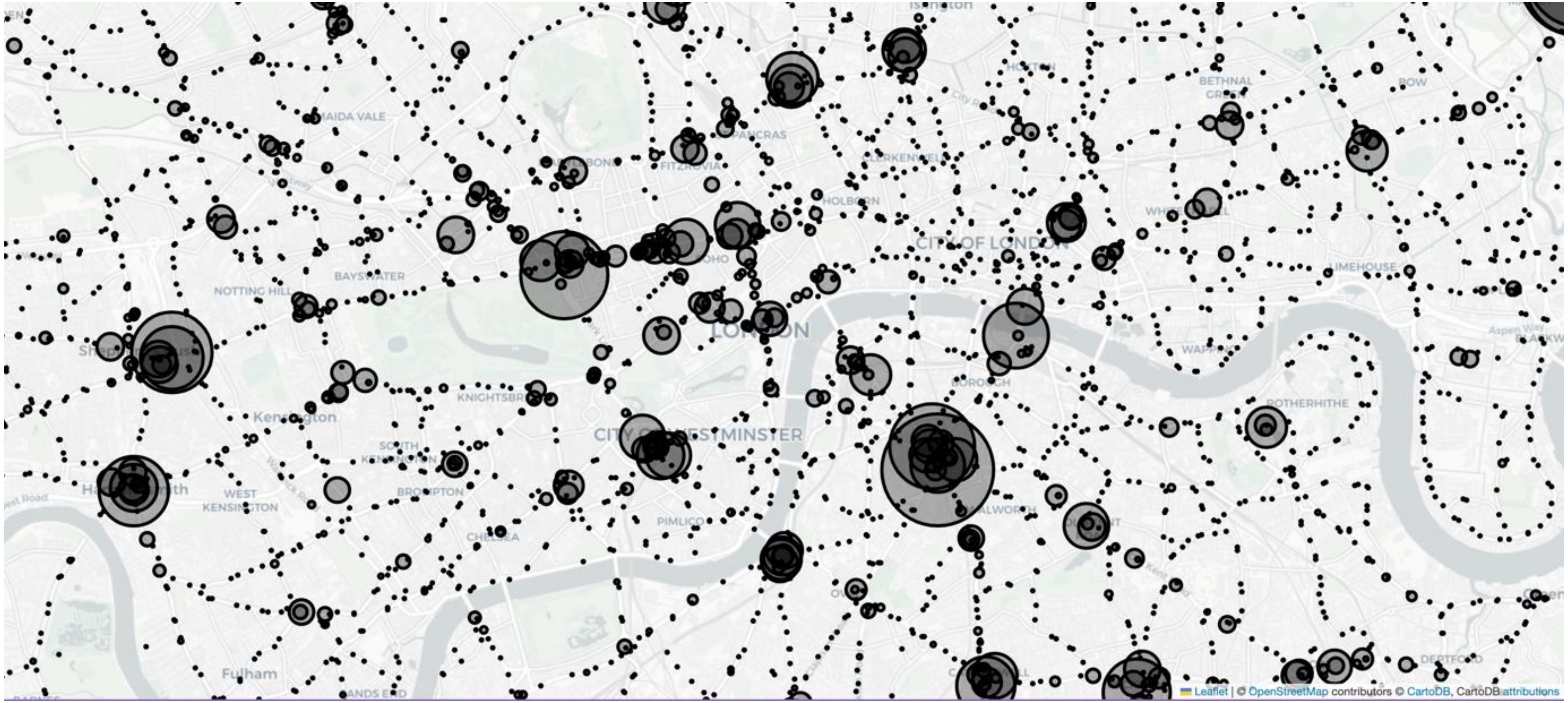
How the compressed features contribute the output?

Post-hoc explanation: Feature Importance



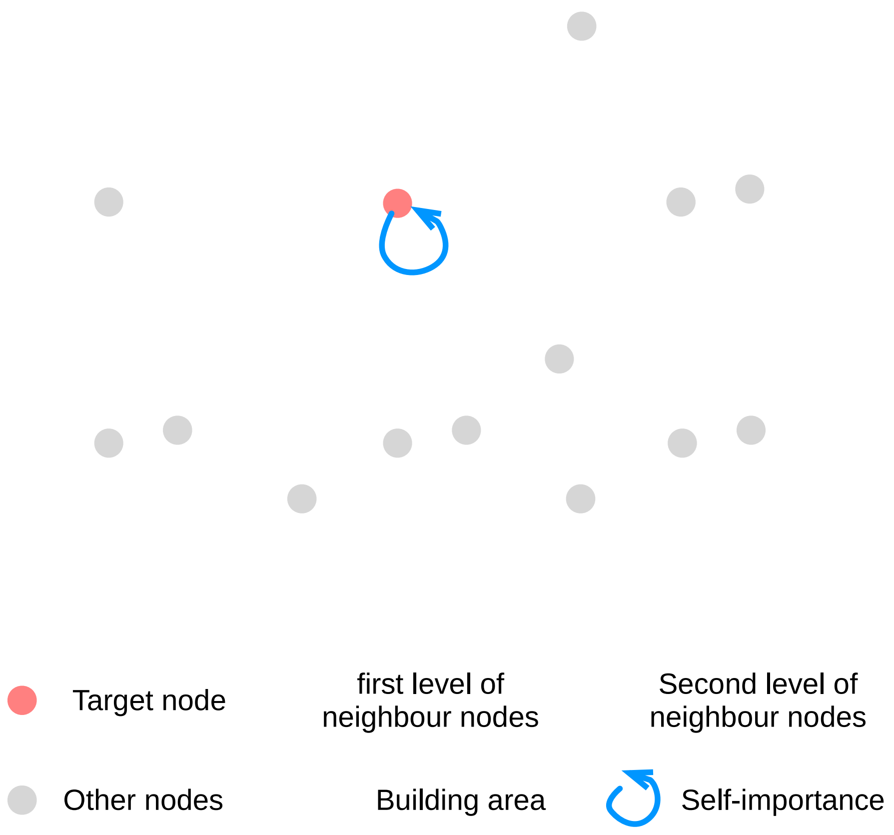
Z1 and Z3 have higher importance;
Z2 and Z5 have lower importance

C3 ‘Importance’ in Nodes level



The spatial distribution of ‘importance’ for a land-use identification case in London (using GraphSage)

C4 'Importance' in edges level



The histogram of self-importance vs neighbour importance

Case Two : Post-hoc Explanation for Traffic Anomaly Detection

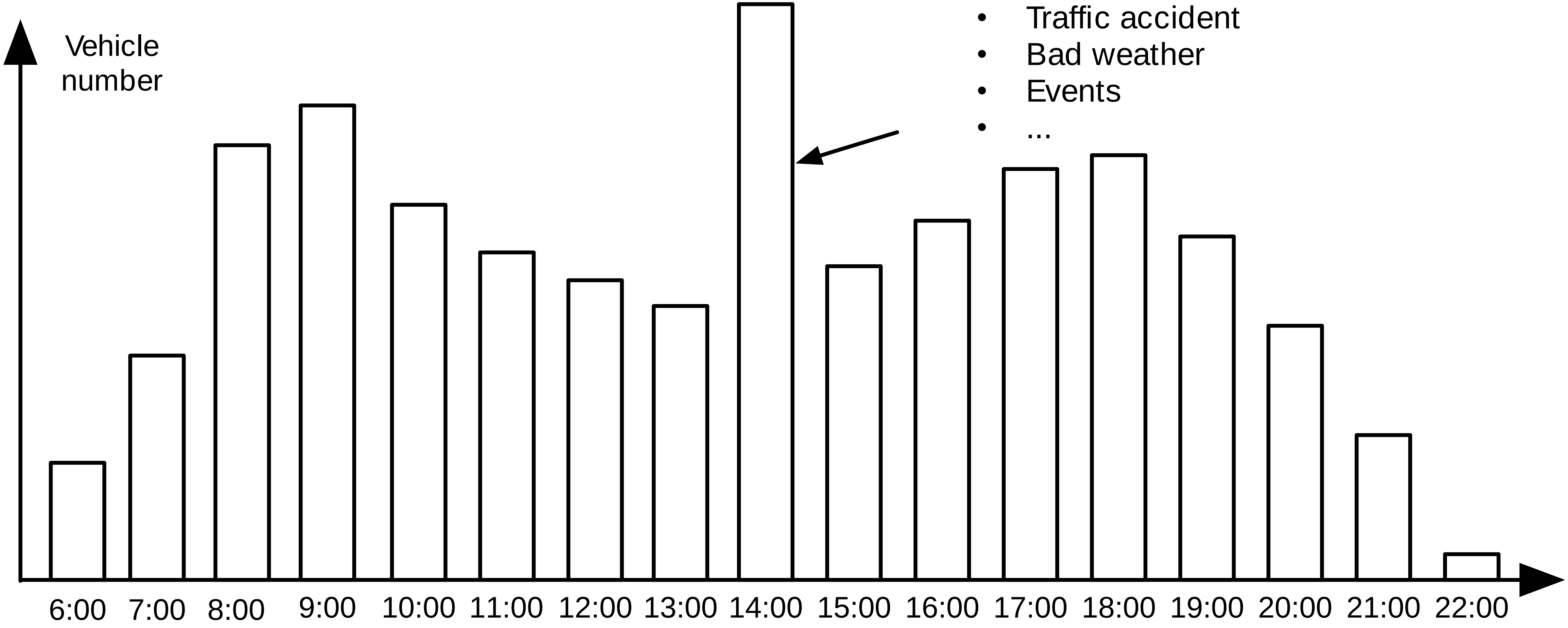
Zhai, X.*, Guo, F., Sivakumar, A., (2023). Towards Interpretable Traffic Anomaly Detection: A Spatiotemporal Self-Supervised Network With Multi-Probing Tasks. The 102th annual meeting of Transport Research Board, Washington D.C, USA.

What is Traffic Anomaly?

- Urban transport systems are often impacted by anomalies such as **traffic accidents** and **inclement weather**, which can create negative impacts on traffic safety and service quality.

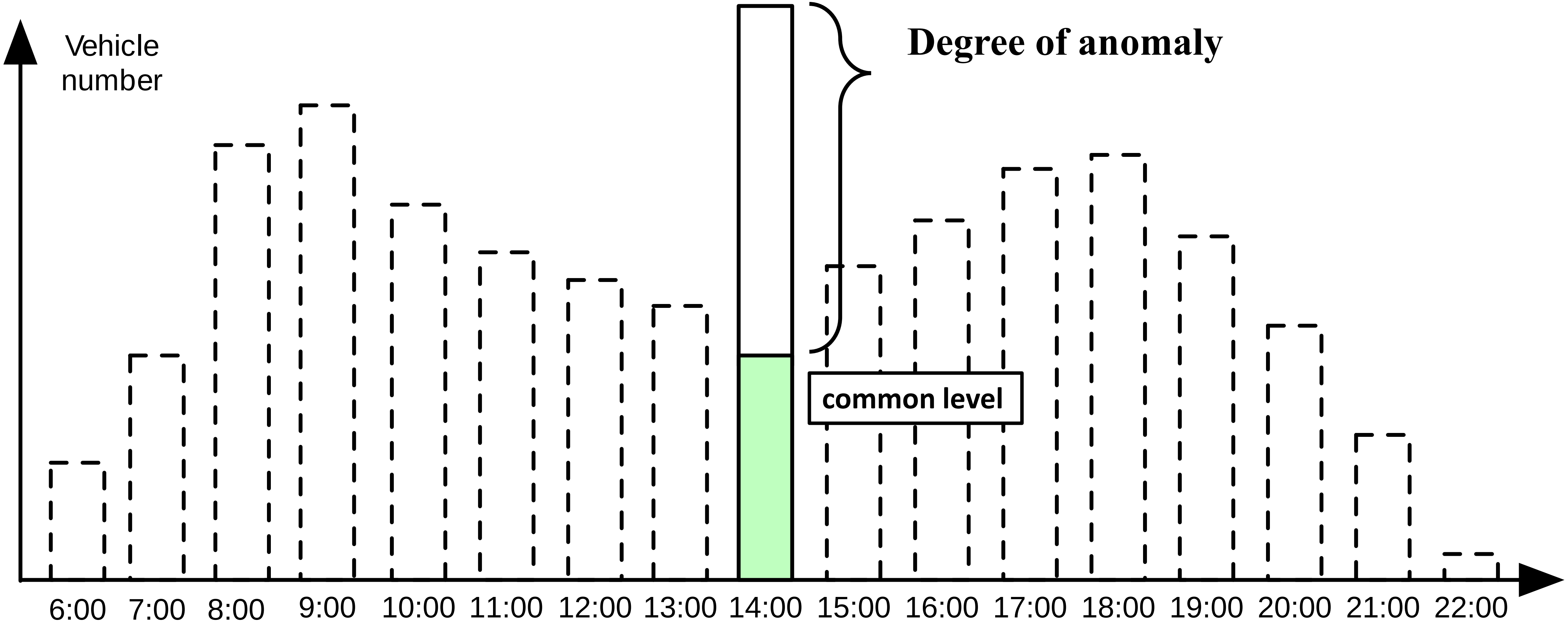
- **Traffic Anomaly Detection (TAD)** aims at the accurate detection of traffic anomalies, plays a critical role in **Active strategies** and **Incidents managements.**

What is Traffic Anomaly?



- Traffic accident
- Bad weather
- Events
- ...

What is Traffic Anomaly?

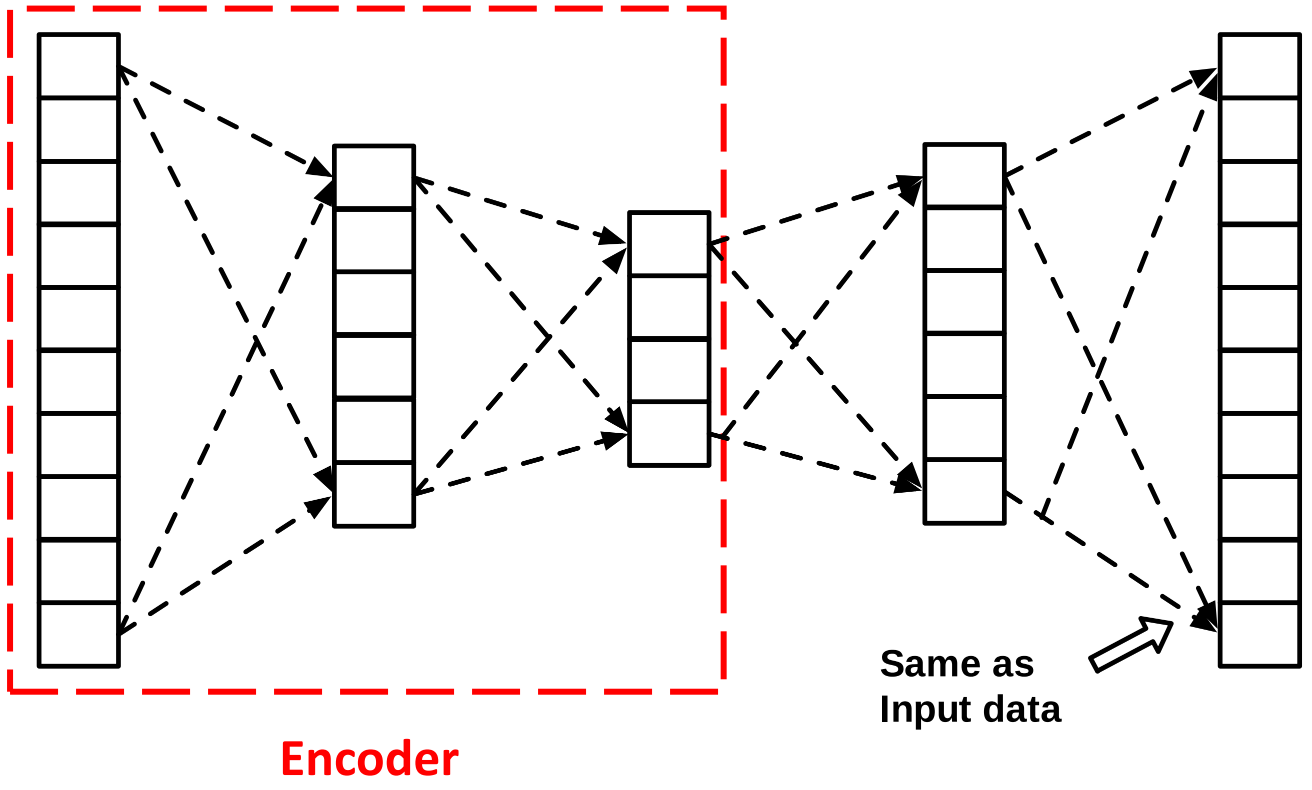


How to detect Traffic Anomaly?

TAD (self-supervised-based) focuses on learning **common patterns** and measuring the anomaly degree based on samples' deviation from the learned common patterns.

- This common pattern is learned by two processes:

Encoder compresses input matrix into a series of lower-dimensional representations.

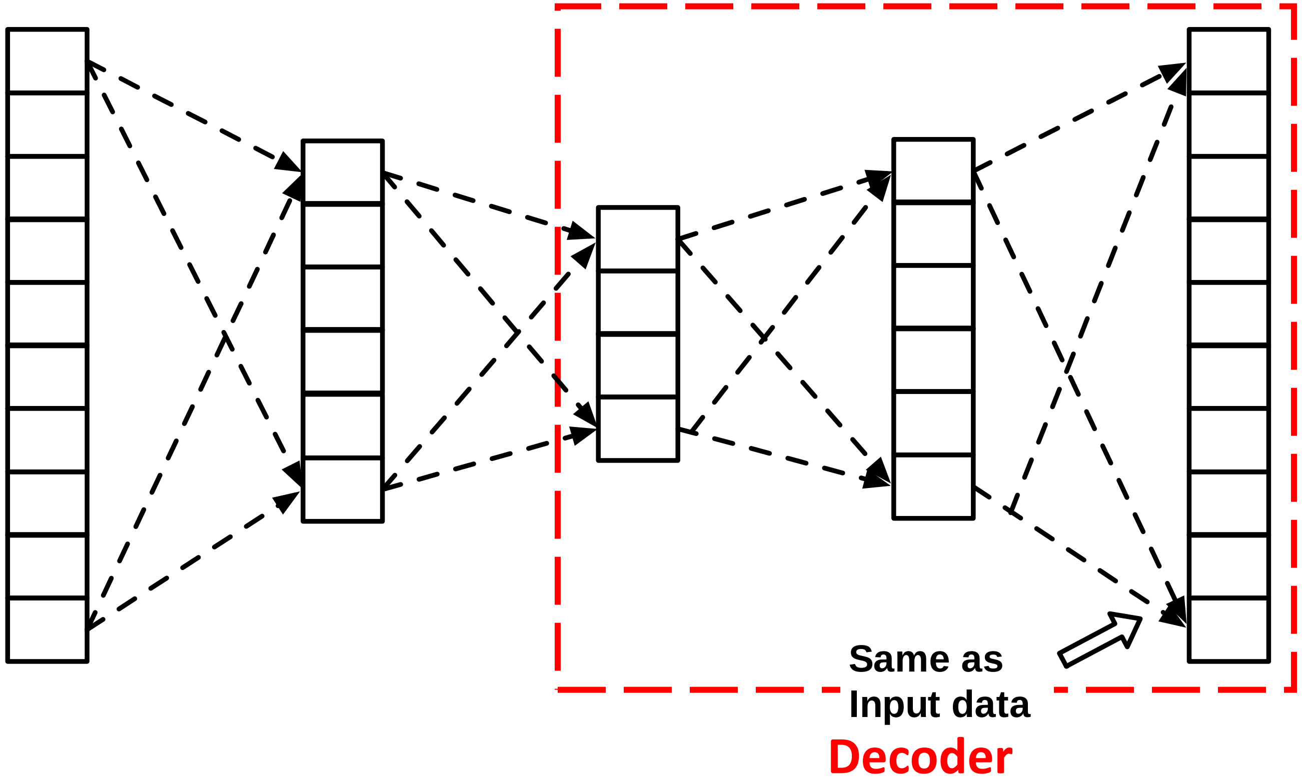


How to detect Traffic Anomaly?

TAD (self-supervised-based) focuses on learning common patterns and measuring the anomaly degree based on samples' deviation from the learned common patterns.

- This common pattern is learned by two processes:

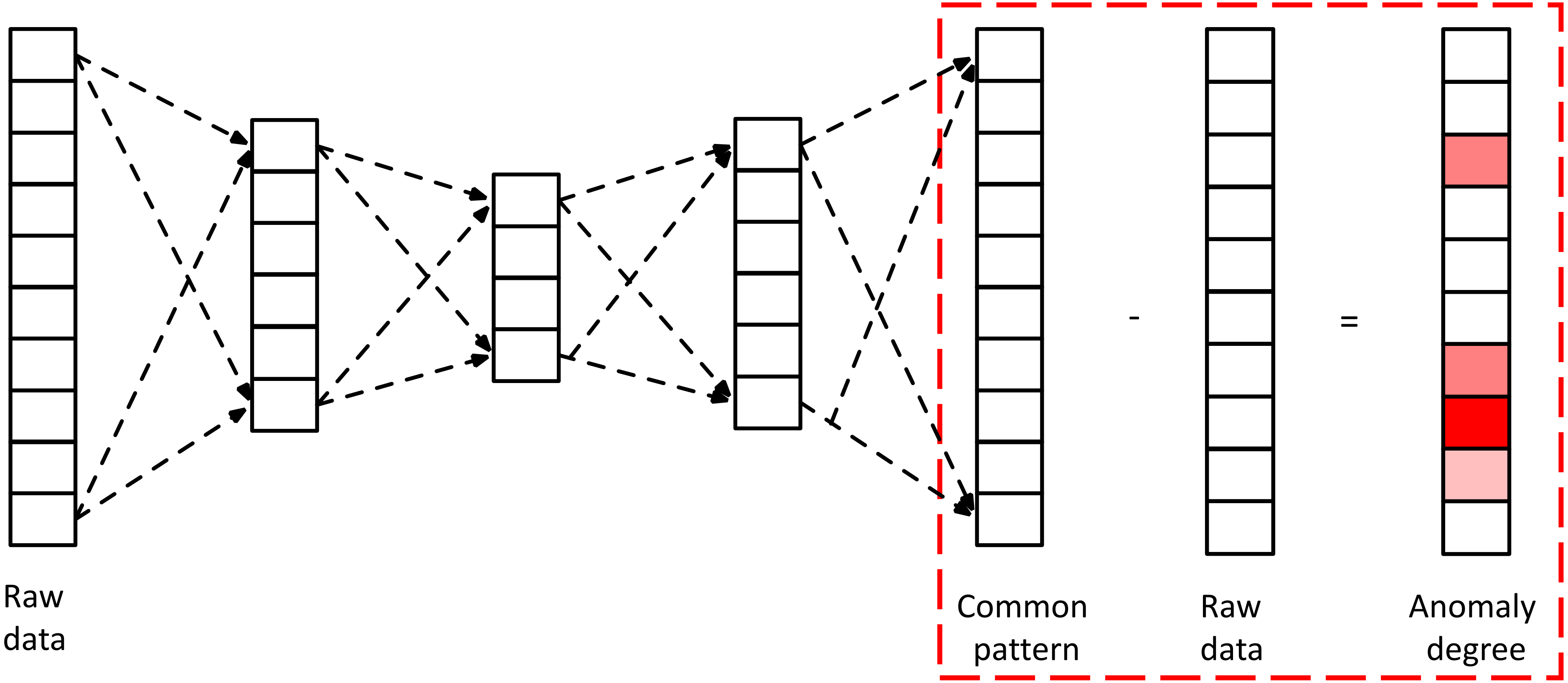
Decoder reconstructs these representations back to the input matrix.



How to detect Traffic Anomaly?

TAD (self-supervised-based) focuses on learning common patterns and measuring the anomaly degree based on samples' deviation from the learned common patterns.

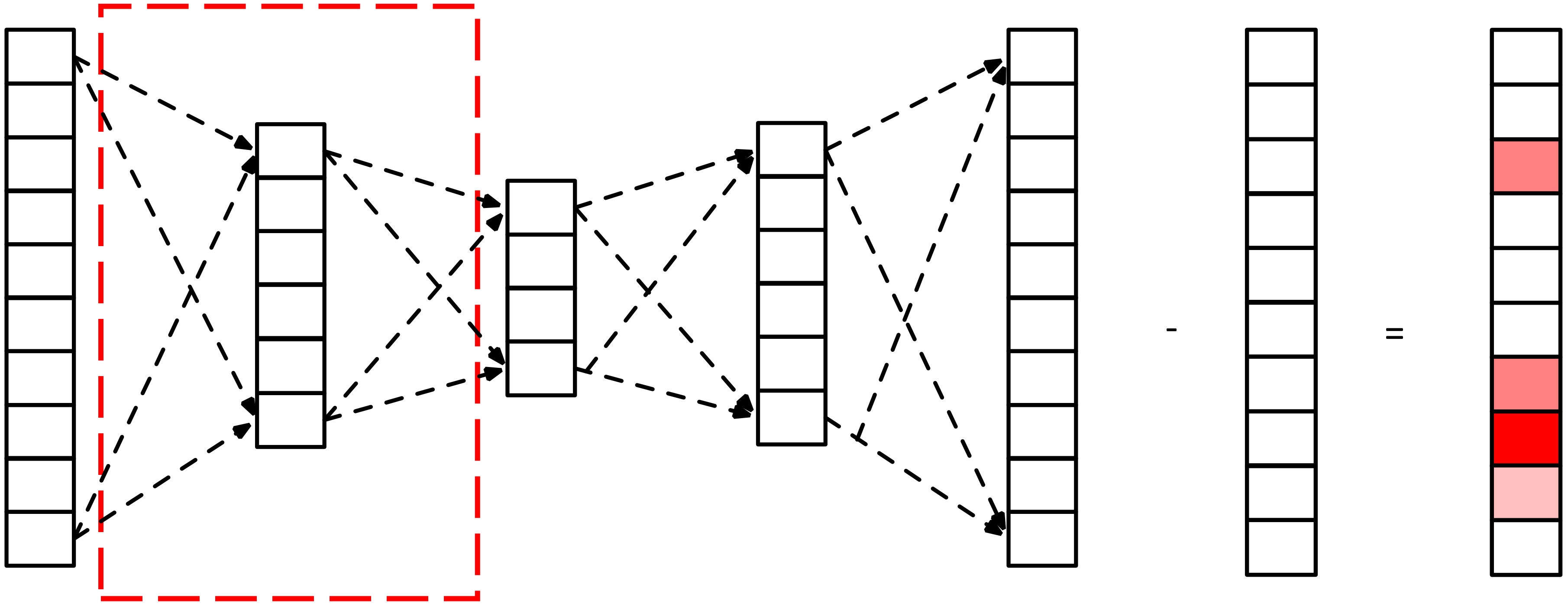
The absolute error distribution represents the degree of anomaly.



How to detect Traffic Anomaly?

TAD (self-supervised-based) focuses on learning common patterns and measuring the anomaly degree based on samples' deviation from the learned common patterns.

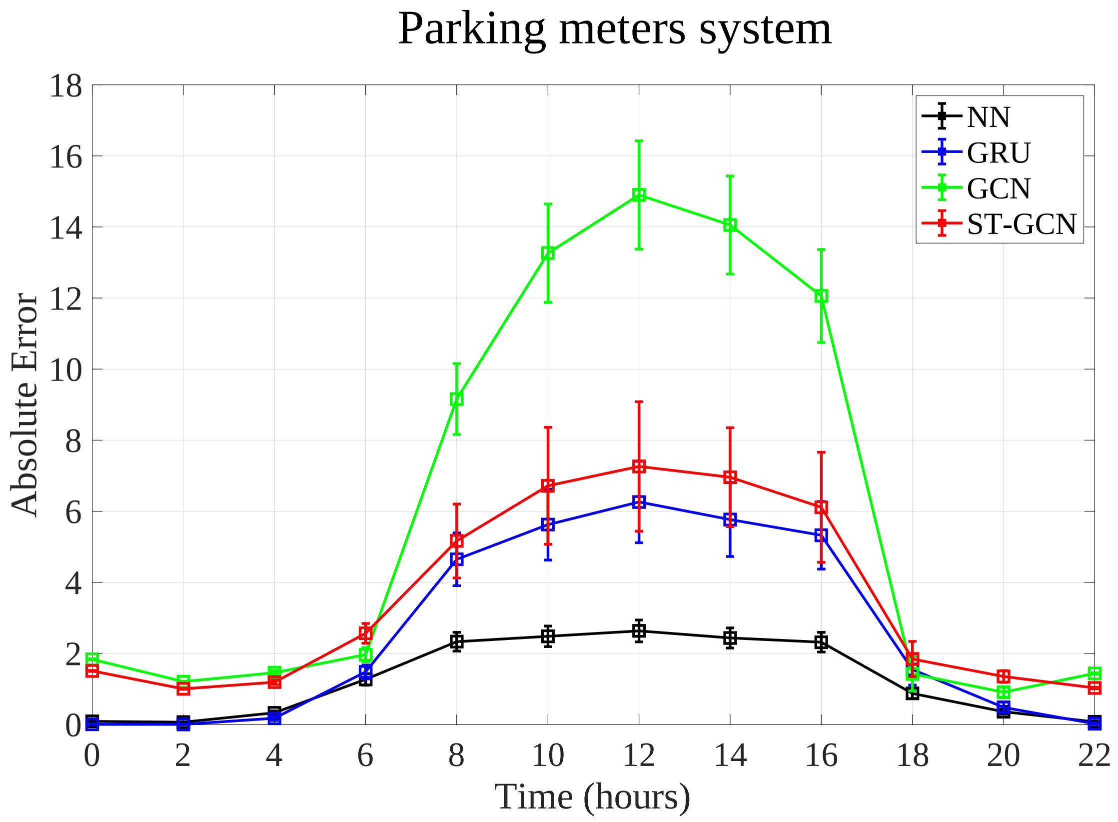
The absolute error distribution represents the degree of anomaly.

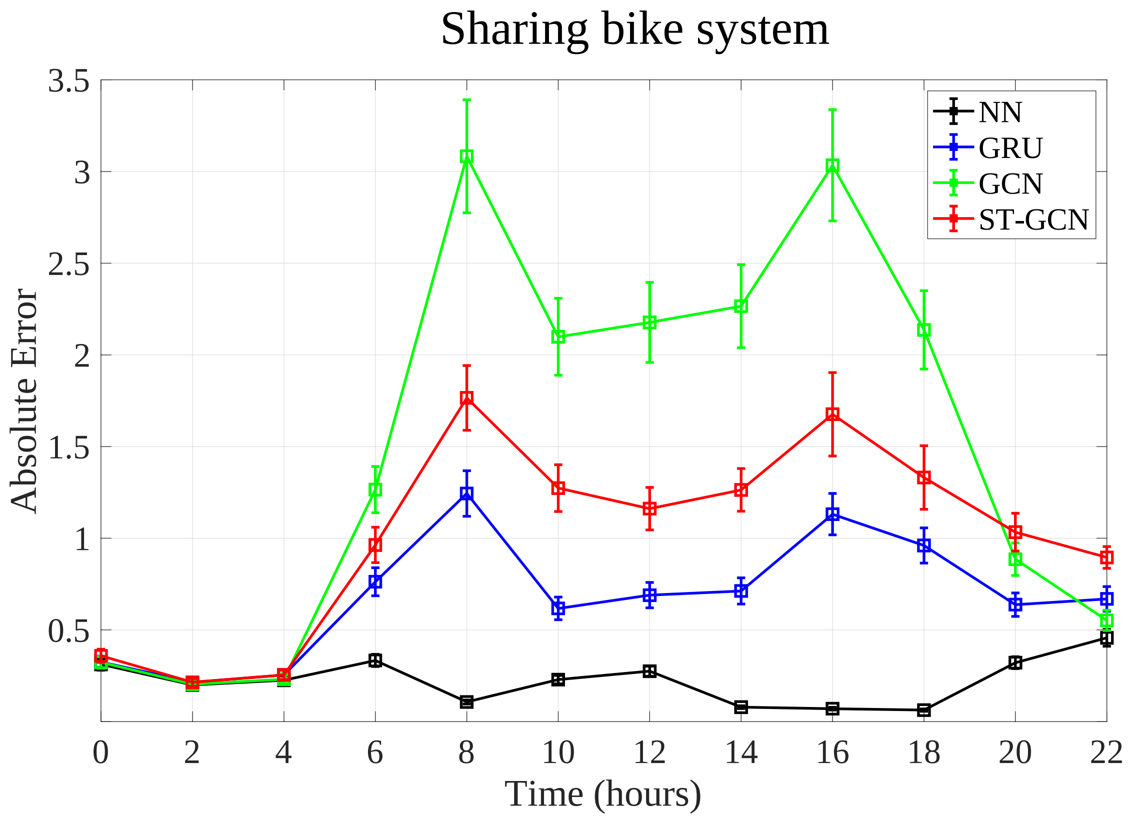


What deep learning structure should be used?

Motivations and Objectives

4 types of **encoders** are tested: general neural network(NN), temporal model(GRU), spatial model(GCN), and spatiotemporal model (ST-GCN)



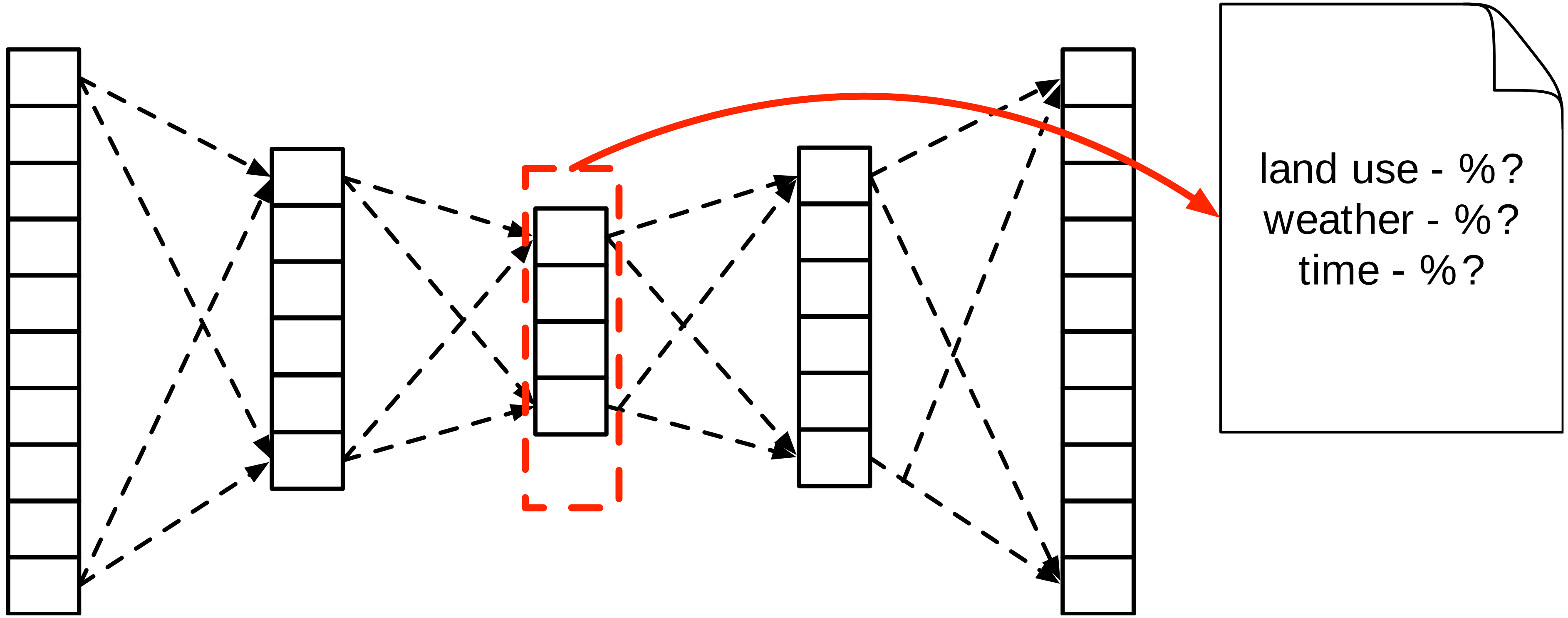


Which deep learning structure is better?

Motivations and Objectives

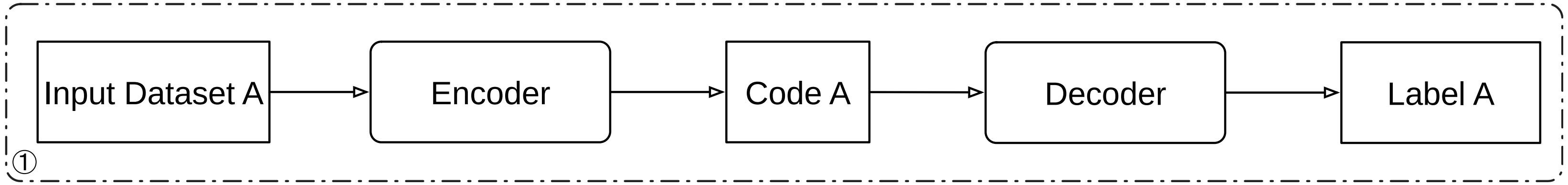
My objectives

- To quantitively explain what and how well **Reference Information** have been learned in the hidden processes of a bunch of DL-based TAD models.
- **Reference Information** could be land use intensities and types, weather, time.



XAI Methodology: Probing Classifiers

1. Train the test model and fix the structure and parameters.

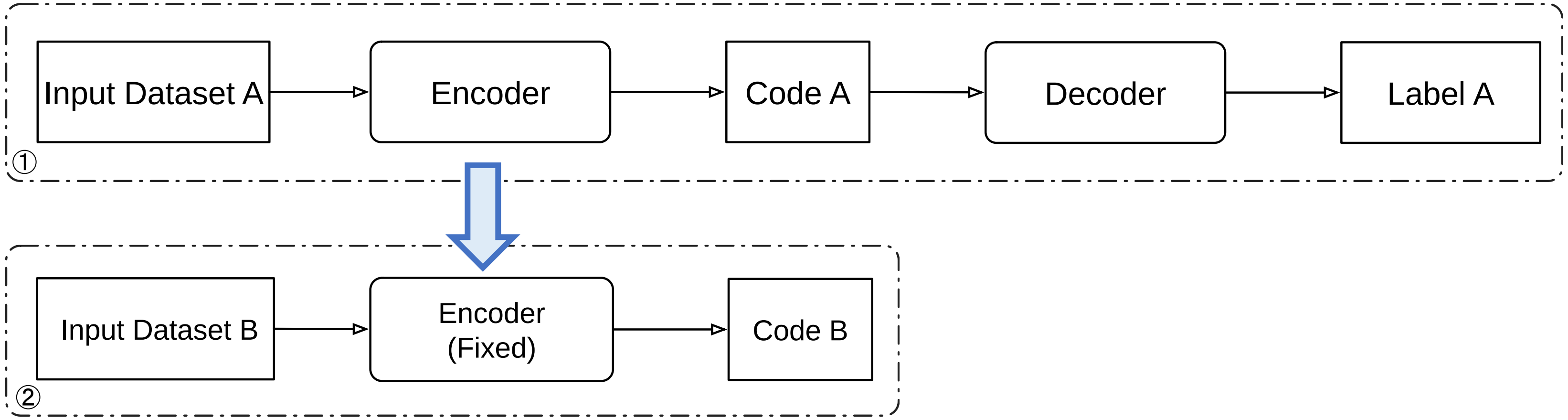


Belinkov, Y., 2022. Probing classifiers: Promises, shortcomings, and advances. *Computational Linguistics*, *48*(1), pp.207-219.

XAI Methodology: Probing Classifiers

1. Train the test model and fix the structure and parameters.

2. Calculate the testing layer-- Code B using another input (with same data source).



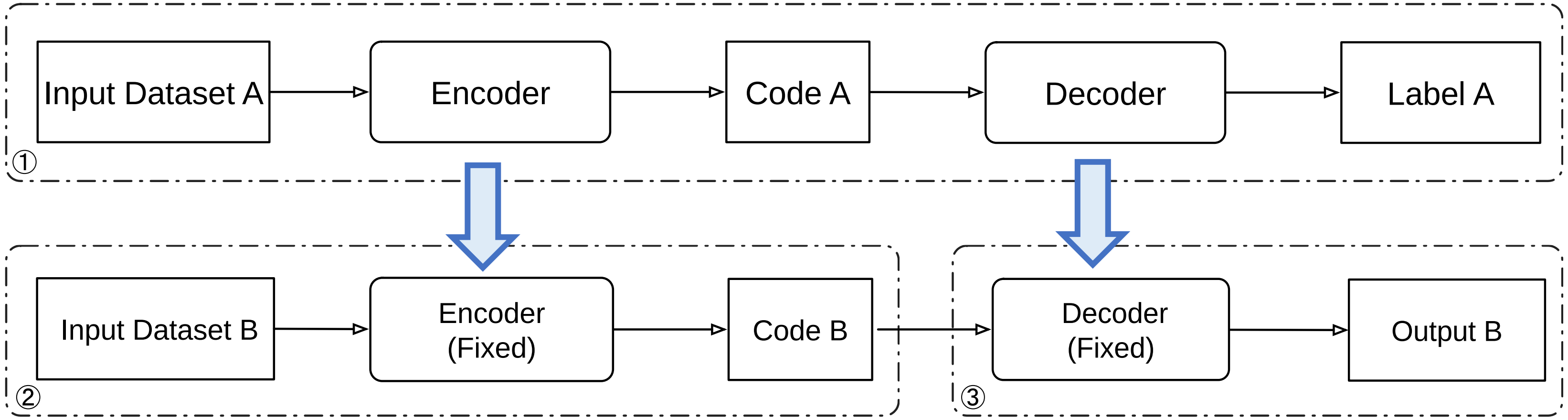
Belinkov, Y., 2022. Probing classifiers: Promises, shortcomings, and advances. *Computational Linguistics*, *48*(1), pp.207-219.

XAI Methodology: Probing Classifiers

1. Train the test model and fix the structure and parameters.

2. Calculate the testing layer--Code B using another input (with same data source).

3. Calculate the output B and Accuracy performance



Belinkov, Y., 2022. Probing classifiers: Promises, shortcomings, and advances. *Computational Linguistics*, *48*(1), pp.207-219.

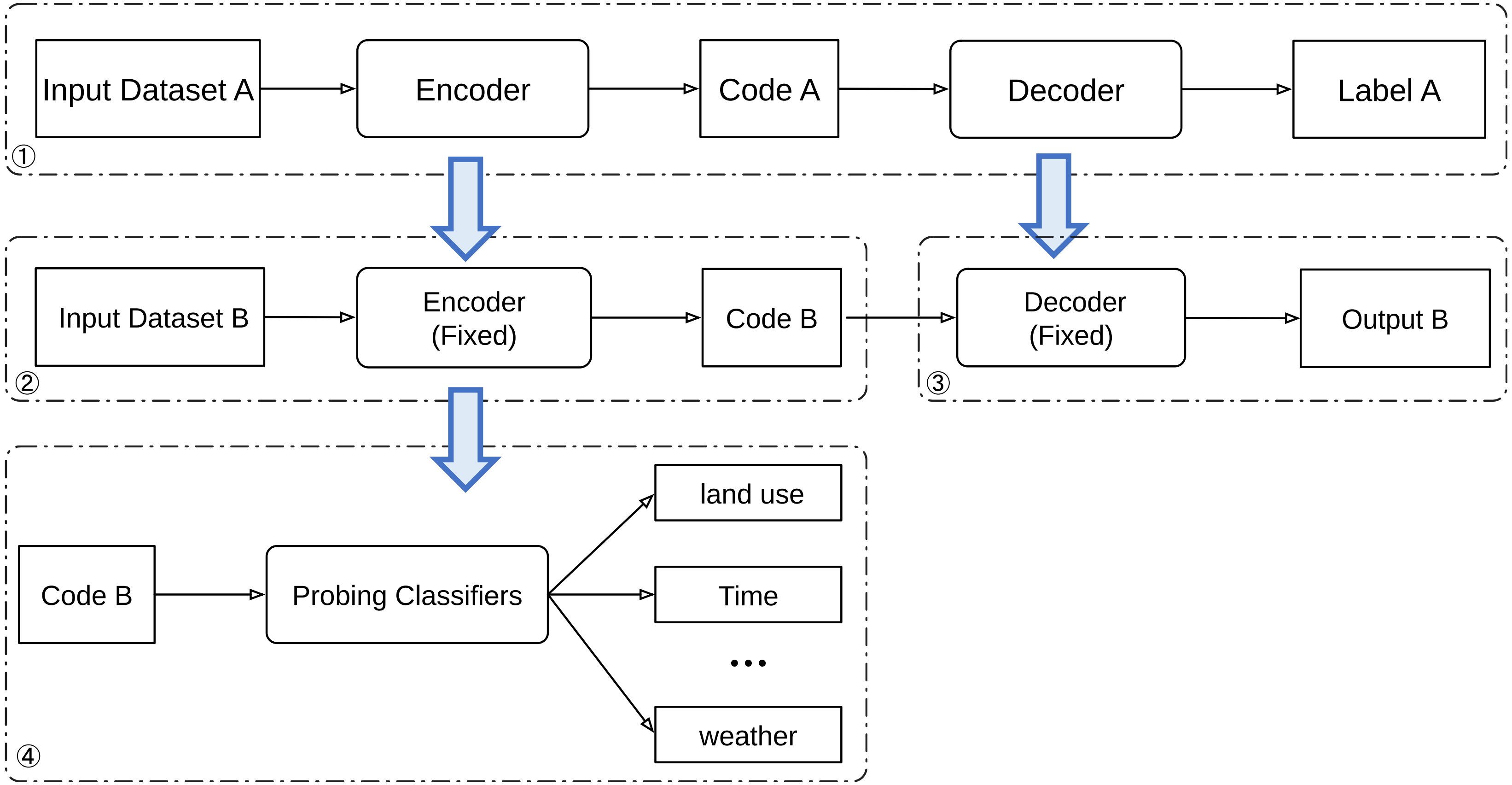
XAI Methodology: Probing Classifiers

1. Train the test model and fix the structure and parameters.

2. Calculate the testing layer--Code B using another input (with same data source).

3. Calculate the output B and Accuracy performance

4. Use Code B as inputs for Probing Classifier with **Reference information** as labels.



Belinkov, Y., 2022. Probing classifiers: Promises, shortcomings, and advances. *Computational Linguistics*, *48*(1), pp.207-219.

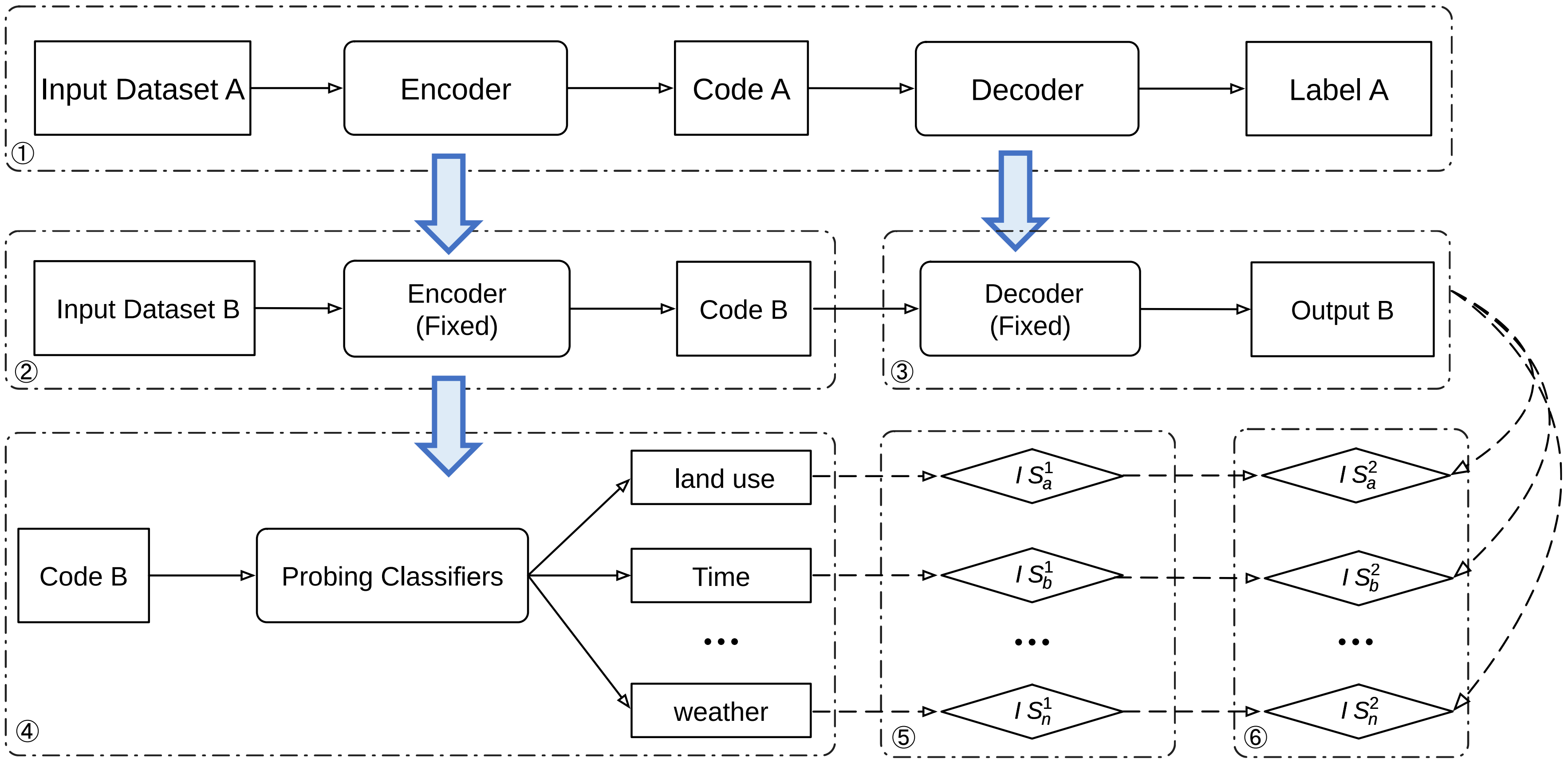
XAI Methodology: Probing Classifiers

1. Train the test model and fix the structure and parameters.

2. Calculate the testing layer--Code B using another input (with same data source).

3. Calculate the output B and Accuracy performance

4. Use Code B as inputs for Probing Classifier with **Reference information** as labels.



5. Calculate explainability score 1: accuracy of probing tasks

6. Calculate explainability score 2:
correlation between score 1 and Output B

The higher accuracy of test layer in probing task, the more reference information is captured in test layer.

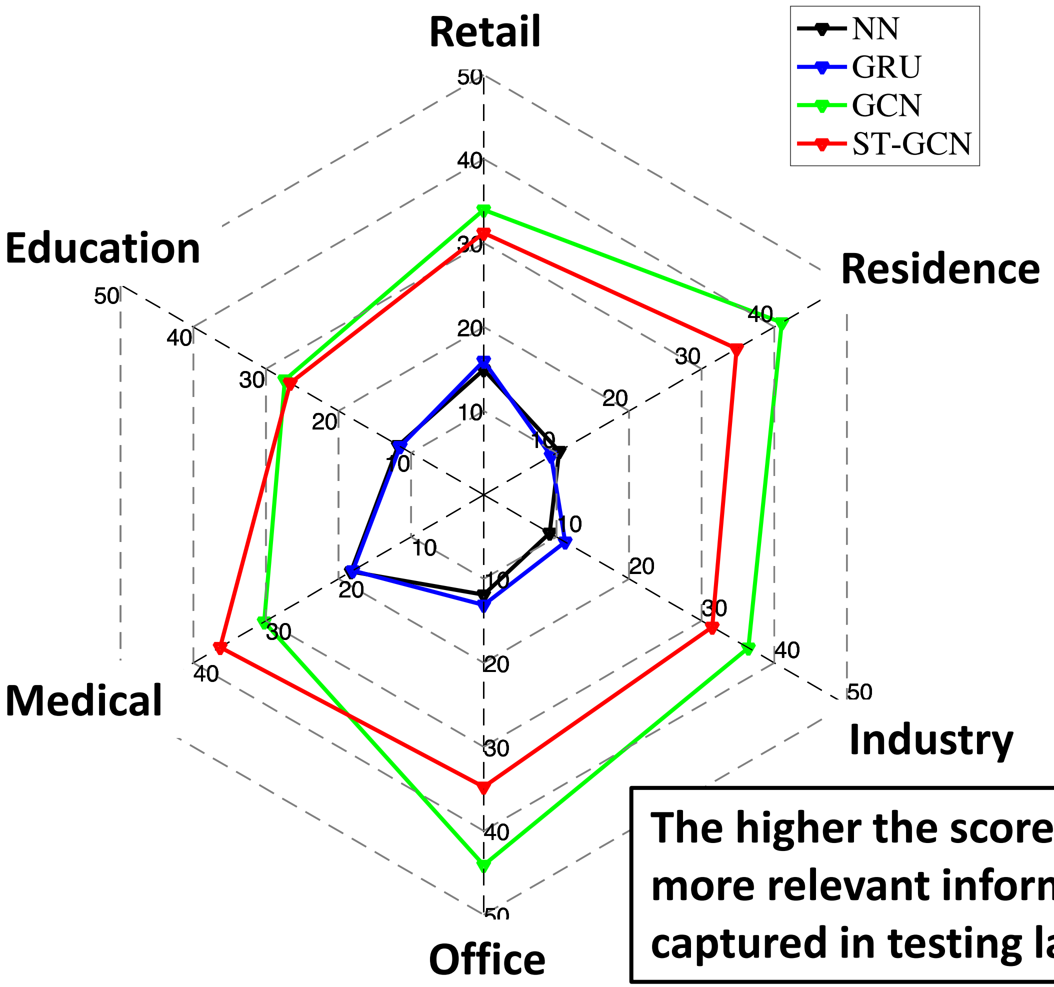
Reference information

Description of reference dataset for probing task

Reference Index	Data Type	Description
Spatial indicators		
CIE	Numerical	Intensity of cultural, institutional, and educational services
MED	Numerical	Intensity of medical service
MIPS	Numerical	Intensity of management, information, and professional services
PDR	Numerical	Intensity of industrial facilities (production, distribution, and repair)
RES	Numerical	Intensity of Residential activities
RET	Numerical	Intensity of retail, and entertainment services
Temporal indicators		
DOW	Categorical	Day of week
WD/WE	Categorical	Weekday or weekend
SEA	Categorical	Season in a year
Weather indicators		
TEM	Numerical	Average temperature from 6:00 to 24:00 (Celsius)
SOW	Numerical	Average speed of wind from 6:00 to 24:00 (mph)
VIS	Numerical	Average visible sight from 6:00 to 24:00 (m)
PREC	Numerical	Average precipitation from 6:00 to 24:00 (mm/h)

Post-hoc explanation case for TAD

➢ **Example of Probing classifiers results in Parking demand prediction**

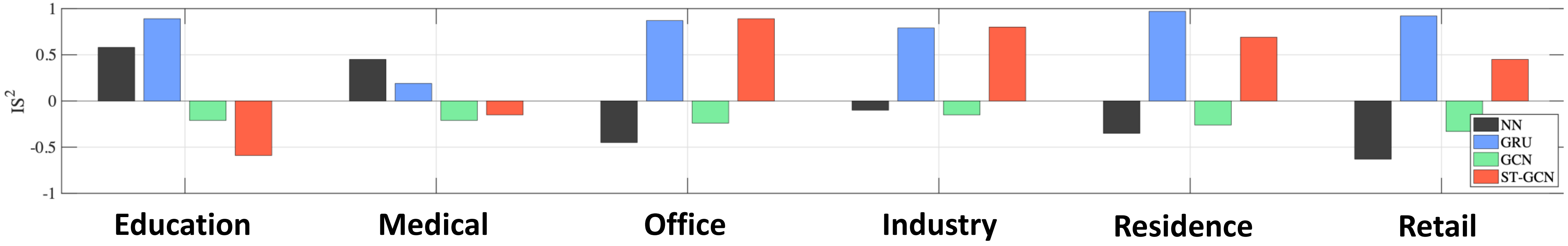


The higher the score, the more relevant information captured in testing layer

Explainability score with using different types of land-use information as **'Reference'**

Post-hoc explanation case for TAD

- **Example of Probing classifiers results in Parking demand prediction**



The higher the score, the more relevant information positively contribute to the outcome

Imperial College London

Thank you!

Date: 14 Dec 2023

Xuehao Zhai

x.zhai20@imperial.ac.uk