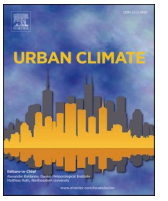
Urban Climate 67 (2026) 102955

Contents lists available at ScienceDirect

Urban Climate

journal homepage: www.elsevier.com/locate/uclim



# Identifying morphological thresholds in spatio-thermal interactions: Benchmarking interpretable AI for urban heat

Mert Akay [a,*], Deniz Erdem Okumus [b]

[a] Department of Human-Centered Design, Delft University of Technology, Delft, the Netherlands
[b] City and Regional Planning Department, Yildiz Technical University, Istanbul, Turkey

ARTICLE INFO

*Keywords:*
Explainable AI
Urban Heat Island
Change-point detection
Shapley additive explanations
Machine learning
Morphological breakpoints

ABSTRACT

Urban heat exhibits complex nonlinear relationships with morphological form, with thermal responses shifting across critical thresholds in building density, canyon geometry, and block configuration. While machine learning enables the detection of these interactions, existing studies predominantly report broad threshold ranges without identifying precise morphological breakpoints or validating findings across different models, thereby limiting translation into regulatory standards. This study benchmarks four tree-based AI algorithms – XGBoost, Gradient Boosting Machine (GBM), Random Forest (RF), and Light Gradient Boosting Machine (LGBM) – for predictive performance and interpretability in Istanbul's urban heat analysis. Shapley Additive Explanations (SHAP) quantify feature contributions across models, while hierarchical change-point detection identifies precise morphological thresholds where thermal effects shift between nonlinear regimes. Results reveal narrow differences in accuracy ($R^2$ = 0.675–0.685), with computational efficiency and interpretability proving more decisive. LGBM trained 4.5 times faster than RF, and XGBoost exhibited the highest morphological sensitivity. Normalised Difference Vegetation Index and building count emerge as dominant thermal drivers across all models. Consensus-based threshold detection, quantified as inter-model standard deviation (SD) across the four algorithms, yielded 80 breakpoints. Three exhibit high cross-model agreement (SD < 0.02), indicating model-invariant regime shifts: verticality at height-to-footprint ratio (H/A) = 0.14 marks the onset of canyon shading effects that offset thermal mass penalties; urban block area thresholds at 2248 $m^2$ and 6883 $m^2$ indicate permeability constraints triggering heat retention. Four thresholds demonstrated moderate consensus (SD 0.02–0.04): building density at ~16 buildings/ha corresponds to nonlinear heat intensification as impervious coverage reaches 49%; floor area ratio at 1.38 marks mid-rise regime shifts. By validating thresholds across models, this study moves beyond approximate value ranges toward robust thresholds applicable to climate-responsive zoning. This represents one of the first empirical contributions to systematically detect morphological breakpoints through cross-model validation.

## 1. Introduction

As cities continue to evolve, the physical transformation of the urban fabric has intensified interactions between urban morphology and multiple disciplinary domains. Among these, urban climate has emerged as a particularly critical interface due to its direct

* Corresponding author at: Faculty of Industrial Design Engineering, TU Delft, Landbergstraat 15, 2628 CE Delft, the Netherlands.
*E-mail addresses:* m.a.akay@tudelft.nl (M. Akay), denizer@yildiz.edu.tr (D. Erdem Okumus).

https://doi.org/10.1016/j.uclim.2026.102955
Received 28 January 2026; Received in revised form 29 April 2026; Accepted 15 May 2026
Available online 22 May 2026

implications for environmental performance, public health, and everyday urban life. A central concern within this nexus lies in how morphological components, such as buildings, plots, blocks, and street networks, influence the spatial distribution of urban heat. These elements do not operate independently; rather, they interact through complex, scale-dependent, and often nonlinear mechanisms, producing thermal patterns that traditional linear modelling approaches cannot adequately capture.

To address this complexity, artificial intelligence (AI)–based machine learning (ML) predictive models have increasingly been adopted in urban climate research. By accommodating high-dimensional data structures and capturing intricate interactions among morphological variables, these AI-driven approaches provide a robust operational foundation for uncovering the nonlinear dynamics underlying the urban climate–morphology nexus (Ding et al., 2025). Within this expanding body of work; the Urban Heat Island (UHI) effect has emerged as one of the most prominent application domains; given its direct implications for zoning regulations; density controls; and climate-sensitive urban design strategies. A growing number of studies employ ML algorithms to examine how combinations of morphological; environmental; and land-use variables contribute to intra-urban heat variability (Sattar et al., 2025; Ullah et al., 2024). Their ability to process large datasets and model nonlinear, multivariate relationships has positioned ML approaches as a key analytical tool for understanding spatial heat patterns in complex urban environments. Within this context, tree-based ensemble models, most notably Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Gradient Boosting Machine (GBM), and Light Gradient Boosting Machine (LGBM), have become commonly employed algorithms in recent urban heat studies (Tanoori et al., 2024; Vaidya et al., 2024; Zhang et al., 2024).

RF has become one of the most extensively adopted approaches for analysing how urban morphology influences land surface temperature (LST) and canopy air temperature (Aslani et al., 2025; Ding et al., 2025). Its widespread use reflects a combination of strong predictive performance, robustness to overfitting, adaptability across spatial resolutions, and the ability to generate outputs that remain interpretable for real-world applications. Empirical studies have successfully employed RF (Liu et al., 2024b; Wang et al., 2025a) to quantify the thermal effects of key morphological variables, including floor area ratio (Chen et al., 2025b); building height (Yao et al., 2022); building coverage ratio (Li et al., 2023); vegetation cover (Wang et al., 2024); and impervious surfaces (Badaro-Saliba et al., 2021; Chen et al., 2022). Moreover, RF has proven particularly capable of handling spatial heterogeneity and fine-resolution inputs, making it well-suited for neighbourhood-level and city-scale heat assessments (Sun et al., 2024; Zhang et al., 2025). In addition, RF's resistance to overfitting and its capacity to rank feature importance render it especially useful (Segal, 2004) in urban datasets characterised by spatial autocorrelation and multicollinearity.

More recently, XGBoost has gained prominence in UHI research, particularly as attention has shifted from prediction alone toward explainability and nonlinear sensitivity analysis (Gu et al., 2024; Okumus and Akay, 2025). XGBoost is well-suited for capturing complex and non-stationary relationships between urban form and temperature dynamics, especially when combined with post-hoc interpretation tools such as Shapley Additive Explanations (SHAP). A growing body of literature employs this combination to investigate how two- and three-dimensional morphological indicators exert nonlinear and scale-sensitive influences on LST, particularly during extreme heat events (Chen et al., 2025a; Liu et al., 2023; Zou et al., 2024). Recent applications have extended SHAP-based interpretability to spatial scale decomposition, cross-city comparisons, and multi-source data integration (Chen and Guo, 2025; Liu et al., 2024a; Liu et al., 2025b; Ming et al., 2024; Sheng et al., 2025). Studies have also examined the role of spatial configurations, including block typologies, polycentric layouts, and land-use patterns, in shaping localised heat intensities (Fan et al., 2024; Lin et al., 2025).

In parallel with RF and XGBoost, GBM and LGBM models have recently been introduced into urban heat research due to their strong performance in modelling nonlinear, multi-feature relationships and their computational efficiency. Studies applying these models demonstrate their capacity to capture spatial and temporal heterogeneity in surface temperature responses using both two- and three-dimensional morphological indicators (B. Liu et al., 2023; Wang et al., 2025b; Yu et al., 2025; Yuan et al., 2024). GBM-based approaches; in particular; have shown promise in addressing non-stationary heat dynamics that pose challenges for traditional statistical methods (Liu et al., 2025a).

Further extending these perspectives, recent studies demonstrate that morphology-UHI relationships are often non-stationary, varying across spatial scales, functional contexts, and zoning regimes (Shen et al., 2024). Within this framework; interpretability has become a central methodological concern; not only enhancing transparency but also enabling the identification of threshold behaviours; points at which incremental morphological changes result in disproportionate thermal responses (Chen et al., 2025a; Wu et al., 2024; Yuan et al., 2024). Very recently, empirical evidence of such thresholds has emerged in blue-green infrastructure cooling onset (Jiang et al., 2025; Shi et al., 2026), built-up density regime shifts (Hou et al., 2026); and terrain-sensitive morphological exploration (Zhang et al., 2026a). This supports a growing methodological shift toward explainable AI in climate-sensitive urban planning and design, where prediction and interpretation should be considered together. For planning practice, such interpretability is critical not only for methodological transparency but also for translating model outputs into spatial regulations, design thresholds, and evidence-based zoning decisions.

Despite such substantial methodological progress, three critical research gaps remain in the literature on AI-based urban heat modelling. First, the vast majority of studies adopt a single-model perspective, implicitly treating the selected algorithm as the most accurate and internally consistent representation of urban heat dynamics. Model outputs are typically evaluated in isolation, without systematic comparison across alternative modelling paradigms (Ding et al., 2025; Li et al., 2025a; Rashtian et al., 2025; Shao et al., 2023; Tang et al., 2025; Tanoori et al., 2024; Yu et al., 2025). Therefore, the field lacks a robust understanding of the trade-offs between predictive accuracy, computational efficiency, and explanatory transparency–an omission that limits both methodological advancement and the translation of model outputs into planning-relevant knowledge. Second, although nonlinearity is widely recognised as a defining characteristic of morphology-heat interactions, existing studies predominantly operationalise this complexity through broad threshold ranges rather than identifying precise morphological breakpoints at which thermal responses shift between

regimes (Chen et al., 2025a; Fan et al., 2024; Lin et al., 2025). This approach offers limited guidance for urban planning and design, where regulatory instruments and spatial interventions require clearly defined and operationalisable parameters and intervals. The absence of exact reference points constrains the incorporation of quantitative findings into zoning regulations, design standards, and climate-responsive planning strategies. Third, even when such thresholds are identified, their reliability is rarely examined across different modelling approaches. Existing studies typically assume that detected breakpoints reflect stable underlying relationships, without testing their consistency across alternative algorithms (Okumus and Akay, 2025; Shen et al., 2024). Accordingly, it remains unclear whether the reported thresholds represent intrinsic properties of urban systems or are influenced by model-specific biases. This lack of cross-model validation introduces a critical source of uncertainty, particularly for planning applications, where regulatory decisions depend on defensible and transferable threshold values. Without assessing robustness, threshold identification risks producing model-dependent interpretations that may not be suitable for policy translation. These three limitations constrain the translation of ML findings into actionable planning instruments, where regulatory codes require defensible, cross-validated breakpoints rather than exploratory value ranges.

Addressing these gaps, this study develops an integrated analytical framework to systematically benchmark four tree-based ML

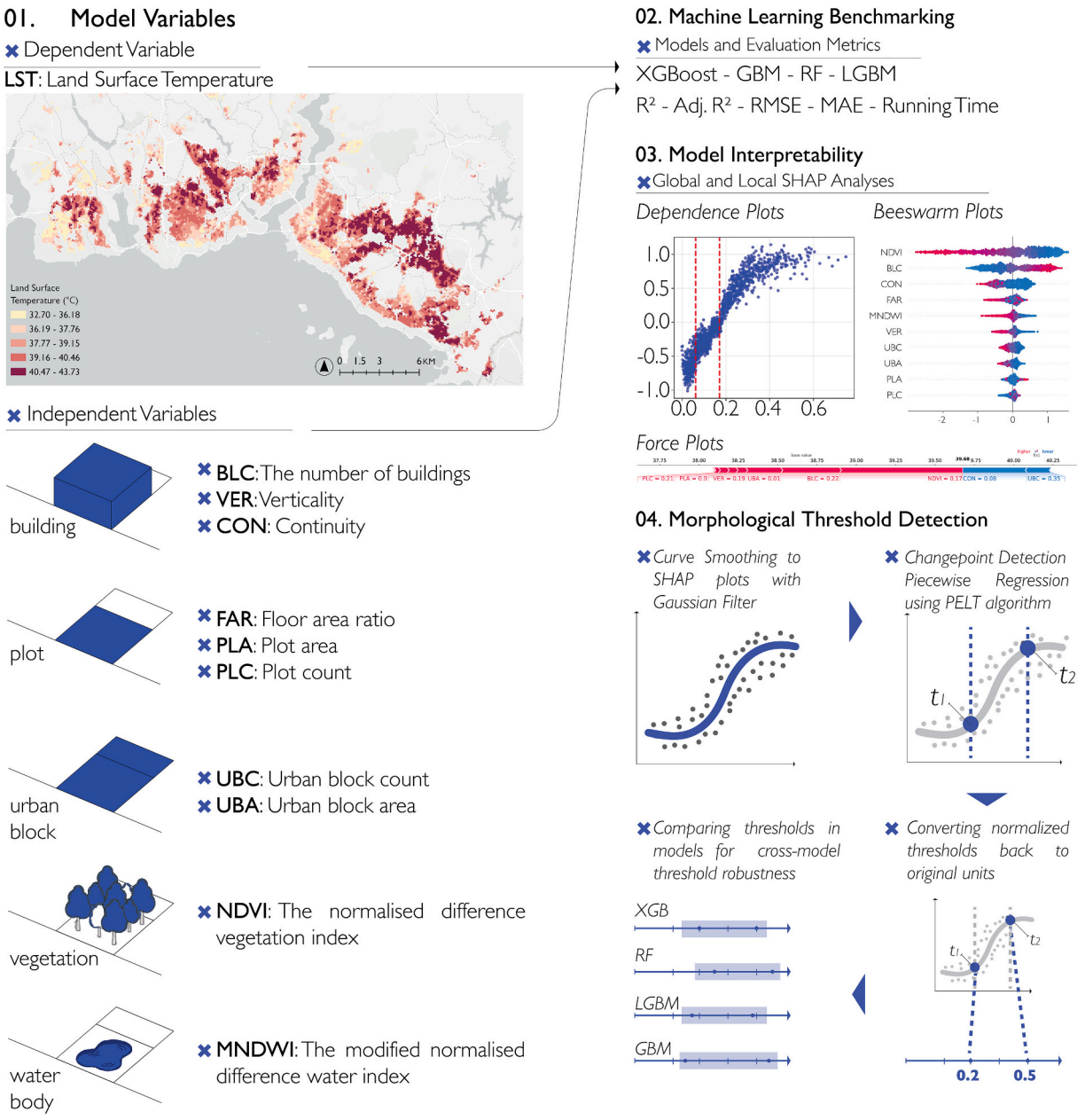



**Fig. 1.** Methodological flow of the research.

algorithms (Random Forest, XGBoost, Gradient Boosting Machine, and LightGBM) for urban heat modelling. Istanbul is employed as the study area due to its pronounced intra-urban thermal heterogeneity, Mediterranean climate characterised by hot summers, and diverse morphological fabric spanning historical cores, rapid suburban expansion, and high-density commercial districts. The framework combines SHAP-based interpretability with hierarchical change-point detection to identify statistically grounded morphological thresholds at which thermal regime shifts occur. Crucially, it introduces a consensus-based robustness classification to evaluate the stability of detected thresholds across models, explicitly distinguishing model-invariant breakpoints from model-dependent ones. This integrated approach enables a systematic comparison of predictive accuracy, computational efficiency, and interpretability, while ensuring that identified thresholds are not only detectable but also reliable and transferable for planning applications.

This research addresses two research objectives: (1) To benchmark four tree-based algorithms in terms of predictive accuracy, computational efficiency, and interpretability for urban heat modelling, explicitly comparing performance-transparency trade-offs that remain largely implicit in single-model studies (2) To identify precise morphological threshold values where thermal effects shift nonlinearly and to quantify cross-model consensus levels, distinguishing robust, model-invariant breakpoints from context-dependent values requiring local calibration. In doing so, this study makes four methodological contributions to the urban climate literature. Rather than relying on a single algorithm, it provides a structured multi-model benchmark that explicitly exposes accuracy-efficiency-interpretability trade-offs across four competing algorithms, a comparison that prior studies largely leave implicit. Building on this, it introduces a threshold detection framework that transforms approximate value ranges into precise, operationalisable morphological breakpoints through hierarchical change-point detection. Crucially, it couples this with a consensus-based robustness classification that distinguishes model-invariant breakpoints from model-dependent ones, providing a transparent basis for calibrating regulatory stringency to empirical confidence. Finally, the integrated framework is designed for replicability, offering a generalisable methodological pipeline applicable beyond urban heat to broader urban climate challenges.

The paper is organised as follows: Section 2 describes the methodological framework, including spatial unit definition, model implementation, SHAP-based interpretability, and hierarchical change-point detection. Section 3 presents model performance benchmarks, feature importance rankings, and consensus-validated thresholds. Section 4 discusses the translation of thresholds into tiered regulatory frameworks and spatial intervention strategies. Section 5 concludes with methodological contributions and planning implications.

## 2. Methodology and materials

This research develops a five-stage analytical framework consisting of: (1) Spatial unit and variable definition, (2) implementation of four tree-based ML algorithms – RF, XGBoost, GBM, and LGBM; (3) applications of SHAP for model interpretability; (4) identification of change-points using Piecewise Regression; and (5) execution of policy translation (Fig. 1).

### 2.1. Spatial unit definition

Istanbul (41.01°N, 28.98°E), Turkey's largest metropolitan area, experiences a Mediterranean climate (Köppen Csa) with hot, dry summers (July–August mean: 28 °C, maximum frequently exceeding 35 °C) (Beck et al., 2018). The city's morphological fabric exhibits substantial heterogeneity, ranging from historic low-rise cores (1–4 stories) and mid-rise residential districts (5–9 stories) to high-density commercial centres (10+ stories), with a mean building height of approximately 15 m. Topography is relatively gentle, with most areas below 50 m elevation (Fig. A1).

For statistical analyses, hexagon-based sampling was adopted as the spatial unit of analysis (Chien et al., 2020; Duan et al., 2023). The administrative boundaries of the Istanbul Metropolitan Municipality were partitioned into hexagons with a side length of 150 m (Okumus and Akay, 2025). A systematic exclusion protocol was then applied to refine the dataset. Initially; hexagons located outside designated residential areas were excluded to reduce the impact of heterogeneity arising from land cover variability. Subsequently; hexagons with an average slope greater than 25%; as well as those situated within 1 km of large water bodies; were removed to mitigate the influence of natural environmental factors on the results (Erdem Okumus and Terzi, 2021). Following this exclusion process, a total of 7826 hexagonal units were retained for subsequent statistical analyses.

### 2.2. Measurement of model variables

The study utilised one dependent and ten independent variables to investigate spatio-thermal interactions across Istanbul. LST mean values were employed as the dependent variable to characterise the thermal environment. The independent variables were categorised into two main groups: (a) morphological metrics, including the number of buildings (BLC), floor area ratio (FAR), plot area (PLA), plot count (PLC), urban block count (UBC), urban block area (UBA), verticality (VER), and continuity (CON); and (b) spectral indices, comprising the normalised difference vegetation index (NDVI) and the modified normalised difference water index (MNDWI).

LST values were derived from NASA's single Landsat-8/9 OLI/TIRS scene acquired on 26 July 2023 between 08:44 and 08:45 am, with 0% land cloud coverage (USGS, 2023a, 2023b); representing summer daytime thermal conditions under stable atmospheric circumstances. A seven-step mono-window algorithm was employed to estimate LST; based on established methodologies (Jiménez-Muñoz et al., 2006, 2009; Sobrino et al., 2004; Xu & Chen, 2024; Guha et al., 2018; USGS, 2019). The procedure involved (1) conversion to top-of-atmosphere radiance values; (2) conversion to top-of-atmosphere brightness temperature; (3) NDVI extraction; (4) calculation of the proportion of vegetation cover; (5) estimation of land surface emissivity; (6) computation of LST values; (7)

validation of LST estimates using meteorological station data (Weather Underground, 2023), and (8) the mean LST value of each hexagon was calculated and employed in statistical analyses.

Morphological metrics include building-level parameters (BLC, VER, CON), plot-level indicators (FAR, PLA, PLC), and block-scale measures (UBC, UBA). BLC quantifies building density, influencing urban heat through increased impervious surface area, reduced albedo, anthropogenic heat generation, and thermal mass accumulation (Tang et al., 2025). FAR captures habitable space intensity and land-use efficiency; mediating heat through building volume-to-land ratios that determine surface-to-volume relationships and heat storage capacity (Ding et al., 2025). VER measures vertical development through height-to-footprint ratios; affecting thermal dynamics via canyon geometry effects including sky view factor reduction; radiative trapping; and wind channelling patterns (Çalışkan et al., 2022). CON reflects building alignment and spatial continuity; influencing airflow circulation and heat retention in the built fabric. Plot-scale variables (PLA; PLC) bridge micro- and meso-scale morphology; conditioning ventilation pathways and microclimate variability through parcel fragmentation patterns (Okumus and Akay, 2025). Block-scale metrics (UBC; UBA) (Akay and Çalışkan, 2025) capture structural configuration; with block size and aggregation affecting urban permeability; internal airflow patterns; and heat dissipation pathways. Spectral indices include NDVI; quantifying vegetation density and health (Rees et al., 2024); and MNDWI; detecting surface water presence (Yu et al., 2025). NDVI values approach +1 for dense vegetation and decrease toward zero for impervious surfaces. MNDWI values exceed zero for water bodies and fall below zero for built-up areas. Both indices were derived from Landsat-8/9 imagery (26 July 2023) at hexagon scale. Computational definitions for all variables are provided in Table 1. These indicators represent the key morphological and biophysical controls governing urban heat dynamics, enabling a comprehensive assessment of how built form and surface characteristics interact to influence thermal behaviour.

### 2.3. Supervised learning with tree-based algorithms

Before model implementation, all input variables were normalised using min-max scaling to a range of [0,1] to ensure comparability across features and optimise model convergence. Subsequently, four supervised tree-based ML algorithms – XGBoost, GBM, RF, and LGBM – were applied to model spatio-thermal interactions by training on ten urban morphological predictors. The model hyperparameters (learning rate, maximum depth, regularisation coefficients, tree count) were optimised via Bayesian search (Wu et al., 2019) with 5-fold cross-validation, conducting 50 iterations per model to identify performance-maximising configurations (Table 4). To train the models, the sample data was divided into 80% training and 20% validation sets. For each model, predictive performance and accuracy were evaluated using the coefficient of determination ($R^2$), the adjusted coefficient of determination (Adj. $R^2$), root mean square error (RMSE), and mean absolute error (MAE). Multicollinearity among the predictor variables was assessed using a correlation matrix (Fig. A2). While moderate correlations were observed between certain variables, tree-based ensemble models are inherently less sensitive to multicollinearity due to their split-based learning structure and random feature selection mechanisms. Therefore, all variables were retained in the modelling process to preserve their distinct contributions to urban thermal dynamics.

As a first model, XGBoost is a highly efficient and scalable implementation of gradient boosting that constructs an ensemble of regression trees through a stage-wise additive model. The algorithm optimises a regularised objective function that combines a loss term with a penalty for model complexity, thus controlling overfitting and enhancing generalisation. XGBoost's objective function at each iteration $t$ combines a differentiable loss function $L$ and a regularisation term $\Omega$ to penalise model complexity:

**Table 1**
Computational definitions of morphological metrics and spectral indices.

| Variables* | Formulas | Definitions |
|---|---|---|
| BLC | $\sum_{i=1}^{n} 1$ | n = Total number of individual building units located within the hexagon. The summation counts each building once within the spatial bounds of the hexagon. |
| FAR | $\frac{Atfj}{Aj}$ | Atfj = the total floor area of the buildings on plot j.<br>Aj = the area of plot j. |
| PLA | $\frac{1}{n}\sum_{i=1}^{n} Ai$ | n = Number of plots within the hexagon.<br>Ai = Area of the $i^{th}$ plot (in $m^2$). |
| PLC | $\sum_{i=1}^{n} 1$ | n = Total number of parcels within the urban block. The summation counts each parcel once within the spatial bounds of the urban block. |
| UBC | $\sum_{i=1}^{n} 1$ | n = Total number of urban blocks within the hexagon. The summation counts each urban block once within the spatial bounds of the hexagon. |
| UBA | $\frac{1}{n}\sum_{i=1}^{n} Ai$ | n = Total number of urban blocks in the hexagon.<br>Ai = Area of the $i^{th}$ urban block within the hexagon. |
| VER | $\frac{h_j}{Af_j}$ | $h_j$ = the height of building *j*.<br>$Af_j$ = the total area of the building footprint for building *j*. |
| CON | $\frac{Nbld}{Nenv}$ | Nbld = the total number of buildings.<br>Nenv = the total number of envelopes in a given area. |
| NDVI | $\frac{NIR - RED}{NIR + RED}$ | NIR = Near-infrared reflectance.<br>RED = Red reflectance. |
| MNDWI | $\frac{Green - SWIR}{Green + SWIR}$ | Green = Reflectance in the green band.<br>SWIR = Reflectance in the shortwave infrared band. |

* BLC: The number of buildings; FAR: Floor area ratio; PLA: Plot area; PLC: Plot count; UBC: Urban block count; UBA: Urban block area; VER: Verticality; CON: Continuity; NDVI: The normalised difference vegetation index; MNDWI: The modified normalised difference water index.

$$(L)^t = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \tag{1}$$

where $f_t \in F$ represents the newly added regression tree, $\hat{y}_i^{(t-1)}$ is the prediction from the previous iteration, and $\Omega(f_t) = \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2$ controls the complexity of the tree (with T as the number of leaves and $w_j$ the weight of the leaf j).

GBM is a widely used ensemble learning method that builds predictive models by sequentially adding weak learners–typically decision trees–to minimise a specified loss function. Each new decision tree is trained to correct the residual errors made by the ensemble of previously grown trees. This iterative process allows GBM to gradually improve model performance by reducing bias and capturing complex, nonlinear relationships in the data. GBM is particularly well-suited for structured datasets where interactions among variables are not explicitly defined. The core idea of GBM is to optimise a differentiable loss function L (y, ŷ) by gradient descent, where the prediction $\hat{y}^{(t)}$ at iteration $t$ is updated as:

$$\hat{y}^{(t)} = \hat{y}^{(t-1)} + \eta f_t(x) \tag{2}$$

where, $f_t(x)$ is the newly trained regression tree that fits the negative gradient of the loss function, and η is the learning rate controlling the contribution of each tree.

RF is an ensemble learning algorithm that constructs multiple decision trees during the training phase and aggregates their outputs to produce the final prediction, typically by averaging in regression tasks. Unlike boosting algorithms that build trees sequentially, RF trains trees independently in parallel, using a random subset of both the data and the features for each tree. This randomisation enhances model robustness, reduces overfitting, and improves generalisation, particularly when dealing with high-dimensional or noisy datasets.

Given a training dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}$, where $x_i \epsilon R^p$ represents the input features (morphological variables) and $y_i \epsilon R$ denotes the target variable (LST), the RF constructs a collection of T decision trees $\{h_i(x)\}_{t=1}^{T}$. Each tree $h_i(x)$ is trained on a bootstrapped sample $\mathcal{D}_t$ drawn with replacement from the original dataset $\mathcal{D}$. At each split in the tree, a random subset of features of size $m < p$ is selected, from which the best split is chosen to reduce a node impurity criterion–typically mean squared error (MSE) in regression tasks. This dual randomisation–in both sampling and feature selection–promotes diversity among individual trees, thereby enhancing the generalisation capability of the ensemble. The overall prediction for a new observation x is calculated as the average of the individual predictions from all trees in the ensemble:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \tag{3}$$

This averaging mechanism reduces variance and mitigates overfitting, which are common issues in single decision tree models. Importantly, since each tree captures different patterns within the dataset due to inherent randomness, the ensemble prediction becomes more stable and less sensitive to noise.

LGBM is a highly efficient gradient boosting framework that uses a leaf-wise tree growth strategy rather than the level-wise approach used in traditional boosting methods. This allows LGBM to achieve faster training speeds and lower memory usage while maintaining high accuracy, particularly on large datasets. LGBM is designed to handle large-scale data with categorical features and missing values efficiently, which makes it well-suited for urban climate studies that involve diverse geospatial and morphological variables.

Formally, given a training dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}$, where $x_i \epsilon R^p$ denotes the feature vector (urban morphology) and $y_i \epsilon R$ is the continuous response variable (LST), LGBM builds an ensemble of additive functions in a stage-wise manner. At iteration $t$, the objective is to minimise a regularised loss function:

$$L^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \tag{4}$$

where l is a differentiable loss function (such as squared error), $\hat{y}_i^{(t-1)}$ is the prediction from the previous iteration, $f_t(x)$ is the newly added regression tree, and $\Omega(f_t)$ is a regularisation term that penalises model complexity, typically based on the number of leaves and their weights. To efficiently optimise this objective, LGBM applies a second-order Taylor approximation of the loss function:

$$L^{(t)} \approx \sum_{i=1}^{n}\left[g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i)\right] + \Omega(f_t) \tag{5}$$

where $g_i = \frac{\partial l(y_i, \hat{y}_i)}{\partial \hat{y}_i}$ and $h_i = \frac{\partial^2 l(y_i, \hat{y}_i)}{\partial \hat{y}_i^2}$ are the first- and second-order gradients of the loss function, respectively. By leveraging histogram-based feature binning and gradient-based one-sided sampling, LGBM further reduces computational overhead while preserving model accuracy.

The four tree-based ML models–XGBoost, GBM, RF, and LGBM–share a common foundation in decision tree learning but differ significantly in their boosting strategies, computational efficiency, and handling of data complexity. XGBoost and GBM are boosting algorithms that build trees sequentially to correct the errors of previous models, with XGBoost incorporating advanced regularisation

to reduce overfitting. In contrast, RF constructs multiple trees independently using bootstrapped samples and random feature selection, which enhances robustness but can dilute interpretability. LGBM, while also a boosting method, uses a leaf-wise split strategy that accelerates training and improves efficiency on large datasets. These algorithmic distinctions influence not only predictive accuracy but also the models' suitability for handling urban-scale spatial heterogeneity and feature interactions (Table 2).

Comparing these models was essential not only to assess the trade-offs between accuracy and interpretability but also to evaluate the stability of derived relationships and thresholds across different algorithmic structures. Given the growing demand for explainable AI in urban planning, selecting a model that strikes a balance between performance and interpretability is crucial. By evaluating all models using the same performance metrics and SHAP-based interpretability, this study establishes a methodological benchmark for integrating ML into urban heat research.

### 2.4. Shapley additive explanations

To enhance interpretability, SHAP was applied across all models, enabling the identification and comparison of key variable contributions to urban heat predictions (Lundberg et al., 2018). This approach facilitated a detailed exploration of the spatial drivers of urban heat and supported the detection of critical thresholds through Piecewise Regression. By quantifying the marginal contribution of each feature to the prediction, SHAP provided a rigorous basis for explaining the outputs of complex tree-based models. Formally, for a model prediction f(x) with feature vector $x \in R^p$, the SHAP value for feature j is defined as:

$$\varnothing_j(f,x) = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!(p-|S|-1)!}{p!} \left[ f_{(S \bigcup \{j\})}\left(x_{S \bigcup \{j\}}\right) - f_S(x_S) \right] \tag{6}$$

where N = {1,2, …,p} is the set of all features, S is a subset of features excluding j, and $f_S(x_S)$ denotes the model trained on only the feature subset S. The combinatorial weighting ensures that contributions are fairly averaged across all possible feature orderings. The SHAP framework satisfies desirable properties such as local accuracy, consistency, and additivity, making it particularly suitable for urban climate research where interpretability is essential for actionable policy recommendations. In this study, SHAP values provided insights into the influence of urban morphological and network-related variables on LST, offering a comprehensive and transparent interpretation of model behaviour.

### 2.5. Piecewise regression (change-point detection)

Piecewise regression was employed to identify critical thresholds in SHAP-feature relationships where the marginal effect of morphological predictors on LST undergoes regime shifts. The statistical detectability of a breakpoint does not guarantee its reliability; therefore, threshold identification was framed not only as a segmentation task but also as a cross-model validation process. This approach allows assessment of whether detected breakpoints represent stable underlying relationships or are influenced by algorithm-specific biases. The method was applied to SHAP dependence plots, where SHAP values (representing feature impact on LST) were plotted against normalised feature values (0–1 scale), and these trajectories were smoothed using a Gaussian filter ($\sigma = 10$) to reduce local noise while preserving systematic trends. Formally, let y denote the SHAP value for a given feature, and X denote the normalised feature value (morphological predictor). A simple one-breakpoint piecewise linear regression is defined as:

$$y_i = \begin{cases} \beta_0 + \beta_1 X_i + \epsilon_i \; if & X_i \le T \\ \beta_0 + \beta_1 T + \beta_2 (X_i - T) + \epsilon_i \; if & X_i > T \end{cases} \tag{7}$$

where T is the estimated breakpoint, β0, β1, β2 are regression coefficients, and $\epsilon_i$ is the error term. More generally, for K breakpoints $T_1$, $T_2$, …, $T_K$, the model is expressed as a continuous, piecewise linear function with slope changes occurring at each breakpoint. The parameters $T_K$ are estimated by minimising the residual sum of squares.

In this research, change-points were detected using a hierarchical three-strategy approach that prioritises robustness. First, the Pruned Exact Linear Time (PELT) algorithm (Killick et al., 2012) was attempted with penalty parameters ranging from 1 to 5. PELT optimally segments sequences by minimising a penalised cost function, balancing fit quality against segmentation complexity. The algorithm was configured with a rank-based cost model and adaptive minimum segment size (max (30.10% of observations)) to prevent spurious breakpoints in sparse regions. If PELT failed to converge or detect valid breakpoints, Binary Segmentation with an L2 cost model was applied as a secondary method, constraining detection to two breakpoints ($n_{bkps} = 2$). In rare cases where both rupture-based methods failed, a quantile-based fallback divided features into tertiles, providing exploratory thresholds pending manual validation. For each detected breakpoint, importance was quantified as the absolute difference in Spearman correlation coefficients (ρ)

**Table 2**
Comparison of tree-based algorithms.

| Model | Strategy | Speed | Key Advantage |
|---|---|---|---|
| XGBoost | Sequential Boosting | Medium | Granular sensitivity |
| GBM | Sequential Boosting | Medium | Balanced performance |
| RF | Parallel Ensemble | Slowest | Robustness to noise |
| LGBM | Leaf-wise Boosting | Fastest | Computational efficiency |

between adjacent segments, specifically, between feature values and SHAP values within windows extending ±5% of observations from the breakpoint. This metric captures the magnitude of slope change, where larger differences indicate more pronounced regime shifts. To avoid spurious thresholds at feature distribution extremes, breakpoints falling below the 15th or above the 85th percentile received a penalty factor (0.3×), reducing their importance scores. From all detected breakpoints, the top two per feature were retained and designated as primary (T1, largest slope change) and secondary (T2, second largest) thresholds, provided they exhibited sufficient separation (>0.05 in normalised units) to represent distinct regimes.

Since all features were normalised during pre-processing using min–max scaling, which transforms values into the [0,1] range while retaining their original minimum and maximum values, the threshold values identified through change-point detection initially correspond to standardised units. To ensure interpretability and practical relevance, the normalised thresholds were subsequently rescaled back to their original feature scales, using the inverse of the min–max transformation,

$$\mathrm{T}_{\mathrm{original}} = \mathrm{T}_{\mathrm{normalised}} \times (T_{max} - T_{min}) + T_{min} \tag{8}$$

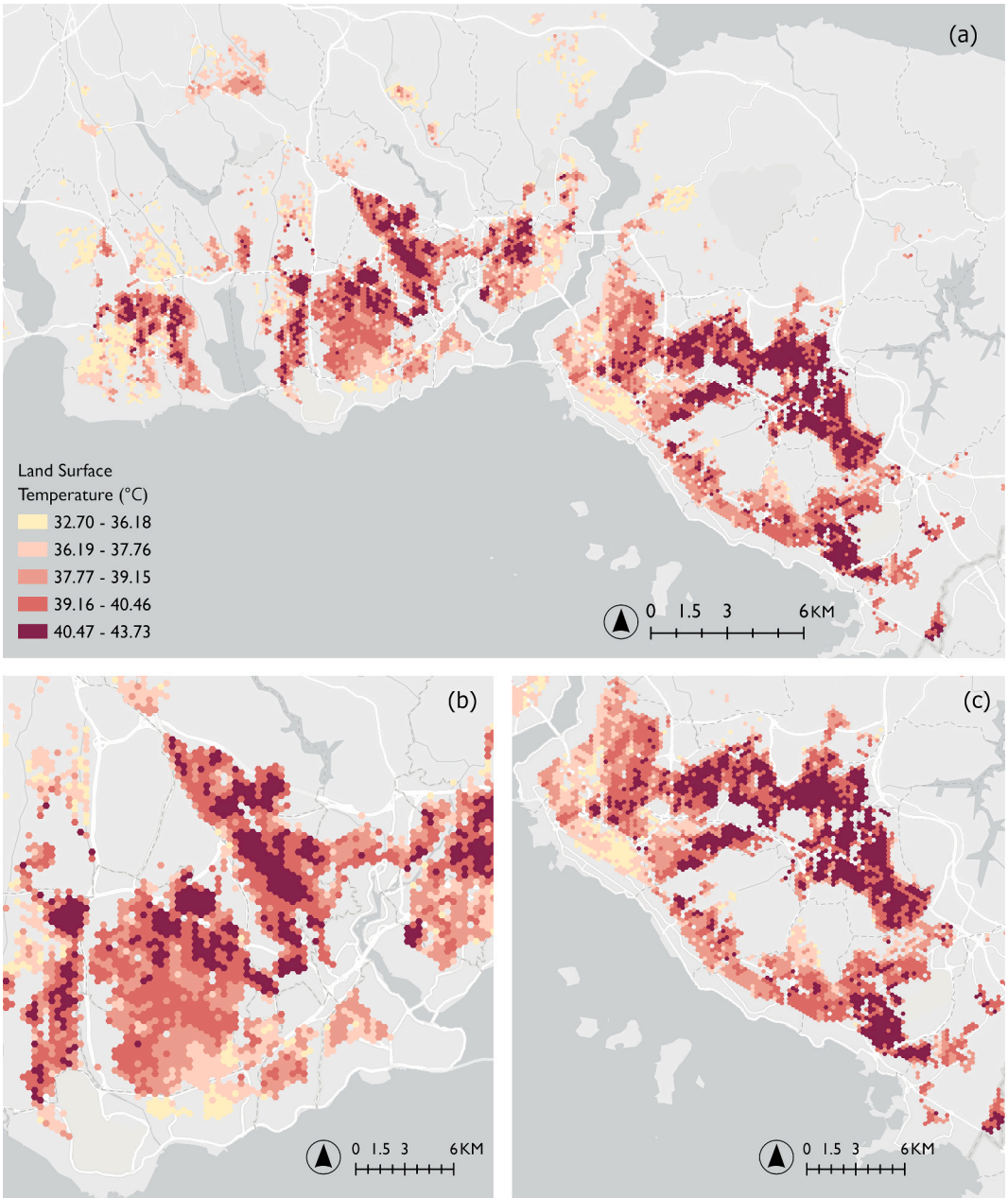



**Fig. 2.** Heat patterns across Istanbul are represented by hexagonal cells (a), (b), and (c) provide greater detail on subregional variations.

thereby allowing the thresholds to be expressed in real spatial terms. This procedure enabled the interpretation of breakpoints in their actual units rather than in abstract normalised values.

To assess threshold robustness, breakpoints detected across four models for each feature were systematically compared to quantify cross-model consensus. This approach enables distinguishing model-invariant thresholds from those sensitive to algorithmic structure, thereby reducing the risk of interpreting model-specific biases as generalisable relationships. Thresholds were classified into three tiers based on inter-model standard deviation of normalised values: (1) High-consensus (SD $<$ 0.02) indicates model-invariant breakpoints suitable for universal regulation, (2) moderate-consensus (0.02 $\leq$ SD $<$ 0.04) suggests robust yet context-sensitive thresholds, and (3) low-consensus (SD $>$ 0.04) reflects model-dependent detection requiring local calibration. These SD cutoffs were established based on empirical model convergence analysis: High consensus (SD $<$ 0.02) indicates threshold variation below 2% of the normalised feature range, reflecting strong cross-model agreement; moderate consensus (0.02 $\leq$ SD $<$ 0.04) reflects systematic but manageable variation (2–4% of feature range); low consensus (SD $>$ 0.04) exceeds 4%, suggesting model-specific detection requiring local validation. This consensus-based classification explicitly addresses threshold robustness to model selection and sampling variability, enabling planners to calibrate policy stringency to empirical confidence, distinguishing defensible regulatory benchmarks from exploratory findings.

As a result, model-derived thresholds could be directly linked to the underlying urban and environmental phenomena, strengthening both their interpretability and operational relevance. By combining SHAP values with piecewise regression, the analysis revealed the most influential morphological variables and identified precise thresholds at which their marginal effects on urban heat intensify or diminish. Importantly, this framework evaluates threshold identification not only in terms of statistical detectability but also in terms of cross-model consistency, ensuring that the derived thresholds are both robust and transferable for planning applications. Together, these methodological steps provided concrete, actionable insights that bridge the gap between statistical modelling and planning practice, supporting evidence-based interventions for climate-sensitive urban design.

## 3. Results

### 3.1. Heat patterns in hexagons

Heat patterns across Istanbul's hexagons show that LST values in Istanbul vary substantially, ranging from approximately 32.7 °C to 43.7 °C (Fig. 2). Cooler clusters (32.7–36.2 °C), represented in lighter shades, are concentrated in the northern peripheries. These areas, characterised by dense vegetation and substantial evapotranspiration, illustrate the cooling influence of natural land cover. The intermediate temperature range (36.2–37.8 °C) is observed in semi-urbanised transition zones, reflecting mixed land-use patterns with residual green spaces. By contrast, extensive hotspots are visible in the core and southern sectors of the city, with LST values frequently exceeding 39 °C, and in some cases surpassing 40.5 °C. The most intense thermal zones, reaching above 43 °C, are concentrated in heavily urbanised districts along both sides of the Bosporus, in the western expansion zones, and in industrial or high-density residential areas. These regions, marked by high impervious surface coverage, limited vegetation, and compact urban morphology, clearly demonstrate the intensification of the UHI.

The spatial distribution of heat patterns reflects Istanbul's topographic and landcover heterogeneity (Fig. A1). Northern cooler zones correspond to elevated, vegetated peripheries, while southern and Bosphorus-adjacent hotspots align with low-lying, densely built districts. The western districts exhibit large, contiguous patches of elevated temperatures (≥40 °C), which align with areas of rapid urban sprawl and industrial activity. The eastern sector, while also showing hotspots, reveals a more fragmented distribution, with cooler patches embedded within the urban fabric, likely due to the presence of residual forests. Overall, the heat pattern in

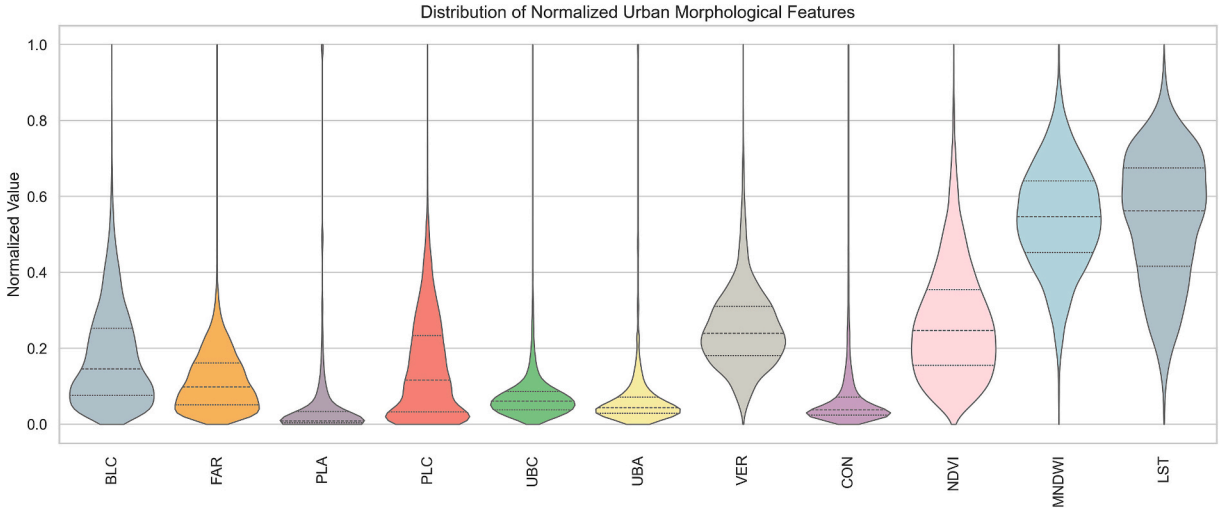



**Fig. 3.** Violin plot of the dependent variables based on normalised values.

Istanbul emphasises a pronounced intra-urban thermal disparity of over 10 °C between the coolest and hottest zones. The correspondence between morphological intensity and LST concentration underscores the critical role of land-use regulation, green–blue infrastructure, and spatial planning in mitigating urban heat.

### 3.2. Exploratory statistics of model variables

Model variables demonstrated substantial spatial heterogeneity across the study area (Fig. 3, Table 3). At the building level, BLC showed a wide range from 2 to 565 buildings (mean: 101.93), indicating pronounced variation in building density. VER exhibited values between 0.01 and 0.56 (mean: 0.15), reflecting heterogeneity in vertical development patterns, while CON ranged from 0.75 to 18.5 (mean: 1.92), suggesting considerable diversity in built fabric continuity.

Plot-related parameters also demonstrated broad variability. FAR ranged from 0.11 to 9.26 (mean: 1.15), while PLA extended from 61.5 to 58,460 $m^2$ (mean: 2837.86), highlighting the presence of both small and large parcels. PLC varied from 1 to 639 plots (mean: 95.65). Block-scale variables showed similarly wide ranges: UBC from 1 to 266 blocks (mean: 20.34) and UBA from 264 to 58,460 $m^2$ (mean: 4160.54). Environmental variables revealed apparent spatial differences. NDVI ranged from 0.04 to 0.39 (mean: 0.14), while MNDWI varied from −0.24 to −0.03 (mean: −0.12), reflecting variation in vegetation cover and surface water presence.

### 3.3. Performance evaluation of models

Four tree-based ML algorithms were benchmarked to evaluate both predictive accuracy and computational efficiency for modelling urban heat dynamics. Model hyperparameters were systematically optimised through Bayesian search (J. Wu et al., 2019) to maximise predictive performance while minimising overfitting. All models demonstrated strong predictive capabilities, with coefficient of determination ($R^2$) values ranging from 0.675 to 0.685 (Table 4). GBM achieved the highest accuracy ($R^2 = 0.685$, RMSE = 1.06 °C, MAE = 0.81 °C), followed closely by RF ($R^2 = 0.682$) and XGBoost ($R^2 = 0.677$). LGBM exhibited marginally lower performance ($R^2 = 0.675$) but demonstrated superior computational efficiency, requiring only 45.4 s for training compared to RF's 204.9 s – a 4.5× speedup. The adjusted $R^2$ values remained within 0.003 of their respective $R^2$ scores across all models, indicating minimal overfitting despite the high-dimensional feature space.

Critically, the narrow performance band ($\Delta R^2 = 0.010$) across models suggests that predictive accuracy is not the primary differentiator for urban heat applications. Instead, the benchmark reveals trade-offs between training efficiency and interpretability overhead. While LGBM achieved the fastest training time, its SHAP computation time (38.4 s) was comparable to that of XGBoost (29.7 s), whereas RF required substantially longer (401.3 s), representing an order-of-magnitude increase in explainability cost. These computational differences become particularly critical when scaling to city-wide analyses that require iterative model updates or real-time decision support.

### 3.4. Feature importance and SHAP analysis

The contribution of morphological parameters to urban heat predictions was systematically evaluated across four tree-based ML models using complementary feature importance metrics, including Gain and SHAP values, and, for XGBoost, Cover and Weight (Table 5). These metrics reveal a high degree of cross-model consistency in feature ranking, while also exposing meaningful differences in how individual algorithms attribute and distribute feature effects.

Across all models, NDVI consistently emerged as the most influential predictor of LST, exhibiting the highest SHAP values (ranging from 0.67 in RF to 0.82 in XGBoost). This persistent dominance confirms the central role of vegetation in moderating urban thermal conditions and highlights the effectiveness of surface greening strategies at both plot and neighbourhood scales. The stability of NDVI's negative SHAP signature across models further indicates that its cooling effect is robust to model architecture and learning strategy.

**Table 3**
Descriptive statistics of dependent and independent variables.

| Variables ($n = 7826$) | Unit | Variable Code | Mean | SD | Median | Min | Max |
|---|---|---|---|---|---|---|---|
| **Dependent Variable** | | | | | | | |
| Land Surface Temperature | °C | LST | 38.68 | 1.89 | 38.90 | 32.70 | 43.74 |
| **Independent Variables** | | | | | | | |
| Number of Buildings | count | BLC | 101.93 | 72.91 | 84 | 2 | 565 |
| Verticality | – | VER | 0.15 | 0.06 | 0.14 | 0.01 | 0.56 |
| Continuity | – | CON | 1.92 | 1.4 | 1.43 | 0.75 | 18.5 |
| Floor Area Ratio | – | FAR | 1.15 | 0.73 | 1.01 | 0.11 | 9.26 |
| Plot Area | number | PLA | 2837.86 | 7780.44 | 562.89 | 61.5 | 58,460.04 |
| Number of Plots | number | PLC | 95.65 | 83.01 | 75 | 1 | 639 |
| Number of Urban Blocks | number | UBC | 20.34 | 15.24 | 17 | 1 | 266 |
| Urban Block Area | number | UBA | 4160.54 | 5483.67 | 2800.82 | 264.49 | 58,460.02 |
| Normalised Difference Vegetation Index | – | NDVI | 0.14 | 0.05 | 0.13 | 0.04 | 0.39 |
| Modified Normalised Difference Water Index | – | MNDWI | −0.12 | 0.03 | −0.12 | −0.24 | −0.03 |

- variables without physical units.
SD: Standard deviation.

**Table 4**
Comparative performances of ML models.

| Model | Model Parameters | $R^2$ | Adj. $R^2$ | RMSE | MAE | Running time (seconds) |
|---|---|---|---|---|---|---|
| GBM | {'learning_rate': 0.01, 'max_depth': 11, 'max_features': 'sqrt', 'min_samples_leaf': 4, 'min_samples_split': 3, 'n_estimators': 380, 'subsample': 0.5} | 0.685 | 0.682 | 1.0601 | 0.8148 | Training times (BayesSearch + fit):104.2 SHAP + breakpoint times: 74.0 |
| XGBoost | {'colsample_bytree': 1.0, 'gamma': 0, 'learning_rate': 0.02152851229123662, 'max_depth': 6, 'min_child_weight': 1, 'n_estimators': 500, 'reg_alpha': 0, 'reg_lambda': 10, 'subsample': 0.5} | 0.677 | 0.675 | 1.073 | 0.818 | Training times (BayesSearch + fit):48.7 SHAP + breakpoint times: 29.7 |
| RF | {'max_depth': 18, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 500} | 0.682 | 0.679 | 1.065 | 0.8232 | Training times (BayesSearch + fit): 204.9 SHAP + breakpoint times: 401.3 |
| LGBM | {'colsample_bytree': 0.6591300944497073, 'learning_rate': 0.041609215111176115, 'max_depth': 8, 'min_child_samples': 12, 'n_estimators': 477, 'num_leaves': 63, 'reg_alpha': 10, 'reg_lambda': 10, 'subsample': 0.5885927226612261} | 0.675 | 0.673 | 1.075 | 0.8222 | Training times (BayesSearch + fit):45.4 SHAP + breakpoint times: 38.4 |

Notes:
$R^2$ and Adj. $R^2$ indicate predictive accuracy; RMSE and MAE quantify prediction error in °C
Training times include Bayesian hyperparameter optimisation + model fitting - SHAP + breakpoint times reflect explainability computation overhead
All models trained on 7826 hexagonal units (80% train, 20% validation)
Optimal hyperparameters identified through 50-iteration Bayesian search with 5-fold cross-validation.

BLC ranked second in importance across all models, with particularly strong attribution in XGBoost, where it recorded the highest Gain and Weight values, reflecting frequent selection during tree construction. The predominance of positive SHAP values for BLC indicates systematic heat intensification with increasing building density. As a building-scale indicator, its prominence highlights the sensitivity of urban heat patterns to localised structural concentration, especially in dense urban fabrics. Block-scale morphological indicators, notably CON and FAR, formed a secondary tier of influential predictors. Both variables demonstrated moderate-to-high SHAP values, underscoring the role of morphological compactness and vertical intensity in shaping microclimatic conditions. Their largely positive SHAP distributions suggest that increased spatial continuity and vertical massing tend to exacerbate heat accumulation, likely due to constrained airflow and enhanced material continuity at the block level.

MNDWI and VER exhibited moderate yet stable importance, reflecting their secondary but contextually significant roles. MNDWI, indicating water presence, maintained consistent SHAP values (0.15 in GBM and XGBoost), suggesting a modest but reliable cooling contribution. However, its SHAP distributions were more diffuse than those of NDVI or BLC, indicating that the thermal influence of surface water varies with spatial context and model structure. In GBM and LGBM, lower MNDWI values tended to cluster at positive SHAP values, implying thermal accumulation under reduced surface water presence, whereas this relationship appeared weaker and more context dependent in RF and XGBoost.

Plot-scale subdivision indicators (PLC, PLA) and block-scale area metrics (UBC, UBA) consistently ranked lower in importance, with SHAP values typically around or below 0.10. Nevertheless, these variables displayed notable inter-model variability. For instance, PLC showed substantially higher Gain values in XGBoost compared to RF (Gain: 0.10 in RF vs. 5.13 in XGBoost), suggesting that fine-grained land subdivision characteristics are intermittently captured by more complex ensemble architectures. Despite their lower rankings, the persistent inclusion of these features across models indicates subtle yet non-negligible contributions to urban heat dynamics, likely mediated through interactions with dominant morphological drivers rather than direct effects.

SHAP beeswarm plots (Fig. 4) provide further insight into the directionality and distribution of feature effects. NDVI exhibited a clear and stable negative SHAP pattern across all models, confirming its consistent cooling influence. In contrast, BLC and CON showed predominantly positive SHAP values, reflecting heat intensification associated with increasing building density and spatial continuity. MNDWI presented more dispersed SHAP patterns, reinforcing its context-dependent influence. In GBM and LGBM, lower MNDWI values clustered at positive SHAP values, indicating thermal accumulation associated with reduced surface water presence, whereas this relationship was weaker and more context dependent in RF and XGBoost. VER and FAR occupied an intermediate, context-dependent position, with narrower SHAP ranges than dominant features but broader distributions than subdivision metrics, indicating mixed and occasionally opposing thermal responses at extreme values. PLC, PLA, UBA, and UBC consistently showed narrow SHAP distributions across models, indicating a weak direct influence on LST. Larger parcels in RF and XGBoost were occasionally associated with slight negative SHAP values, suggesting marginal thermal benefits despite limited standalone explanatory power.

Model-specific differences were also evident in SHAP distributions. XGBoost produced the widest SHAP ranges for dominant predictors (NDVI, BLC), reflecting its granular sensitivity to morphological heterogeneity, consistent with its higher Weight and Cover scores (Table 5). LGBM similarly exhibited pronounced nonlinear responses, with extended SHAP tails beyond ±1.5, for key variables. In contrast, RF showed systematically compressed SHAP distributions across all variables, a pattern attributable to its ensemble-averaging mechanism, which smooths individual tree predictions and yields more conservative feature attributions.

**Table 5**
Feature importance scores across models.

| | Gain | SHAP | Cover | Weight |
|---|---|---|---|---|
| **GBM** | | | | |
| NDVI | 0.23 | 0.80 | – | – |
| BLC | 0.17 | 0.48 | – | – |
| CON | 0.09 | 0.28 | – | – |
| MNDWI | 0.07 | 0.15 | – | – |
| VER | 0.07 | 0.14 | – | – |
| FAR | 0.08 | 0.13 | – | – |
| PLC | 0.09 | 0.11 | – | – |
| PLA | 0.09 | 0.09 | – | – |
| UBC | 0.04 | 0.08 | – | – |
| UBA | 0.05 | 0.07 | – | – |
| **XGBoost** | | | | |
| NDVI | 28.06 | 0.82 | 489.29 | 3146.00 |
| BLC | 20.10 | 0.56 | 363.13 | 3770.00 |
| CON | 6.05 | 0.31 | 394.66 | 2728.00 |
| FAR | 6.13 | 0.15 | 323.66 | 3391.00 |
| MNDWI | 5.75 | 0.15 | 322.11 | 2946.00 |
| VER | 5.55 | 0.13 | 342.21 | 2634.00 |
| UBC | 3.49 | 0.13 | 288.80 | 1729.00 |
| UBA | 4.43 | 0.10 | 423.77 | 1871.00 |
| PLA | 4.16 | 0.08 | 282.68 | 2355.00 |
| PLC | 5.13 | 0.07 | 332.69 | 1426.00 |
| **RF** | | | | |
| NDVI | 0.24 | 0.67 | – | – |
| BLC | 0.19 | 0.46 | – | – |
| CON | 0.09 | 0.21 | – | – |
| PLC | 0.10 | 0.13 | – | – |
| FAR | 0.07 | 0.12 | – | – |
| VER | 0.07 | 0.11 | – | – |
| PLA | 0.10 | 0.09 | – | – |
| MNDWI | 0.06 | 0.09 | – | – |
| UBC | 0.03 | 0.05 | – | – |
| UBA | 0.05 | 0.04 | – | – |
| **LGBM** | | | | |
| NDVI | 6445 | 0.78 | – | – |
| BLC | 6066 | 0.50 | – | – |
| CON | 6529 | 0.28 | – | – |
| MNDWI | 7611 | 0.15 | – | – |
| FAR | 7152 | 0.15 | – | – |
| VER | 7117 | 0.14 | – | – |
| PLC | 4922 | 0.11 | – | – |
| UBC | 4026 | 0.10 | – | – |
| PLA | 5535 | 0.09 | – | – |
| UBA | 5773 | 0.09 | – | – |

Notes:
Feature importance quantified via multiple metrics: Gain (average improvement in loss function from splits on the feature), SHAP (mean absolute Shapley value indicating marginal contribution to predictions), Cover (total number of observations affected by splits on the feature, XGBoost only), and Weight (frequency of feature selection in tree splits, XGBoost only). Features ranked by SHAP value within each model. Higher values indicate a stronger influence on LST predictions.

A cross-model comparison of mean absolute SHAP values (Fig. 5) further quantifies both consensus and divergence in feature attribution. NDVI exhibited the highest mean SHAP across all models (0.67–0.83), with XGBoost showing maximum sensitivity, followed sequentially by GBM, LGBM, and RF. This consistent primacy, despite varying model architectures, underscores vegetation's universal importance in urban heat mitigation. BLC ranked second across all models, though with greater inter-model variability (0.46–0.56). XGBoost again attributed the highest importance, reflecting its granular sensitivity to structural density, while RF's more conservative estimate suggests its ensemble-averaging mechanism attenuates individual building-level signals.

The greatest inter-model divergence emerged for CON, where XGBoost, LGBM, and GBM attributed substantially higher importance ($SHAP_{mean} = 0.30$) than RF (0.21). This split likely reflects boosting-based algorithms' superior capacity to capture interaction effects, specifically, how building alignment compounds with reduced airflow and material continuity to amplify heat retention at the block scale. RF's lower attribution suggests its tree-averaging mechanism obscures such multi-feature synergies. FAR, MNDWI, and VER, by contrast, exhibited stable mid-tier importance, with mean SHAP values clustering tightly across models (0.11–0.15). This consistency, in contrast to CON's divergence, indicates that these features exert predictable, modulating influences that condition rather than drive urban heat dynamics. Their stable attribution across diverse learning mechanisms suggests straightforward, monotonic relationships with LST.

The cross-model SHAP comparison establishes XGBoost as the most sensitive model, consistently assigning the highest importance

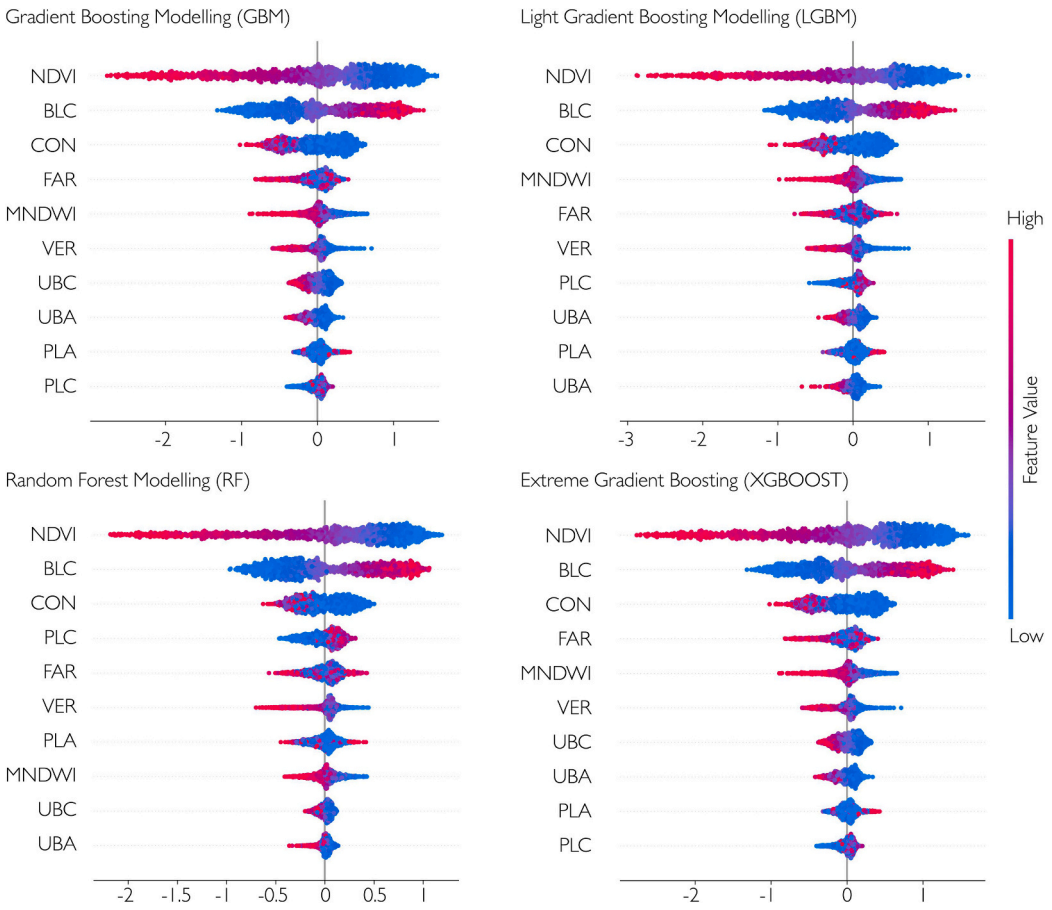



**Fig. 4.** SHAP Beeswarm Plots across four ML models. Each subplot displays SHAP value distributions for all features to visualise the feature importance and impact directionality. Points represent individual observations, with horizontal position indicating SHAP value (left = heat mitigating/cooling effect, right = heating effect) and colour denoting normalised feature magnitude (red = high, blue = low). Features ordered by mean absolute SHAP value (top = most important). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

scores, particularly for morphological features with complex spatial dependencies. GBM and LGBM exhibit comparable sensitivity profiles, though with slightly attenuated magnitudes. RF, conversely, systematically understates feature importance across the board, a pattern attributable to its ensemble-averaging mechanism, whereby predictions from numerous shallow, decorrelated trees are aggregated, resulting in smoother feature attributions compared with boosting algorithms that sequentially refine residuals.

### 3.5. *Threshold detection via piecewise regression*

Breakpoints in SHAP–feature relationships were identified via piecewise regression applied to each feature–model pair, to extract interpretable morphological thresholds relevant for planning and regulation. SHAP dependence values were ordered along normalised feature ranges (0–1) and smoothed using a Gaussian filter to suppress local noise while preserving systematic trends. The PELT algorithm (penalty = 1) then detected change-points in these smoothed trajectories. Up to two breakpoints per feature were retained and ranked by slope change magnitude, denoted as the primary (T1) and secondary (T2) thresholds. Thresholds identified in normalised space were converted to original units using variable-specific min-max transformations. Cross-model consensus was quantified as mean ± standard deviation across XGBoost, GBM, LGBM, and RF, and classified as High (SD <0.02), Moderate (0.02–0.04), or Low (>0.04) consensus (Table 6).

NDVI and MNDWI exhibited detectable primary thresholds with low cross-model consensus, alongside more variable secondary thresholds (Fig. 6). $T1_{NDVI}$ ranged from 0.148 to 0.274 across models (mean: 0.227 ± 0.057 normalised; 0.117 ± 0.019 original units), delineating the minimum vegetation density required for measurable evapotranspiration cooling. Below this threshold (~0.12 NDVI), surfaces function thermally as impervious heat sinks with minimal vegetative mitigation. However, the high inter-model variability (SD = 0.057) suggests that the precise inflection point is model-dependent, likely reflecting differences in how algorithms partition vegetation-temperature relationships. $T2_{NDVI}$ exhibited even greater variability (0.148–0.465 normalised; mean: 0.258 ± 0.148), also classified as low-consensus. LGBM's $T2_{NDVI}$ value was excluded as a statistical outlier, likely capturing a distinct saturation regime at

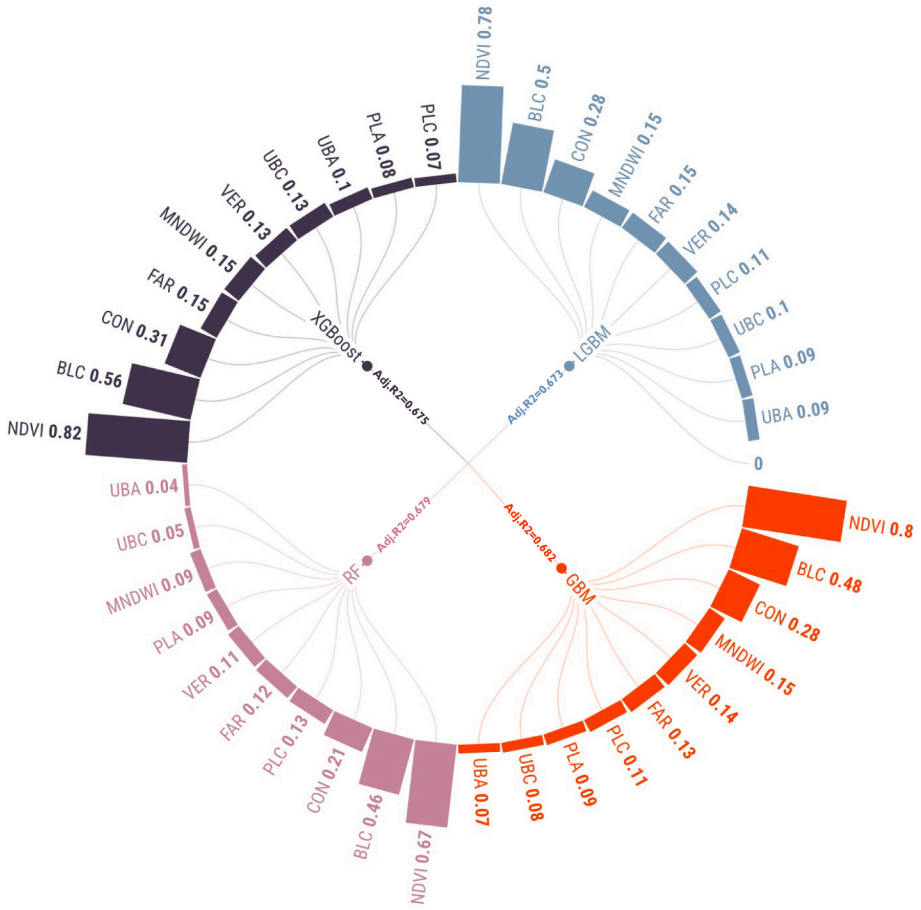



**Fig. 5.** Cross-model comparison of mean absolute SHAP values. Radial bar chart displaying mean absolute SHAP values for each feature across four models. Each feature occupies one radial segment, with four bars representing model-specific attributions. Bar length indicates feature importance magnitude.

higher vegetation densities. MNDWI displayed a parallel pattern with a low-consensus primary threshold (0.602 ± 0.042 normalised; −0.108 ± 0.009 original units) marking the transition where surface water presence initiates localised cooling, and a more dispersed secondary threshold (0.555 ± 0.099; −0.131 ± 0.021; low-consensus) reflecting model-specific sensitivities to moisture gradients.

Built-form indicators yielded multiple high- and moderate-consensus thresholds, reflecting their stable and interpretable relationships with urban heat (Fig. 6). VER displayed a high-consensus $T1_{VER}$ (0.238 ± 0.013; 0.140 ± 0.008H/A ratio), identifying the vertical development threshold above which canyon effects–shading and channelled ventilation–begin to offset heat retention from increased building mass. This threshold exhibited near-perfect cross-model agreement (SD = 0.013), providing a robust benchmark for regulating building height-to-footprint ratios in heat-vulnerable districts. $T2_{VER}$ exhibited low-consensus (0.161 ± 0.149; 0.098 ± 0.089H/A ratio), reflecting variable geometry-dependent responses.

At the block scale, UBA demonstrated exceptional stability with two high-consensus thresholds: $T1_{UBA} = 0.034 \pm 0.014$ (2248 ± 820 $m^2$) and $T2_{UBA} = 0.113 \pm 0.012$ (6883 ± 698 $m^2$). The primary threshold marks the block size beyond which reduced permeability begins amplifying heat retention, while the secondary threshold (SD = 0.012, the lowest across all features) indicates that large urban blocks uniformly intensify thermal stress once a critical size is exceeded, likely due to diminished internal ventilation and reduced sky view factors. UBC demonstrated moderate-consensus at its primary threshold ($T1_{UBC}$: 18.8 ± 5.9 blocks; SD = 0.022), indicating that block count per hexagon exerts systematic but algorithmically variable influence on heat patterns. $T2_{UBC}$ exhibited low-consensus (26.5 ± 11.2 blocks; SD = 0.042), suggesting secondary effects sensitive to local block configuration.

BLC showed strong cross-model agreement at its primary threshold: $T1_{BLC} = 0.158 \pm 0.026$ (90.9 ± 14.6 buildings; moderate-consensus), with three models (XGBoost, LGBM, GBM) converging exactly at 0.171 (98.3 buildings per hexagon). This threshold marks the density inflection point beyond which building count-driven heat accumulation intensifies nonlinearly. While achieving moderate rather than high consensus (SD = 0.026), the tri-model convergence and clear physical interpretation support its use as a density management benchmark in advisory planning frameworks. $T2_{BLC}$ showed greater variability (0.104 ± 0.049; 60.5 ± 27.6 buildings; low-consensus), suggesting context-dependent lower-density effects.

FAR presented two different thresholds: $T1_{FAR} = 0.091 \pm 0.054$ (0.94 ± 0.49 in original FAR values, low consensus) exhibited high

**Table 6**
Consensus morphological thresholds across ML models.

| Feature | ID | XGB | LGBM | GBM | RF | Mean ± SD (normalised) | Mean ± SD (original units) | Consensus |
|---|---|---|---|---|---|---|---|---|
| NDVI | T1 | 0.274 | 0.213 | 0.273 | 0.148 | 0.227 ± 0.057 | 0.117 ± 0.019 | Low |
| NDVI | T2 | 0.148 | 0.465* | 0.148 | 0.269 | 0.258 ± 0.148 | 0.131 ± 0.050 | Low |
| BLC | T1 | 0.171 | 0.171 | 0.171 | 0.119 | 0.158 ± 0.026 | 90.9 ± 14.6 bldg | Moderate |
| BLC | T2 | 0.119 | 0.171 | 0.062 | 0.064 | 0.104 ± 0.049 | 60.5 ± 27.6 bldg | Low |
| FAR | T1 | 0.098 | 0.164 | 0.037 | 0.063 | 0.091 ± 0.054 | 0.94 ± 0.49 | Low |
| FAR | T2 | 0.164 | 0.105 | 0.123 | 0.162 | 0.139 ± 0.029 | 1.38 ± 0.27 | Moderate |
| PLA | T1 | 0.011 | 0.092 | 0.096 | 0.006 | 0.051 ± 0.047 | 2804 ± 2466 $m^2$ | Low |
| PLA | T2 | 0.092 | 0.004 | 0.092 | 0.092 | 0.070 ± 0.044 | 4168 ± 2563 $m^2$ | Low |
| PLC | T1 | 0.257 | 0.022 | 0.257 | 0.158 | 0.174 ± 0.113 | 135 ± 74 parcels | Low |
| PLC | T2 | 0.022 | 0.152 | 0.033 | 0.224 | 0.108 ± 0.095 | 70 ± 60 parcels | Low |
| UBC | T1 | 0.094 | 0.038 | 0.068 | 0.068 | 0.067 ± 0.022 | 18.8 ± 5.9 blocks | Moderate |
| UBC | T2 | 0.038 | 0.094 | 0.125 | 0.128 | 0.096 ± 0.042 | 26.5 ± 11.2 blocks | Low |
| UBA | T1 | 0.053 | 0.019 | 0.034 | 0.031 | 0.034 ± 0.014 | 2248 ± 820 $m^2$ | High |
| UBA | T2 | 0.122 | 0.112 | 0.122 | 0.097 | 0.113 ± 0.012 | 6883 ± 698 $m^2$ | High |
| VER | T1 | 0.25 | 0.25 | 0.228 | 0.225 | 0.238 ± 0.013 | 0.140 ± 0.008 | High |
| VER | T2 | 0.151 | 0.385 | 0.08 | 0.027 | 0.161 ± 0.149 | 0.098 ± 0.089 | Low |
| CON | T1 | 0.148 | 0.024 | 0.08 | 0.027 | 0.070 ± 0.057 | 1.99 ± 1.01 | Low |
| CON | T2 | 0.034 | 0.083 | 0.021 | 0.081 | 0.055 ± 0.032 | 1.59 ± 0.57 | Moderate |
| MNDWI | T1 | 0.625 | 0.539 | 0.621 | 0.622 | 0.602 ± 0.042 | −0.108 ± 0.009 | Low |
| MNDWI | T2 | 0.539 | 0.688 | 0.538 | 0.453 | 0.555 ± 0.099 | −0.131 ± 0.021 | Low |

T# indicates threshold ranking by importance within each model (T1 = primary, T2 = secondary)
*LGBM NDVI_T2 (0.465) excluded from consensus calculation as statistical outlier
Consensus categories: High (Std < 0.02), Moderate (0.02–0.04), Low (>0.04)
Original units: bldg = buildings, $m^2$ = square meters
BLC T1 shows tri-model converge: XGB = LGBM = GBM at 0.171 (98.3 buildings)
Complete threshold inventory with importance scores available in Table A1.

inter-model variability, while $T2_{FAR} = 0.139 \pm 0.029$ (1.38 ± 0.27, moderate consensus) showed more stable detection. In this sense, $T2_{FAR}$ near FAR = 1.4 provides empirical support for regulatory distinctions between low-rise sprawl and mid-rise urban fabric, aligning with conventional planning benchmarks while grounding them in thermal performance data.

Plot-scale subdivision and continuity metrics predominantly exhibited low-consensus thresholds, indicating model-dependent detection (Fig. 6). PLA and PLC displayed high standard deviations for both T1 and T2 (PLA: 0.051 ± 0.047 and 0.070 ± 0.044; PLC: 0.174 ± 0.113 and 0.108 ± 0.095). CON showed $T1_{CON}$ with low-consensus (0.070 ± 0.057; 1.99 ± 1.01 original units) but $T2_{CON}$ achieved moderate-consensus (0.055 ± 0.032; 1.59 ± 0.57), suggesting that secondary continuity effects may exhibit more systematic thermal influence than primary effects. The predominantly low-consensus classification for plot-scale metrics likely stems from weak signal-to- noise ratios and complex interaction effects: parcel configuration and spatial continuity may exert thermal influence primarily through mediation of dominant drivers (e.g., how plot size conditions vegetation distribution or building density arrangements) rather than direct effects. In Fig. 6, these features display less sharply defined inflection points and more model-dependent breakpoint locations compared to UBA, BLC, or VER, providing visual confirmation of their algorithmic sensitivity.

## 4. Discussion

### *4.1. From models to policy: Bridging predictive AI and urban planning*

Machine learning models have demonstrated considerable predictive power in urban climate research (Ghorbany et al., 2024); however, their translation into planning practice remains constrained by a persistent gap between algorithmic accuracy and actionable insight. This study addresses this challenge through a dual-lens methodological framework. First, four tree-based algorithms were benchmarked to assess robustness, computational efficiency, and interpretability trade-offs. Second, SHAP-based explainability was integrated with piecewise regression to identify interpretable morphological thresholds with potential regulatory relevance.

The multi-model benchmark revealed a narrow performance band ($R^2 = 0.675$–$0.685$), indicating that predictive accuracy alone provides limited guidance for model selection in applied urban contexts (Li et al., 2025a; Okumus and Akay, 2025; Rashtian et al., 2025; Tang et al., 2025). Instead, key differences emerged in computational efficiency and interpretability overhead. While LGBM trained substantially faster than RF, SHAP computation times varied by an order of magnitude across models, with RF requiring considerably greater processing time than XGBoost or LGBM. Such disparities become operationally relevant when scaling analyses to city-wide applications involving iterative scenario testing or decision support workflows (Lu et al., 2023). Among the evaluated models; XGBoost demonstrated the highest sensitivity to morphological heterogeneity; capturing granular feature interactions (Huang et al., 2025) that were partially smoothed in other algorithms. This positions XGBoost as the optimal choice for research applications demanding maximal explanatory depth, while GBM and LGBM offer pragmatic alternatives for planning workflows prioritising a balance between computational efficiency and accuracy.

Beyond model comparison, coupling SHAP with change-point detection enabled the transition from global feature importance to

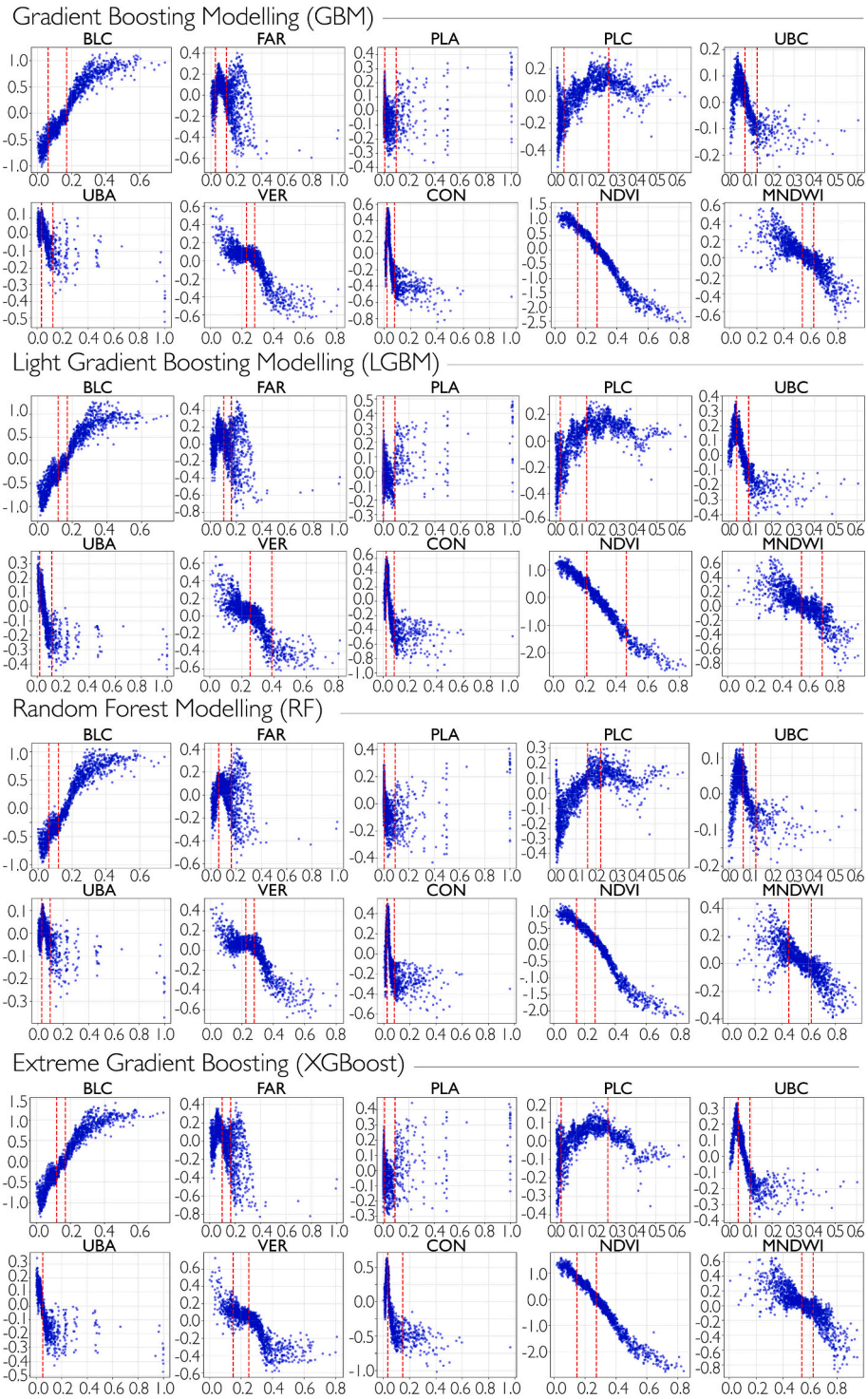



**Fig. 6.** SHAP–feature response curves with detected thresholds across four ML models. Each panel shows the relationship between the normalised observed value of a morphological feature (x-axis) and its SHAP contributions to LST prediction (y-axis). Vertical red lines mark the primary (T1) and secondary (T2) breakpoints identified by the PELT algorithm in the smoothed SHAP–feature curves. Threshold values corresponding to these breakpoints are summarised in Table 6. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

spatially explicit thresholds, a critical step toward operational utility. Rather than stating that vegetation or density influences urban heat in general terms, this framework quantifies where regime shifts occur. For example, NDVI>0.117 marks the onset of measurable evapotranspiration-driven cooling, while BLC > 90.9 buildings per hexagon (approximately 16 buildings per hectare) correspond to a nonlinear intensification of heat accumulation.

The NDVI threshold at 0.117 aligns with empirical findings identifying vegetation cooling onset at approximately NDVI 0.15 in built-up urban contexts (Zhang et al., 2026b), though the precise inflection point varies with climate zone, season, and vegetation type. In Mediterranean contexts with hot, dry summers, this threshold likely reflects the minimum vegetation density required to sustain measurable evapotranspiration under high solar radiation and limited moisture availability. Below this threshold, sparse or stressed vegetation provides negligible cooling, while above it, canopy shading and transpiration effects become thermally significant.

By classifying thresholds according to cross-model consensus (High: SD < 0.02; Moderate: 0.02–0.04; Low: >0.04), the study further distinguishes robust, model-invariant breakpoints suitable for universal regulation from context-dependent values requiring local calibration. This tiered threshold framework directly addresses limitations in prior urban heat studies that often report broad value ranges without identifying precise transition points (Chen et al., 2025a; Fan et al., 2024; Lin et al., 2025; Shen et al., 2024). For instance, while earlier studies noted that vegetation coverage or building density exhibit nonlinear relationships with LST (B. Liu et al., 2023; Yuan et al., 2024; Zou et al., 2024); few have quantified exact transition points with cross-model validation (Okumus and Akay, 2025). In the context of Istanbul, where intra-urban thermal disparities exceed 11 °C under sustained densification pressures (Table 3), such precision is not merely methodological refinement but a prerequisite for operational relevance. It enables planning guidance to move beyond general principles toward empirically grounded and locally interpretable thresholds that can inform zoning ordinances, design standards, and environmental impact assessments.

### 4.2. Translating thresholds into regulatory levels

The consensus-based classification of morphological thresholds provides a structured framework for aligning empirical confidence with appropriate planning instruments. Thresholds were grouped into three regulatory tracks, reflecting increasing degrees of uncertainty and corresponding policy flexibility.

The three-tier classification aligns regulatory instruments with empirical confidence, addressing a critical gap in evidence-based planning: distinguishing between findings suitable for mandatory enforcement versus those requiring contextual judgment. High-consensus thresholds (SD < 0.02) exhibit model-invariance, warranting direct incorporation into mandatory zoning codes and design standards citywide or within designated climate-sensitive zones. Moderate-consensus thresholds (0.02 ≤ SD < 0.04) inform advisory guidelines and performance- based standards, allowing flexibility in implementation where equivalent thermal performance can be demonstrated through context-specific design solutions. Low-consensus thresholds (SD > 0.04) serve as exploratory signals and sensitivity indicators, guiding targeted local studies rather than immediate regulatory action. This framework reduces risks of both under-regulation (ignoring robust findings) and over-prescription (mandating uncertain thresholds), enabling calibrated policy responses proportionate to empirical support. Also, it ensures that the proposed strategies are grounded in methodologically validated relationships, while maintaining appropriate caution regarding temporal and spatial generalisability.

#### 4.2.1. Group 1: High-consensus thresholds (universal regulatory benchmarks)

High-consensus thresholds (SD < 0.02) exhibited exceptional model agreement, indicating robust, model-invariant breakpoints suitable for direct incorporation into zoning codes and design standards.

*VER* exhibited high cross-model consensus at T1 = 0.238 ± 0.013 in normalised units (0.140 ± 0.008H/A ratio), identifying the point at which canyon-related effects, particularly shading and channelled ventilation, begin to offset the thermal penalties associated with increased building mass. This threshold corresponds to canyon geometries where building height approximates 1.4× street width. Below H/A = 0.14, canyon surfaces receive prolonged direct solar radiation with limited mutual shading, maximising daytime heat absorption through exposed façades and street surfaces. Above this threshold, taller buildings create afternoon shade on opposite façades and pavement while maintaining sufficient sky openness to avoid nighttime radiative trapping, explaining the counterintuitive cooling benefit of moderate vertical development. This finding aligns with canyon geometry studies in Mediterranean, subtropical, and temperate climates reporting that moderate to high aspect ratios (H/W > 1) provide superior thermal comfort during warm periods through enhanced shading (Chatzidimitriou and Axarli, 2017; Lachir, 2025; Park et al., 2024), though the precise optimal value varies with latitude, solar angle, street orientation, and seasonal priorities.

*UBA* exhibited two high-consensus thresholds, demonstrating exceptional model stability across block scales. The primary threshold (T1 = 0.034 ± 0.014 in normalised units; 2248 ± 820 $m^2$ in original units) marks the initial block size inflection point beyond which reduced permeability begins amplifying heat retention. This threshold provides guidance for maximum block dimensions in new developments and redevelopment zones, suggesting that blocks exceeding approximately 0.22 ha warrant enhanced permeability measures such as internal courtyards, pedestrian passages, or green corridors to maintain adequate ventilation pathways. The secondary threshold (T2 = 0.113 ± 0.012 in normalised; 6883 ± 698 $m^2$ in original units) achieved high-consensus, indicating that large urban blocks uniformly amplify heat retention once exceeding approximately 0.7 ha, likely reflecting reduced permeability and limited internal ventilation pathways. The primary threshold at 0.22 ha reflects the critical point where block permeability begins

constraining airflow circulation and heat dissipation. Smaller blocks maintain urban porosity through frequent street intersections and internal passages, facilitating convective heat exchange with ambient air. Beyond this threshold, reduced street frontage per unit area and longer internal travel distances creates ventilation dead zones where heat becomes trapped within the block interior. The secondary threshold at 0.7 ha marks severe permeability degradation, where block size eliminates effective cross-ventilation entirely. These findings resonate with block-scale morphology studies documenting nonlinear thermal responses to building density and permeability constraints in compact urban contexts (Gao et al., 2023; Li et al., 2025b), and with thermal comfort research confirming that block openness and internal ventilation pathways systematically govern pedestrian-level heat accumulation (J. Zhang et al., 2022), suggesting that block permeability represents a climate-relevant thermal mechanism across diverse urban settings, though optimal threshold dimensions depend on prevailing wind patterns and street network density.

#### 4.2.2. Group 2: Moderate-consensus thresholds (context-specific guidelines)

Moderate-consensus thresholds ($0.02 \leq SD < 0.04$) demonstrated robust cross-model agreement but exhibited sufficient variability to warrant context-sensitive application rather than universal mandates.

*BLC* showed strong cross-model agreement at its primary threshold ($T1 = 0.158 \pm 0.026$ in normalised units; $90.9 \pm 14.6$ buildings in original units), with three models (XGBoost, LGBM, GBM) converging exactly at 0.171 (98.3 buildings per hexagon). While achieving moderate rather than high consensus due to RF divergence ($SD = 0.026$), the tri-model convergence provides robust guidance for density management. For Istanbul's 5.85-ha hexagons, this translates to approximately 16 buildings per hectare. This threshold marks the density inflection point beyond which building count-driven heat accumulation intensifies nonlinearly. At this density (~16 buildings/ha), impervious surface coverage reaches approximately 49% of hexagon area, approaching the critical threshold where anthropogenic heat emissions, reduced surface albedo, and thermal mass accumulation begin compounding nonlinearly.

The threshold value is notably lower than density inflection points reported in comparable compact European contexts, where building density and impervious surface fraction exhibit nonlinear LST escalation at substantially higher development intensities (Gao et al., 2023; Roy et al., 2025), likely reflecting Istanbul's urban fabric with a mean building height around 15 m and higher solar radiation loads characteristic of the Mediterranean climate, which amplify heat accumulation at lower structural densities relative to temperate or high-density Asian contexts (Li et al., 2025b). In lower-rise contexts, horizontal expansion of impervious surfaces dominates thermal impacts, whereas in high-rise cities, vertical development can trigger canyon cooling effects that partially offset density penalties, raising the effective threshold. From a planning perspective, this threshold can inform advisory density targets in heat-vulnerable districts, implemented through performance-based standards that allow flexibility while ensuring thermal outcomes. Developments approaching or exceeding this level should incorporate enhanced mitigation measures such as increased vegetation, reflective materials, or façade-based cooling strategies, with equivalent thermal performance demonstrated through microclimate modelling or post-occupancy monitoring.

FAR's secondary threshold ($T2 = 0.139 \pm 0.029$ in normalised units; FAR≈1.38 in original units) achieved moderate consensus, marking a density regime shift at higher intensities where additional mitigation measures become increasingly important. This threshold provides empirical support for performance-based density standards in mid-rise contexts, suggesting that developments exceeding FAR≈1.4 warrant enhanced cooling strategies through green roofs, permeable surfaces, or strategic shading elements. The primary FAR threshold (T1) exhibited low consensus and is discussed in Group 3.

At the block scale, UBC's primary threshold ($T1 = 0.067 \pm 0.022$ in normalised units; $18.8 \pm 5.9$ blocks in original units) provides indicative guidance for urban fabric configuration and subdivision policy. This threshold highlights the role of block aggregation in shaping permeability and ventilation potential; however, its practical application requires careful calibration to local street network patterns, historical fabric constraints, and existing morphological conditions.

CON's secondary threshold ($T2 = 0.055 \pm 0.032$ in normalised units; $1.59 \pm 0.57$ in original units) achieved moderate consensus, suggesting that secondary continuity effects may exhibit more systematic thermal influence than primary effects. This threshold can inform guidelines for street network connectivity and building façade alignment in climate-sensitive planning, though local validation is recommended before regulatory application.

#### 4.2.3. Group 3: Low-consensus thresholds (exploratory, local calibration)

Low-consensus thresholds ($SD > 0.04$) exhibited substantial inter-model variability, indicating context-dependent or interaction-mediated effects rather than universal breakpoints.

PLA, PLC, and the primary thresholds of CON, NDVI, MNDWI, and FAR (T1) all fell into this category, with standard deviations frequently exceeding 50% of mean values. These features likely exert thermal influence primarily through modulation of dominant drivers (e.g., how plot size conditions vegetation distribution or building arrangements) rather than strong standalone effects. Instead of direct regulatory encoding, low-consensus thresholds should inform context-specific investigations when combined with high-consensus drivers. For instance, in districts approaching the BLC threshold (90 buildings/hexagon), PLC may influence whether that density produces uniform heat accumulation or spatially differentiated microclimates. Similarly, NDVI and MNDWI thresholds, while exhibiting model-dependent detection, can guide exploratory green-blue infrastructure strategies when paired with local

microclimate validation. Targeted local calibration studies can further refine these relationships before any policy application.

This three-tier classification provides a structured framework for translating ML-derived thresholds into planning instruments by aligning regulatory ambition with empirical confidence. Tier 1 thresholds represent robust, model-invariant breakpoints that warrant direct incorporation into evidence-informed zoning codes and design standards, either citywide or within designated climate-sensitive zones. Tier 2 thresholds inform advisory guidelines and performance-based standards, allowing flexibility in implementation where equivalent thermal performance can be demonstrated through context-specific design solutions. Tier 3 thresholds serve as exploratory signals and sensitivity indicators, guiding targeted local studies rather than immediate regulatory action. By explicitly linking thresholds to consensus levels, this framework addresses a critical gap in evidence-based urban planning, distinguishing between findings suitable for mandatory regulation and those requiring contextual judgment, while reducing the risks of both under-regulation and over-prescription.

### *4.3. Spatial heterogeneity and intervention strategies*

While threshold benchmarks provide general guidance on heat-sensitive urban morphology, their operational relevance depends on how these conditions are spatially distributed across the urban landscape. Observation-level SHAP force plot (Fig. 7) addresses this challenge by decomposing predicted thermal outcomes for individual hexagons into feature-specific contributions. This perspective reveals not only which morphological features influence urban heat but also where and to what extent they do so, enabling intervention strategies to be tailored to local morphological contexts rather than applied uniformly across the city. Force plot decomposition of five representative hexagons spanning Istanbul's thermal gradient reveals distinct morphological regimes associated with different LST conditions. Unlike global feature importance metrics, which summarise average effects across all observations, force plots expose pronounced intra-urban heterogeneity. The same morphological feature, such as building density, may contribute to cooling in one location through canyon-related shading and ventilation effects, while intensifying heat elsewhere through thermal mass accumulation, depending on surrounding urban form and feature interactions.

#### *4.3.1. Group 1: Very cool zones (LST = 35.1 °C, −4.0 °C from base prediction)*

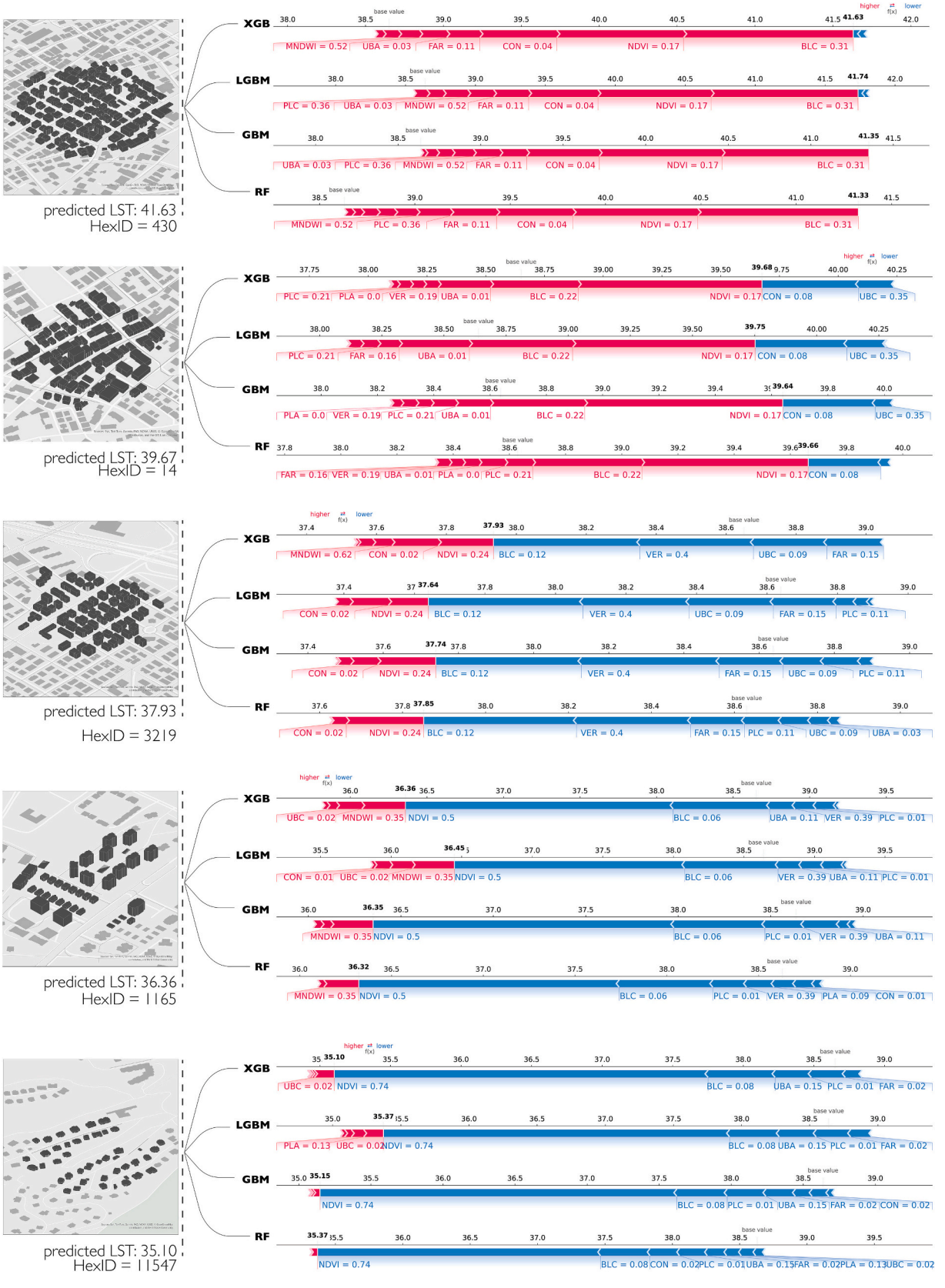
The coolest hexagons exhibit morphological configurations characterised by strong cooling synergies. Vegetation dominates thermal mitigation, with NDVI values (0.35) far exceeding the activation threshold, resulting in substantial cooling contributions (−1.8 °C). Simultaneously, building count remains well below the density inflection point (BLC = 0.05; T1 = 0.171), limiting heat accumulation from impervious surfaces and structural mass. Water proximity (MNDWI = 0.68, above T1 = -0.108) provides additional modest cooling (−0.4 °C), exemplifying multi-factor mitigation synergies. These hexagons represent optimal morphological templates for climate-sensitive design. Spatially, Very Cool zones are concentrated in Istanbul's northern peripheries and Bosporus-adjacent districts (Fig. 2), where historical preservation, topographic constraints, and lower development pressures have maintained high vegetation coverage. From a planning perspective, these areas function as thermal refuges. Protective zoning approaches that limit further densification may help preserve their cooling capacity, while their favourable thermal conditions suggest suitability for heat-sensitive land uses such as public gathering spaces, elderly care facilities, and emergency cooling centres during heat waves (Erdem Okumus, 2025). Enhancing connectivity between these zones through green corridors may further support city-scale ventilation and cooling flows.

#### *4.3.2. Group 2: Cool zones (LST = 36.4 °C, −1.1 °C from base prediction)*

Cool hexagons exhibit net cooling, though through different morphological pathways than the Very Cool zones. While NDVI remains above the threshold (0.23), its cooling contribution (−0.8 °C) is tempered by moderate building density (BLC = 0.14). The distinguishing feature is elevated MNDWI (0.62), indicating the proximity or presence of blue infrastructure, which contributes an additional −0.5 °C cooling. These patterns demonstrate that different combinations of threshold-compliant features can yield comparable thermal outcomes, with partial deficiencies in one variable compensated by strengths in another. Spatially, Cool zones often form transitional belts between vegetated peripheries and denser urban cores, as well as linear corridors along waterfront districts. In planning terms, these areas are particularly well-suited to integrated blue-green infrastructure strategies. Opportunities for increasing vegetation are constrained by existing buildings or paved surfaces, water-based interventions, like bioswales, retention ponds, and permeable pavements with subsurface storage may provide effective alternatives. The observed cooling associated with combined vegetation and water effects suggests that such morphological synergies warrant prioritisation in investment and design decisions.

#### *4.3.3. Group 3: Moderate zones (LST = 37.9 °C, +0.2 °C from base prediction)*

Moderate hexagons occupy a critical position near key morphological thresholds. Building density (BLC = 0.16) approaches but does not exceed the inflection point (T1 = 0.171), while vegetation levels (NDVI = 0.11) remain just below the activation threshold (T1 = 0.117). This near equilibrium produces offsetting heating and cooling contributions, resulting in minimal deviation from the city-wide mean temperature. These zones are particularly sensitive to intervention. Relatively small morphological adjustments (i.e., modest increases in vegetation) can shift thermal regimes disproportionately. For instance, increasing NDVI by 0.01–0.02, achievable through targeted street tree planting (approximately 3–5 trees per hexagon based on canopy estimates), would likely cross the




*(caption on next page)*

**Fig. 7.** Morphological heat decomposition via SHAP force plots based. SHAP force plots visualise feature-level contributions to predicted LST for five hexagons representing natural break categories from Fig. 2 (Very Cool–HexID 11,547: 32.7–36.2 °C; Cool–HexID 1165: 36.2–37.8 °C; Moderate–HexID 3219: 37.8–39.0 °C; Hot–HexID 14: 39.0–40.5 °C; Very Hot–HexID 430: >40.5 °C). Left column: Maps displaying 3D-built fabric (black) within hexagonal units. Right column: XGBoost SHAP force plots starting from base prediction (38.68 °C, dataset mean), where each horizontal bar represents a morphological or spectral feature. Bar width = contribution magnitude; red = heating effect (rightward), blue = heat mitigating/cooling effect (leftward). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

activation threshold and generate measurable cooling benefits. Conversely, allowing further densification beyond 90 buildings per hexagon risks pushing these areas beyond the density inflection point, triggering nonlinear heat intensification. From a planning perspective, these areas are well-suited to preventative, performance-based approaches aimed at stabilising thermal conditions before unfavourable trajectories become locked in.

#### *4.3.4. Group 4: Hot zones (LST = 39.7 °C, +0.6 °C from base prediction)*

Hot hexagons exhibit threshold violations beginning to compound. Building density (BLC = 0.22) exceeds the inflection point by 30%, contributing +1.2 °C heating, while vegetation levels (NDVI = 0.08) fall well below the activation threshold, providing insufficient cooling (−0.3 °C) to offset densification effects. Additional heating arises from spatial continuity (CON = 0.12), suggesting that building arrangements further amplify thermal retention by limiting airflow and increasing material continuity. These zones are concentrated in industrial corridors, high-density residential districts and commercial centres characterised by extensive impervious surfaces. In such contexts, advisory measures alone are unlikely to be sufficient. Retrofit-oriented strategies (i.e., green roofs, façade greening, and intensified street tree programmes) are likely to play a central role, particularly where ground-level space is constrained. Incentive-based instruments, including tax reductions or development bonuses, may support broader adoption. New developments in these areas must be assessed against performance criteria demonstrating net thermal improvement, with force plot–based diagnostics used to substantiate mitigation effectiveness.

#### *4.3.5. Group 5: Very hot zones (LST = 41.6 °C, +1.2 °C from base prediction)*

Very Hot hexagons represent acute thermal vulnerability hotspots, exhibiting severe threshold violations across multiple features. Building density (BLC = 0.31) nearly doubles the inflection point, contributing +2.1 °C of heating, while vegetation is critically deficient (NDVI = 0.07), providing negligible cooling. Elevated spatial continuity (CON = 0.15) and floor area ratio (FAR = 0.18) further compound heating through increased thermal mass and reduced permeability. These zones cluster in Istanbul's densest urban districts, including the historic core, rapidly developed peripheral areas, and industrial waterfronts. From a planning standpoint, these areas require a combination of immediate relief measures and longer-term structural transformation. Short-term responses may include temporary cooling infrastructure and surface treatments to reduce heat absorption. Medium-term strategies could focus on aggressive vegetation enhancement and mandatory green roof retrofits, while longer-term redevelopment approaches may seek to reduce excessive building counts, introduce permeability corridors, and strategically increase verticality to leverage canyon-related cooling effects while maintaining overall density controls (Erdem Okumus and Terzi, 2022). Given the magnitude of thermal deviation; careful attention to social equity and displacement risks is essential to ensure that adaptation measures benefit existing residents (Erdem Okumus, 2025).

### *4.4. Limitations and future research*

This study is shaped by a set of methodological choices that warrant careful consideration. First, the threshold detection relies on multi-modal comparison of a single temporal snapshot capturing peak summer conditions. To enhance discussions of cross-model robustness, future work should also test temporal stability based on seasonal heat variations. Second, as the modelling approach is inherently correlational, the identified thresholds indicate regime shifts in relationships rather than causal effects. Third, while the consensus-based framework evaluates robustness across different modelling approaches, it does not explicitly assess sensitivity to data sampling variability. Future research could incorporate bootstrap-based stability analysis to further examine the consistency of identified thresholds under different data compositions, spatial resolutions, and contextual factors. Spatial transferability to cities with different climates, urban fabrics, or thermal regimes requires local recalibration, though the methodological framework and consensus logic remain generalisable. Future work integrating quasi-experimental designs or physics-based microclimate simulations could strengthen both the causal understanding and geographic transferability of the proposed thresholds.

## 5. Conclusion

This study demonstrated how ML–based urban heat modelling can be translated into planning-relevant knowledge through systematic model benchmarking and consensus-based threshold detection. Four tree-based algorithms exhibited only marginal differences in predictive accuracy ($R^2$ = 0.675–0.685), with operational considerations – computational efficiency, interpretability

overhead, and sensitivity to spatial heterogeneity – proving more decisive than accuracy alone in applied planning contexts. LGBM is trained 4.5 times faster than RF, making it well-suited for large-scale or iterative workflows, whereas the greater morphological sensitivity of XGBoost supports research applications that prioritise explanatory depth and interpretability. This finding suggested that algorithm selection for urban climate applications should balance prediction-interpretation trade-offs rather than pursue marginal accuracy gains, with practical implications extending beyond Istanbul to cities employing ML for climate-responsive planning.

By integrating SHAP with change-point detection, we identified 80 morphological thresholds with varying levels of cross-model consensus. Three thresholds achieved high cross-model agreement (SD $< 0.02$): verticality at H/A $= 0.14$, and urban block area at 2248 $m^2$ and 6883 $m^2$. These model-invariant breakpoints support direct incorporation into evidence-informed zoning codes. Four thresholds demonstrated moderate consensus (SD 0.02–0.04): building count at 90.9 buildings/hexagon (~16/ha), FAR at 1.38, block count at 18.8 blocks, and continuity at 1.59, informing performance-based standards and context-sensitive guidance rather than fixed mandates. The remaining 13 thresholds exhibited low consensus, highlighting the importance of local calibration and interaction effects. This consensus-based classification provides a transparent basis for tiered regulatory frameworks calibrated to empirical confidence.

The identified thresholds captured critical regime shifts in morphology-heat relationships with distinct physical mechanisms. The verticality threshold (H/A $= 0.14$) corresponds to canyon geometries where building height approximates 1.4× street width, marking the onset of afternoon shading effects that begin to offset thermal mass penalties while maintaining sufficient sky openness to avoid nighttime radiative trapping. The building density threshold (~16 buildings/ha) reflected the point where impervious surface coverage reaches approximately 49% of hexagon area, triggering nonlinear intensification as anthropogenic heat emissions, reduced albedo, and thermal mass accumulation begin compounding. Urban block area thresholds at 0.22 ha and 0.7 ha indicate progressive permeability constraints: smaller blocks maintain urban porosity through frequent street intersections, while larger blocks create ventilation dead zones where heat becomes trapped within block interiors. These physically grounded thresholds enable planners to understand not merely where thermal effects intensify, but why – strengthening the empirical foundation for climate-responsive design standards.

The threshold values identified reflect Istanbul's specific morpho-climatic context and require careful interpretation when transferring to other cities. The building density threshold is notably lower than values likely applicable to compact European cities or high-density Asian contexts, likely reflecting the higher solar radiation loads characteristic of Mediterranean climates. In high-rise cities, vertical development can trigger canyon cooling effects that partially offset density penalties, potentially raising the effective threshold. Similarly, the verticality threshold (H/A $= 0.14$) may vary with latitude, solar angle, and prevailing street orientations, though canyon geometry studies across Mediterranean and subtropical climates report comparable optimal aspect ratios (1.2–1.6). Block size thresholds appeared more climate-invariant, with permeability mechanisms operating similarly across diverse contexts, though optimal dimensions depend on prevailing wind patterns and street network density. While specific threshold values require local recalibration, the methodological framework – multi-model benchmarking, SHAP-based attribution, hierarchical change-point detection, and consensus classification – remains generalisable globally by substituting local morphological data and climate conditions.

This work advances urban climate research through four key contributions. First, it moves beyond single-model traditions by benchmarking four algorithms for accuracy-efficiency-interpretability trade-offs, exposing performance-transparency tensions that remained implicit in prior work. Second, it reframes nonlinearity through consensus-validated thresholds across independent models, transforming approximate value ranges into precise morphological breakpoints. Third, it integrates SHAP with piecewise modelling to identify where and how morphological influences shift between thermal regimes, producing nuanced explanatory maps of spatio-thermal interactions. Finally, the framework demonstrates replicability beyond urban heat, offering a generalisable methodological pipeline for translating explainable AI into evidence-based planning instruments applicable to broader urban climate challenges where robust and interpretable decision-support approaches are required.

## CRediT authorship contribution statement

**Mert Akay:** Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Writing – review & editing, Writing – original draft. **Deniz Erdem Okumuş:** Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization, Writing – review & editing, Writing – original draft.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Appendices

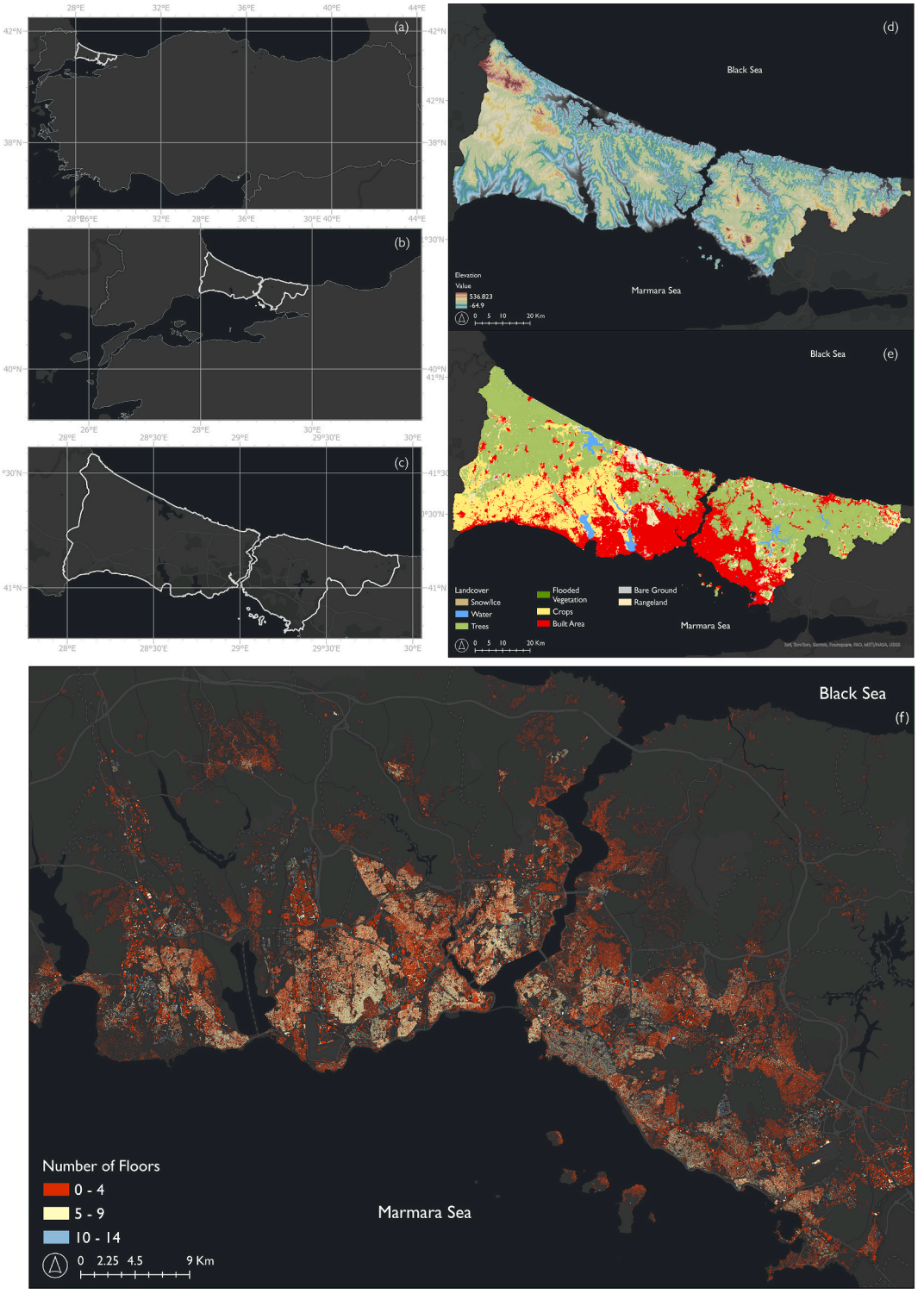



**Fig. A1.** The location of Istanbul within Türkiye in different scales (a-c), elevation distribution across the metropolitan area (d) (Copernicus, 2025); landcover map (e) (ESRI, 2023), and building floor numbers (f).

**Table A1**
Threshold transformation of each parameter from model values to actual values.

| Feature | Model | Min | Max | Threshold 1 | Actual Threshold 1 | Importance 1 | Threshold 2 | Actual Threshold 2 | Importance 2 |
|---|---|---|---|---|---|---|---|---|---|
| BLC | XGB | 2.00 | 565.00 | 0.17 | 98.27 | 1.24 | 0.12 | 69.00 | 1.11 |
| FAR | XGB | 0.11 | 9.26 | 0.10 | 1.01 | 1.64 | 0.16 | 1.61 | 0.62 |
| PLA | XGB | 61.50 | 58,460.04 | 0.01 | 703.88 | 1.20 | 0.09 | 5434.17 | 0.08 |
| PLC | XGB | 1.00 | 639.00 | 0.26 | 164.97 | 1.82 | 0.02 | 15.04 | 1.01 |
| UBC | XGB | 1.00 | 266.00 | 0.09 | 25.91 | 0.59 | 0.04 | 11.07 | 0.21 |
| UBA | XGB | 264.49 | 58,460.02 | 0.05 | 3348.85 | 1.92 | | | |
| VER | XGB | 0.01 | 0.56 | 0.25 | 0.15 | 0.59 | 0.15 | 0.09 | 0.41 |
| CON | XGB | 0.75 | 18.50 | 0.03 | 1.35 | 0.29 | 0.15 | 3.38 | 0.17 |
| NDVI | XGB | 0.04 | 0.39 | 0.27 | 0.14 | 1.11 | 0.15 | 0.09 | 0.82 |
| MNDWI | XGB | −0.24 | −0.03 | 0.63 | −0.11 | 1.23 | 0.54 | −0.13 | 1.10 |
| BLC | RF | 2.00 | 565.00 | 0.12 | 69.00 | 0.81 | 0.06 | 38.03 | 0.66 |
| FAR | RF | 0.11 | 9.26 | 0.06 | 0.69 | 1.67 | 0.16 | 1.59 | 0.77 |
| PLA | RF | 61.50 | 58,460.04 | 0.01 | 411.89 | 1.63 | 0.09 | 5434.17 | 0.19 |
| PLC | RF | 1.00 | 639.00 | 0.16 | 101.80 | 0.49 | 0.22 | 143.91 | 0.37 |
| UBC | RF | 1.00 | 266.00 | 0.07 | 19.02 | 1.22 | 0.13 | 34.92 | 0.31 |
| UBA | RF | 264.49 | 58,460.02 | 0.03 | 2068.55 | 1.34 | 0.10 | 5909.46 | 0.42 |
| VER | RF | 0.01 | 0.56 | 0.28 | 0.16 | 1.53 | 0.23 | 0.13 | 1.34 |
| CON | RF | 0.75 | 18.50 | 0.03 | 1.23 | 0.50 | 0.08 | 2.19 | 0.37 |
| NDVI | RF | 0.04 | 0.39 | 0.15 | 0.09 | 0.23 | 0.27 | 0.13 | 0.21 |
| MNDWI | RF | −0.24 | −0.03 | 0.62 | −0.11 | 0.93 | 0.45 | −0.14 | 0.26 |
| BLC | LGBM | 2.00 | 565.00 | 0.12 | 69.00 | 1.61 | 0.17 | 98.27 | 1.60 |
| FAR | LGBM | 0.11 | 9.26 | 0.11 | 1.07 | 0.95 | 0.16 | 1.61 | 0.63 |
| PLA | LGBM | 61.50 | 58,460.04 | 0.00 | 295.09 | 0.52 | 0.09 | 5434.17 | 0.04 |
| PLC | LGBM | 1.00 | 639.00 | 0.02 | 15.04 | 1.29 | 0.15 | 97.98 | 0.70 |
| UBC | LGBM | 1.00 | 266.00 | 0.04 | 11.07 | 1.09 | 0.09 | 25.91 | 0.97 |
| UBA | LGBM | 264.49 | 58,460.02 | 0.02 | 1370.21 | 0.46 | 0.11 | 6782.39 | 0.31 |
| VER | LGBM | 0.01 | 0.56 | 0.25 | 0.15 | 0.80 | 0.39 | 0.22 | 0.32 |
| CON | LGBM | 0.75 | 18.50 | 0.02 | 1.18 | 0.71 | 0.08 | 2.22 | 0.27 |
| NDVI | LGBM | 0.04 | 0.39 | 0.21 | 0.11 | 1.46 | 0.47 | 0.20 | 0.19 |
| MNDWI | LGBM | −0.24 | −0.03 | 0.54 | −0.13 | 0.92 | 0.69 | −0.10 | 0.76 |
| BLC | GBM | 2.00 | 565.00 | 0.17 | 98.27 | 1.29 | 0.06 | 36.91 | 0.57 |
| FAR | GBM | 0.11 | 9.26 | 0.12 | 1.24 | 1.52 | 0.04 | 0.45 | 1.04 |
| PLA | GBM | 61.50 | 58,460.04 | 0.01 | 411.89 | 1.97 | 0.10 | 5667.76 | 0.16 |
| PLC | GBM | 1.00 | 639.00 | 0.26 | 164.97 | 1.68 | 0.03 | 22.05 | 0.38 |
| UBC | GBM | 1.00 | 266.00 | 0.07 | 19.02 | 1.19 | 0.13 | 34.13 | 0.35 |
| UBA | GBM | 264.49 | 58,460.02 | 0.03 | 2243.14 | 1.04 | 0.12 | 7364.35 | 0.36 |
| VER | GBM | 0.01 | 0.56 | 0.28 | 0.16 | 1.09 | 0.23 | 0.14 | 0.75 |
| CON | GBM | 0.75 | 18.50 | 0.08 | 2.17 | 0.27 | 0.02 | 1.12 | 0.16 |
| NDVI | GBM | 0.04 | 0.39 | 0.27 | 0.14 | 0.81 | 0.15 | 0.09 | 0.62 |
| MNDWI | GBM | −0.24 | −0.03 | 0.62 | −0.11 | 1.06 | 0.54 | −0.13 | 0.70 |

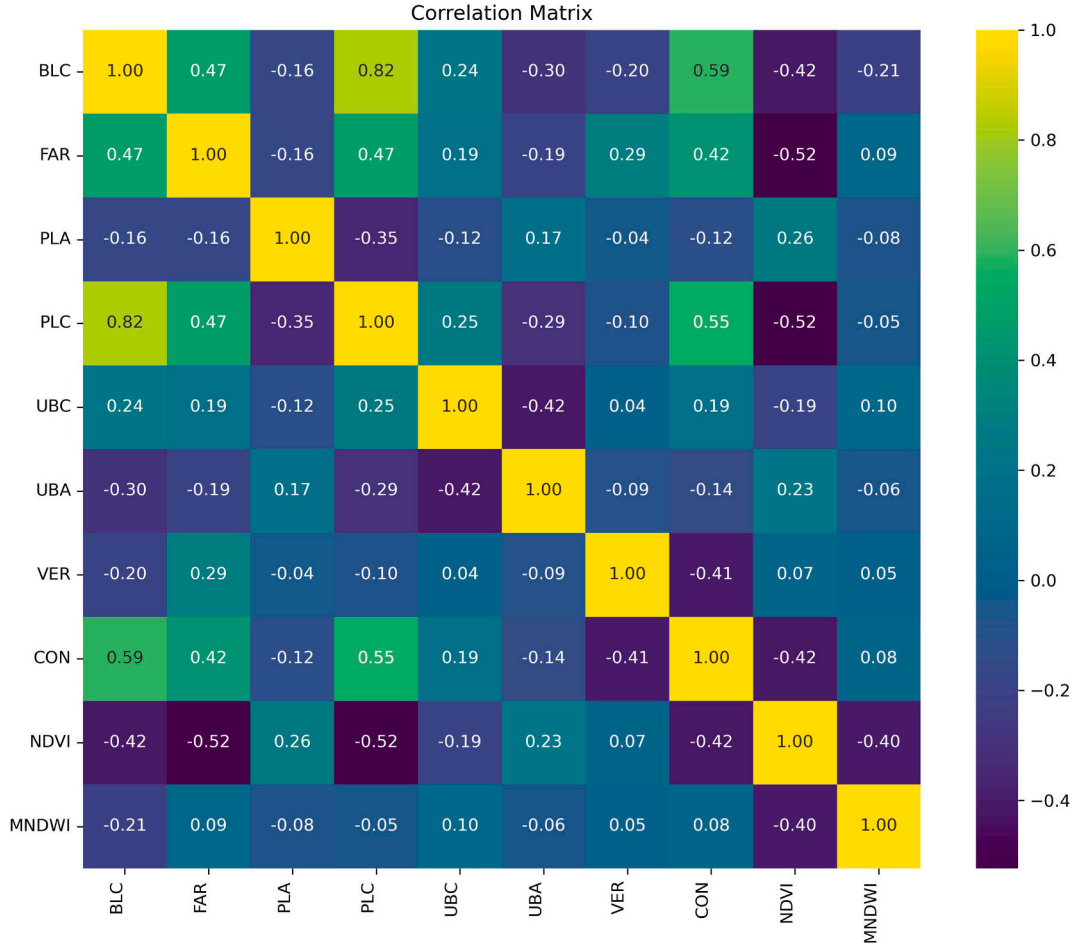



**Fig. A2.** Pearson correlation matrix for morphological predictors (n = 7826 hexagons). Values range from −1 (perfect negative correlation) to +1 (perfect positive correlation).

## Data availability

Data will be made available on request.

## References


Akay, M., Çalışkan, O., 2025. A parametric approach to plot-based urban design: a climate-responsive algorithmic control for the generation of urban block. Urban Des. Int. 1–22.

Aslani, A., Sereshti, M., Sharifi, A., 2025. Urban heat island mitigation in Tehran: district-based mapping and analysis of key drivers. Sustain. Cities Soc. 125, 106338.

Badaro-Saliba, N., Adjizian-Gerard, J., Zaarour, R., Najjar, G., 2021. LCZ scheme for assessing urban Heat Island intensity in a complex urban area (Beirut, Lebanon). Urban Clim. 37. https://doi.org/10.1016/j.uclim.2021.100846.

Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F., 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. Sci. Data 5 (1), 180214.

Çalışkan, O., Mashhoodi, B., Akay, M., 2022. Morphological indicators of the building fabric: towards a metric typomorphology. J. Urban Des. 1–30.

Chatzidimitriou, A., Axarli, K., 2017. Street canyon geometry effects on microclimate and comfort; a case study in Thessaloniki. Procedia Environ. Sci. 38, 643–650.

Chen, L., Guo, G., 2025. Exploring the nonlinear interactions and threshold effects of urban building morphology and green space on land surface temperature in high-density areas: a cross-city comparative study. Sustain. Cities Soc. 136, 107069.

Chen, G., Shi, Y., Wang, R., Ren, C., Ng, E., Fang, X., Ren, Z., 2022. Integrating weather observations and local-climate-zone-based landscape patterns for regional hourly air temperature mapping using machine learning. Sci. Total Environ. 841. https://doi.org/10.1016/j.scitotenv.2022.156737.

Chen, Y., Ma, W., Shao, Y., Wang, N., Yu, Z., Li, H., Hu, Q., 2025a. The impacts and thresholds detection of 2D/3D urban morphology on the heat island effects at the functional zone in megacity during heatwave event. Sustain. Cities Soc. 118, 106002.

Chen, Y., Wang, Y., Luo, X., Pei, L., Li, W., 2025b. Nonlinear force analysis of urban morphology and building heat emission based on multi-scale microclimate prediction. Energ. Buildings 346, 116146.

Chien, Y.-M.C., Carver, S., Comber, A., 2020. Using geographically weighted models to explore how crowdsourced landscape perceptions relate to landscape physical characteristics. Landsc. Urban Plan. 203, 103904.

Copernicus, 2025. Copernicus DEM - Global and European Digital Elevation Model. https://dataspace.copernicus.eu/explore-data/data-collections/copernicus-contributing-missions/collections-description/COP-DEM.
Ding, L., Xiao, X., Wang, H., 2025. Temporal and spatial variations of urban surface temperature and correlation study of influencing factors. Sci. Rep. 15 (1), 914.
Duan, Y., Yuan, C., Mao, X., Zhao, J., Ma, N., 2023. Influence of the built environment on taxi travel demand based on the optimal spatial analysis unit. PLoS One 18 (10), e0292363.
Erdem Okumus, D., 2025. Decoding nonlinear dynamics in urban heat vulnerability: a multi-model framework for socio-spatial responses to extreme heat. Sustain. Cities Soc. 135, 106995. https://doi.org/10.1016/j.scs.2025.106995.
Erdem Okumus, D., Terzi, F., 2021. Evaluating the role of urban fabric on surface urban heat island: the case of Istanbul. Sustain. Cities Soc. 73. https://doi.org/10.1016/j.scs.2021.103128.
Erdem Okumus, D., Terzi, F., 2022. Reconsidering urban densification for microclimatic improvement: planning and design strategies for Istanbul. ICONARP Int. J. Architect. & Plann. 10 (2), 660–687.
ESRI, 2023. ESRI Sentinel-2 10-Meter Land Use/Land Cover. https://livingatlas.arcgis.com/landcoverexplorer/#mapCenter=-62.54228%2C-33.91918%2C11.00&mode=step&timeExtent=2017%2C2024&year=2023.
Fan, R., Wu, Y., Chen, Q., Wang, Y., Li, L., Shi, D., Xu, R., Xia, Y., Cheng, Y., 2024. Characterizing urban heat islands in karst areas-the case of Kunming and Guiyang in Southwest China. Front. Built Environ. 10. https://doi.org/10.3389/fbuil.2024.1457295.
Gao, Y., Zhao, J., Han, L., 2023. Quantifying the nonlinear relationship between block morphology and the surrounding thermal environment using random forest method. Sustain. Cities Soc. 91, 104443.
Ghorbany, S., Hu, M., Yao, S., Wang, C., 2024. Towards a sustainable urban future: a comprehensive review of urban Heat Island research technologies and machine learning approaches. In: Sustainability (Switzerland). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/su16114609 (Vol. 16, Number 11).
Gu, X., Wu, Z., Liu, X., Qiao, R., Jiang, Q., 2024. Exploring the nonlinear interplay between urban morphology and nighttime thermal environment. Sustain. Cities Soc. 101, 105176.
Guha, S., Govil, H., Dey, A., Gill, N., 2018. Analytical study of land surface temperature with NDVI and NDBI using Landsat 8 OLI and TIRS data in Florence and Naples city, Italy. European J. Remote Sens. 51 (1), 667–678.
Hou, F., Taib, N., Marzbali, M.H., Abdullah, A., 2026. Threshold effects of urban built-up and green space morphology on seasonal daytime–night-time land surface temperature. Sustain. Cities Soc. 140,, 107269.
Huang, C., Liu, K., Ma, T., Xue, H., Wang, P., Li, L., 2025. Analysis of the impact mechanisms and driving factors of urban spatial morphology on urban heat islands. Sci. Rep. 15 (1), 18589.
Jiang, X., Ma, T., Liang, J., Wang, Y., 2025. The greener the cooler? Nonlinear thresholds in the cooling effects of urban blue-green spaces and implications for sustainable planning in the Yangtze River Delta. Urban Clim. 64, 102679.
Jiménez-Muñoz, J.C., Sobrino, J.A., Gillespie, A., Sabol, D., Gustafson, W.T., 2006. Improved land surface emissivities over agricultural areas using ASTER NDVI. Remote Sens. Environ. 103 (4), 474–487.
Jiménez-Muñoz, J.C., Sobrino, J.A., Plaza, A., Guanter, L., Moreno, J., Martínez, P., 2009. Comparison between fractional vegetation cover retrievals from vegetation indices and spectral mixture analysis: Case study of PROBA/CHRIS data over an agricultural area. Sens 9 (02), 768–793.
Killick, R., Fearnhead, P., Eckley, I.A., 2012. Optimal detection of changepoints with a linear computational cost. J. Am. Stat. Assoc. 107 (500), 1590–1598.
Lachir, A., 2025. Assessing the impact of urban canyon geometry on outdoor thermal comfort: a case study in Marrakech, Morocco. Environ. Sci. & Sustain. Dev. 10 (1), 61–72.
Li, X., Yang, B., Liang, F., Zhang, H., Xu, Y., Dong, Z., 2023. Modeling urban canopy air temperature at city-block scale based on urban 3D morphology parameters– a study in Tianjin, North China. Build. Environ., 230 https://doi.org/10.1016/j.buildenv.2023.110000.
Li, S., Wong, M.S., Zhu, R., Shi, G., Yang, J., 2025a. Impacts of land surface temperature and ambient factors on near-surface air temperature estimation: a multisource evaluation using SHAP analysis. Sustain. Cities Soc. 122, 106257.
Li, S., Zhang, J., Xiao, Q., Fang, K., Ha, S., Liu, H., Teng, M., Wu, X., Wu, C., 2025b. Ameliorating canopy urban heat island in block level: a holistic investigation into the radius, form, and typologies. Build. Environ. 282, 113281.
Lin, L., Zhao, Y., Zhao, J., Wang, D., 2025. Comprehensively assessing seasonal variations in the impact of urban greenspace morphology on urban heat island effects: a multidimensional analysis. Sustain. Cities Soc. 118, 106014.
Liu, B., Guo, X., Jiang, J., 2023. How urban morphology relates to the urban Heat Island effect: a multi-Indicator study. Sustainability 15 (14), 10787.
Liu, Q., Wang, J., Bai, B., 2024a. Unveiling nonlinear effects of built environment attributes on urban heat resilience using interpretable machine learning. Urban Clim. 56, 102046.
Liu, Y., Zhang, W., Liu, W., Tan, Z., Hu, S., Ao, Z., Li, J., Xing, H., 2024b. Exploring the seasonal effects of urban morphology on land surface temperature in urban functional zones. Sustain. Cities Soc. 103. https://doi.org/10.1016/j.scs.2024.105268.
Liu, H., Li, M., Zhan, Q., Ma, Z., He, B.-J., 2025a. Homogeneity and heterogeneity of diurnal and nocturnal hotspots and the implications for synergetic mitigation in heat-resilient urban planning. Comput. Environ. Urban. Syst. 117. https://doi.org/10.1016/j.compenvurbsys.2024.102241.
Liu, X., Wang, S., Tang, G., 2025b. Understanding nonlinear and spatially heterogeneous effects of urban residential morphology on land surface temperature: integrating SOM, XGBoost-SHAP, and GWR models. Sustain. Cities Soc. 136, 107100.
Lu, Y., Chen, Q., Yu, M., Wu, Z., Huang, C., Fu, J., Yu, Z., Yao, J., 2023. Exploring spatial and environmental heterogeneity affecting energy consumption in commercial buildings using machine learning. Sustain. Cities Soc. 95. https://doi.org/10.1016/j.scs.2023.104586.
Lundberg, S.M., Erion, G.G., Lee, S.-I., 2018. Consistent Individualized Feature Attribution for Tree Ensembles. ArXiv Preprint ArXiv:1802.03888.
Ming, Y., Liu, Y., Li, Y., Song, Y., 2024. Unraveling nonlinear and spatial non-stationary effects of urban form on surface urban heat islands using explainable spatial machine learning. Comput. Environ. Urban. Syst. 114, 102200.
Okumus, D.E., Akay, M., 2025. Quantitative assessment of non-stationary relationship between multi-scale urban morphology and urban heat. Build. Environ. 272, 112669.
Park, K., Jun, C., Baik, J., Kim, H.-J., 2024. Urban canyon design with aspect ratio and street tree placement for enhanced thermal comfort: a comprehensive thermal comfort assessment accounting for gender and age in Seoul, Republic of Korea. Buildings 14 (8), 2517.
Rashtian, Z., Tabatabaei Manesh, M., Tahsildoost, M., Zomorodian, Z.S., 2025. Data-driven real-time visualization of urban heat islands using mean radiant temperature for urban design. Energ. Buildings 333. https://doi.org/10.1016/j.enbuild.2025.115470.
Rees, G., Hebryn-Baidy, L., Belenok, V., 2024. Temporal variations in land surface temperature within an urban ecosystem: a comprehensive assessment of land use and land cover change in Kharkiv, Ukraine. Remote Sens. 16 (9), 1637.
Roy, S.D., Kuffer, M., Wang, J., 2025. Exploring the influence of building morphology on surface temperatures: a multi-city analysis in Europe. Build. Environ. 282, 113274.
Sattar, T., Mirza, N.F., Javed, M.A., Nasar-u-Minallah, M., Malik, S., 2025. Changing pattern of urban landscape and its impact on thermal environment of Lahore; implications for climate change and sustainable development. Environ. Monit. Assess. 197 (2). https://doi.org/10.1007/s10661-024-13559-1.
Segal, M.R., 2004. Machine Learning Benchmarks and Random Forest Regression.
Shao, L., Liao, W., Li, P., Luo, M., Xiong, X., Liu, X., 2023. Drivers of global surface urban heat islands: surface property, climate background, and 2D/3D urban morphologies. Build. Environ. 242. https://doi.org/10.1016/j.buildenv.2023.110581.
Shen, Y., Kong, W., Fei, F., Chen, X., Xu, Y., Huang, C., Yao, J., 2024. Stereoscopic urban morphology metrics enhance the nonlinear scale heterogeneity modeling of UHI with explainable AI. Urban Clim. 56, 102006.
Sheng, T., Zhang, Z., Qian, Z., Ma, P., Xie, W., Zeng, Y., Zhang, K., Sun, Z., Yu, J., Chen, M., 2025. Examining urban agglomeration heat island with explainable AI: an enhanced consideration of anthropogenic heat emissions. Urban Clim. 59, 102251.

Shi, Y., Shi, Y., Zhou, X., Zhai, W., Tang, J., Zhou, S., 2026. Unveiling nonlinear dynamics in urban blue-green space cooling: spatiotemporal supply-demand matching via integrated machine learning in Hangzhou. Urban Clim. 65, 102792.
Sobrino, J.A., Jiménez-Muñoz, J.C., Paolini, L., 2004. Land surface temperature retrieval from LANDSAT TM 5. Remote Sens. Environ. 90 (4), 434–440.
Sun, L., Xie, C., Qin, Y., Zhou, R., Wu, H., Che, S., 2024. Study on temperature regulation function of green spaces at community scale in high-density urban areas and planning design strategies. Urban For. Urban Green., 101 https://doi.org/10.1016/j.ufug.2024.128511.
Tang, G., Du, X., Wang, S., 2025. Impact mechanisms of 2D and 3D spatial morphologies on urban thermal environment in high-density urban blocks: a case study of Beijing's Core area. Sustain. Cities Soc. 123, 106285.
Tanoori, G., Soltani, A., Modiri, A., 2024. Machine learning for urban Heat Island (UHI) analysis: predicting land surface temperature (LST) in urban environments. Urban Clim. 55. https://doi.org/10.1016/j.uclim.2024.101962.
Ullah, S., Khan, M., Qiao, X., 2024. Examining the impact of land use and land cover changes on land surface temperature in Herat city using machine learning algorithms. GeoJournal 89 (5). https://doi.org/10.1007/s10708-024-11217-0.
USGS, 2019. Landsat 8 Data Users Handbook, 8, Nasa, 2019, p. 97. Version 5.0.
USGS, 2023. Earthexplorer. https://earthexplorer.usgs.gov/scene/metadata/full/5e81f14f59432a27/LC81800312023207LGN00/.
Vaidya, M., Keskar, R., Kotharkar, R., 2024. Quantifying the effect of surrounding spatial heterogeneity on land surface temperature based on local climate zones using mutual information. Sustain. Cities Soc. 107. https://doi.org/10.1016/j.scs.2024.105455.
Wang, C., Zhang, H., Ma, Z., Yang, H., Jia, W., 2024. Urban morphology influencing the urban heat island in the high-density city of Xi'an based on the local climate zone. Sustainability (Switzerland) 16 (10). https://doi.org/10.3390/su16103946.
Wang, L., Li, R., Jia, J., Zhai, Y., Tian, Y., Xu, D., Chen, Y., Zhang, X., Ren, Z., Ye, Z., Ye, Z., Yao, Y., 2025a. Integrating morphology and vitality to quantify seasonal contributions of urban functional zones to thermal environment. Sustain. Cities Soc. 120. https://doi.org/10.1016/j.scs.2025.106136.
Wang, Z., Zhou, R., Yu, Y., 2025b. The impact of urban morphology on land surface temperature under seasonal and diurnal variations: marginal and interaction effects. Build. Environ. 272. https://doi.org/10.1016/j.buildenv.2025.112673.
Weather Underground, 2023. https://www.wunderground.com.
Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, H., Deng, S.-H., 2019. Hyperparameter optimization for machine learning models based on Bayesian optimization. J. Electron. Sci. & Technol. 17 (1), 26–40.
Wu, H., Ming, Y., Liu, Y., 2024. Investigating the influence of morphologic and functional polycentric structures on urban heat island: a case of Chongqing, China. Sustain. Cities Soc. 114, 105790.
Xu, H., Chen, B., 2024. Remote sensing of the urban heat island and its changes in Xiamen City of SE. J. Environ. Sci. 16, 276–281.
Yao, X., Zhu, Z., Zhou, X., Shen, Y., Shen, X., Xu, Z., 2022. Investigating the effects of urban morphological factors on seasonal land surface temperature in a "furnace city" from a block perspective. Sustain. Cities Soc. 86. https://doi.org/10.1016/j.scs.2022.104165.
Yu, H., Zhang, C., Yu, L., 2025. Do different urban morphologies have consistent patterns on urban Heat Islands in different cities and scales? A local climate zone exploration. IEEE J. Sel. Top. Appl. Earth Observ. & Rem. Sens. 18, 2429–2443. https://doi.org/10.1109/JSTARS.2024.3517837.
Yuan, B., Zhou, L., Hu, F., Wei, C., 2024. Effects of 2D/3D urban morphology on land surface temperature: contribution, response, and interaction. Urban Clim. 53. https://doi.org/10.1016/j.uclim.2023.101791.
Zhang, J., Li, Z., Hu, D., 2022. Effects of urban morphology on thermal comfort at the micro-scale. Sustain. Cities Soc. 86, 104150.
Zhang, Y., Teoh, B.K., Zhang, L., 2024. Multi-objective optimization for energy-efficient building design considering urban heat island effects. Appl. Energy 376. https://doi.org/10.1016/j.apenergy.2024.124117.
Zhang, H., Gao, J., Zhao, J., Guo, F., Bai, J., Wang, Z., Zhu, P., 2025. Applicability of local climate zones in assessing urban heat risk-a survey of coastal city. Cities 164, 106068.
Zhang, H., Shu, B., Wei, Y., Liu, Y., 2026a. Terrain and season-dependent impacts of urban morphology on land surface temperature: a machine learning study of three Chinese cities. Urban Clim. 67, 102855.
Zhang, T., Xu, R., Ye, J., 2026b. Spatial heterogeneity of the relationship between NDVI and LST under urban land use patterns—a case study of Shanghai (2000–2024). Environ. Monit. Assess. 198 (2), 171.
Zou, B., Fan, C., Li, J., 2024. Quantifying the influence of different block types on the urban heat risk in high-density cities. Buildings (2075-5309) 14 (7).