Supplementary materials for

Estimating Building-Scale Monthly CO_2 Emissions of Different Functions: A Case Study of England

This part presents preprocessing procedures for selected datasets employed in this research, along with representative figures and tables illustrating key results.

The POI data used in this study were obtained from OpenStreetMap (OSM) and its open-source platform Overpass Turbo, from which over 700,000 records were extracted via API. To improve classification accuracy, we filtered out points of lesser importance or with minimal contribution to building classification (e.g., water towers), and subsequently reclassified the remaining points into 13 distinct categories. At the same time, to mitigate the impact of data uncertainty, we conducted strict filtering and manual inspection of the data. Additionally, the road network, urban center points, and water vector data were all sourced from the official OSM repository (https://download.geofabrik.de/, accessed February 20, 2025).

Table S1. Reclassification of the building functional labels used for this study and their descriptions.

	Building function	Description
1	Commercial	Buildings associated with commercial activities, such as retail and restaurant
2	Education	Buildings associated with educational purposes, such as schools and training centers
3	Industrial	Buildings associated with industrial production activities
4	Public	Buildings associated with public services and activities, such as churches and community center
5	Residential	Buildings associated with domestic purposes and encompass dwellings such as houses and apartments

Table S2. Reclassification of POI used for this study and their descriptions, which were sourced from OpenStreetMap (OSM) and Overpass Turbo, totaling over 700,000 records.

	POI Types	Description
1	Commercial	Locations related to commercial activities
2	Transport	Locations related to transportation for traffic flow, travel convenience, and logistics transportation
3	Public	Locations related to public functions and community activities
4	Sports	Locations related to sports activities
5	Live	Locations related to living purposes
6	Governmental	Locations related to government entities
7	Medical	Locations related to healthcare services, offering a variety of medical care
8	Scenery	Locations related to natural beauty and tourist attractions
9	Financial	Locations related to financial services
10	Company	Locations related to businesses and corporate entities
11	Cars	Locations related to the automotive services or car enthusiasts
12	Educational	Locations related to educational institutions and activities
13	Industrial	Locations related to entities for industrial production activities

Table S3. Features for classification and their calculation formulas, POI data were processed via kernel density analysis.

Shape Index of building footprint			POI	Distance
Shape	**Variable**	**Notation of Equation**		
Original Gragh	perimeter	PERIM_GE	Industrial Educational Car Company Financial Scenery Hospital Government Residential Sports Public Transport Commercial (500/1000m)	Disroad Discity Disbuilding Diswater
	Area	AREA_GEO		
	Fractality	$1-\frac{\log(\text{AREA_GEO})}{2*\log(\text{PREIM_GEO})}$		
	Cook JC	$\frac{\text{PERIM_GEO}}{\sqrt[4]{\text{AREA_GEO}}}-1$		
	Number of Vertices	PNT_COUNT		
Minimum Bounding Rectangle	Area	MBR_area		
	Perimeter	MBR_peri		
	Width	MBG_Width		
	Length	MBG_Length		
	Elongation	$\frac{\text{MBG_Length}}{\text{MBG_Width}}$		
Convex Hull	Compatchness	$\frac{4\pi*\text{AREA_GEO}}{\text{PERIM_GEO}^2}$		
	Concavity	$\frac{\text{AREA_GEO}}{\text{Convex Hull Area}}$		

Table S4. Results of Obtaining comprehensive night-time light using different algorithms

Month	Linear		XGBoost		Random Forest		DCT		GB		ANNs	
	R^2	MRSE	R^2	MRSE	R^2	MRSE	R^2	MRSE	R^2	MRSE	R^2	MRSE
July	0.768	4.049	0.748	4.223	**0.783**	**3.914**	0.629	5.106	0.764	4.084	0.780	3.944

Table S5. Classification results of different algorithms under various combinations of variables

variable combination		Number of variables	XGBoost		Random Forest		DCT	
			Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa
G1	Building Morphology	12	**0.776**	0.502	0.765	0.474	0.768	0.469
G2	Distance + Building Morphology	16	**0.828**	0.641	0.801	0.564	0.799	0.575
G3	POI	26	**0.836**	0.626	0.733	0.264	0.730	0.286
G4	POI + Distance	30	**0.873**	0.736	0.750	0.348	0.758	0.446
G5	POI+ Building Morphology	38	**0.875**	0.741	0.805	0.567	0.797	0.562
G6	POI + Distance + Building Morphology	42	**0.902**	0.802	0.819	0.602	0.820	0.618

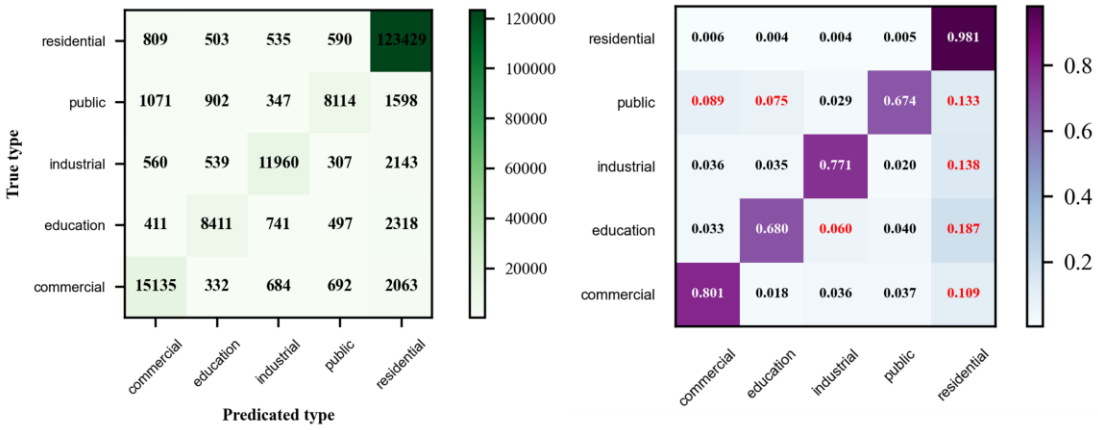


Figure S1. Confusion matrix and normalized confusion matrix of the model using the XGBoost algorithm

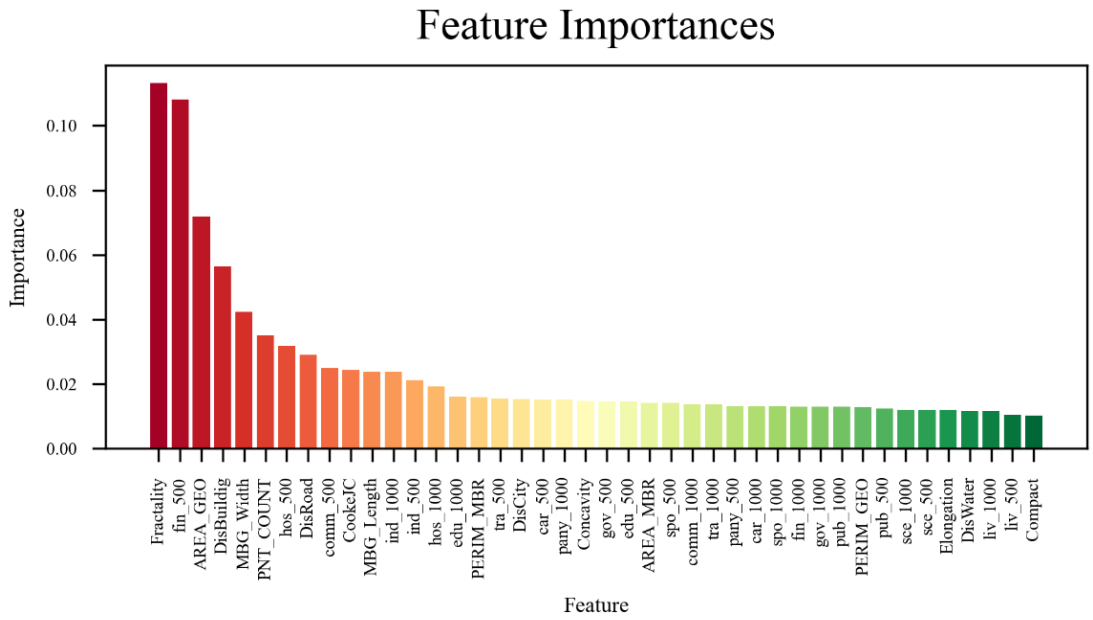


Figure S2. Feature importance plot of the model using the XGBoost algorithm