ACM SIGSPATIAL 2024

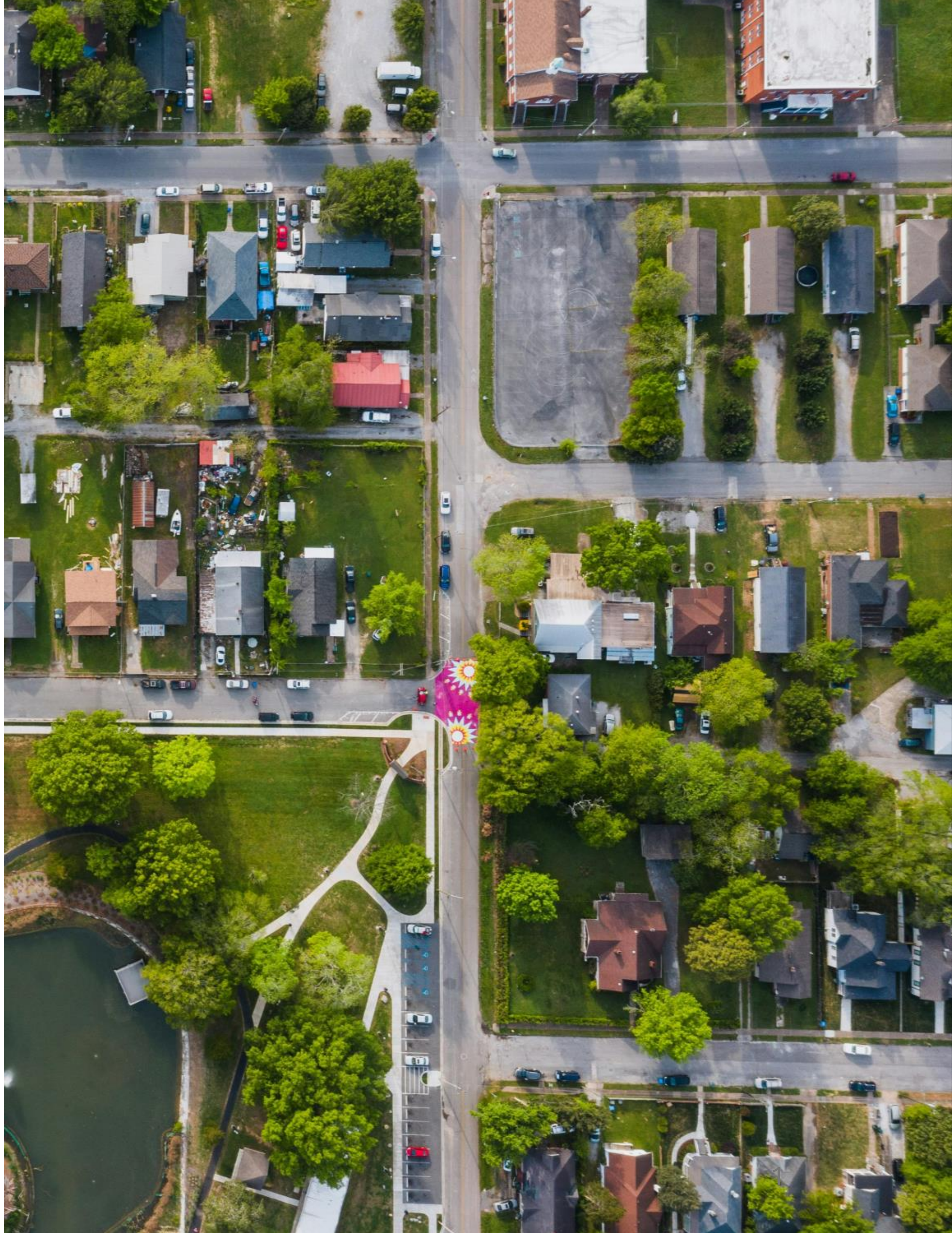
Session 6: Prediction and forecasting

GraphParcelNet: Predicting Parcel-Level Imperviousness from Geospatial Vector Data using Graph Neural Networks

Lapone Techapinyawat
Wenlu Wang
Mehrube Mehrubeoglu
Hua Zhang

TEXAS A&M UNIVERSITY CORPUS CHRISTI | COLLEGE OF ENGINEERING & COMPUTER SCIENCE

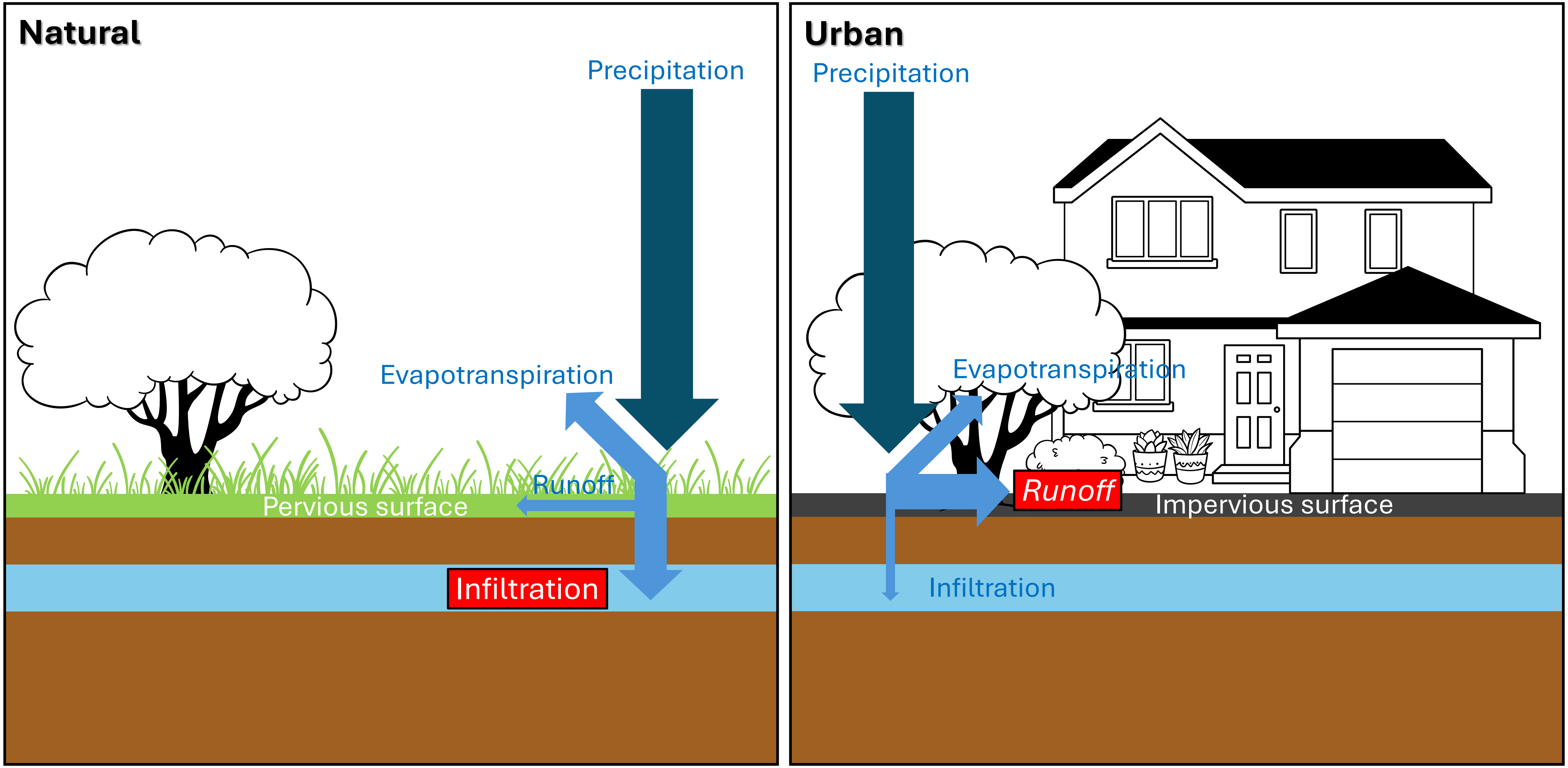
NSF



Pervious and Impervious Surfaces in Urban Areas

(Image ref: Binyamin Mellish)

How Cities Shape the Movement of Water



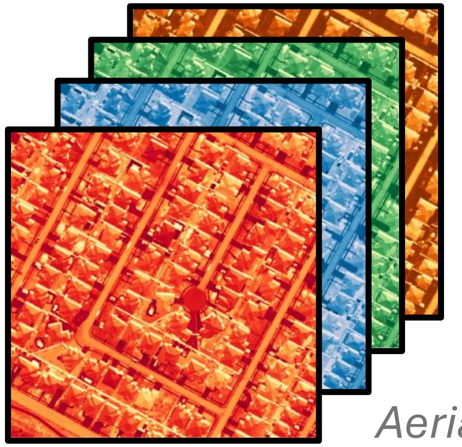
Equivalent residential units(ERUs) VS Tiered

Case study – City of Corpus Christi, Texas

Source: https://experience.arcgis.com/experience/07f98a4be33844668470fdad51fdf869

From Raster to Vector: Introducing GraphParcelNet

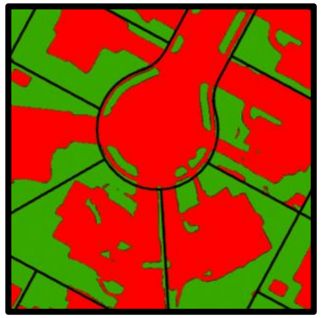
Traditional procedure



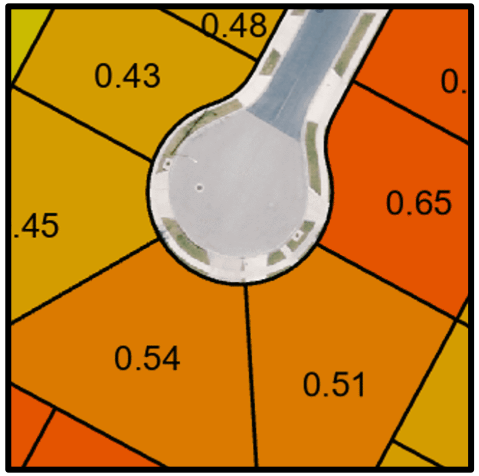
Aerial images

Pixel classification

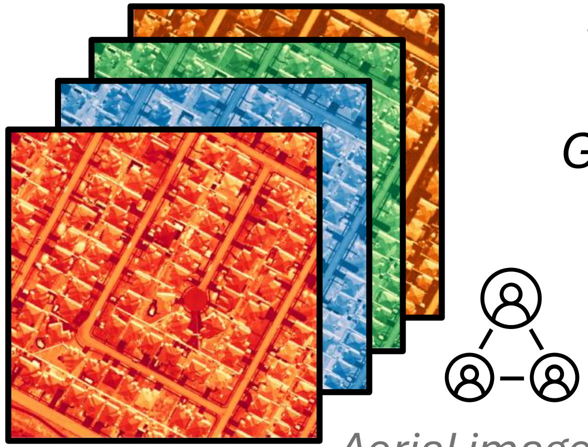
Raster



Spatial statistics

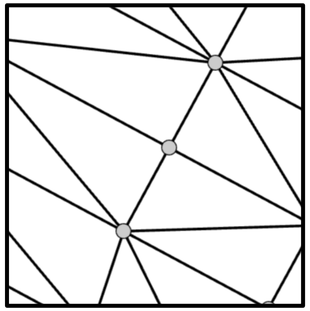


Proposed procedure



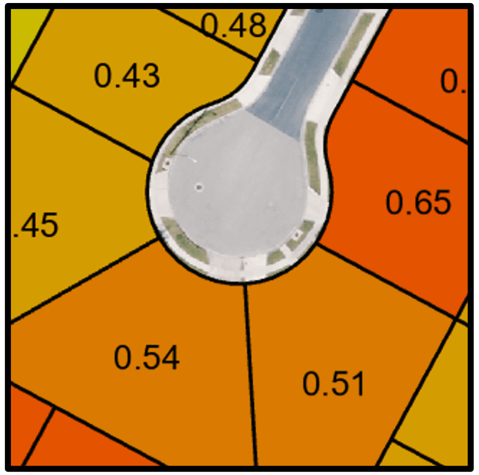
Aerial images + Demographic

Spatial statistics, Vectorization, Graph construction



Graph neural network

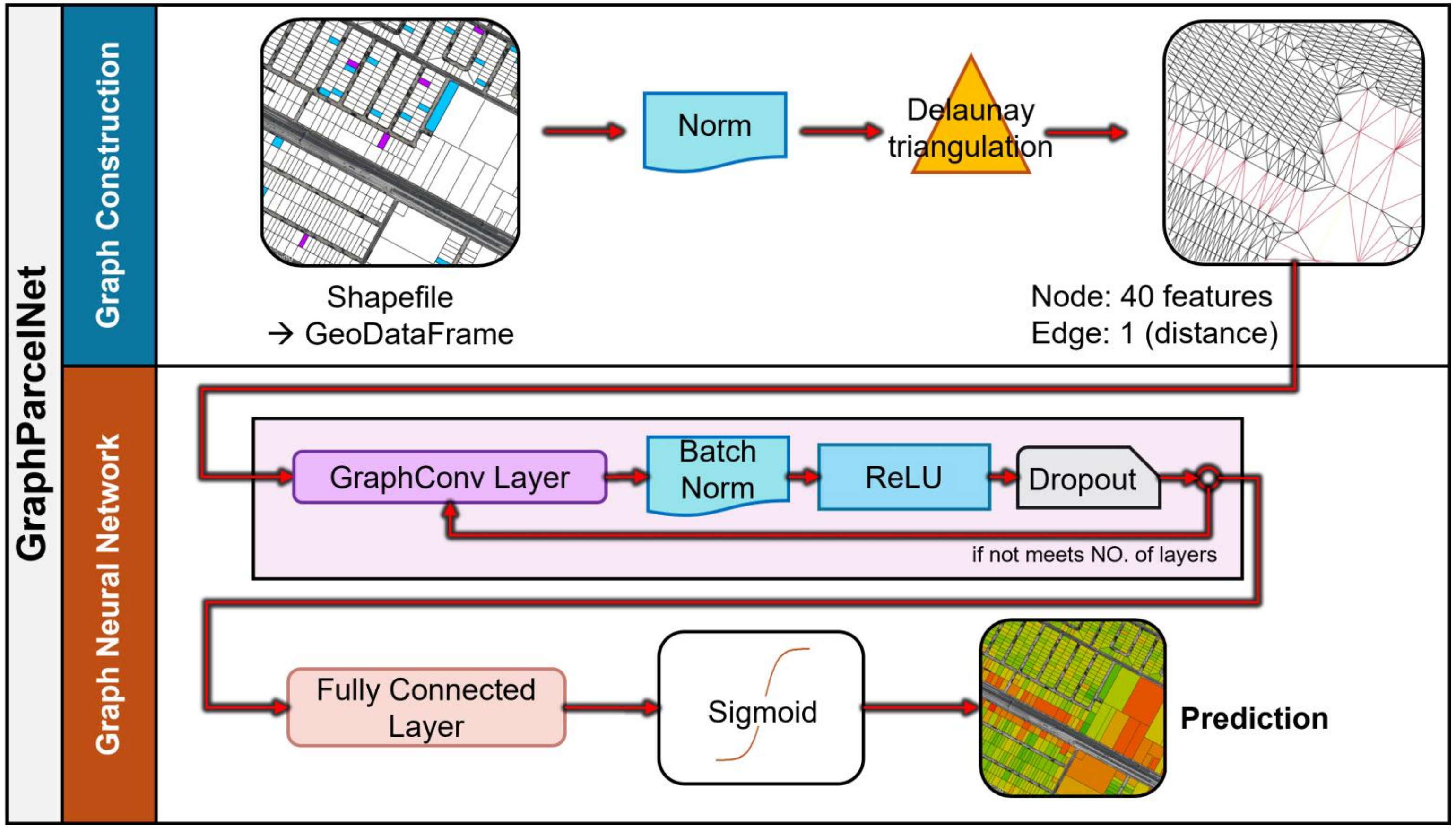
Vector



Proposed method

Phase 1: Graph Construction

- Input features from aerial imagery and ACS demographic data.
- Features are normalized (MinMax scaling from -1.0 to 1.0).
- Parcel system is represented as a graph G=(N,E) where:
  - **N**: Nodes (individual parcels).
  - **E**: Edges (spatial relationships).
- **Delaunay triangulation** used to define edges (distance as features).
- Data divided into training, validation, and testing subsets.



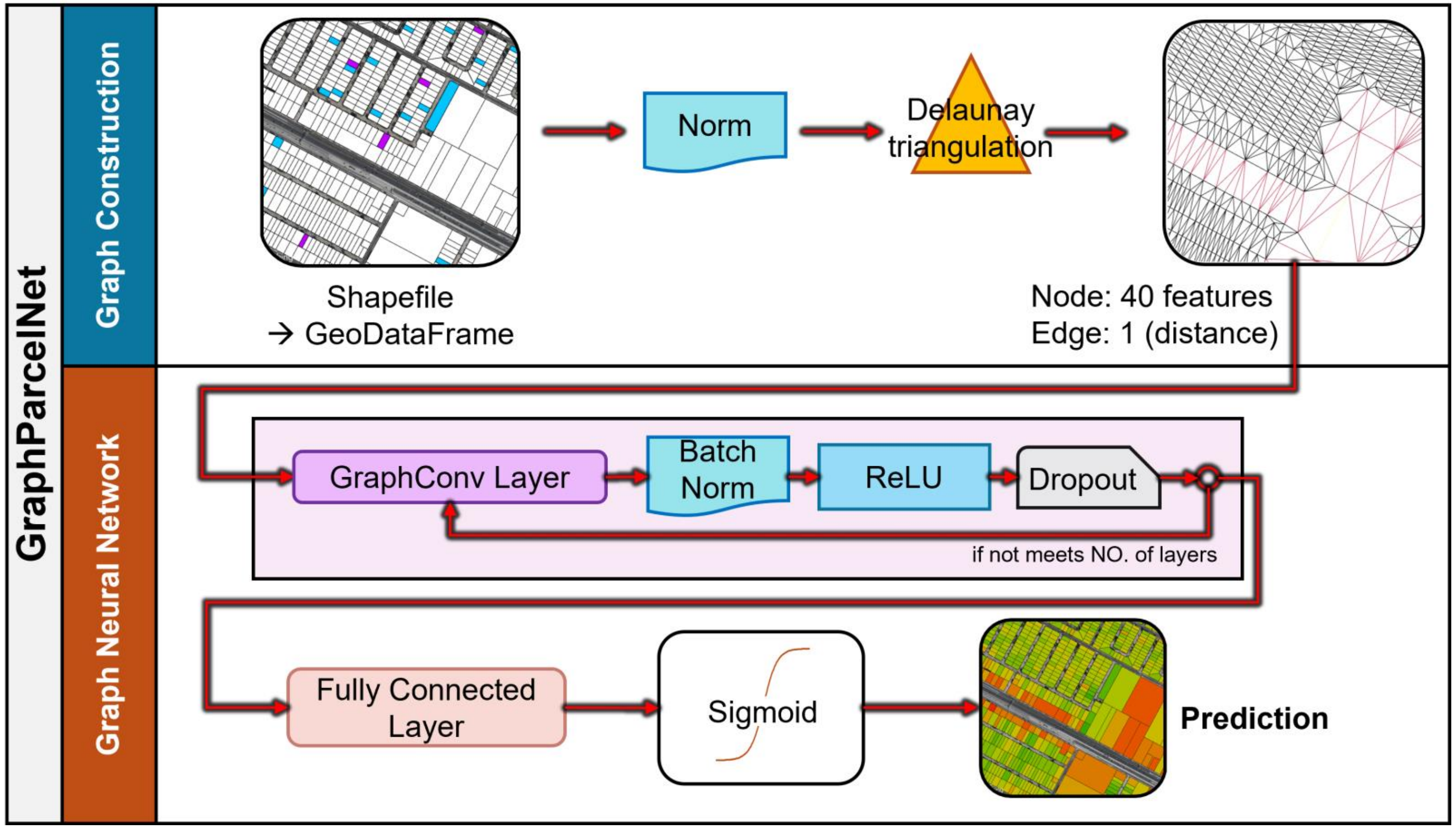
Proposed method

Phase 2: Graph Neural Network

- GraphConv Layer: Aggregates node features, capturing dependencies between parcels.
- Residual block: ReLU activation, batch normalization, dropout.
- Fully Connected Layer
- Sigmoid function ensures output is [0.00-1.00]

$$\mathrm{GNN} = \sigma\left(\mathrm{DBN}(\mathrm{GraphConv}(H^{(k)}))\Big|_{k=1}^{n}\right)$$

Final Output:
- Prediction of impervious surface ratio using **MAE** as the loss function.



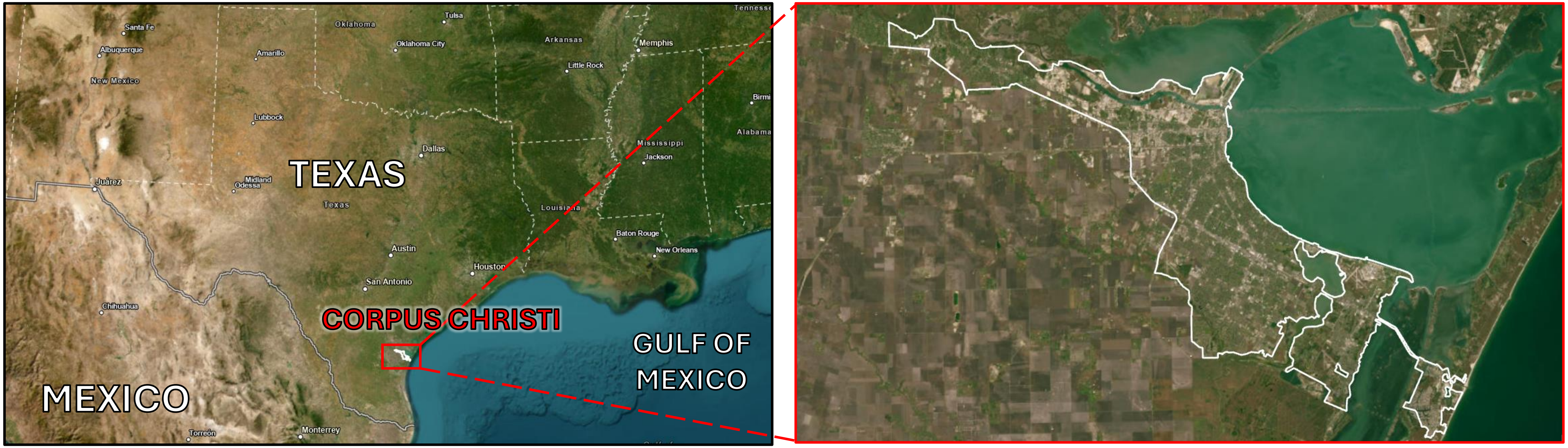
Experiments

Datasets

Location: Corpus Christi, Texas, USA

Data Overview:

- Total parcels: 103,828



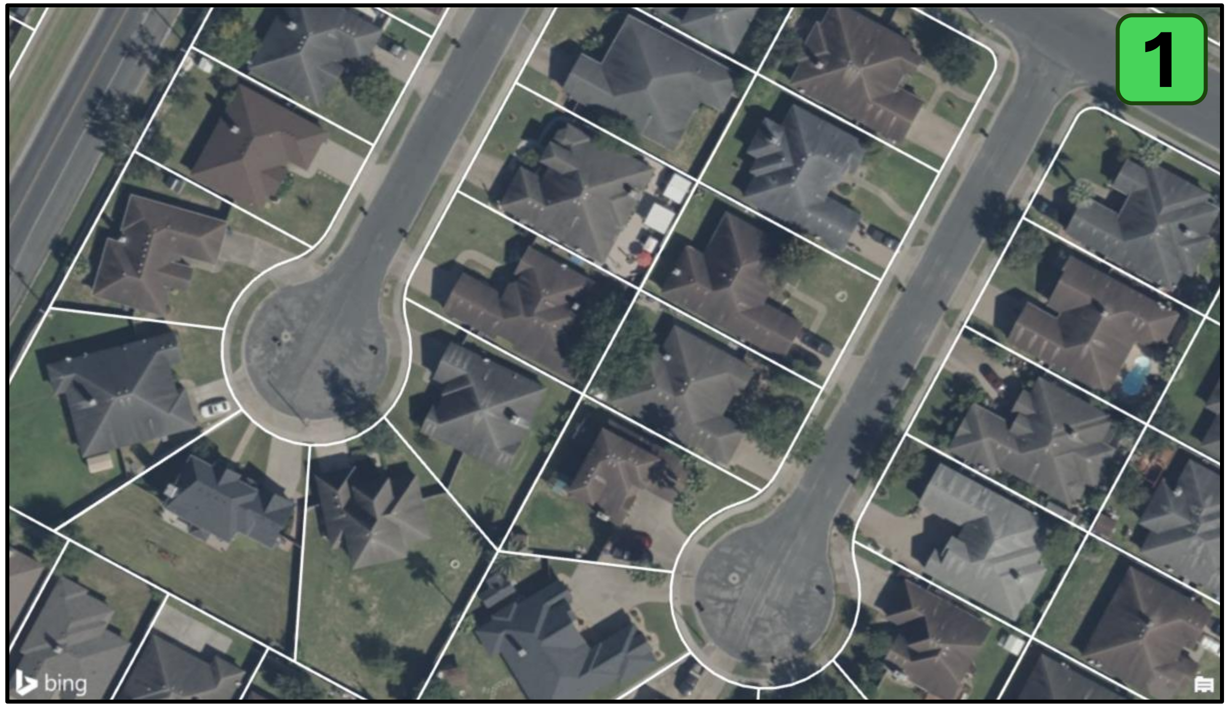
Experiments

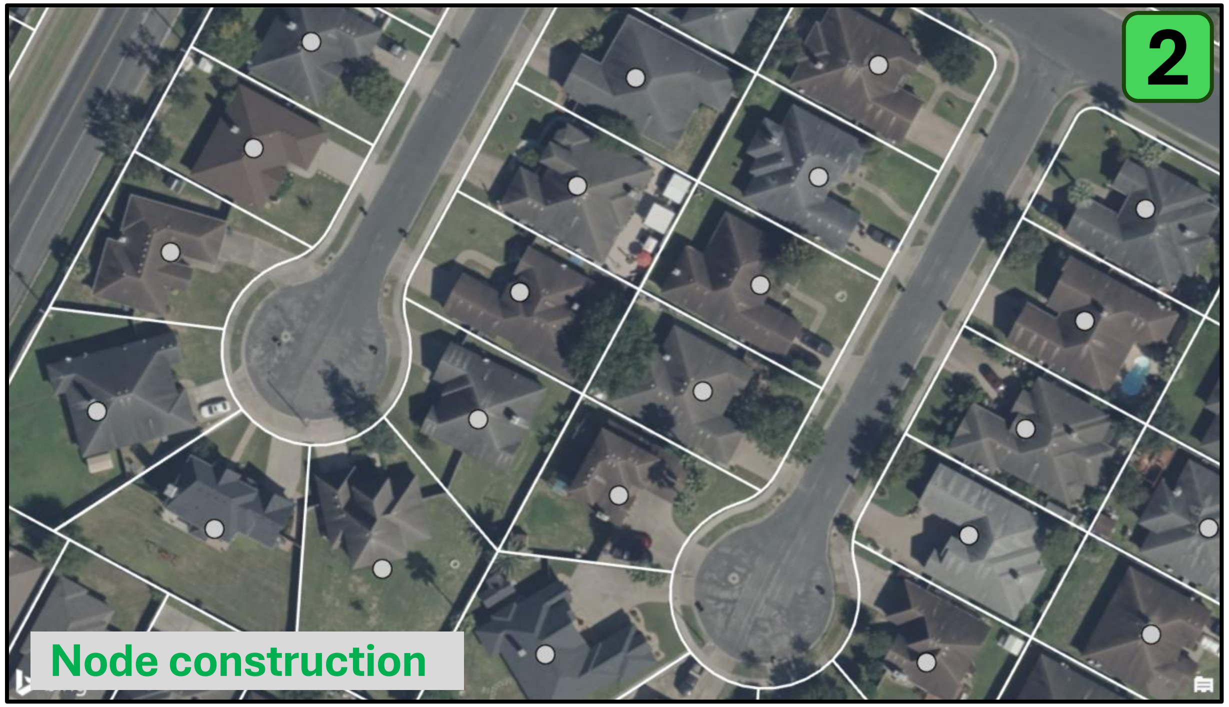
Datasets

Location: Corpus Christi, Texas, USA
Node Features: Spectral indices, Demographic, Land parcel data
Target Feature: Impervious surface ratio [0.00 - 1.00]

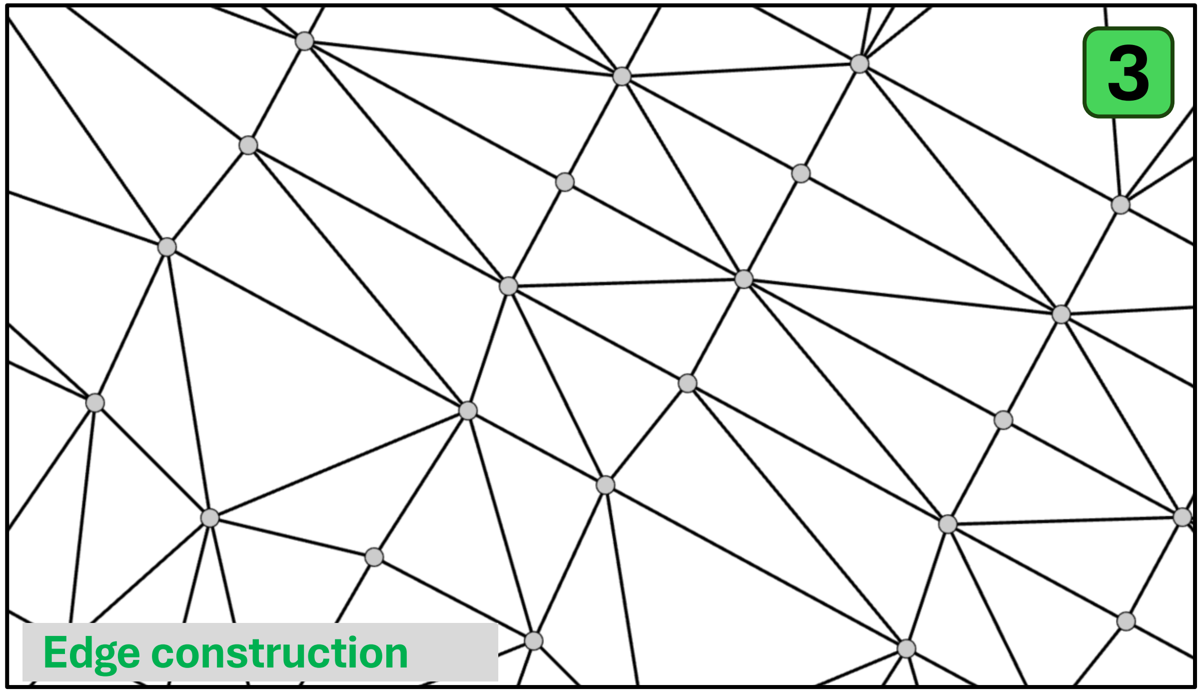
Data Overview:

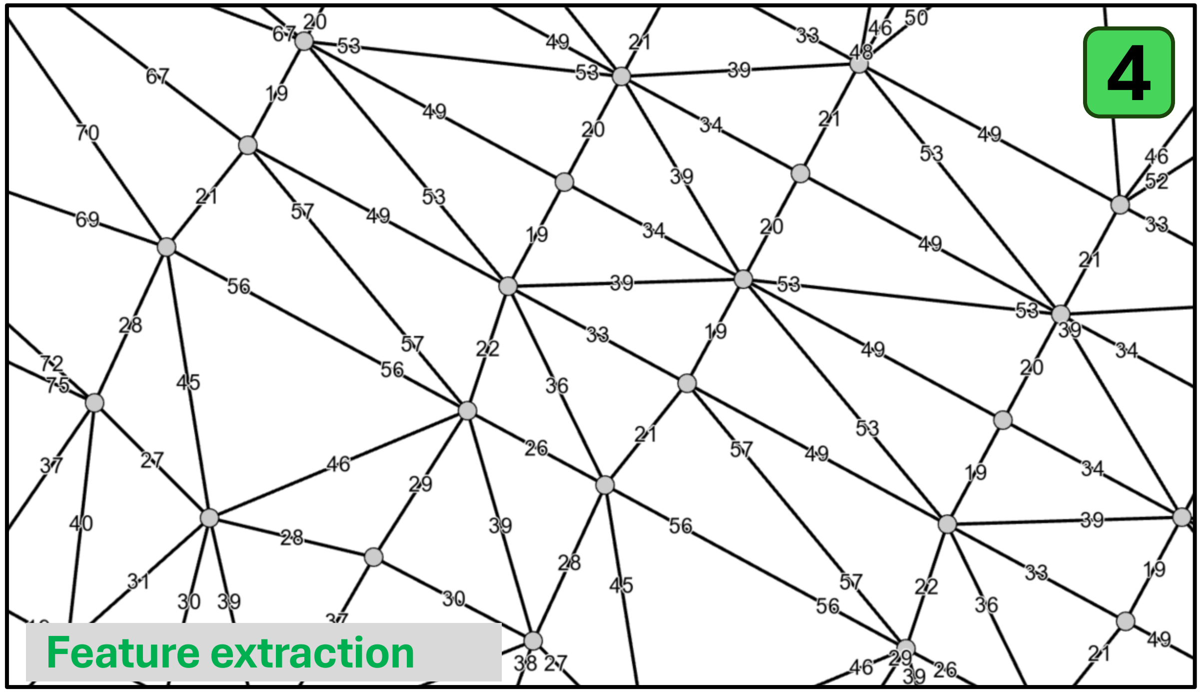
- Total parcels: 103,828
- Training set: 5,000 parcels
- Validation set: 1,000 parcels
- Test set: 97,828 parcels





Experiments



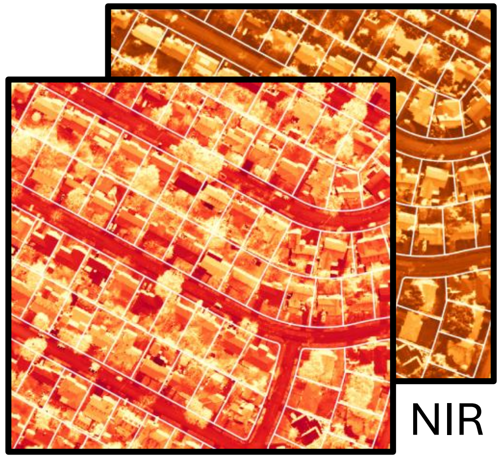


Experiments

Node features and their descriptions (32 spectral indices+ 8 demographic features)

Features	Variants	Description
NDVI	Mean, Std, Min, Max, Range, Sum, Median, Pct90	Vegetation index derived from visible and near-infrared light reflected by vegetation.
NDWI		Water index for moisture/liquid water content of soil and vegetation.
MTVI2		Enhanced vegetation index to correct soil and atmospheric influences.
VARI		Index for visualizing vegetation in RGB imagery.

Example: Mean of NDVI

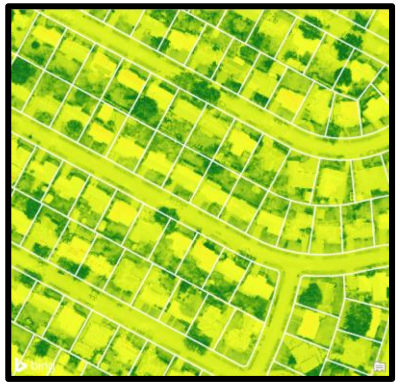


NIR

Red

NDVI

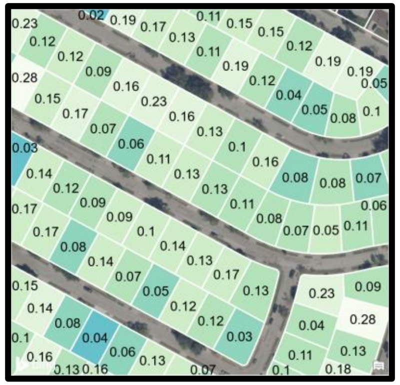
$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + \mathrm{RED}}$$



-1 1

Spatial statistics

Mean

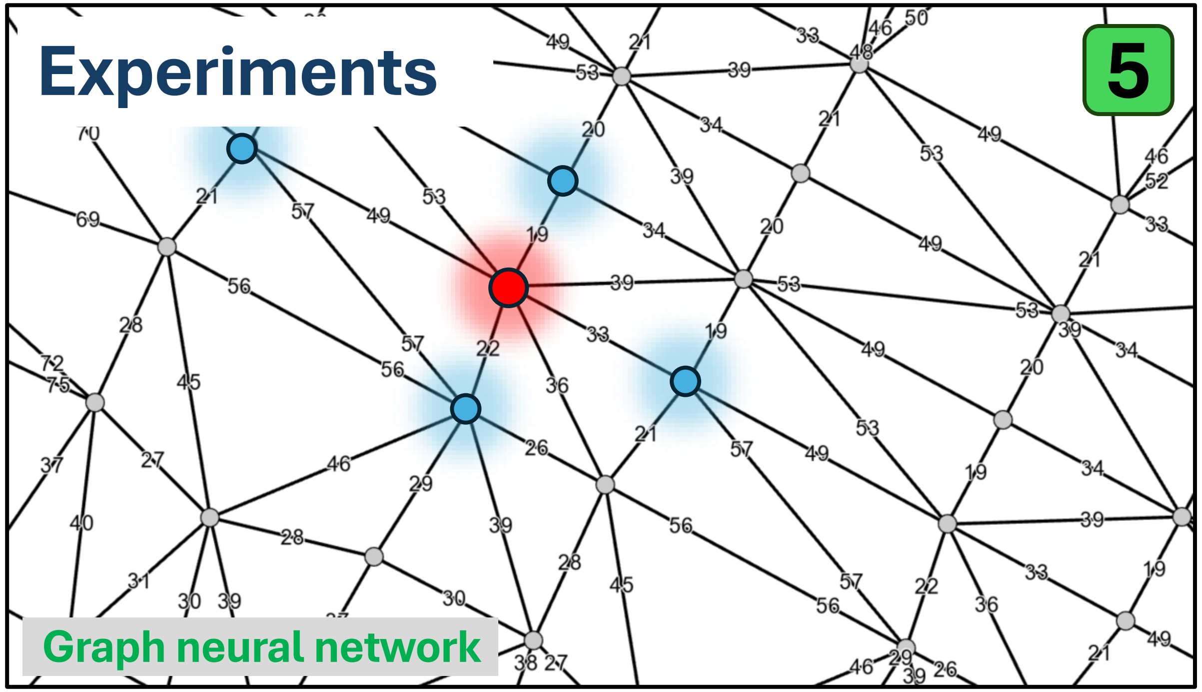


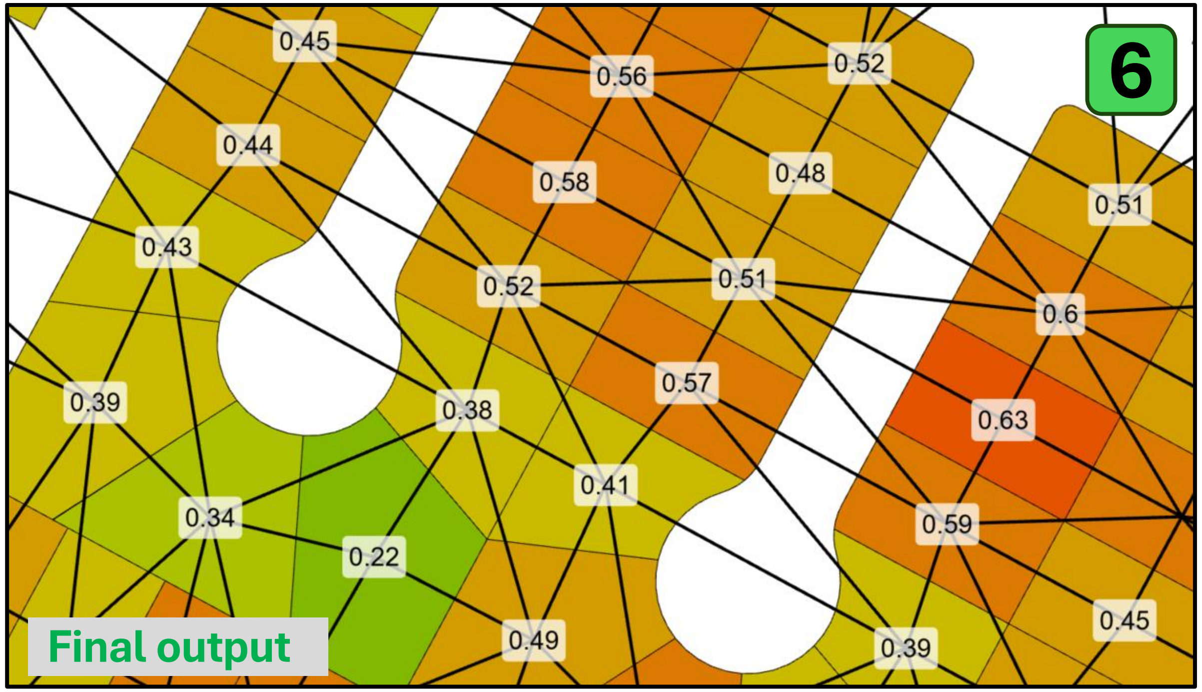
Experiments

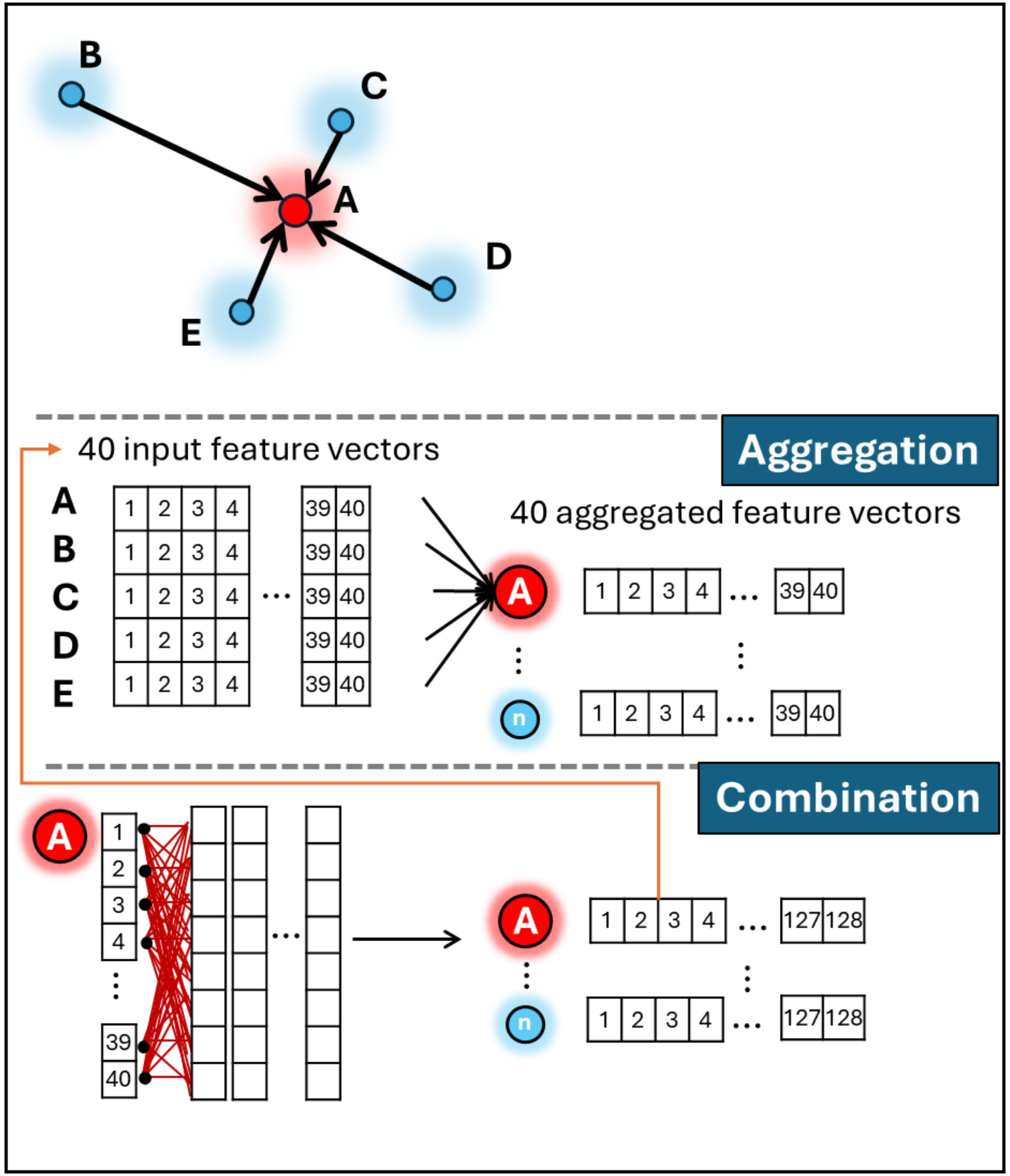
Node features and their descriptions (32 spectral indices+ 8 demographic features)

Features	Variants	Description
NDVI	Mean, Std, Min, Max, Range, Sum, Median, Pct90	Vegetation index derived from visible and near-infrared light reflected by vegetation.
NDWI		Water index for moisture/liquid water content of soil and vegetation.
MTVI2		Enhanced vegetation index to correct soil and atmospheric influences.
VARI		Index for visualizing vegetation in RGB imagery.
Demographics	Census track, Population density, Average home price, Average family income	Demographic information at block group level.
Coordinates	X, Y	Projected coordinates of the parcel.
Shape	Length, Area	Geometric measurements of the parcel shape.

Experiments







Evaluation (Training and Validation results)

Optimal Configuration:

- **Learning Rate**: 0.001
- **Layers**: 2 GraphConv
- **Hidden Dimension**: 128
- **Dropout Rate**: 0.5
- **Weight Decay**: 0.00001

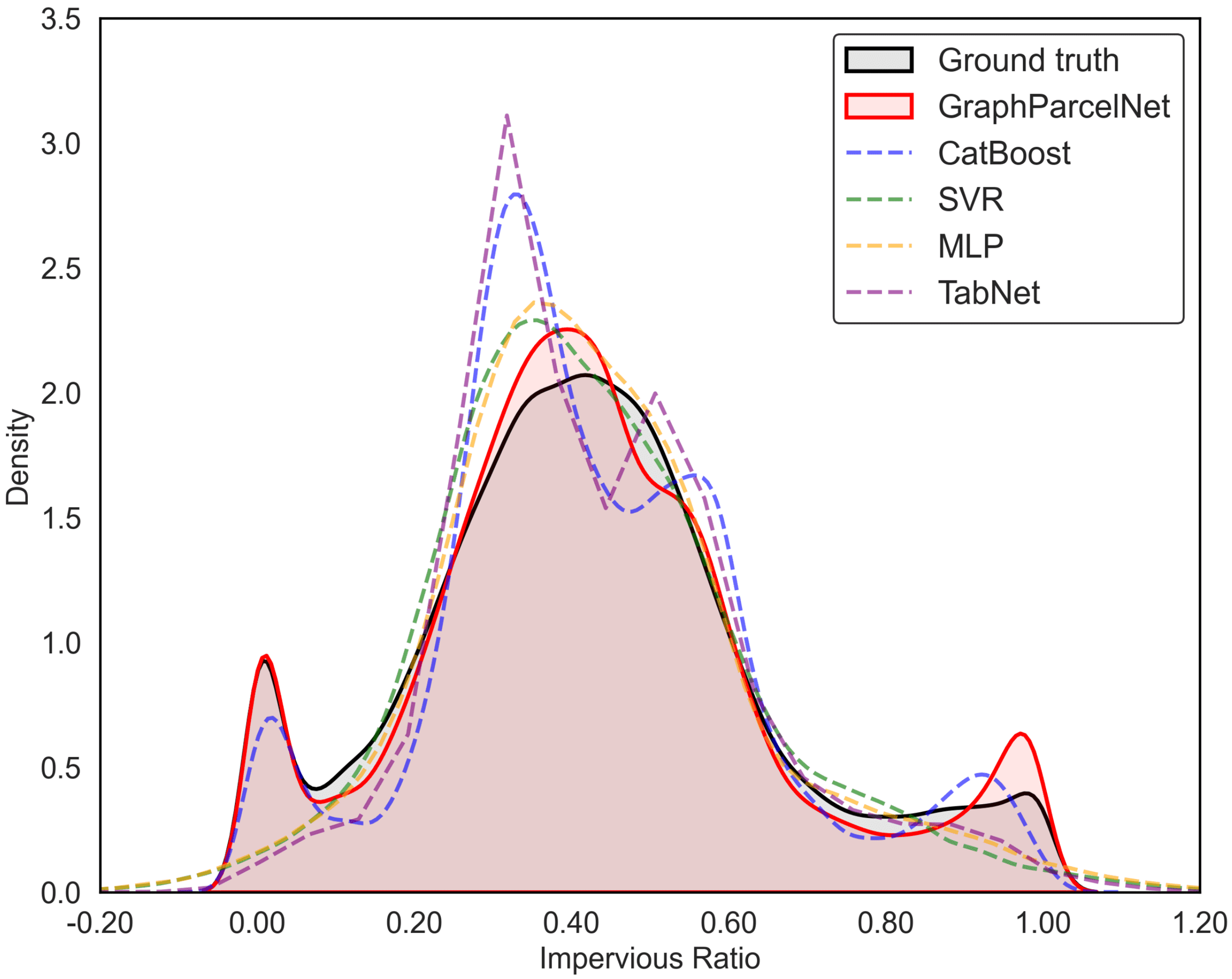
Performance (Averaged over 10 iterations):

- **Training Loss**: 0.0486
- **Validation Loss** (MAE): 0.0548
- Validation MSE: 0.0070

Model	Validation Loss MAE
TabNet	0.0806
MLP	0.0731
SVR	0.0732
CatBoost	0.0573
GraphParcelNet	**0.0548**

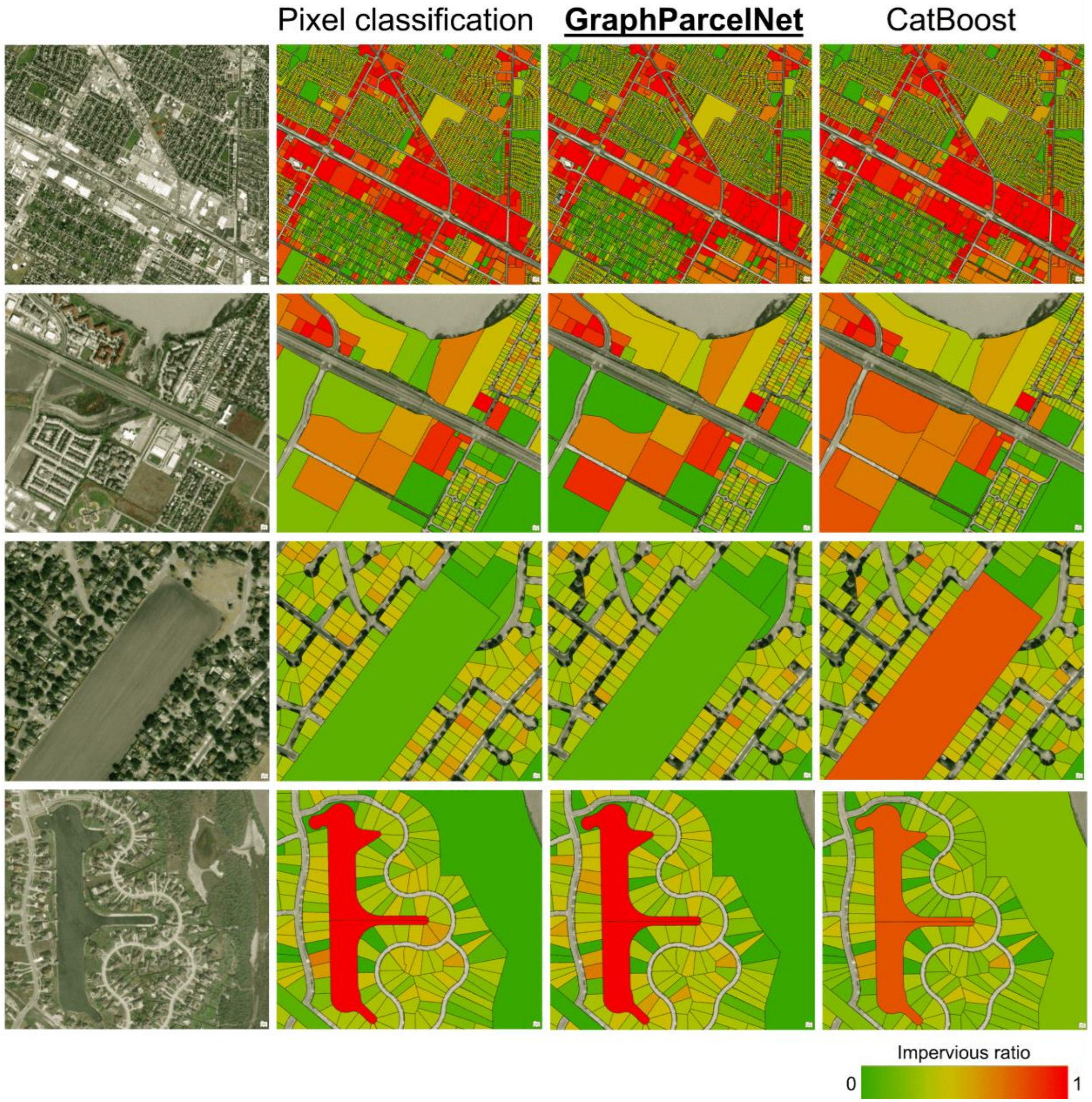
Evaluation (Testing result, 97,828 parcels)

Metric	TabNet	MLP	SVR	CatBoost	GraphParcelNet
MAE					✓
MSE				✓	
Median AE					✓
Quantile Loss				✓	✓
EMD					✓
KS Statistic					✓
Min					✓
Max					✓
Mean			✓		
Median		✓			✓
Std					✓
Mode					✓
IQR					✓
Skewness				✓	
Kurtosis					✓
CV					✓



(EMD: Earth Mover's Distance, KS Statistic: Kolmogorov-Smirnov Statistic, IQR: Interquartile Range, CV: Coefficient of Variation)

Evaluation (Testing result, 97,828 parcels)



Example 1:

- Both models capture high imperviousness in commercial zones (buildings, parking lots).

Example 2:

- Accurate residential predictions by both models.
- CatBoost: Misclassifies undeveloped land (orange) as impervious.

Example 3:

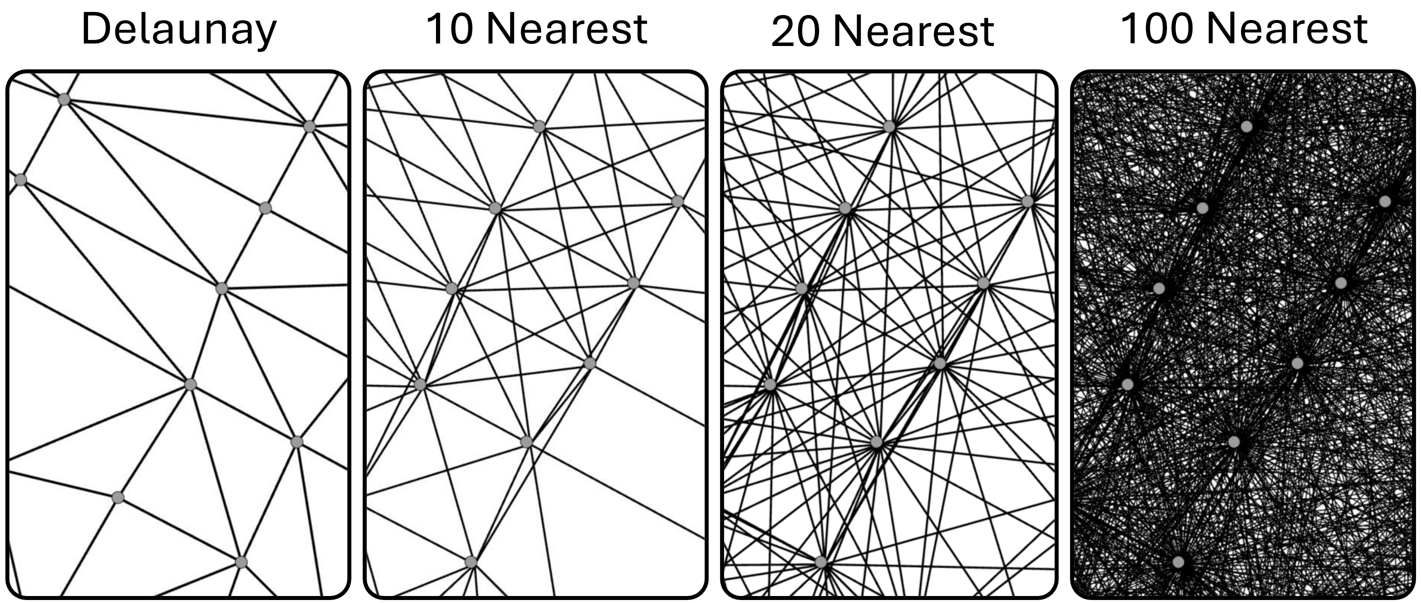
- CatBoost: Overestimated imperviousness of large undeveloped parcel, unrealistic values.

Example 4:

- GraphParcelNet: Correctly predicts 1.00 impervious ratio for retention pond.

Ablation Study

- Using alternative algorithms to construct graphs



312 km^2 study area size

Model	Graph Construction		Training/Validation		Testing		Evaluation	
	Time (s)	Memory (MB)	Time (s)	Memory (MB)	Time (s)	Memory (MB)	MAE	MSE
Vector Domain								
1. Delaunay	**132.20**	1254.69	**33.81**	**1127.00**	**0.07**	**1166.72**	**0.0548**	**0.0070**
2. 10 Nearest Neighbors	134.09	**1229.54**	64.18	1358.91	0.20	**1476.41**	0.0570	0.0076
3. 20 Nearest Neighbors	134.38	1240.91	107.53	1214.88	0.36	1426.82	0.0563	0.0073
4. 100 Nearest Neighbors	**141.10**	**1380.72**	**434.13**	**1423.29**	**1.53**	1400.41	**0.0608**	**0.0084**
Raster Domain								
Pixel Classification [15 cm pixel size]	-	-	36000	-	439200	-	-	-

Conclusions

1. **Efficiency:** GraphParcelNet significantly reduces prediction time from days to seconds, making vector-based prediction more efficient than traditional raster methods.

2. **Scalability:** Delaunay triangulation ensures fast, memory-efficient processing for large urban areas.

3. **Performance:** While closely matching pixel classification in accuracy, GraphParcelNet surpasses regression models such as CatBoost, MLP, SVR, and TabNet in most metrics.

4. **Spatial Insight:** GNNs effectively capture complex spatial relationships between parcels, improving prediction reliability and accuracy.

5. **Versatility:** The framework can be applied to other geospatial tasks, offering potential beyond impervious surface predictions and stormwater management.

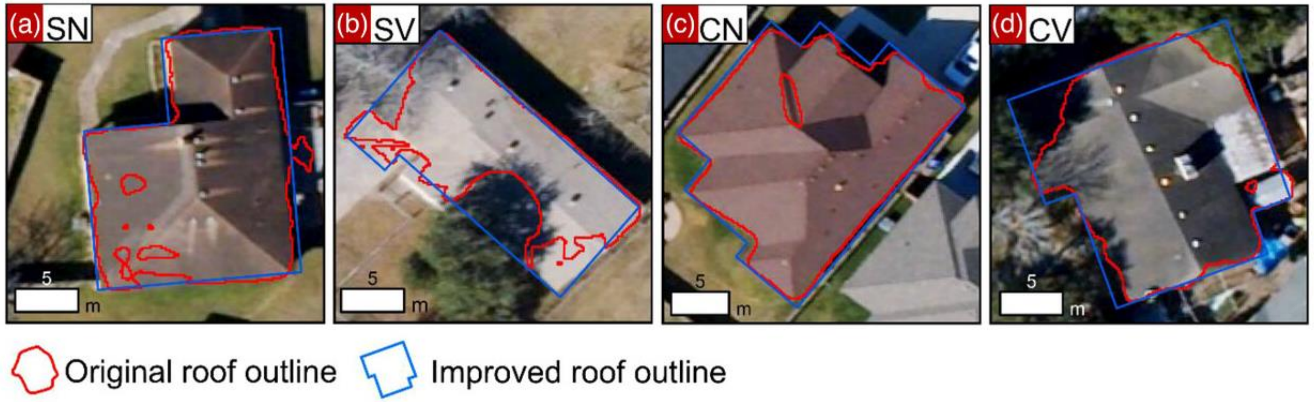
COMPUTER-AIDED CIVIL AND INFRASTRUCTURE ENGINEERING

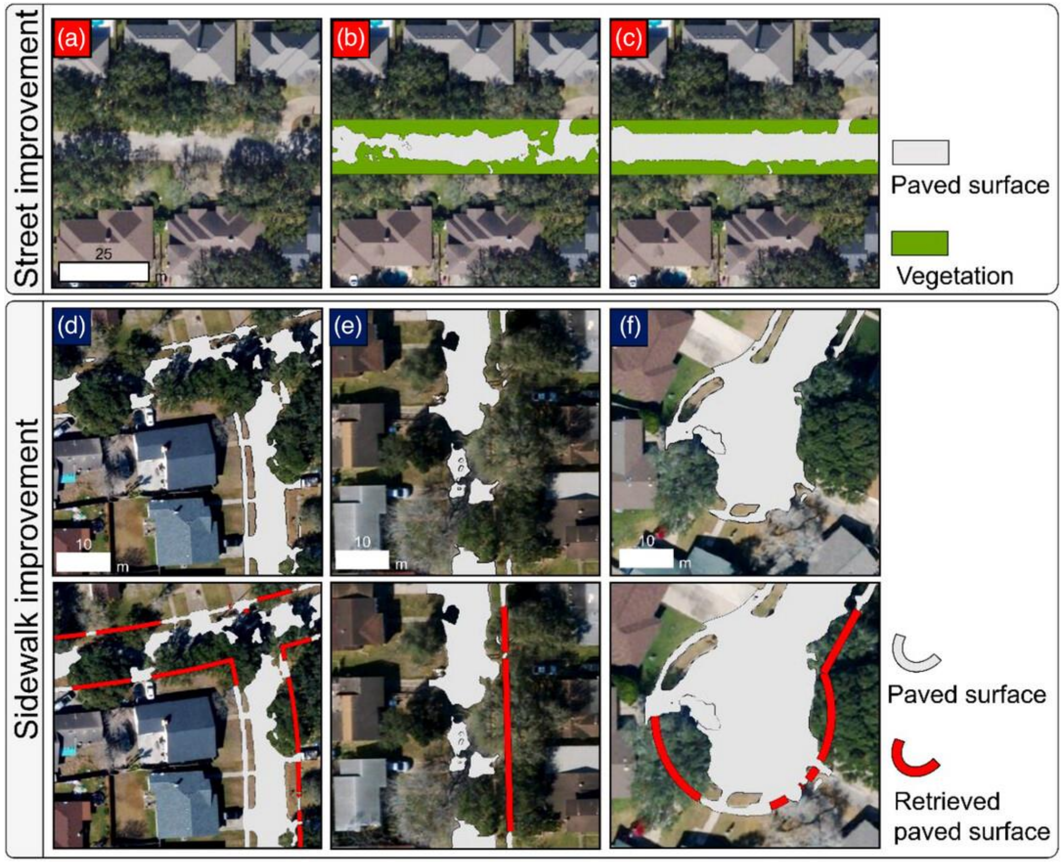
INDUSTRIAL APPLICATION | Open Access

Integrated urban land cover analysis using deep learning and post-classification correction

Lapone Techapinyawat, Aaliyah Timms, Jim Lee, Yuxia Huang, Hua Zhang

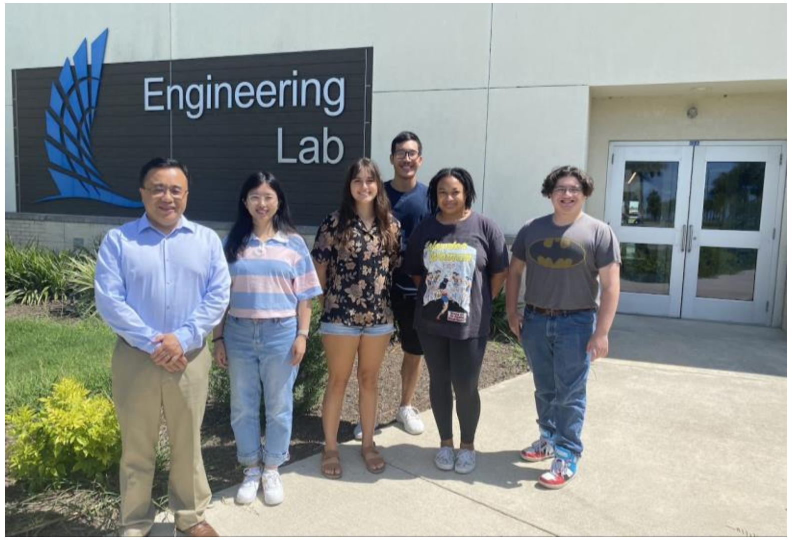
First published: 27 May 2024 | https://doi.org/10.1111/mice.13277 | Citations: 1





AGU24 Washington, D.C. | 9–13 December 2024 WHAT'S NEXT FOR SCIENCE

Abstract Number H21A-08
Session Number: H21A
Date & Time: Tuesday, 10 December 2024, 09:40 – 09:50 EST
Location: Convention Center, 140 A-B





Thank-you

More information:

https://laponet.github.io
www.linkedin.com/in/laponet
www.wesalab.com