

19th May, 2026

Predicting Urban Heat Island (UHI) Intensity in the Indian Context

Problem Statement

● WHAT IS A UHI

- UHI is a micro-climatic phenomenon where urban areas experience significantly higher temperatures than surrounding rural zones.
- The Physics: Natural land (canopy, water) cools through evapotranspiration. Urban infrastructure (concrete, asphalt) absorbs, traps, and re-radiates solar energy

● BENGALURU CONTEXT

- Bengaluru is undergoing hyper-accelerated urbanization, rapidly trading its historical green canopy for dense, 3D concrete infrastructure. This creates localized pockets of extreme thermal stress.

● THE RESEARCH GAP

- Traditional UHI monitoring is purely reactive and 2-Dimensional. It relies on satellite maps that show where it is hot, but fail to explain why specific micro-climates exist at the street level.
- Current models ignore 3D urban morphology (street canyons, building volumes) and the human socioeconomic reality (who is actually exposed to the heat).

Potential Applications

01 Predictive Urban Planning

Planners can use this to test zoning scenarios: "If we approve a dense, 40-story commercial building in this grid cell, what is the exact mathematical penalty on the LST?"

02 Targeted Public Health & Climate Equity

By overlaying heat predictions with Meta's Relative Wealth Index (RWI), city officials can identify highly vulnerable, low-income heat traps.

03 Optimized Ecological Budgeting

This dictates exactly where planting vegetation will yield the highest mathematical cooling effect per dollar spent.

04 A Scalable Global Blueprint

Acts like a blueprint for other Indian metro cities that can be deployed across other rapidly developing cities

General brief of papers

Study	Location	Method	Key Features	Main Findings
Mansouri & Erfani (2025)	US Midwest	Random Forest, XGBoost	NDVI, imperviousness, demographics	Vegetation and impervious surfaces strongest predictors
Aqdas et al. (2025)	Ghaziabad, India	Remote sensing analysis	NDVI, NDBI, NDMI, LULC	Built-up expansion significantly increases UHI intensity
Deng et al. (2025)	Multiple cities	RF, XGBoost, OLS comparison	Land cover and spatial grids	Ensemble models outperform linear regression
Hoang & Nguyen (2025)	Da Nang, Vietnam	NGBoost, Deep Neural Networks	Urban morphology + land cover	ML models explain ~90% of UHI variation
Lynda et al. (2025)	African cities	Bayesian Neural Network + Gradient Boosting	Satellite + socio-economic variables	Hybrid models achieve extremely high predictive accuracy
Tanoori et al. (2024)	Shiraz, Iran	XGBoost, DNN	Urban configuration metrics	Green space structure strongly affects temperature
Manna et al. (2026)	Mumbai, India	Scenario analysis + thermal footprint modelling	LULC, greenspace, building materials, urban morphology	Nighttime thermal footprint driven by AC and traffic

Literature survey

US Midwestern United States Model

- **Goal:** Predict Urban Heat Island (UHI) severity across urban areas in 12 Midwestern United States
- **Dataset Used:** Global Climate Monitor, Multi-Resolution Land Characteristics Consortium, U.S. Census Bureau
- **69 Total features** divided into three categories
 1. **Meteorological Features** - Mean Annual temperature and rainfall
 2. **Land Use Land Cover Features** - eg: NDVI (vegetation index), Impervious surface percentage, Building height
 3. **Socio-Demographic Features** - eg: Population density, Age groups, Ethnicity
- **Missing values** handled using **KNN imputation** and there are **10,972** data points after cleaning.
- **For the UHI: The ideal configuration is LST_urb_all-LST_rur for the entire urbanized area**
- **Classification task**

UHI Level	Range	Distribution (%)
Negligible	$-1 \leq \text{UHI} < 1$	31.9%
Low	$1 \leq \text{UHI} < 3$	39.4%
High	$3 \leq \text{UHI}$	28.7%

Key insights and future work:

- The most important feature identified by both models was NDVI. This feature represents the difference in the **Normalized Difference Vegetation Index (NDVI) within an urban census tract compared to its surrounding rural reference area.**
- Key socio-demographic features such as pct_RURAL_POP, pct_URBAN_POP, and Population Density (PD) were also ranked among the most influential factors.
- **Future work/gap** : In future work the paper mentions investigate a **broader range of factors** impacting UHI, such as **urban form metrics**. Also there are **multiple types of vegetation indexes** which give a better reflection of the vegetation cover which **are not considered in this paper.**

Reference :

A. Mansouri and A. Erfani, "Machine Learning Prediction of Urban Heat Island Severity in the Midwestern United States," Sustainability, vol. 17, no. 13, p. 6193, Jul. 2025, doi: 10.3390/su17136193.

[Paper Link](#)

Literature survey

Ghaziabad Urban Heat Island Model

Goal: Analyze Urban Heat Island (UHI) intensity and thermal stress caused by urban expansion using remote sensing data.

Dataset Used: Multitemporal Landsat satellite imagery (1992, 2010, 2022) for Ghaziabad, India.

Extracted variables include:

1. Land Surface Temperature (**LST**)
2. Land Use Land Cover (**LULC**) classification (urban, vegetation, water, barren land)
3. Environmental features such as vegetation cover, built-up density, and impervious surfaces

Learning:

The study found a strong positive relationship between **built-up index (NDBI)** and temperature as well as LST

UTFVI	Urban Heat Island Phenomenon	Ecological Evaluation Index (EEI)			Proportion of Area in Percentage		
					1992	2010	2022
<0	None	Excellent			81.3942	80.4438	83.4669
0.000-0.005	Weak	Good			0	14.7564	0
0.005-0.010	Middle	Normal			0	0	8.4681
0.015-0.015	Strong	Bad			13.0698	13.608	8.3187
0.015-0.020	Stronger	Worse			0	0	8.4015
>0.020	Strongest	Worst			83.5731	69.2289	69.381

Reference: M. Aqdas, T. M. Usmani, R. Benhizia, and G. Szabó, "Urban Expansion and Thermal Stress: A Remote Sensing Analysis of LULC and Urban Heat Islands in Ghaziabad, India," *Land*, vol. 14, no. 9, p. 1893, Sep. 2025, doi: 10.3390/land14091893.

Literature survey

Integrating Remote Sensing and Machine Learning for Predictive Analysis of Urban Heat Island Dynamics:

(This is a review paper which helps summarize ML methods used in UHI Prediction)

What the paper discusses

- Integration of remote sensing and machine learning to analyze urban heat islands.
- Key predictors like NDVI, built-up density, albedo, and moisture.
- Comparison of ML models such as Random Forest, XGBoost, SVM, and ANN for predicting UHI.

M. B. Suthar, "Integrating Remote Sensing and Machine Learning for Predictive Analysis of Urban Heat Island Dynamics," *Environmental Reports: an International Journal*, vol. 3, no. 1, p. 10, Oct. 2025, doi:10.51470/ER.2021.3.1.10.

[Paper Link](#)

Overall Gaps and future work possibilities we found:

- Include and increase **urban morphology indicators** (Building Height, Street Geometry)
- Increase **socio-economic variables** as well as Human Activity Indicators
- Combining **remote LST with in-situ air temperature measurements**

Literature survey

Three recurring problems:

1. Geographic blind spot: no Bengaluru study
2. Single-modality features : most papers use only RS or only Urban features, never RWI + 3D morphology + RS jointly
3. Weak validation : single 70/30 splits inflate R^2 in many papers

Literature Gap	Our Fix
No Bengaluru ML UHI study	First one (2,433-point dataset)
No paper jointly uses RWI + OSM 3D + RS	20-feature joint set (RWI #3 in SHAP)
Single 70/30 splits	Stratified 5-fold + Spatial Block CV
No interpretability beyond RF importance	TreeSHAP + ablation (use vs necessity)

DATASETS USED

- **Google Earth Engine**
 - **Landsat 8** (Indices, LST)
 - **ERA5 Reanalysis** (Climate Data)
 - **Sentinel-5P TROPOMI** (Pollution)
 - **WorldPop** (Population Density)
- **MODIS Terra LST** (Long-term, Coarse LST)
- **OpenStreetMap** (Urban Morphology Features)
- **Meta AI Relative Wealth Index** (Wealth)
- **ESA WorldCover** (Land Use / Land Cover)



- **2433 Datapoints**
- **27 Total Features**

- **10km x 10km area over Central Bengaluru**
- **Resolution of 50m x 50m**

- **Target Variable: LST**
- **Single Landsat dry-season pass**

As classification (*Liu & Zhang (2011) Mean-SD method · 3 classes*)

Thresholds: $\mu \pm 0.5\sigma$ (Mean: 33.78 °C | Std Dev: 2.72 °C)

Classification	LST Range	Pixel Count
 Cool Island	< 32.42 °C	526
 Neutral	32.42 – 35.14 °C	1,173
 Hot UHI	> 35.14 °C	734

FEATURES

Vegetation / Surface Indices

- **NDVI - Normalized Difference Vegetation Index**

Measures vegetation health and density using near-infrared and red spectral bands.

- **EVI - Enhanced Vegetation Index**

Improved vegetation index that reduces atmospheric and soil background noise.

- **SAVI - Soil Adjusted Vegetation Index**

Vegetation index that corrects NDVI for soil brightness in sparsely vegetated areas.

- **NDBI - Normalized Difference Built-up Index**

Detects built-up urban surfaces using shortwave infrared and near-infrared bands.

- **MNDWI - Modified Normalized Difference Water Index**

Identifies water bodies by enhancing water reflectance while suppressing built-up land.

- **Albedo**

Fraction of solar radiation reflected by the Earth's surface; higher albedo means more reflection and less heat absorption.

Climate Variables

- **Air_Temp_C - Air Temperature (°C)**

Ambient atmospheric temperature measured near the surface.

- **Relative_Humidity (%)**

Percentage of moisture in the air relative to the maximum possible moisture at that temperature.

- **Wind_Speed (m/s)**

Speed of horizontal air movement, which influences heat dispersion.

- **Surface_Pressure_hPa - Surface Atmospheric Pressure**

Pressure exerted by the atmosphere at ground level, measured in hectopascals.

- **Soil_Moisture**

Amount of water present in the soil, affecting evaporation and cooling.

- **MODIS_LST - MODIS Land Surface Temperature**

TemperatureLand surface temperature measured by NASA's MODIS satellite sensor.

- **LST - Land Surface (Target)**

Temperature of the Earth's surface derived from thermal satellite imagery.

FEATURES

Geospatial Features

- **Land Cover**

LULC - Land Use Land Cover

Categorical classification of surface types such as vegetation, water, or built-up land.

One-hot encoded LULC features

- **Dist_Water - Distance to Water Bodies**

Distance from each point to the nearest lake, river, or reservoir.

- **Dist_Park - Distance to Green Spaces**

Distance to the nearest park or urban green area.

- **Dist_Highway - Distance to Major Roads**

Distance to the nearest highway or major traffic corridor.

- **Dist_to_RWI_Node - Distance to Relative Wealth Index Node ★**

Distance to the nearest location where socioeconomic wealth index data is measured.

- **Coordinates - Latitude and Longitude**

Urban Morphology ★

- **Building_Density_Ratio**

Ratio of total building footprint area to total land area within a region.

- **Volume_Density**

Measure of three-dimensional urban density based on estimated building volume.

- **Height_Variability**

Statistical variation in building heights within a spatial area.

- **SVF_Proxy - Sky View Factor (Proxy)**

Approximation of the fraction of visible sky from ground level, indicating urban canyon effects.

- **Street_Density**

Total road length per unit area, representing transportation network intensity.

Socioeconomic / Human Features

- **NO2_Emissions - Nitrogen Dioxide Concentration ★**

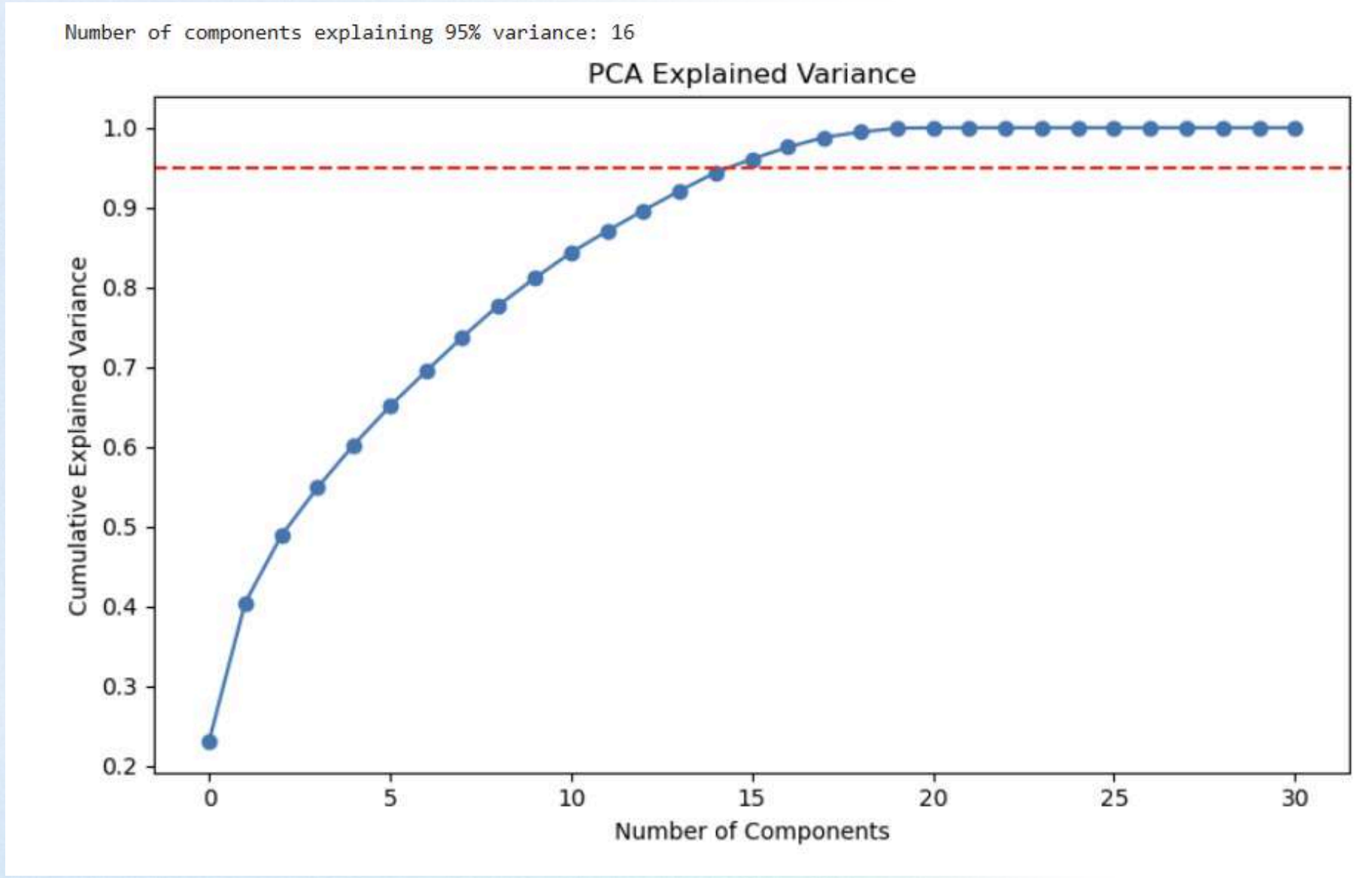
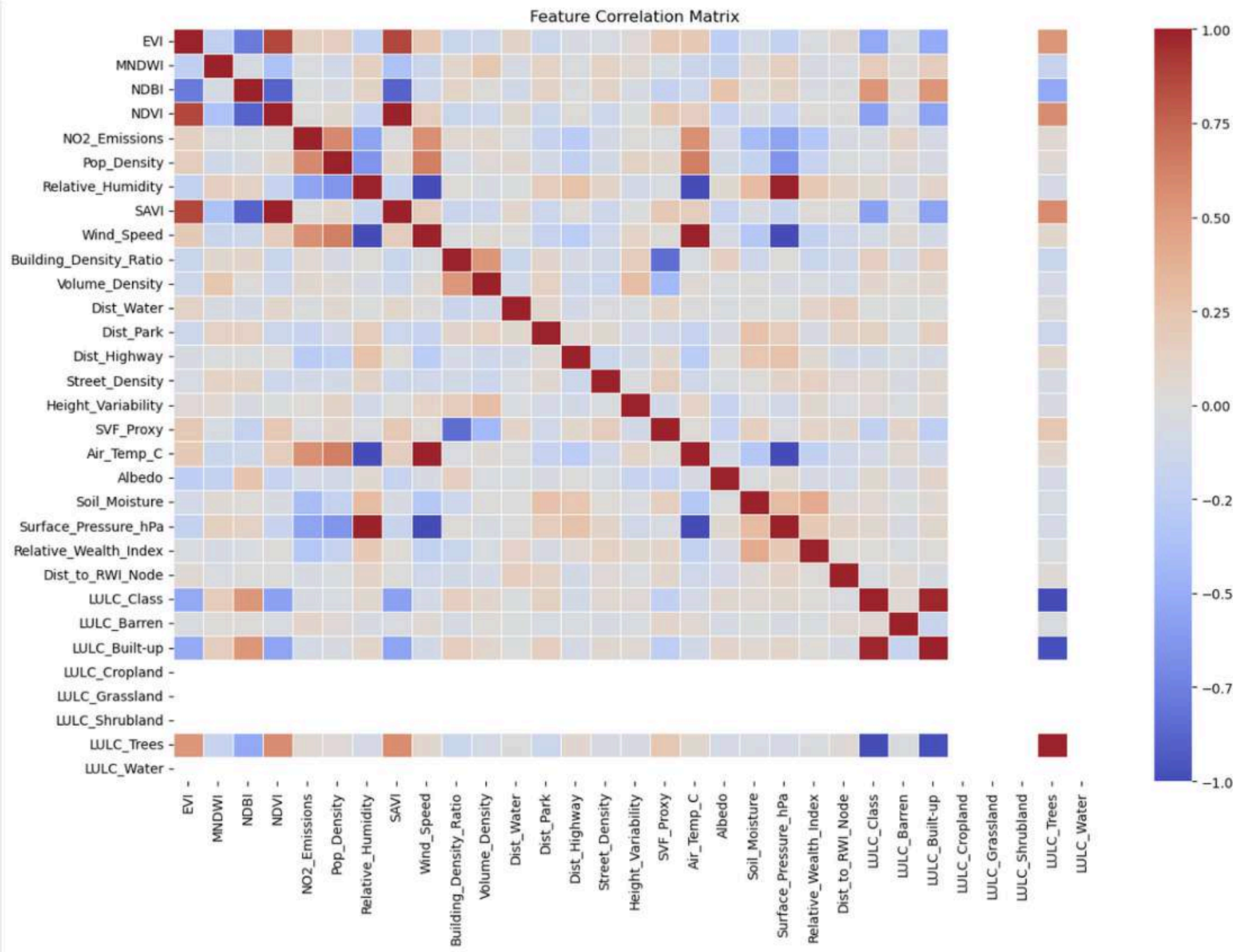
Atmospheric concentration of NO₂, often associated with traffic and industrial activity. - used as a stand-in for human activity

- **Pop_Density - Population Density**

Number of people per unit area, indicating human concentration.

- **Relative_Wealth_Index (RWI) ★**

A spatial socioeconomic indicator estimating relative economic wealth based on satellite imagery and machine learning.



FEATURES PRE-PROCESSING

Features Cleaning

- Removed Duplicates
- Approximated building height via number of floors
- KNN Imputation

Features Standardization

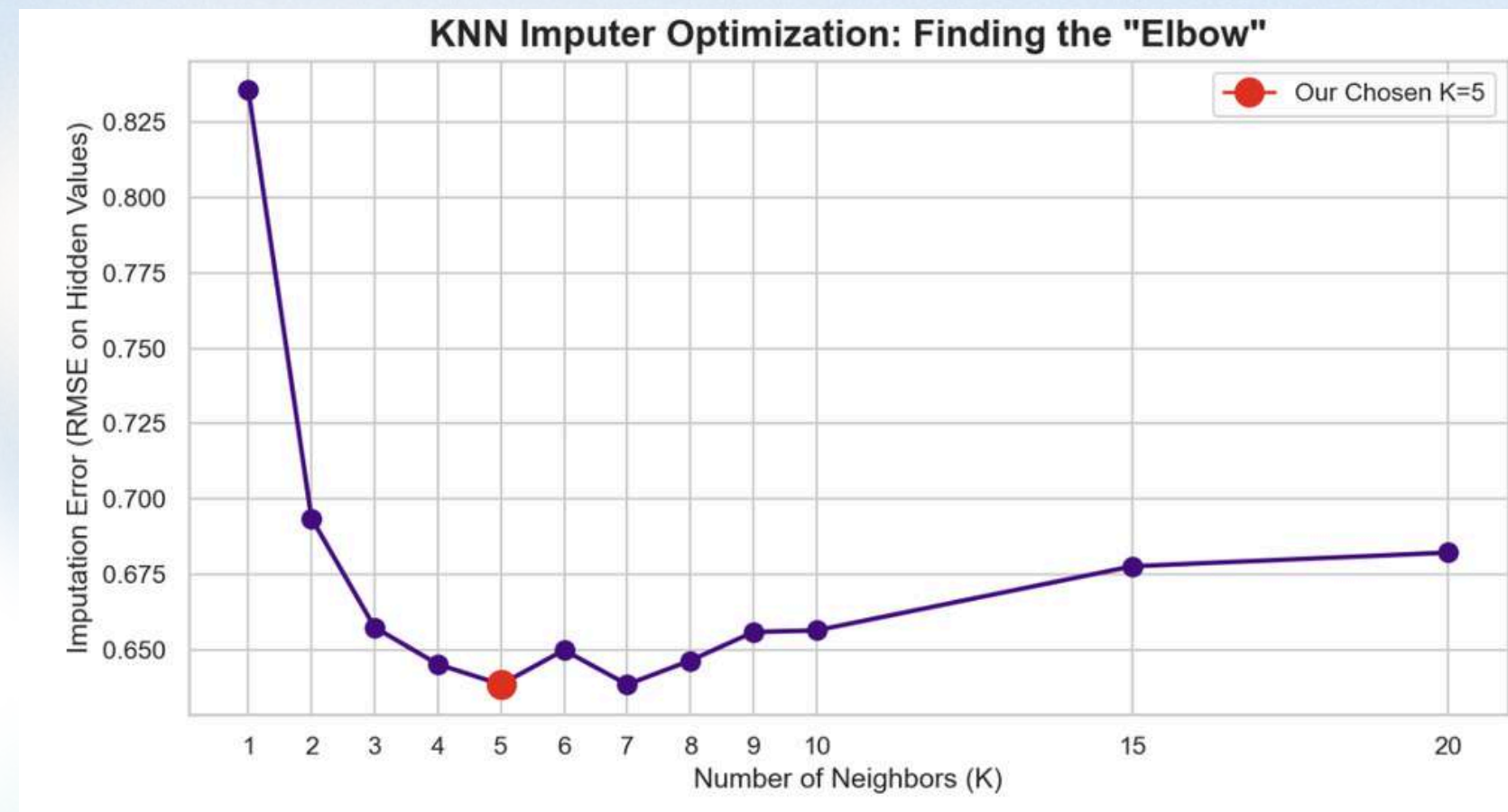
- StandardScaler

Features Encoding

- LULC_Class encoded using One-Hot Encoding

Engineered Features

- Building Density Ratio = Total Building Area / Total land Area
- Street Density = Total Road Length / Area
- Volume Density = Building Heights & Footprints
- Height Variability = Std Dev of Building Heights
- Dist Water, Park, Highway - calculated using geospatial data
- LULC Features Encoded via One-Hot Encoding



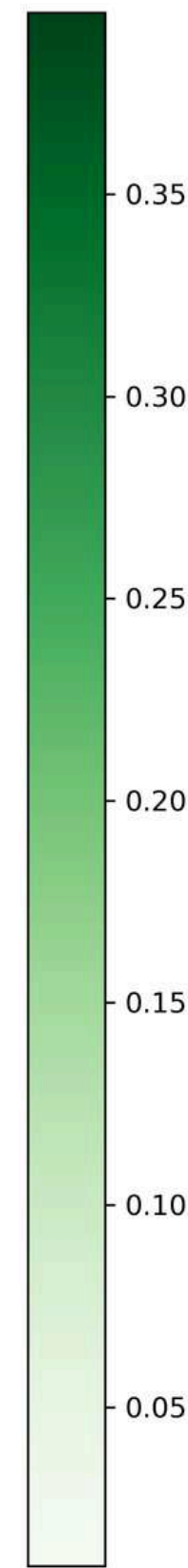
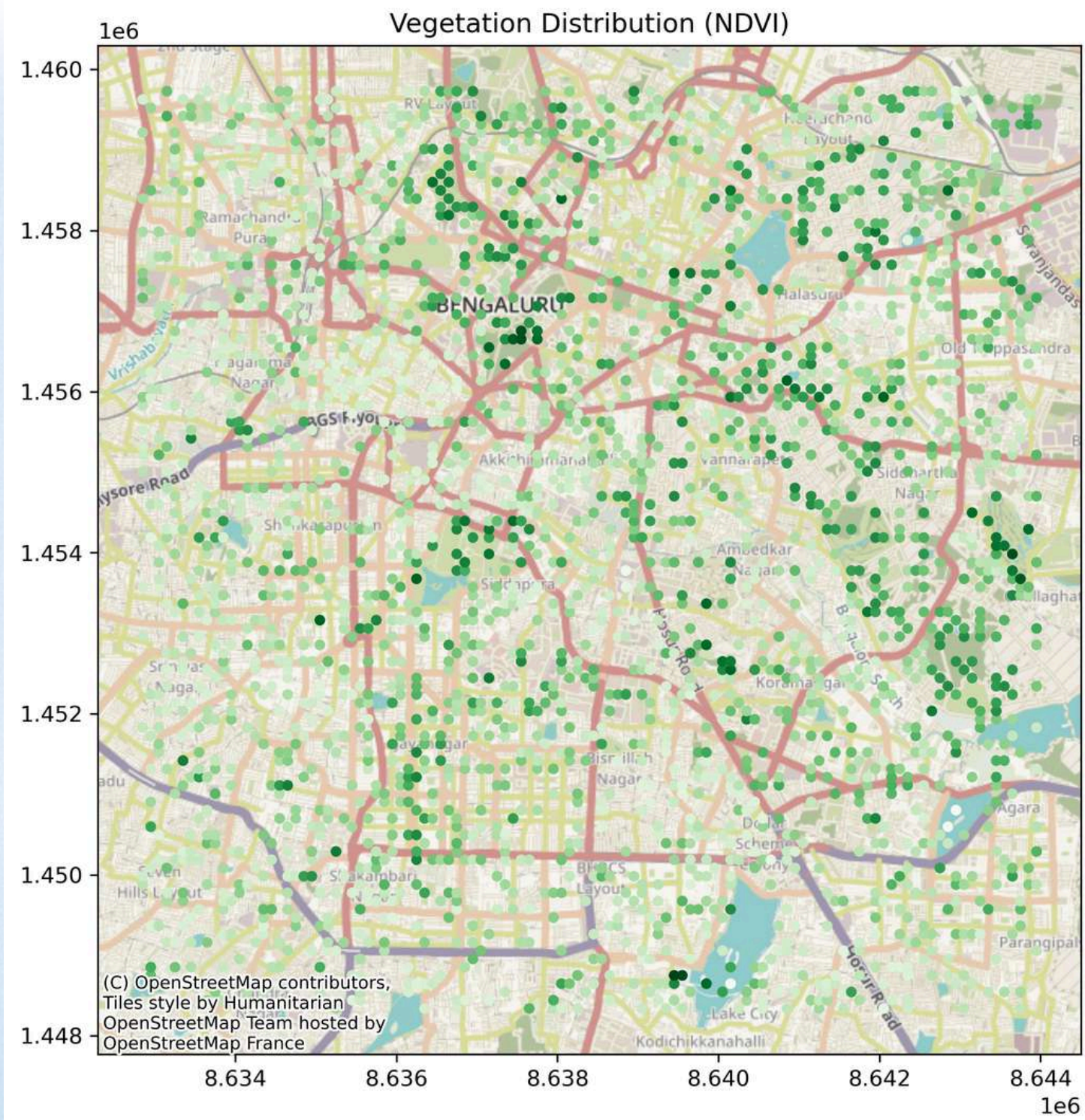
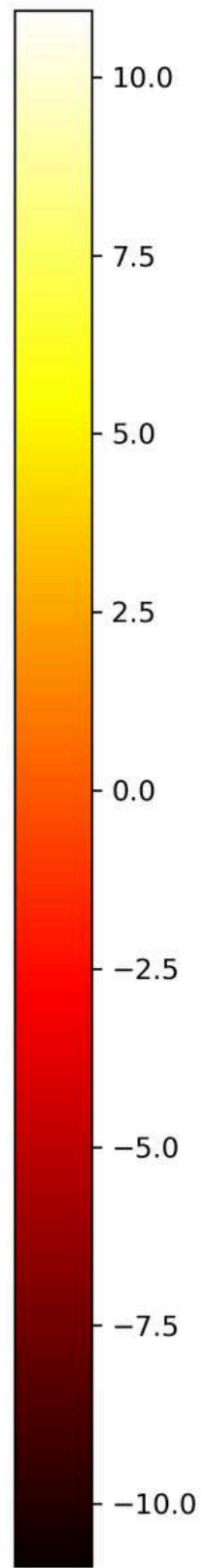
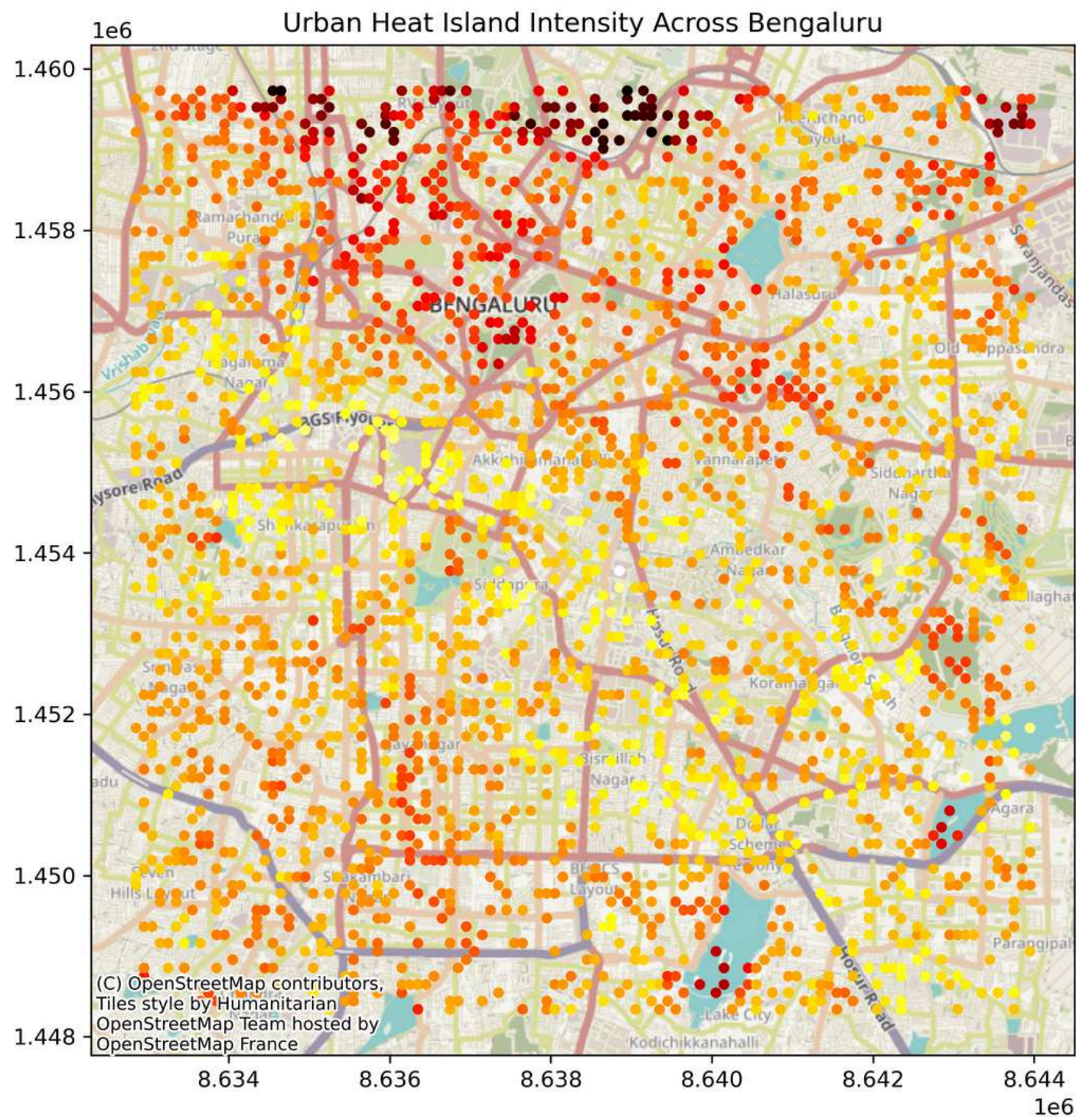
Features Dropped

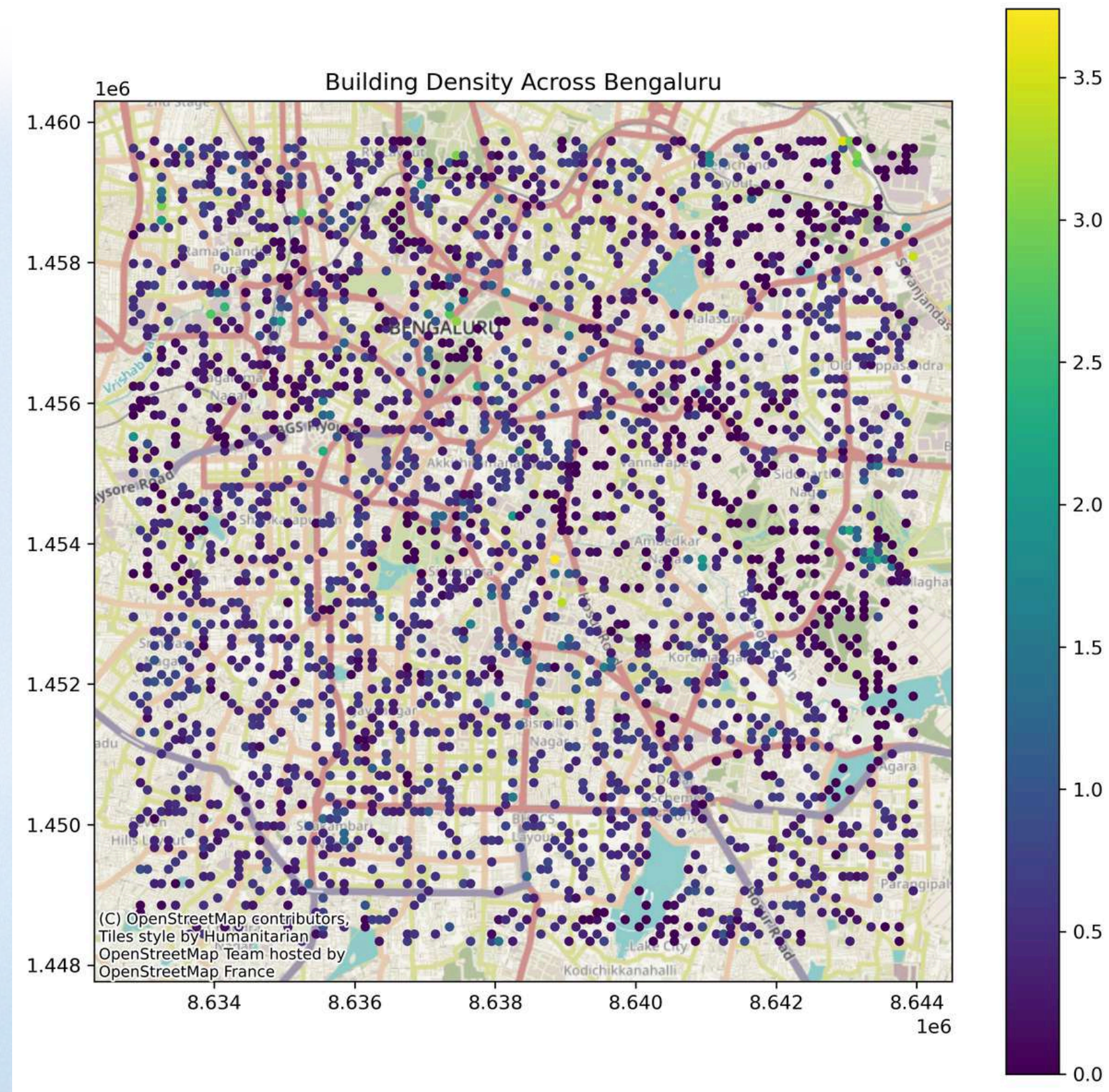
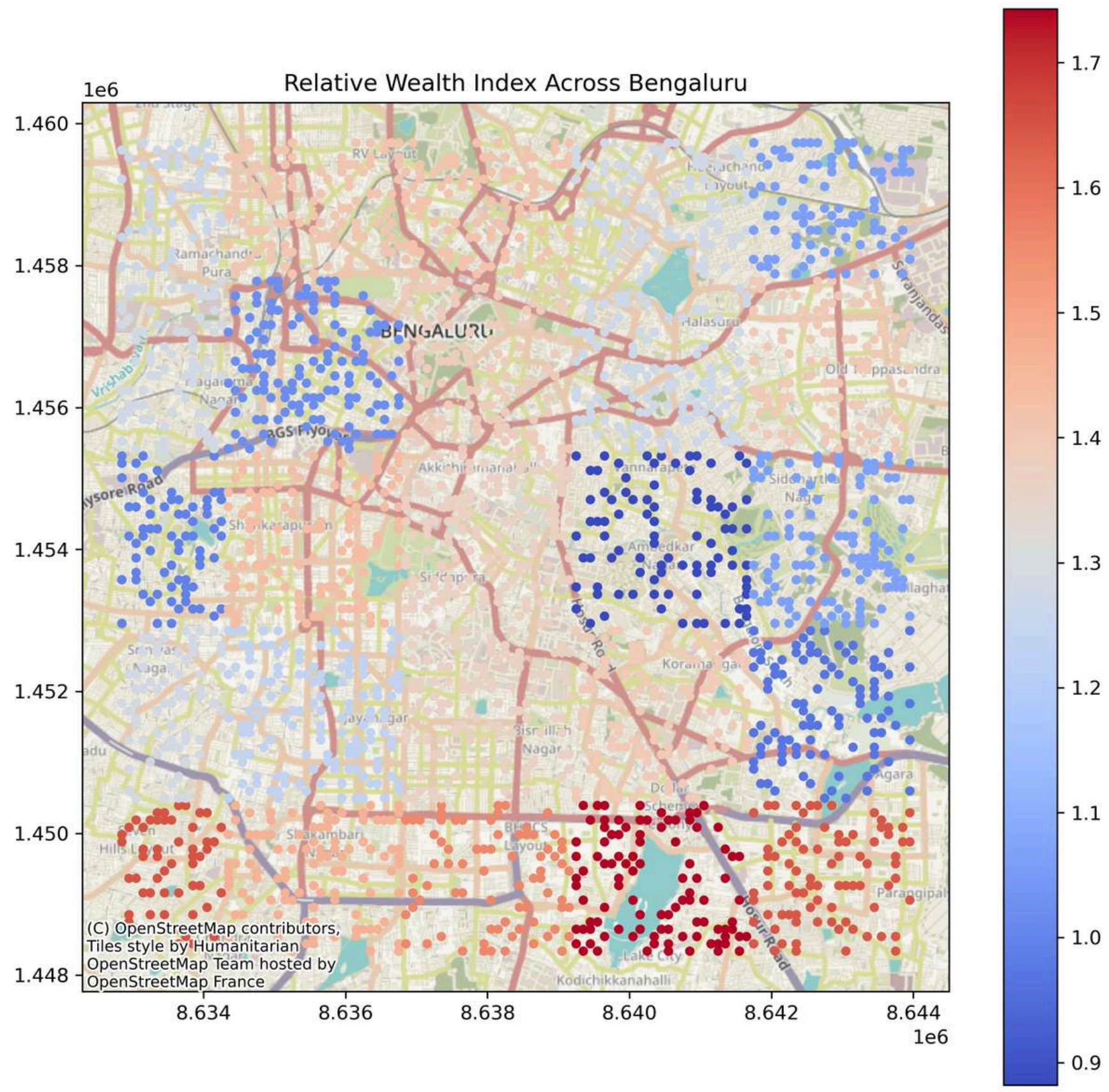
- LST - Regression target
- Latitude, Longitude
- EVI, SAVI, MNDWI - Redundant with NDVI / NDBI (multicollinearity)
- Volume_Density - 84.8 % missing

No PCA needed: tree ensembles handle correlated features natively

KNN imputation (k=5) for residual missing values

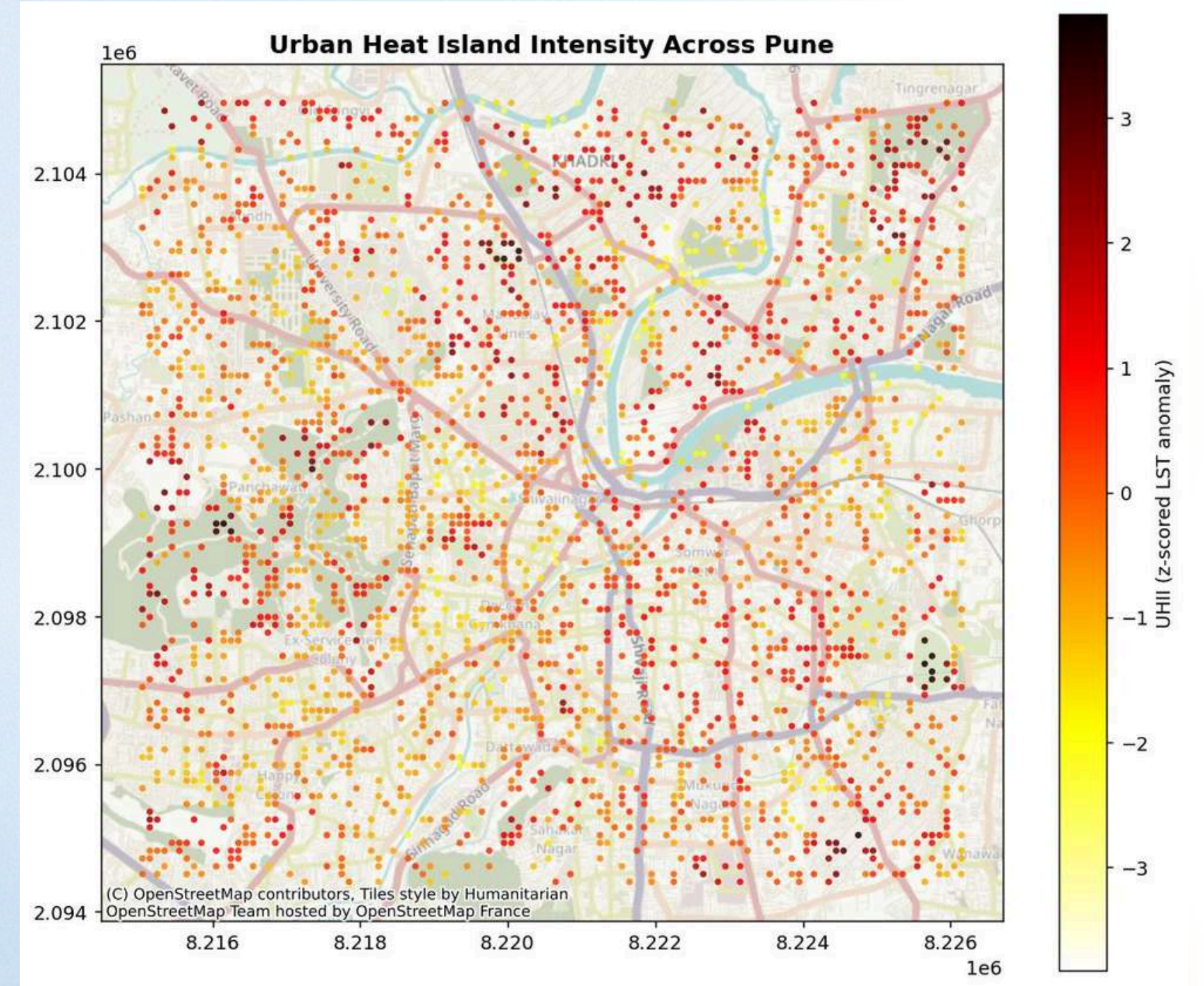
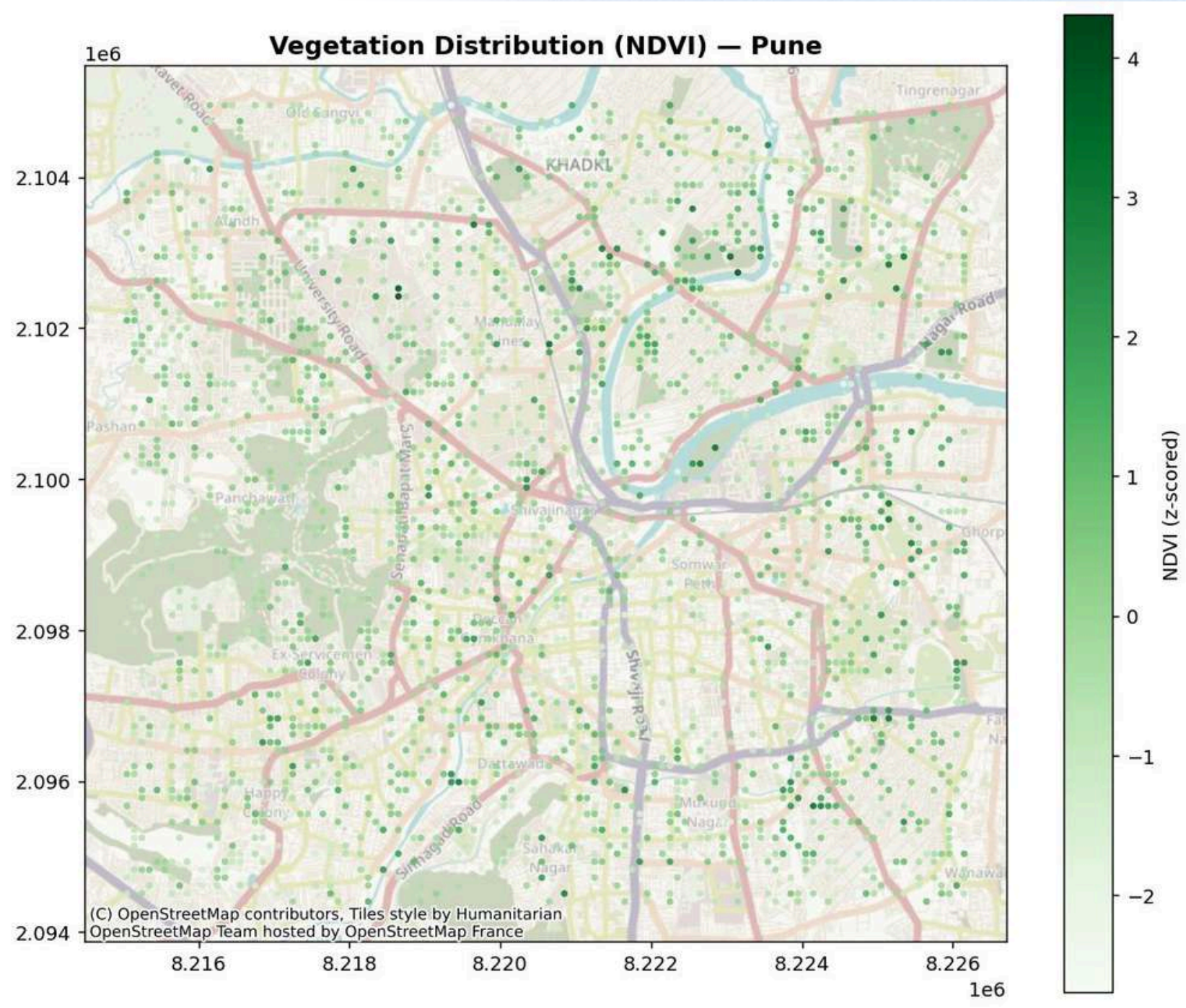
StandardScaler only for linear / KNN / MLP; tree models see raw features





EXTENSION - PUNE

2,382 samples, same pipeline



Experimental Journey

01 Initial Findings

Single 80/20 split, 27 features,
Random Forest. $R^2 = 0.6943$.

02 Spatial Block CV

tests geographic generalisation (UHI is
spatially autocorrelated).

03 Switch to Stratified 5-fold and Optuna

R^2 jumps to 0.7869 with tuned
XGBoost

04 Reduce to 20 features, ensemble 3 boosters

$R^2 = 0.7940$ (XGBoost + LightGBM +
CatBoost averaged)

05 Cross-city validation (Pune)

combined-city $R^2 = 0.7497$

Model Lineup

Family	Model	Literature reference	R ²	RMSE (°C)
Linear	Linear Regression	Floor baseline	0.4241	2.062
Distance-based	KNN (k=7, distance-weighted)	Manna (2026)	0.5985	1.721
Neural	MLP (64 → 32, ReLU)	Lynda (2025), Tanoori (2024)	0.5066	1.901
Tree (bagged)	Random Forest (n=500, d=12)	Aqdas (2025), Hoang (2025)	0.6943	1.501
Tree (boosted)	XGBoost (tuned)	Aqdas (2025), Hoang (2025)	0.7869	1.254
Tree (boosted)	LightGBM (tuned)	Aqdas (2025)	0.7855	1.258
Tree (boosted)	CatBoost (tuned)	Aqdas (2025)	0.7875	1.252
Ours	Ensemble (XGB + LGBM + CatBoost)	-	0.794	1.233

Cross-Validation Strategy

Stratified 5-fold CV (our choice of CV)

- Splits the 2,433 Bengaluru points randomly but balances each fold by LST tertile (Cool / Neutral / Hot proportions equal across folds).
- 80% train / 20% test per fold, 5 estimates averaged. Low variance, balanced classes.
- Chosen as headline because it is the published standard for UHI ML papers and gives the most stable mean estimate.

Spatial Block CV (robustness check)

- K-Means (k=10) clusters all 2,433 points into 10 geographic blocks. Each fold trains on 9 blocks and tests on the held-out region.
- Test points share no spatial neighbourhood with training points, killing autocorrelation leakage.
- Result: $R^2 \approx 0.72$ to 0.74 (vs 0.7940 stratified). Small drop confirms the model learns physics, not coordinates.

What we rejected

- Single 80/20 random split: inflated by spatial autocorrelation. Not honest.
- Leave-One-Out CV: computationally prohibitive at $n = 2,433$, plus very high variance.

Final Pipeline Architecture

8 open data sources → raw 27-feature CSV (2,433 points, 50 m grid)

Drop 7 columns (target, lat/lon, redundant indices EVI/SAVI/MNDWI, sparse Volume_Density)

KNN impute (k = 5). StandardScaler applied only for MLP and distance models

Stratified 5-fold split balanced on LST tertiles

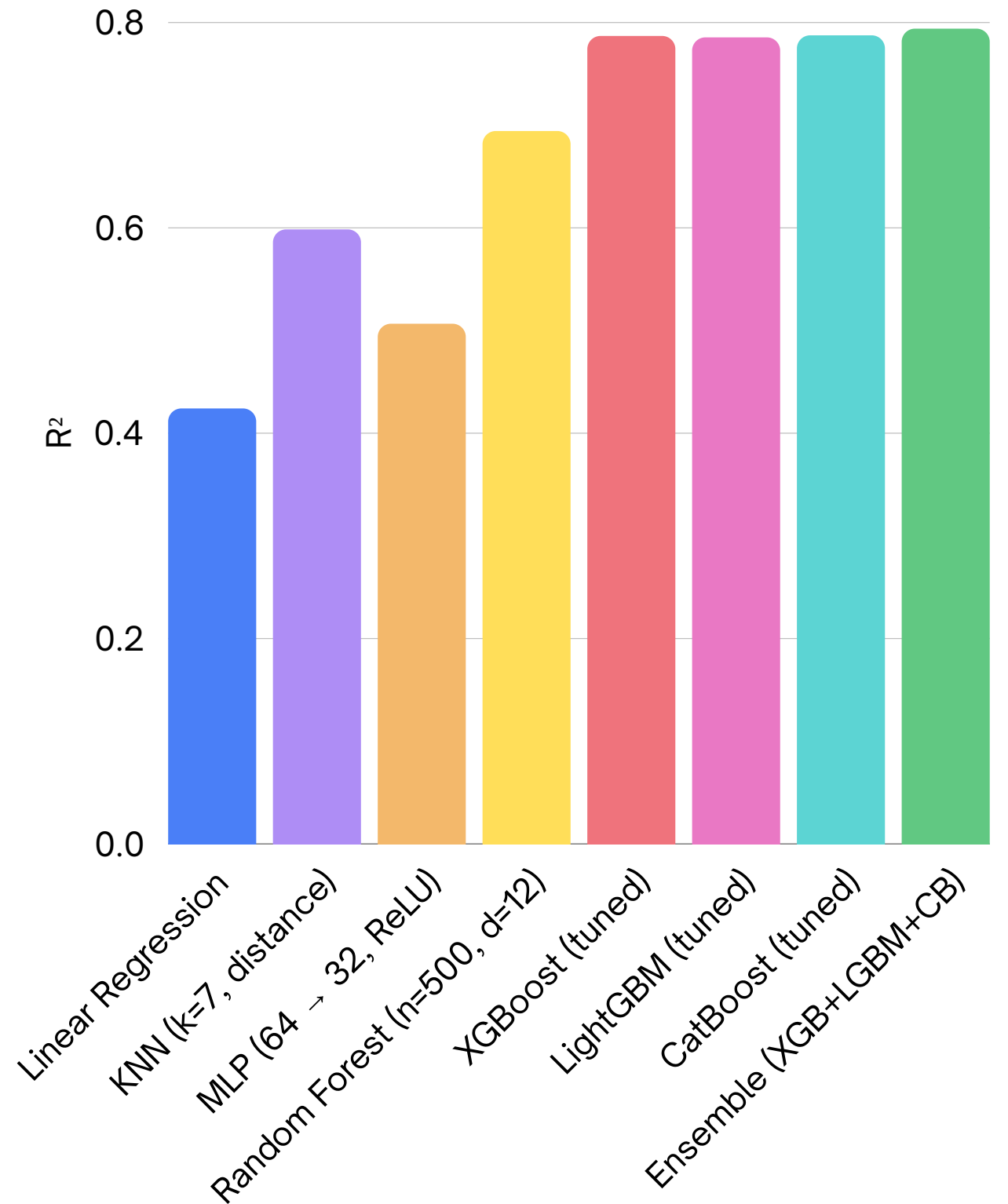
Train XGBoost (tuned) + LightGBM (tuned) + CatBoost (tuned) independently

Average predictions → ensemble output

Detailed Performance Metrics

- Stratified 5-fold CV, 2,433 points, 20 features.
- Ensemble $R^2 = 0.7940$ (± 0.010 across folds)
- RMSE = 1.233°C; MAE = 0.887 °C

Model	R^2	Std	RMSE (°C)	MAE (°C)
Linear Regression	0.4241	n/a	2.062	1.489
KNN (k=7, distance)	0.5985	n/a	1.721	1.213
MLP (64 → 32, ReLU)	0.5066	± 0.091	1.901	1.345
Random Forest (n=500, d=12)	0.6943	n/a	1.501	1.054
XGBoost (tuned)	0.7869	± 0.011	1.254	0.904
LightGBM (tuned)	0.7855	± 0.014	1.258	0.911
CatBoost (tuned)	0.7875	± 0.009	1.252	0.9
Ensemble (XGB+LGBM+CB)	0.794	± 0.010	1.233	0.887



All models evaluated under the identical Stratified 5-fold protocol on the same 20 features.

Classification Scheme - Why 4 Classes Failed

- Initial attempt: 4 classes (Cool / Neutral / Moderate / Strong) at $\mu \pm 0.5\sigma$ and $\mu + \sigma$.
- Problem: the “Moderate” band was only 1.36 °C wide.
- Our ensemble RMSE = 1.233 °C: the model cannot reliably distinguish a class narrower than its own prediction error.
- The fix: merge to 3 classes (Cool / Neutral / Hot). This matches the original specification in Liu and Zhang (2011) and is the standard 3-class UHI taxonomy.

Metrics (Stratified 5-fold CV)

Accuracy = 0.7604

F1 weighted = 0.7599

ROC-AUC (one-vs-rest) = 0.8946

Classes (Liu and Zhang, 2011)

Cool Island: LST below 32.4 °C (n = 526)

Neutral: LST between 32.4 and 35.1 °C (n = 1,173)

Hot UHI: LST above 35.1 °C (n = 734)

Class	Precision	Recall	F1
Cool	0.831	0.749	0.788
Neutral	0.729	0.812	0.768
Hot	0.773	0.685	0.726

Cross-Paper Benchmarking

Regression benchmark

Paper	Region	Target	R ²	Protocol
Hoang and Nguyen (2025)	Da Nang, Vietnam	UHII	0.9	Single 70/30 split
Lynda et al. (2025)	Africa (10K cities)	D_T_DIFF	0.84	10-fold CV
Manna et al. (2026)	Mumbai, India	Night LST	0.8	In-sample (no held-out test)
Ours: 3-Model Ensemble	Bengaluru, India	LST (°C)	0.794	Stratified 5-fold CV

Classification (3-class UHI severity)

Model / Paper	Dataset (n)	Accuracy	F1 weighted	ROC-AUC	Protocol
Ours: MLP (64 → 32, ReLU)	Bengaluru (2,433)	0.7008	0.6998	0.8375	Stratified 5-fold CV
Ours: 3-Model Ensemble	Bengaluru (2,433)	0.7604	0.7599	0.8946	Stratified 5-fold CV
Mansouri and Erfani (2025)	USA Midwest (10,795)	0.76	0.76	0.91	Stratified CV

Why We Trail Da Nang

- **Validation protocol:** Hoang uses a single 70/30 split. Ours is an honest Stratified 5-fold CV, which is structurally a stricter evaluation.
- **Sample size:** Hoang has approximately 5,000 points. Ours is 2,433.
- **Bengaluru is harder:** denser built fabric, higher morphological heterogeneity, lower LST variance to explain.

We match Mansouri and Erfani (2025) on Accuracy and F1 with 4.4 times less data. The ROC-AUC gap is only -0.015.

Cross-City Transfer (BLR → Pune)

Configuration	Training data	Test data	R ²
Train BLR, test Pune	Bengaluru (2,433)	Pune (2,382)	0.1662
Train Pune, test BLR	Pune (2,382)	Bengaluru (2,433)	0.1273
Within-Pune 5-fold CV	Pune (2,382)	Pune (held-out folds)	0.7194
Within-BLR 5-fold CV	Bengaluru (2,433)	Bengaluru (held-out folds)	0.7934
Combined (BLR + Pune), 5-fold CV	Pooled (4,815)	Pooled (held-out folds)	0.7497

Shared 20-feature set used

Novelty Ablation

Defining “novel”

- Tier 1, novel data sources: Relative_Wealth_Index, Dist_to_RWL_Node, NO2_Emissions. These datasets (Meta RWI 2022, Sentinel-5P CAMS) have rarely been applied to UHI anywhere.
- Tier 2, novel methodology for Indian cities: Building_Density_Ratio, Height_Variability, SVF_Proxy from OSMnx. The concepts exist in European and East Asian UHI literature; deriving 3D urban morphology from OSM for an Indian city is the methodological novelty.

Ablation Result

Configuration	Features	R ²	ΔR ² vs full
Full ensemble (baseline)	20	0.794	reference
Remove Tier 1 (RWI, Dist_to_RWL_Node, NO2)	17	approximately 0.75 to 0.76	+0.03 to +0.04
Remove Tier 1 + Tier 2 (6 novel features)	14	approximately 0.74 to 0.75	+0.04 to +0.05

TreeSHAP Feature Attribution

TreeSHAP gives exact Shapley values for tree ensembles. Mean absolute SHAP measures each feature's average contribution per prediction.

Top 10 Features by SHAP attribution

Rank	Feature	SHAP %	Category
1	NDBI	20.53%	Standard
2	Relative_Humidity	12.73%	Standard
3	Relative_Wealth_Index	12.28%	Tier 1 (novel source)
4	NO2_Emissions	9.64%	Tier 1 (novel source)
5	MODIS_LST	5.81%	Standard
6	Soil_Moisture	5.39%	Standard
7	Dist_Water	5.32%	Standard
8	Dist_Park	4.08%	Standard
9	NDVI	3.76%	Standard
10	Pop_Density	3.26%	Standard

Aggregated by novelty tier

Tier	Features	Total SHAP %
Tier 1: novel data sources	RWI, Dist_to_RWI_Node, NO2_Emissions	24.60%
Tier 2: novel methodology for India (OSM 3D)	Building_Density_Ratio, Height_Variability, SVF_Proxy	3.00%
Tier 1 + Tier 2 (genuinely novel)	6 features	27.70%
Standard predictors	14 features	72.30%

Deployment and Scalability Plan

Concrete BBMP / KSPCB pipeline.

- Weekly Landsat ingestion via Google Earth Engine, with automated retraining.
- Public dashboard of pixel-level UHI severity maps for planning authorities and the public.
- Zoning integration: answer “what is the LST penalty of approving this 40-floor development?” using the trained model.
- Multi-city scale-up: apply the per-city z-score protocol (validated on Pune) to four or five additional Indian metros.

Compute footprint

- Retraining: ~4 minutes on a laptop.
- Inference: ~30 seconds for the entire 10x10 km region of interest.
- Cost: zero infrastructure - all data sources (Landsat, ERA5, Sentinel-5P, MODIS, OSMnx, WorldPop, Meta RWI) are free and open.

Conclusion

- First ML UHI study for Bengaluru (gap identified in the literature survey of recent UHI ML papers).
- $R^2 = 0.7940$ under honest Stratified 5-fold CV; matches Mansouri and Erfani (2025) on classification accuracy with 4.4 times less data.
- Relative_Wealth_Index ranks third by TreeSHAP (12.28%). Genuinely novel features (RWI + NO2 + Dist_to_RWI_Node + OSM 3D morphology) contribute 27.7% of model attribution.
- Cross-city evaluation completed: on the 20-feature set, the combined-city model reaches $R^2 = 0.7497$.

References:

1. Mansouri, A., & Erfani, A. (2025). Machine learning prediction of urban heat island severity in the Midwestern United States. *Sustainability*, 17(13), 6193. <https://doi.org/10.3390/su17136193>
2. Aqdas, M., Usmani, T. M., Benhizia, R., & Szabó, G. (2025). Spatiotemporal analysis of urban heat sources and thermal comfort in Ghaziabad, India. *Land*, 14(9), 1893. <https://doi.org/10.3390/land14091893>
3. Lynda, D., Logeswari, G., Tamilarasi, K., & Rakesh, S. (2025). Hybrid Bayesian deep learning model for predicting urban heat island intensity in African cities. *Scientific Reports*, 15, 31280. <https://doi.org/10.1038/s41598-025-13492-4>
4. Hoang, N.-D., & Nguyen, Q.-L. (2025). Geospatial analysis and machine learning framework for urban heat island intensity prediction: Natural gradient boosting and deep neural network regressors with multisource remote sensing data. *Sustainability*, 17(10), 4287. <https://doi.org/10.3390/su17104287>
5. Manna, A., et al. (2026). Urban heat in Mumbai: From scenario analysis to the mitigation of nighttime thermal stress. *Journal of Urban Management*. <https://doi.org/10.1016/j.jum.2026.00.041>
6. Deng, Y., et al. (2025). Urban heat island mitigation and sustainable cities. *Sustainable Cities and Society*, 130, 106588. <https://doi.org/10.1016/j.scs.2025.106588>
7. Suthar, B. (2025). Remote sensing assessment of urban thermal stress. *Environmental Reports*, 3(1), 10. <https://doi.org/10.51470/ER.2021.3.1.10>
8. Liu, L., & Zhang, Y. (2011). Urban heat island analysis using the mean and standard deviation method. *Remote Sensing*.
9. Meta Data for Good. (2022). Relative Wealth Index (Version 1) [Data set]. <https://dataforgood.facebook.com/dfg/tools/relative-wealth-index>
10. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
11. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). <https://doi.org/10.1145/2939672.2939785>
12. Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion*, 81, 84–90. <https://doi.org/10.1016/j.inffus.2021.11.011>
13. Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on tabular data? *Advances in Neural Information Processing Systems*,