ENVIRONMENTAL RESEARCH
COMMUNICATIONS

INCLUSIVE PUBLISHING TRUSTED SCIENCE

Part of the ENVIRONMENTAL RESEARCH SERIES

PURPOSE-LED PUBLISHING™

PAPER • OPEN ACCESS

Machine learning-based modelling of land surface temperature dynamics in response to landscape fragmentation in industrial areas

To cite this article: Sheela *et al* 2025 *Environ. Res. Commun.* **7** 125017

IOP Publishing Environ. Res. Commun. 7 (2025) 125017 https://doi.org/10.1088/2515-7620/ae219c

Environmental Research Communications

OPEN ACCESS

RECEIVED
24 June 2025

REVISED
3 November 2025

ACCEPTED FOR PUBLICATION
19 November 2025

PUBLISHED
12 December 2025

PAPER

Machine learning-based modelling of land surface temperature dynamics in response to landscape fragmentation in industrial areas

Sheela, Medha Jha and Anurag Ohri*

Department of Civil Engineering, Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh -221005, India

* Author to whom any correspondence should be addressed.

E-mail: sheela.rs.civ19@iitbhu.ac.in, mjha.civ@iitbhu.ac.in and aohri.civ@iitbhu.ac.in

Keywords: LST, NDVI, NDBI, albedo, landscape, XGBoost, PCA

Supplementary material for this article is available online

Abstract

Urbanisation and industrialisation often replace vegetation with impervious surfaces, affecting thermal capacity and evaporation rates, and leading to increased land surface temperature (LST). This study develops predictive models for LST using biophysical parameters such as the Normalised Difference Vegetation Index (NDVI), Normalised Difference Built-up Index (NDBI), Enhanced Built-up Index (EBBI), and albedo, along with landscape fragmentation metrics from land use and land cover (LULC) maps for the years 2006, 2011, 2016, 2021, and 2024. We developed a Composite Fragmentation Index (CFI) using Principal Component Analysis (PCA) based on metrics such as the number of patches and cohesion. Random Forest (RF) and XGBoost (XGB) models are trained on this data for a peri-urban region, with XGB showing slight improvement over RF. The models achieved a strong performance, with an R^2 of approximately 0.78 and an RMSE of around 1.7. Feature importance analysis indicates that higher NDVI significantly reduces LST, while increased NDBI and EBBI raise it. Uncertainty, quantified through conformal prediction, showed 90% prediction intervals of approximately ± 2.8 °C. Additionally, Piecewise regression identified an NDVI threshold of 0.32, beyond which LST reductions become more pronounced. Overall, our findings highlight the need to consider both land cover composition and configuration in formulating effective strategies to mitigate heat stress in industrialising areas.

1. Introduction

Monitoring land surface temperature (LST) has become crucial for assessing heat stress under climate change. Rapidly urbanising regions often replace natural landscapes with impervious material, leading to higher surface and air temperatures. The peri-urban areas that are transition zones between urban cores and rural land are increasingly recognised as critical hotspots of environmental change. These areas often suffer rapid fragmentation of LULC into industrial facilities (Motlagh *et al* 2020). Such regions create a unique thermal signature due to low vegetation cover and scattered built-up expansion, along with the waste heat effect of industrial activities (Jiang *et al* 2025, Mustafa and Szydlowski 2020). Therefore, intensified heat disrupts local microclimates and energy balance and poses risks to human health and ecological stability (Kotharkar *et al* 2019).

Several biophysical parameters help in understanding these dynamics through remote sensing techniques. Land Surface Temperature (LST) exhibits a significant interplay with various biophysical parameters, which are indicative of land cover and surface characteristics. Many studies have examined LST-biophysical parameter correlations, such as NDVI, which typically correlates negatively with LST. This is because higher NDVI values indicate increased vegetation cover, which in turn correlates with cooler surface temperatures due to enhanced evapotranspiration (Guha and Govil 2020, 2022). A similar result was observed by Kulsum and Moniruzzaman (2022) and Pandeya *et al* (2022), who found the correlation coefficients between −0.40 and −0.61 using LST and NDVI. Similarly, the Enhanced Vegetation Index (EVI) provides a more sensitive measure of vegetation, also demonstrating a negative relationship with LST (Zhao *et al* 2020). The Soil Moisture Index (SMI) and Modified Normalized Difference Water Index (MNDWI) indicate that areas with higher soil moisture and water bodies are typically cooler, suggesting a negative correlation with LST (Pandey *et al* 2023, Sharma *et al* 2021). Conversely, the NDBI often aligns positively with LST, as urbanized areas exhibit higher heat retention and re-radiate heat (Macarof and Statescu 2017). Studies by Kulsum and Moniruzzaman 2022, Pandeya *et al* 2022 have also found positive correlation coefficients of LST with NDBI, such as, 0.56 to 0.69 (e.g., r = 0.56 in summer, 0.69 in winter in Imphal, India; r = 0.61 in Manipur, India; r = 0.573 in Dhaka, Bangladesh). Indices such as the Urban Index (UI) and the Enhanced Built-up and Bare Soil Index (EBBI) highlight urban areas as areas of heat retention and correlate positively with LST (Ramaiah *et al* 2020). The Normalized Difference Bare Soil Index (NDBaI) reflects the heat absorption properties of bare soil and can also impact LST (Sahoo *et al* 2022). Similarly, albedo measures surface reflectivity, which often correlates inversely with LST, as surfaces with high albedo reflect more solar radiation, resulting in cooler surface temperatures (Debnath *et al* 2025). A study by Miniandi *et al* (2025) also showed that 25% increase in albedo can decrease LST by up to 1.2 °C in high-LST zones. These indices highlight the intricate relationships between land cover dynamics and surface temperature, facilitating a deeper understanding of and more effective management for urban thermal environments.

The spatial configuration of green and built surfaces significantly affects thermal outcomes in urban areas. Metrics from landscape ecology, such as the number of patches (NP), the most extensive patch index (LPI), and the patch cohesion index, provide insight into land cover fragmentation. (Zawadzka *et al* 2021). Generally, larger and less fragmented green patches are associated with lower land surface temperatures (Jiang and Tian 2010). At the same time, greater fragmentation of impervious and bare-land patches correlates with higher LST (Sun *et al* 2022). Research shows mixed results regarding the impact of fragmentation on LST. Some studies indicate that a few large, contiguous green spaces are more effective at cooling than many small patches (Mallick *et al* 2013, Zheng *et al* 2014). For instance, LPI and the aggregation index often correlate negatively with LST, suggesting that larger, aggregated patches lead to cooler temperatures (Bao *et al* 2016, Maimaitiyiming *et al* 2014). Conversely, other studies indicate that a dispersed arrangement of smaller green patches may offer extensive cooling benefits across urban regions, supporting even temperature moderation (Masoudi *et al* 2021). Furthermore, the configuration of impervious surfaces plays a crucial role; extensive urbanised patches tend to increase LST. For instance, Zhou and Wang (2011) also noted that a single massive expanse of paved surface yields hotter conditions than the same paved area broken into separated patches interspersed with vegetation. Clustering of built-up land (characterized by high LPI and cohesion in impervious regions) has been linked to increased urban warmth, whereas a more heterogeneous land-cover mosaic can moderate this effect. The context, including climate and scale, influences the relationship between land cover configuration and temperature (Rahimi *et al* 2021).

In recent years, data-driven land surface temperature modeling (LST) has made significant progress alongside geospatial analyses. Techniques ranging from linear regression to machine learning have been used to predict LST from remote sensing indices and land-cover features. Machine learning models, such as Random Forests (RF) and XGBoost (XGB), offer flexibility in capturing non-linear relationships between predictors (vegetation, built fraction, etc) and LST. For example, one study (Wu *et al* 2022) found that a random forest model outperformed linear regression in explaining LST variation by accounting for complex interactions among factors. These ensemble tree models also facilitate interpretation via SHAP (SHapley Additive exPlanations) to quantify each feature's contribution to predictions. Recent studies have applied machine learning to LST analysis, often with conventional random sampling validation.

Previous studies have examined the relationship between land surface temperature (LST) and landscape fragmentation using individual metrics, such as patch density and edge density. However, these single-metric approaches often yield inconsistent results due to intercorrelation among metrics and varying effects across studies, creating redundancy (Kedron *et al* 2018). For instance, Liu *et al* (2018) It has been highlighted that edge density has shown both negative and positive correlations with LST in different urban contexts. De Montis *et al* (2020) Further emphasis is placed on the importance of creating composite indices that combine several dimensions of fragmentation into a single, comprehensive measure. Therefore, there is a need for composite indices that incorporate multiple dimensions of fragmentation. Techniques such as principal component analysis (PCA) can help reduce multicollinearity, capture the combined effects of patch size, shape, and connectivity, and enhance interpretability. (Song *et al* 2020, Zhang *et al* 2022). While machine learning has excelled in modelling complex biophysical and socio-ecological systems, its integration with landscape fragmentation indices for predicting land surface temperature (LST) remains limited. Most studies have prioritised spectral or biophysical predictors, overlooking the importance of landscape configuration. This gap presents an opportunity to enhance predictive accuracy and policy relevance by combining composite fragmentation indices with

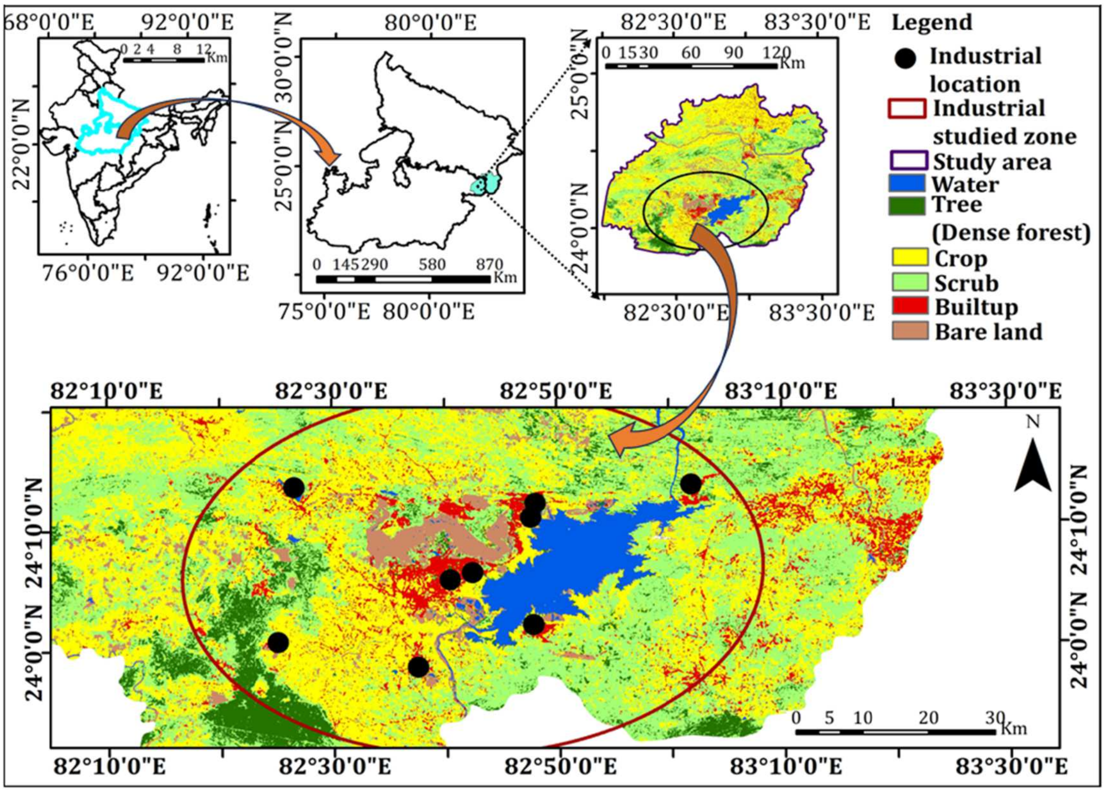

Figure 1. Location map of the study area with industrial sites.

machine learning. To address this, the present study develops a PCA-based Composite Fragmentation Index (CFI) and integrates it with biophysical parameters to improve LST predictions.

The objectives include (1) quantifying spatial and temporal LST variations among land use/land cover (LULC) classes. (2) determining the vegetation threshold related to LST, exploring landscape patterns and their relationship with LST, and (3) developing a predictive model that captures multiple dimensions of heat stress in industrial peri-urban areas.

2. Study area

This study examines the industrial region of Singrauli and Sidhi districts, located in the Vindhya Plateau at the Madhya Pradesh and Uttar Pradesh border, covering approximately 12354 km^2 (figure 1). The area, with elevations ranging from 300 m to over 500 m, features the Govind Vallabh Pant Sagar reservoir on the Rihand River, which is crucial in supplying water to nearby thermal power plants.

The region experiences a tropical monsoonal climate, with temperatures peaking around 41 °C–42 °C in May and frequent heatwaves exceeding 45 °C. About 89% of the annual rainfall (~1100–1200 mm) occurs between June and September.

Historically, this area supported extensive tropical dry deciduous forests rich in biodiversity and vital to indigenous communities. However, industrial development over the past few decades has led to significant deforestation and habitat fragmentation, with only about 5%–10% of the original forest cover remaining. The landscape is now a mix of mines, power plants, villages, agricultural fields, and degraded forest fragments. This transition presents an opportunity to study the effects of land transformation on land surface temperature (LST) and the need for sustainable planning to mitigate thermal stress in vulnerable communities.

3. Methodology

This section describes the steps required to develop and validate the model, which are summarised in figure 2.

3.1. Data sources and preprocessing

This study utilised multi-temporal satellite imagery from the Landsat series over five years: 2006, 2011, 2016, 2021, and 2024. These intervals are chosen to observe significant changes resulting from seasonal variations,

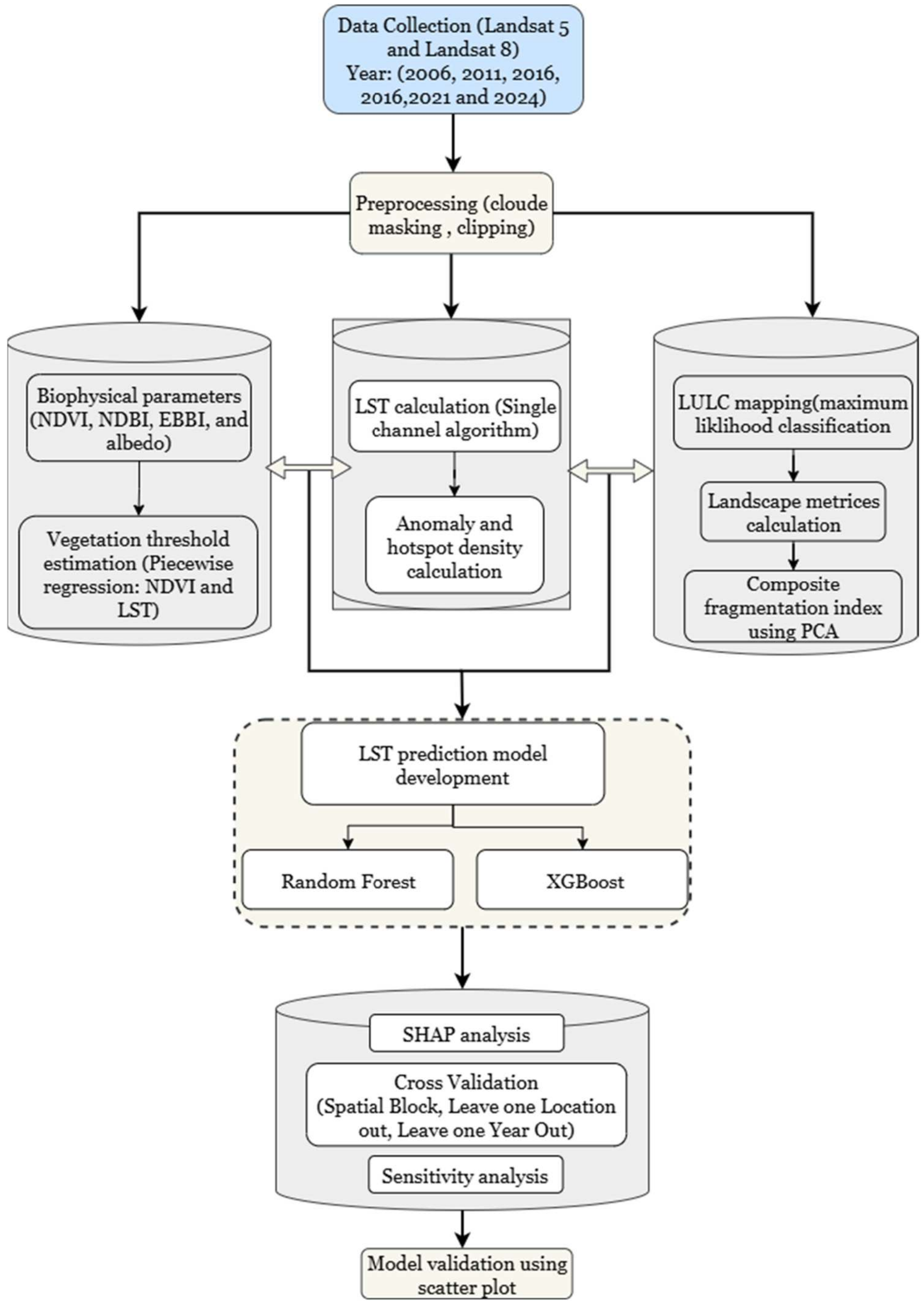

Figure 2. Workflow integrating remote sensing, landscape metrics, and ML for LST modelling.

climatic shifts, and anthropogenic influences, such as industrial impacts. USGS Landsat 5 Level 2, Collection 2, Tier 2, and USGS Landsat 8 Level 2, Collection 2, Tier 2 data have been used. All imagery has a spatial resolution of 30 meters and a Temporal resolution of 16 days. Landsat 5 was selected for 2006 and 2011. At the same time, Landsat 8 data were used for 2016, 2021, and 2024. All images were captured during the same season (the dry-hot season), specifically from April to June, to maintain consistency in seasonal differences. The dry season was chosen due to its reduced cloud cover, facilitating more explicit imagery and more reliable data for analysis. Before analysis, cloud masking was applied to each scene using quality assurance bands. The use of Landsat imagery in this study is attributed to its high resolution and consistent temporal coverage compared to other available sensors like MODIS, which has 1000m resolution for the LST product. Table 1 summarizes the Number of datasets used, along with their sources, resolution, processing level, and applications, for each year of the study.

Table 1. The table summarises the satellite imagery and band combinations used in the study.

Dataset	Spatial/Temporal Resolution	Source	Rows and path	No. of images	Band used	Purpose
Landsat 5	30 m/16 days	USGS Landsat 5 Level 2, Collection 2, Tier 2	142/043,142/044 And 143/043	2006:11 2011:14	Thermal= B6 / Composite band	LST/LULC
					RED = B3, NIR =B4	NDVI
					SWIR = B5, NIR =B4	NDBI
					BLUE=B1, RED=B3, NIR=B4 SWIR=B5	EBBI
					Green=B2, Red=B3, NIR=B4, SWIR=B5, SWIR2=B7	Albedo
Landsat 8	30 m/16 Days	USGS Landsat 8 Level 2, Collection 2, Tier 2	142/043,142/044 And 143/043	2016:17 2021:17 2024:16	Thermal= B10/Composite band	LST/LULC
					RED = B4, NIR =B5	NDVI
					SWIR = B6, NIR =B5	NDBI
					BLUE=B2, RED=B4, NIR=B5 SWIR=B7	EBBI
					Green=B3, Red=B4, NIR=B5, SWIR=B6, SWIR2=B7	Albedo

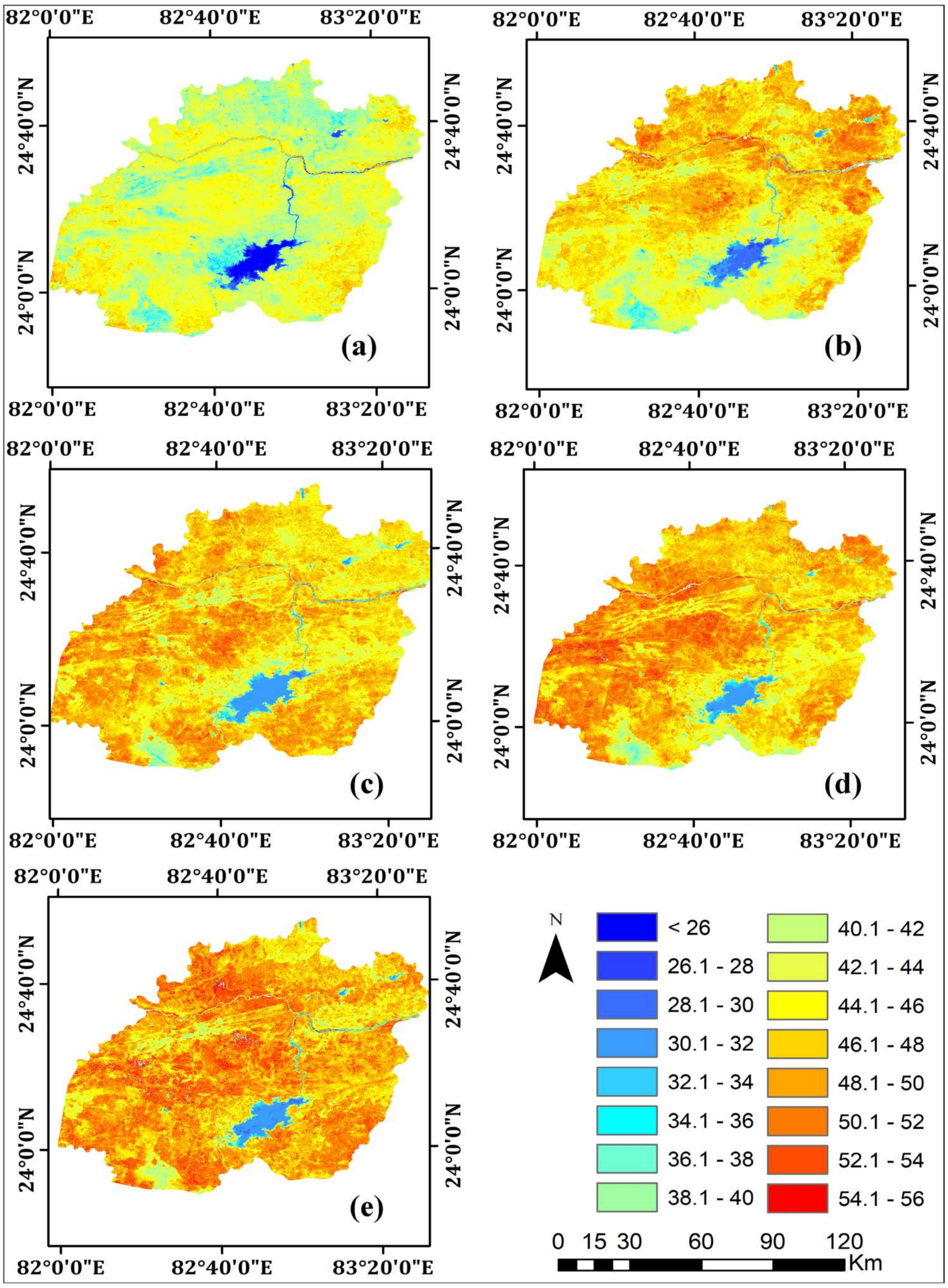

Figure 3. Spatial distribution of Land Surface Temperature (LST) where (a) 2006, (b)2011, (c) 2016, (d) 2021, (e)2024.

3.2. LST derivation

The spatial distribution map of LST was prepared using the USGS Landsat 5 and 8 Collection-2 Level-2 Surface Temperature product data using Google Earth Engine (GEE) for calculation and ArcGIS 10.8 for mapping, as shown in figure 3. After applying the scale, we calculated the mean composite, 0.00341, with a +149 K offset. Landsat 5 products utilised the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS), while Landsat 8 used the Land Surface Reflectance Code (LaSRC). Both collections employed a single-channel algorithm jointly developed by the Rochester Institute of Technology (RIT) and the National Aeronautics and Space Administration's (NASA) Jet Propulsion Laboratory (JPL), which incorporates atmospheric correction and estimates land surface emissivity from the ASTER Global Emissivity Database (GED). To ensure reliability, images from April to June between 2006 and 2024 were filtered for clouds and low-quality retrievals using the QA_PIXEL and QA_ST bitmask bands, and then clipped to the study boundary. Annual mean LST composites were then generated by averaging all valid images within the year. The single-channel method was selected for LST estimation because it is computationally efficient, requires fewer input parameters than more complex algorithms such as the split-window or TES approaches, and has been shown to produce reliable results.

Table 2. Formula table for indices used.

Indices	Formula
NDVI	$\frac{NIR - RED}{NIR + RED}$
NDBI	$\frac{SWIR - NIR}{SWIR + NIR}$
EBBI	$\frac{SWIR - NIR}{\sqrt{SWIR + RED + BLUE}}$
Albedo	Landsat 8: 0.356·B2+0.130·B4+0.373·B5+0.085·B6 +0.072·B7−0.0018 Landsat 5: 0.293·B1+0.274·B3+0.233·B4+0.156·B5 +0.033·B7

We compared Landsat LST with MODIS Terra MOD11A2 v061 daytime LST data for validation. Mean composites were calculated for both sensors from 2006 to 2024. Landsat data were aggregated to the 1-km MODIS grid by computing the within-pixel mean, ensuring only valid pixels from both datasets were retained.

3.2.1. LST anomaly mapping

A normalised anomaly approach was adopted to assess the intensity of temperature deviations relative to the baseline. The normalised LST anomaly was computed using the Z-score method, which expresses anomalies in standard deviation units, thereby facilitating meaningful comparison across space and time. The formula is described in equation (1):

$$LST\,Anomaly\,(i) = \frac{(LST\,(i) - \mu LST)}{\sigma LST} \tag{1}$$

Where LST_i is the land surface temperature for the target year, μLST represents the mean LST of the reference period, and σLST denotes the corresponding standard deviation. Positive anomaly values indicate areas warmer than the baseline mean, while negative values indicate cooler conditions. The magnitude of the normalised anomaly reflects the severity of deviation: values near ±1 indicate mild anomalies, ±2 suggest moderate anomalies, and values beyond ±3 represent extreme deviations. This normalisation accounts for spatial variability and enables robust identification of the study area's thermal hotspots and cold spots.

3.2.2. Hotspot density

Anomaly maps were compared across years to identify persistent thermal hotspots or zones of emerging cooling. LST hotspots were identified as pixels with LST values at or above the 85th percentile each year, as shown below in equation (2). It counts the number of times each location appears as a hotspot in all years. This shows the spatial clustering of thermal extremes (frequent hotspots). Hotspot Count can be understood by the number of years (out of 5) a location appeared in the top 15% of LST values.

$$Hotspot = \begin{Bmatrix} 1 & if & LSTi \geqslant 85th\ percentiles\ of\ all\ LSTi \\ 0 & if & LSTi < 85th\ percentiles\ of\ all\ LSTi \end{Bmatrix} \tag{2}$$

3.3. Spectral indices calculation

The Spearman correlation coefficient was calculated using these indices to determine the correlation with LST. The Spearman correlation coefficient is suitable because it reduces the impact of outliers or non-normal distributions, which are common in remote sensing data. Many spectral indices, such as NDVI and LST, have nonlinear but monotonic relationships.

These indices were selected because of their established correlations with LST (Bala *et al* 2019, Dai *et al* 2018, Guha *et al* 2021). All biophysical parameters are calculated for each pixel using satellite imagery, as outlined in table 2, and the band information is presented in table 1.

3.4. Vegetation threshold detection

To further assess the influence of vegetation cover on land surface temperature, a threshold analysis was conducted to identify the critical NDVI value at which the cooling effect becomes saturated. Piecewise linear regression by Toms and Lesperance, (2003) was applied to the combined LST and NDVI data from all years (2006–2024), allowing for the detection of a breakpoint in the NDVI–LST relationship using equation (3). A piecewise regression model was used to estimate the two linear segments and the inflexion point. This threshold was then used to determine the point at which increasing vegetation had a significant impact on reducing the land surface temperature. The concept is that initially, adding a little vegetation in a largely bare area may not cool the surface significantly until a certain cover threshold is reached. After this threshold is exceeded, additional vegetation leads to substantial cooling (or conversely, beyond a threshold, adding vegetation yields diminishing returns if the maximum cooling has already been achieved). This technique has been effectively utilised in various studies to assess the influence of vegetation cover on LST, providing valuable insights for urban planning and climate adaptation strategies (Li *et al* 2014).

To avoid distortions from obvious non-vegetation outliers (e.g., water bodies or cloud pixels with NDVI < 0), we focused on NDVI $\geqslant 0$ for threshold analysis, since negative NDVI typically indicates no vegetation at all (water or built surfaces), and we are interested in the onset of the vegetation's cooling effect. We then fitted a two-segment linear model, one line for NDVI values below a candidate breakpoint and another for NDVI values above the breakpoint, with continuity enforced at the breakpoint (i.e., the two-line segments meet at the same temperature at the threshold NDVI). We iterated over the possible breakpoints and identified the NDVI value that minimised the piecewise fit's overall residual error (SSE). This approach finds the NDVI at which the LST versus NDVI relationship 'kinks'. Correlation indicates that the relationship is not strictly linear across the full NDVI range; therefore, a segmented model is appropriate.

$$LST = \begin{cases} \beta 1.NDVI + C1,\ NDVI < \theta \\ \beta 2.NDVI + C2,\ NDVI \geqslant \theta \end{cases} \tag{3}$$

where $\beta 1$ is the slope, C1 is the intercept of the line for NDVI, and θ is the NDVI threshold

3.5. LULC classification

The land use land cover map for the years, 2006, 2011, 2016, 2021, and 2024 was created using the supervised maximum likelihood classification (MLC) method in ArcMap 10.8 software, as shown in figure 4. The MLC is based on Bayesian probability theory and considers the mean and variance of each class in all bands. It calculates the probability that a given pixel belongs to a class and assigns it to the class with the highest probability. This model has been used in numerous studies that have shown accurate results (Lu *et al* 2021, Sultana and Satyanarayana 2019). This leads to a more reliable classification. The LULC map is classified into six major classes: water, which includes rivers, ponds, and reservoirs; crop; tree (dense vegetation); scrub (sparse vegetation); built-up (settlement areas and industrial areas); and bare land (including mining areas).

3.5.1. Accuracy table, kappa coefficient

A stratified random sampling approach was used to distribute 300 validation points across six land use and land cover (LULC) classes, ensuring proportional representation with a minimum of 30 points per class. The accuracy assessment included Producer's Accuracy (Recall), User's Accuracy (Precision), the F1-score for each class, and 95% confidence intervals calculated using the binomial method. Ground truth data were obtained from Google Earth, which showed a strong agreement between the classified data and reference data, yielding an overall accuracy of 85.33%. The Kappa coefficient was 0.822, indicating a strong agreement beyond random chance. Table 3 summarises the class-specific accuracy metrics, including 95% confidence intervals and F1-scores. The confusion matrix is also represented in the supplementary file in table 2S.

3.6. Landscape metrics calculation

We used classified Land Use Land Cover (LULC) maps from 2006, 2011, 2016, 2021, and 2024 to calculate fragmentation metrics, employing FRAGSTATS 4.2 software developed by McGarigal and Marks (1995). We focused on landscape-level metrics to minimise bias from dominant classes, as class-level metrics may not generalise effectively. We applied a 900 m moving window and an 8-cell neighbourhood rule to capture neighbourhood patterns. This connectivity method enables a more comprehensive analysis of large patches by grouping diagonally connected cells. The moving window size was selected based on Guo *et al* (2019), It was studied and indicated that it is suitable for thermal studies using 30 m resolution data. For our predictive modelling, we included three key metrics: NP (number of patches), Cohesion, and LPI, as outlined in table 4. These metrics were selected because they capture essential aspects of landscape fragmentation, dominance, and connectivity. Together, they provide a clear and comprehensive representation of landscape configuration while minimising redundancy among highly correlated metrics. These measurements quantify the spatial structure and fragmentation of the landscape, highlighting both the composition (quantity and size of patches) and configuration (spatial arrangement) relevant to thermal dynamics.

3.6.1. Development of the composite fragmentation index (CFI)

To assess the combined effects of various landscape metrics on land surface temperature (LST), we developed a Composite Fragmentation Index (CFI) using Principal Component Analysis (PCA). We selected three key fragmentation indicators: the Cohesion Index (COHESION), the Number of Patches (NP), and the Largest Patch Index (LPI), which represent connectivity, patch density, and patch dominance, respectively. We

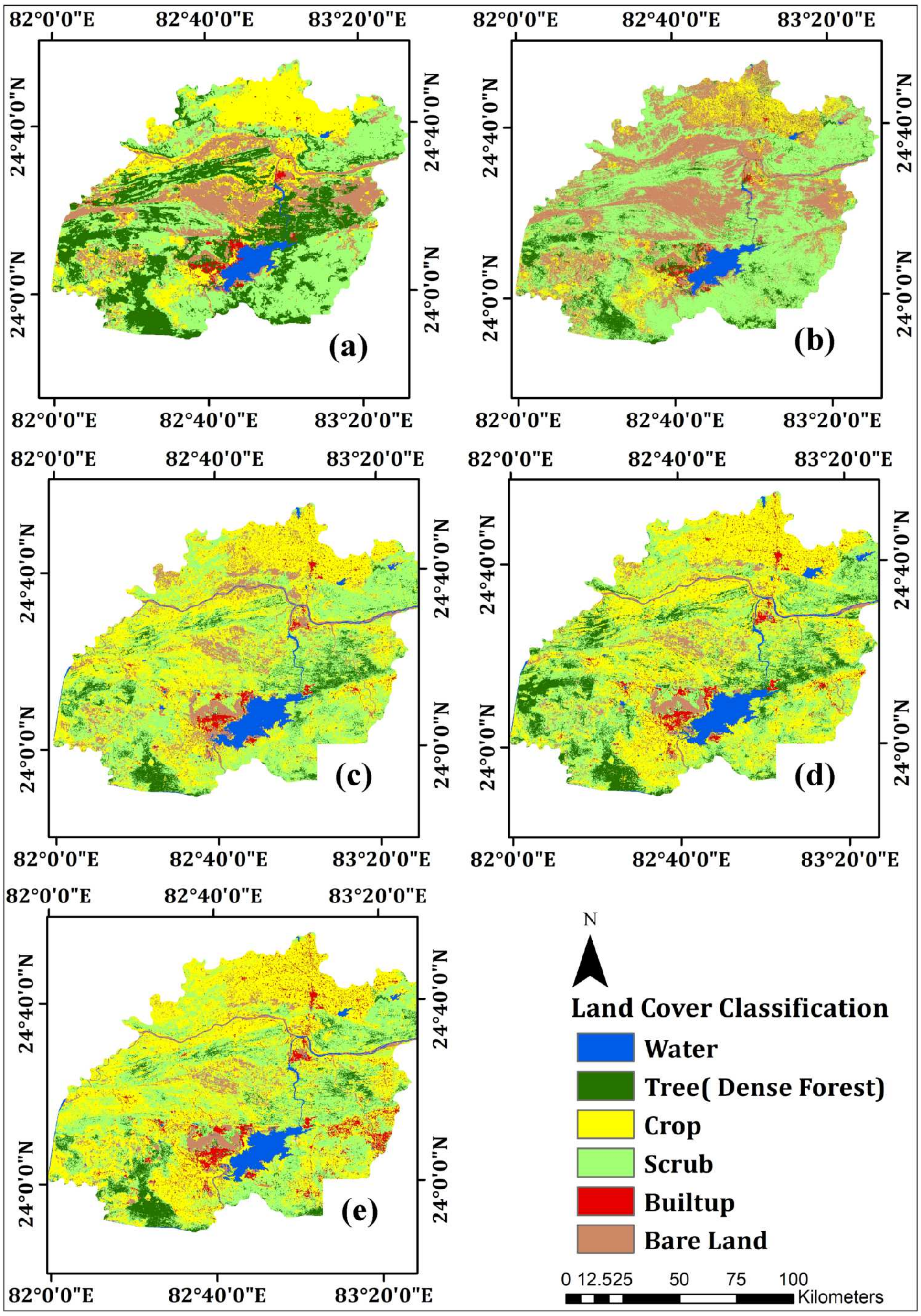

Figure 4. Temporal changes in LULC for multiple years (a) 2006, (b) 2011, (c) 2016, (d) 2021, (e) 2024.

Table 3. Accuracy assessment of LULC classification using confusion matrix-derived metrics.

Class	Sample size	Producer's accuracy (%)	User's accuracy (%)	F1-score	95% CI (±%)
Water	35	91.43	96.97	0.94	±9.2
Tree	60	88.33	86.89	0.87	±7.8
Crop	70	82.86	84.06	0.84	±8.4
Scrub	40	77.5	77.5	0.78	±10.5
Built-up	50	86.0	86.0	0.86	±9.1
Bare Land	45	86.67	82.98	0.85	±9.6

normalised and standardised the metrics using a z-score transformation to ensure comparability. PCA was applied to reduce dimensionality and identify components that explain the most variance. We oriented the principal component axes so that higher scores indicate greater fragmentation intensity, reflected in lower

Table 4. Description and formulas of landscape metrics used for fragmentation analysis.

Metric	Abbreviation	Formula	Description
Largest Patch Index	LPI	$(a_{max} / A) \times 100$	The percentage of the landscape is comprised of the largest patch.
Cohesion Index	COHESION	$[1-(\Sigma p_{ij}/\Sigma(p_{ij}\sqrt{a_{ij}}))] \times (1-1/\sqrt{A})^{-1} \times 100]$	Measures the physical connectedness of patches.
Number of Patches	NP	$NP = n$	Total number of discrete patches.

a_{max}: Area of the largest patch, A: Total landscape area, a_i: Area of patch, p_{ij}: Perimeter of patch, n: Total number of patches per landscape

Table 5. Hyperparameter settings for the XGBoost model used in LST prediction.

Model	Hyperparameter	Value/Setting
Random Forest (RF)	Number of trees (*n_estimators*)	500
	Maximum depth (*max_depth*)	None (nodes expanded until pure)
	Random seed (*random_state*)	42
	Number of parallel jobs (*n_jobs*)	−1 (all cores)
XGBoost (XGB)	Number of trees (*n_estimators*)	600
	Maximum tree depth (*max_depth*)	6
	Learning rate (*eta*)	0.05
	Subsample fraction (*subsample*)	0.85
	Column sampling (*colsample_bytree*)	0.85
	L2 regularisation (*reg_lambda*)	1.0
	Random seed (*random_state*)	42
	Number of parallel jobs (*n_jobs*)	−1 (all cores)
	Tree construction method (*tree_method*)	Hist-based algorithm ('hist')

cohesion, more patches, and smaller LPI. The CFI was then normalised to a scale of 0 to 1, with higher values indicating more fragmented landscapes.

3.7. Machine learning based prediction model

To predict land surface temperature (LST), we employed ensemble tree-based models, Random Forest (RF) and Extreme Gradient Boosting (XGBoost), using biophysical indices (NDVI, NDBI, Albedo, EBBI) and a PCA-derived Composite Fragmentation Index (CFI) as predictors. In each training split, model hyperparameters were optimised exclusively using random search. The seclected hyperparameter setting for both the model is listed in table 5.

3.7.1. Model validation, accuracy, and sensitivity

Model performance was evaluated through three cross-validation methods: Leave-One-Year-Out (LOYO) for temporal generalizability, spatial block cross-validation to reduce spatial autocorrelation bias, and Leave-One-Location-Out (LOLO) for extrapolation to new areas. Accuracy was measured using the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) to provide insights into explained variance and prediction error. Conformal prediction intervals (90% and 95%) were derived from out-of-fold residuals to assess predictive uncertainty.

A one-at-a-time (OAT) sensitivity analysis varied each predictor variable by ±10% to observe changes in predicted land surface temperature (LST). The ±10% perturbation was chosen as it provides a balanced range that is sufficiently large to capture meaningful shifts in predictor variables while remaining within the bounds of ecologically realistic variation. A stratified performance assessment was conducted to prevent performance metrics from being skewed by dominant land cover types, such as bare land. This approach gave a more detailed understanding of the model's robustness and highlighted the added value of incorporating CFI as a predictor.

4. Result

4.1. LST anomaly and hotspot density mapping.

The Land Surface Temperature (LST) analysis from 2006 to 2024, as depicted in the boxplots (figure 5), shows a clear upward trend in surface heating across the study area. The median LST increased from approximately 42 °C in 2006 to nearly 48 °C in 2024. This rise is accompanied by increases in the upper quartile and maximum

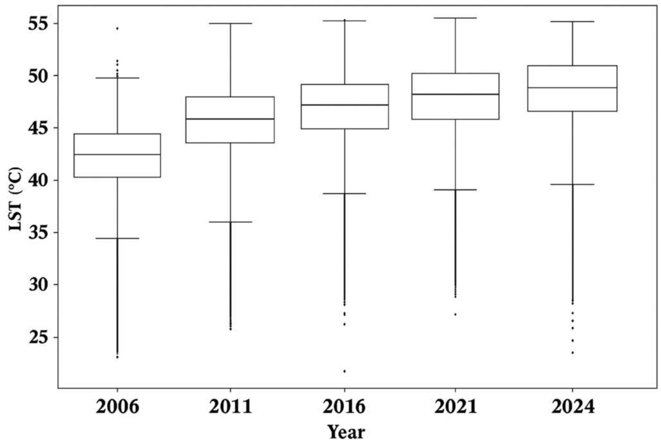

Figure 5. Box plots representing the distribution of LST values for the years 2006, 2011, 2016, 2021, and 2024.

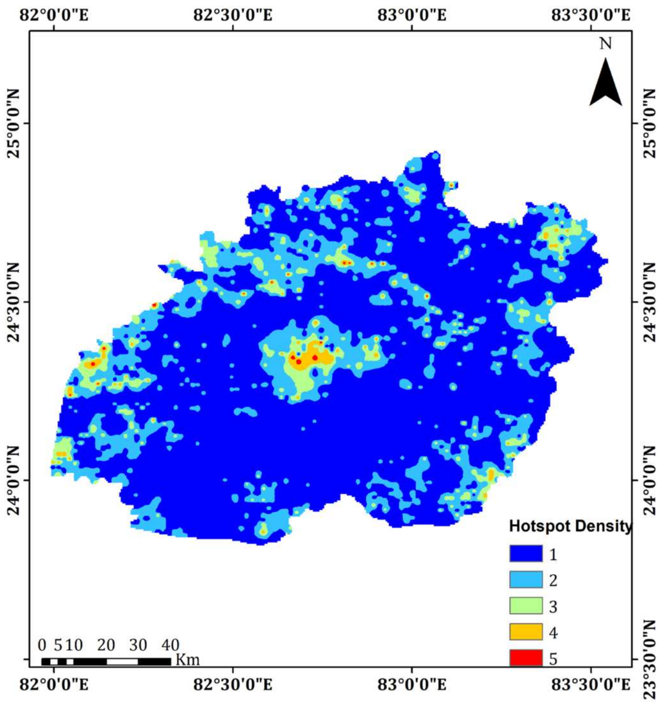

Figure 6. Hotspot density map showing the clustering of high-temperature hotspots within the study region.

values, underscoring an increased frequency of extreme heat events. Over the year, the interquartile range has gradually shifted upward, indicating a reduction in cooler surface zones and a concentration of higher LST values.

The validation result of Landsat-derived LST with MODIS LST also confirmed the reliability of the outcome, with an R^2 value ranging between 0.84 and 0.89, as reported in detail in the supplementary file in table 1S.

The anomaly map shown in the supplementary file (figure 1(S)) indicates that the LST around the industrial area is greater than the average temperature for that particular year. Also, it increases over the year, suggesting the impact of industrial activity. The hotspot density map (figure 6) reveals localized clusters of high-density hotspots (classes 3–5), particularly around major thermal facilities and bare land, indicating persistent thermal stress zones. These spatial patterns highlight the influence of reduced vegetation cover in driving localised surface heating.

4.2. Land use land cover dynamics

The relative proportions of each LULC class over time are shown in figure 7. Between 2006 and 2024, the crop class increased significantly (+24.52%), whereas tree (–17.13%) and bare land (–11.79%) experienced notable

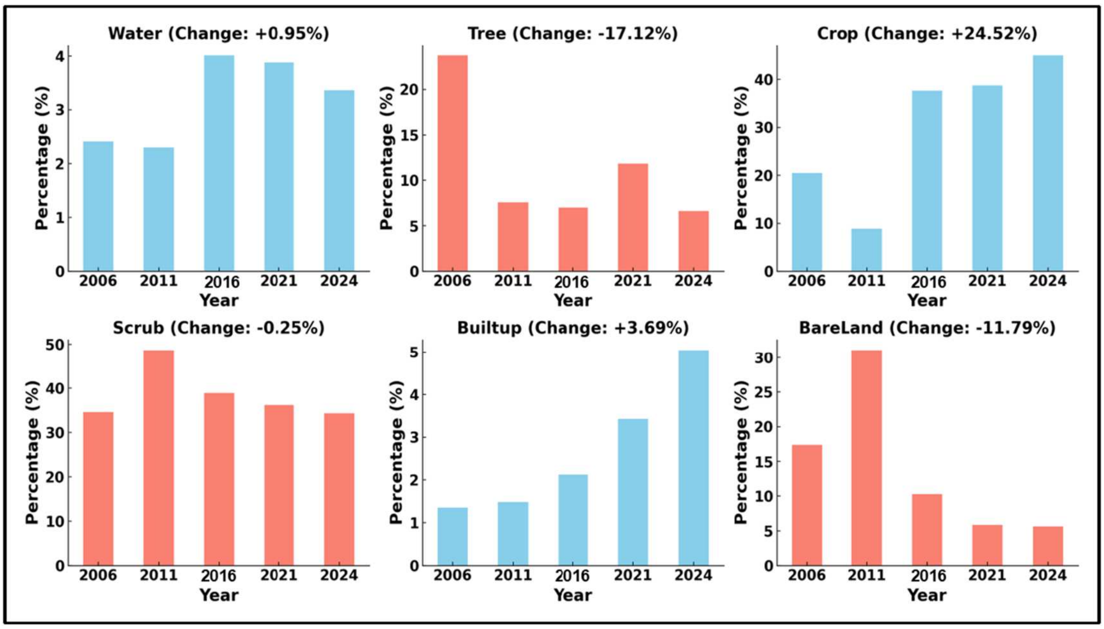

Figure 7. Bar charts showing percentage changes in LULC classes between 2006 and 2024.

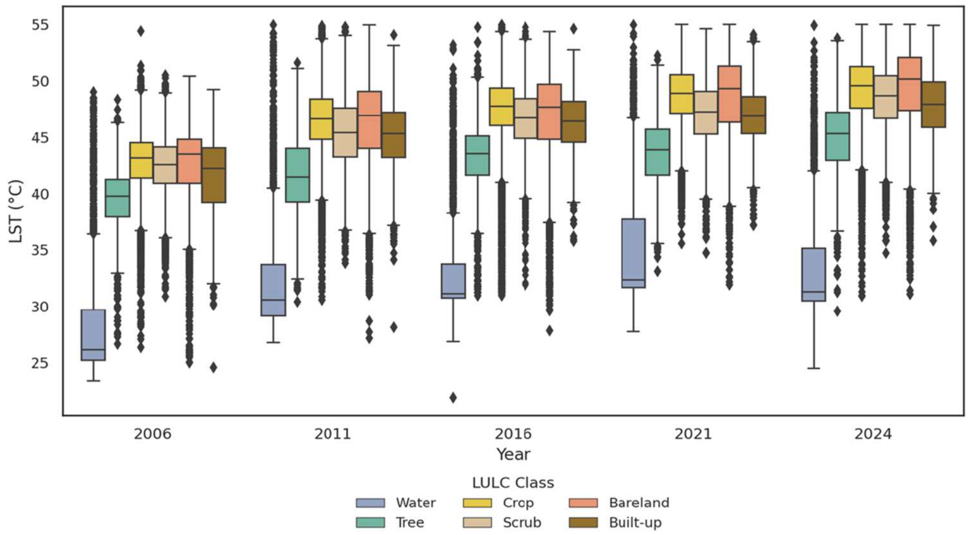

Figure 8. Box plots of LST distribution across different LULC classes for 2006, 2011, 2016, 2021, and 2024.

reductions. The Built-up area expanded by +3.69%, indicating gradual urban encroachment. This shift from vegetated cover to impervious or semi-impervious surfaces strongly correlates with increased LST and thermal hotspot activity in those areas.

To evaluate the influence of land cover type on LST, figure 8 presents a class-wise comparison of LST (°C) over the study years. The Bare land and built-up classes consistently exhibited the highest median LST (>45 °C) across all years, reflecting frequent extremes. These classes also exhibited broader distributions, suggesting higher variability and more frequent extremes. Crop lands register moderately high LST, whereas Tree and Water classes consistently maintain the lowest surface temperatures.

Figure 9 presents the class-wise box plots of the LST anomaly over the years. The early years fluctuate around zero, but anomalies remain predominantly positive in the latter part of the record. In recent years, LST has exhibited significantly above-average growth, with standardized anomalies exceeding +1. Notably, anomaly values for bare land, Built-up, and Crop classes consistently exceeded 0 in the later years, indicating persistent above-average warming in these zones. In contrast, the Water and Tree classes showed relatively lower or

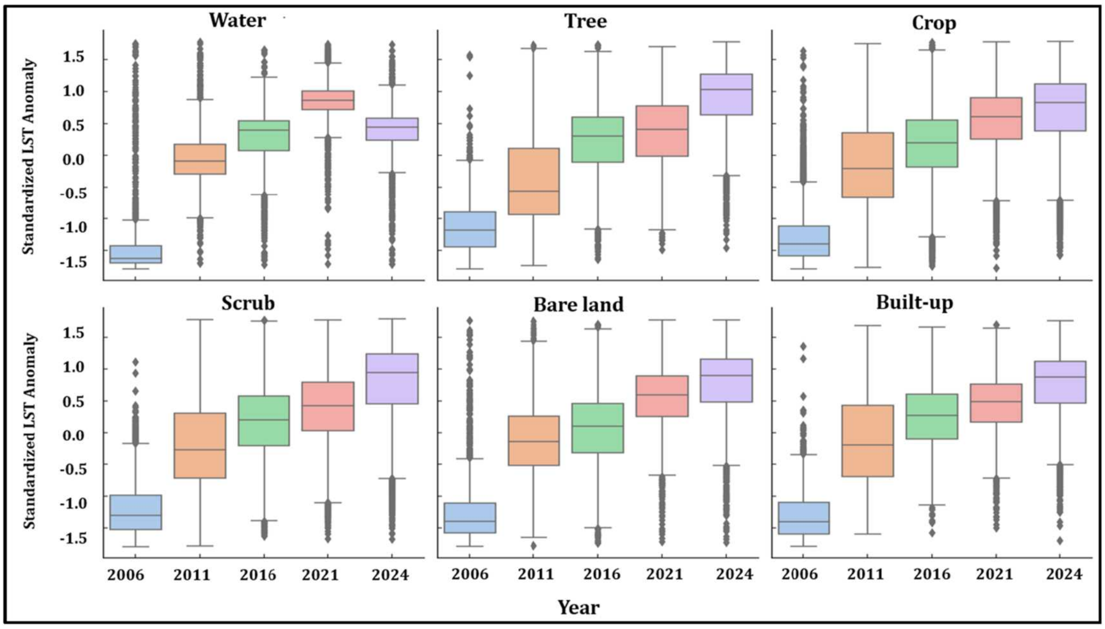

Figure 9. Box plots depicting normalised LST anomalies (Z-scores) for different LULC classes over five time points (2006–2024).

near-neutral anomaly values, confirming their thermal buffering capacity. This sustained warming signal is consistent with prior observations in the region.

4.3. LST pattern around the industries

The spatiotemporal analysis of land surface temperature (LST) around the industrial zone (figure 10(a)) revealed a significant increase in surface heating from 2006 to 2024. In 2006, temperatures were moderate, ranging from 30 to 38 °C. However, by 2016 and subsequent years, large areas transitioned into high-temperature zones exceeding 44 °C, particularly around industrial clusters. The boxplot analysis (figure 10(b)) supports this trend, showing that the median LST rose from approximately 42 °C in 2006 to around 48 °C in 2024. Additionally, upper whisker values surpassed 52 °C in the later years, indicating a rising frequency and intensity of thermal extremes. This assessment of LST, focused on industrial activity, was conducted to capture the localised thermal impact of concentrated anthropogenic and industrial processes.

4.4. Spearman's correlation of LST with biophysical indices

The correlation analysis revealed significant relationships between land surface temperature (LST) and biophysical indices from 2006 to 2024, as shown in table 6. LST had a strong negative correlation with the Normalized Difference Vegetation Index (NDVI), ranging from –0.59 to –0.66, with the strongest correlation in 2016 at –0.656, highlighting the cooling effect of vegetation. In contrast, the Normalised Difference Built-up Index (NDBI) showed a strong positive correlation with LST, ranging from +0.62 to +0.65, peaking in 2021 at +0.648, indicating built-up areas contribute to warming. Albedo had a moderate positive correlation with LST (0.43 to 0.48), rising to +0.476 in 2024, while the Enhanced Built-up and Bare Soil Index (EBBI) also showed a moderate association (0.39–0.46), reflecting the influence of bare and impervious surfaces on thermal dynamics.

4.5. Vegetation threshold value

A piecewise (segmented) regression was fitted to multi-year NDVI versus LST data (excluding extreme NDVI values of <0.05 or >0.9) to detect any changes in the vegetation–temperature relationship. As expected, NDVI and LST exhibited a strong inverse relationship (higher greenness → cooler surface), consistent with global patterns. The piecewise fit revealed a precise breakpoint at NDVI = 0.32 (figure 11). Below this threshold, the fitted line had a steep negative slope. This relationship indicates that vegetation significantly reduces LST only when NDVI exceeds a threshold of 0.32. Below this point, NDVI has a minimal impact on temperature

4.6. Composite fragmentation index

The PCA-based analysis, utilizing the most suitable landscape metrics, such as NP, Cohesion, and LPI, indicates that the first two components (PC1 and PC2) account for 70.1% of the total variance. Weights of 0.554 and

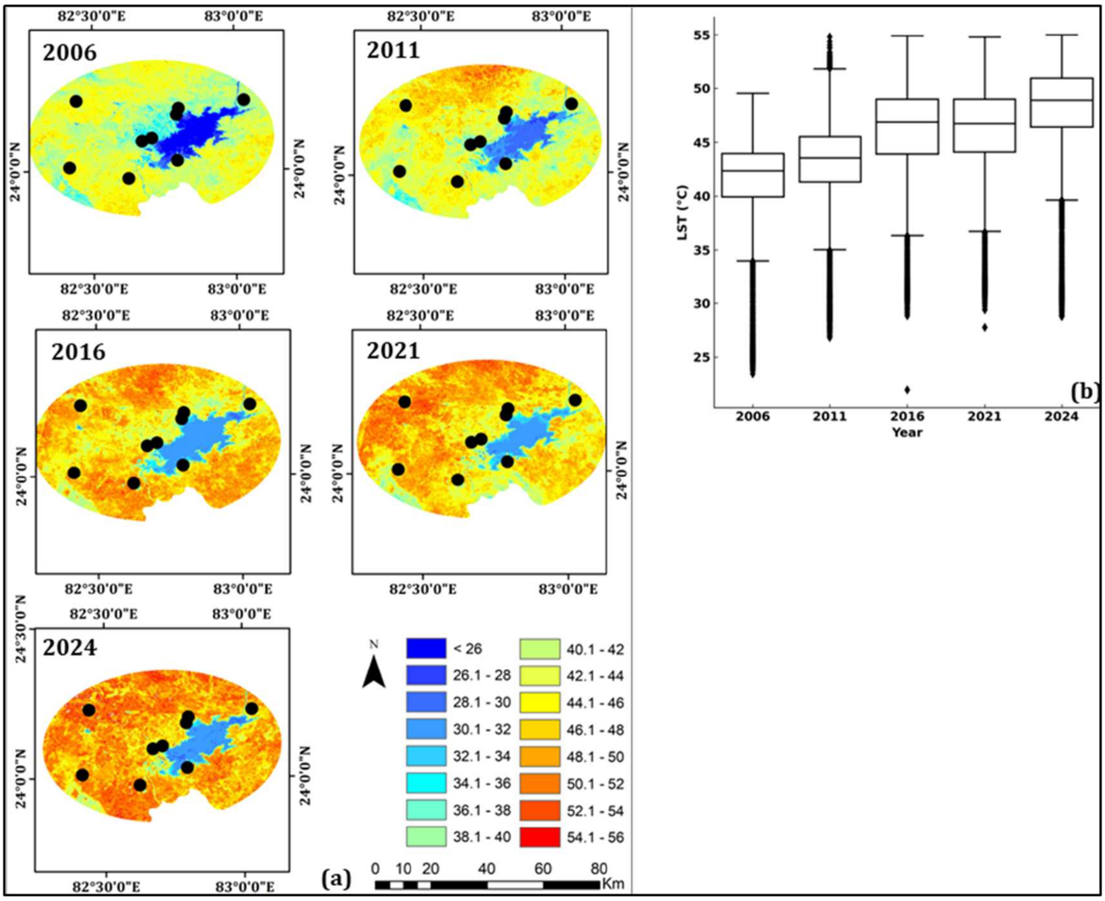

Figure 10. (a) Spatial distribution of LST in the industrial cluster zone for 2006, 2011, 2016, 2021, and 2024, highlighting intensification of heat around industrial sites (black dots). (b) Temporal changes in LST are summarised using box plots.

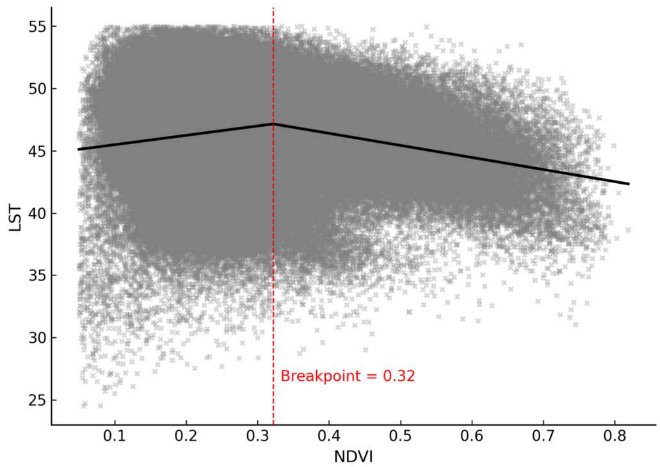

Figure 11. A piecewise regression line highlights the breakpoint at NDVI = 0.32 (red dashed line), indicating a threshold in vegetation.

Table 6. Spearman's correlation analysis between LST and NDVI, NDBI, EBBI, and Albedo for the years 2006, 2011, 2016, 2021, and 2024.

Year	NDVI	NDBI	Albedo	EBBI
LST2006	–0.631	+0.622	+0.485	+0.436
LST2011	–0.630	+0.628	+0.434	+0.457
LST2016	–0.656	+0.621	+0.435	+0.396
LST2021	–0.594	+0.648	+0.427	+0.424
LST2024	–0.630	+0.628	+0.476	+0.456

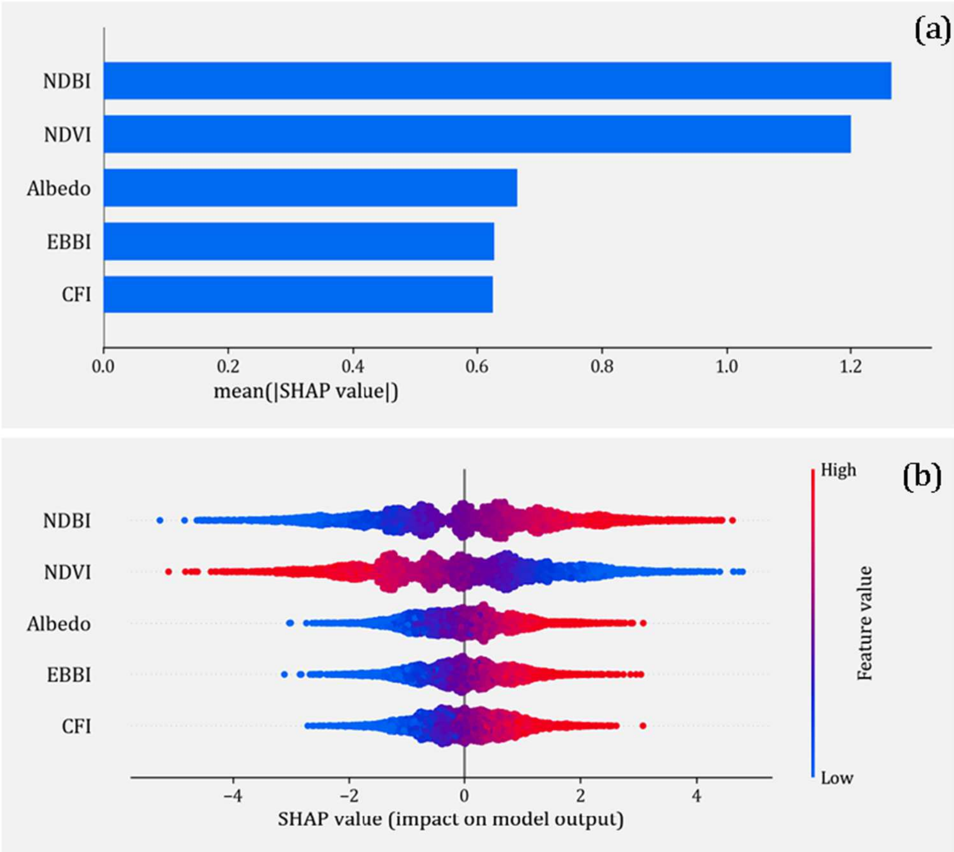

Figure 12. SHAP results for the XGB model: (a) mean SHAP values ranking predictor importance; (b) beeswarm plot showing direction of feature impacts on LST.

0.446 were assigned to PC1 and PC2, respectively, to formulate the Composite Fragmentation Index (CFI) as shown in equation (4).

$$\mathrm{CFI} = 0.554 \times \mathrm{PC1} + 0.446 \times \mathrm{PC2} \tag{4}$$

Higher CFI scores indicate more fragmented landscapes, while lower scores suggest more cohesive land cover structures.

4.7. Predictive model result

4.7.1. Feature selection

The summary bar chart of mean absolute SHAP values (figure 12(a)) quantifies the overall importance of features. NDBI (1.23) and NDVI (1.15) emerged as the dominant predictors, followed by Albedo (0.67), EBBI (0.64), and CFI (0.63). This ranking confirms that both landscape composition (NDVI, NDBI, Albedo, EBBI) and configuration (CFI) contribute to model predictions, with built-up expansion and vegetation cover exerting the strongest influence on thermal variation.

The SHAP beeswarm plot (figure 12(b)) illustrates the distribution and direction of each predictor's influence on LST. High NDBI values (red) are strongly associated with positive SHAP scores, confirming the warming role of built-up areas. In contrast, high NDVI values (red) shift towards negative SHAP scores, indicating that vegetation-driven cooling is the dominant factor. Albedo exhibits a moderate cooling effect, EBBI reflects warming from bare surfaces, and CFI demonstrates a weaker yet consistent tendency for higher fragmentation to increase LST.

4.7.2. Model performance metrics

Both Random Forest and XGBoost demonstrated strong performance across the temporal and spatial validation schemes presented in table 7. In the Leave-One-Location-Out (LOLO) setting, XGBoost achieved an average accuracy of $R^2 = 0.77$, with RMSE = 1.68 and MAE = 1.32 errors, slightly outperforming Random Forest, which recorded $R^2 = 0.76$ and RMSE = 1.70. Accuracy was marginally lower under the spatial block cross-validation (CV), where XGBoost achieved an R^2 of 0.76, and under Leave-One-Year-Out (LOYO), with a

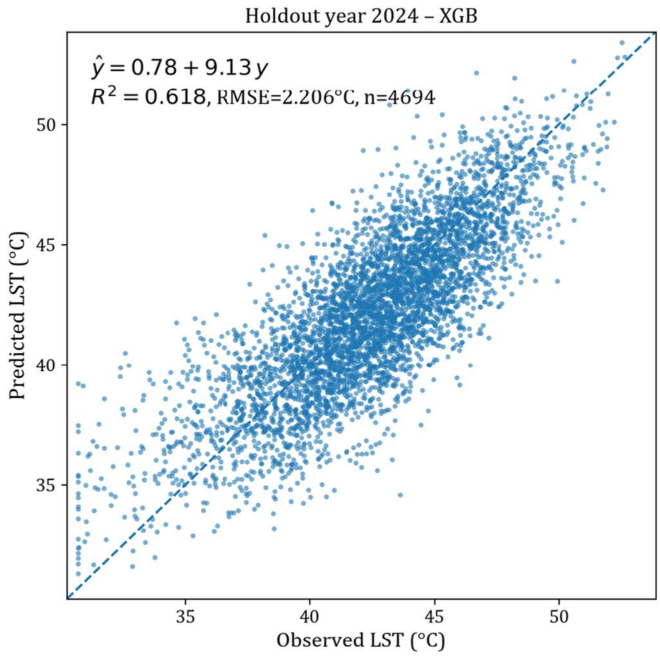

Figure 13. Scatter plot comparing observed and XGBoost-predicted LST values for 2024.

Table 7. Model performance metrics for LST prediction using the XGBoost algorithm.

Model	Scheme	R^2	RMSE	MAE
RF	LOYO	0.750	1.740	1.389
RF	Spatial Block	0.757	1.703	1.344
RF	LOLO	0.769	1.708	1.346
XGB	LOYO	0.755	1.721	1.373
XGB	Spatial Block	0.763	1.681	1.326
XGB	LOLO	0.775	1.685	1.329

validation R^2 of 0.75. The hold-out validation using data from 2024 is illustrated in a scatter plot comparing observed and predicted Land Surface Temperature (LST), as shown in figure 13.

4.7.3. Sensitivity analysis

The sensitivity analysis result presented in figure 14 further supported these findings. Shows that a ±10% perturbation of NDVI and NDBI led to the largest mean prediction shifts (0.68–0.70 °C), while albedo (0.43 °C), EBBI (0.41 °C), and CFI (0.12 °C) produced comparatively smaller responses.

4.7.4. Residual diagnostics

The residual diagnostics, as shown in figure 15, indicate a well-behaved error structure, with residuals distributed symmetrically around zero and the majority confined within ±2 °C, reflecting the strong predictive accuracy of the XGB model.

4.7.5. Conformal prediction analysis

The conformal prediction analysis provided probabilistic bounds around model estimates. For Random Forest, the half-widths of the prediction intervals were ±2.87 °C at the 90% confidence level and ±3.44 °C at the 95% confidence level. XGBoost yielded slightly narrower intervals of ±2.83 °C (90%) and ±3.41 °C (95%). These results indicate that in 90%–95% of cases, actual LST values are expected to lie within approximately ±3 °C of model predictions, with XGBoost offering marginally more reliable uncertainty calibration.

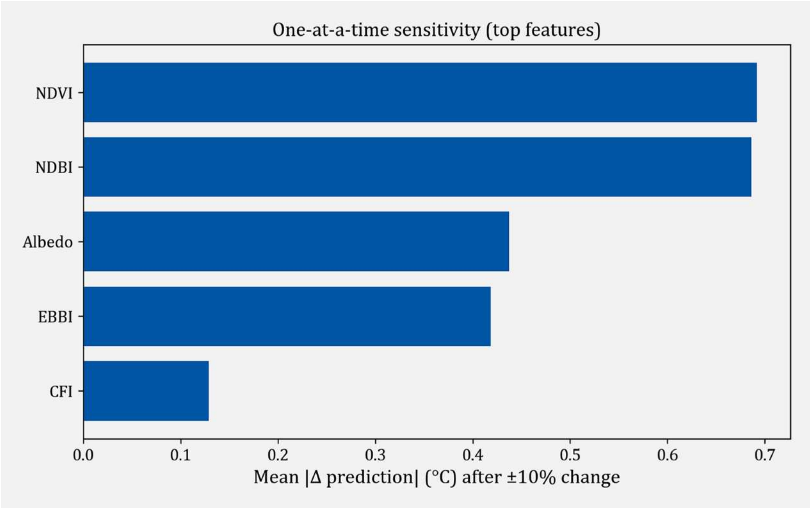

Figure 14. Sensitivity analysis showing the mean prediction change after ±10% perturbation of each predictor.

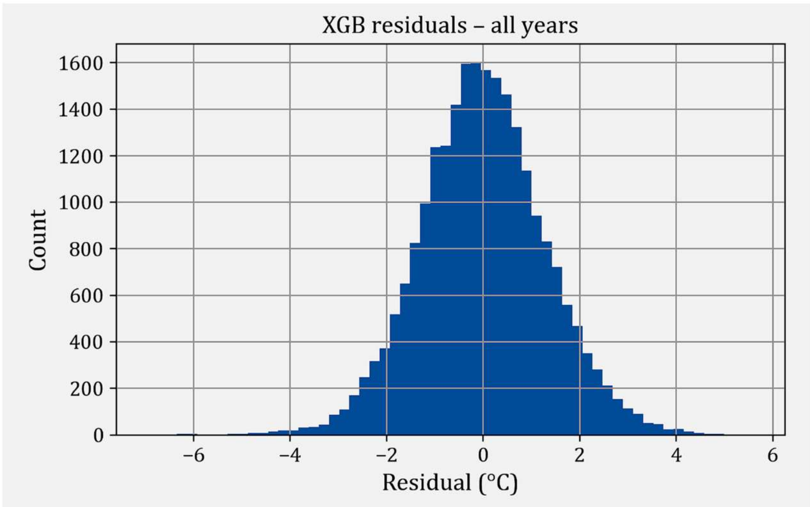

Figure 15. Distribution of XGB residuals across all years, centred near zero, with most within ±2 °C.

Table 8. Stratified performance assessment using an industrial block.

Cross Validation	R^2	RMSE	MAE
LOYO	0.680	1.99	1.58
Spatial Block	0.702	1.96	1.54
LOLO	0.699	1.95	1.53

4.7.6. Stratified performance assessment

The stratified performance assessment in table 8 for the industrial block yielded R^2 values between 0.68 and 0.70, corresponding RMSE values of 1.95–1.99, and MAE values of 1.53–1.58. In comparison, the overall model performance across the study area showed higher R^2 values of 0.755–0.775 and lower errors (RMSE: 1.68 °C–1.72 °C; MAE: 1.33 °C–1.37 °C) under LOYO, LOLO, and Spatial Block cross-validation.

5. Discussion

5.1. LST trends in industrial landscapes

The analysis shows an apparent increase in land surface temperature (LST) in the study area from 2006 to 2024, with thermal hotspots emerging in built-up and bare land areas. This aligns with the industrial heat island effect, where rapid industrialisation leads to local warming (Jiang *et al* 2025, Singh *et al* 2023). A similar study has observed that industrial clusters act as sub-heat islands embedded within broader urban heat patterns (Mao *et al* 2020). In this study, the interquartile range (IQR) in figure 10(b) widened slightly over time, suggesting an increase in variability in the thermal response, likely due to the expansion of power plants, mines, and settlements at the expense of vegetated land. This can be observed in LULC dynamics which represents a significant expansion of cropland (+24.52%) and built-up areas (+3.69%), accompanied by substantial losses in tree cover (–17.13%) and bare land (–11.79%) as shown figure 7. The significant expansion of cropland and built-up areas, along with the substantial loss of tree cover and bare land, suggests rapid urbanization and agricultural intensification in the industrial area. This transformation is likely due to population growth, economic development, and increased demand for housing and food production. The expansion of the built-up regions is directly attributed to industrial growth, infrastructure development, and urban sprawl. Tree cover loss could result from deforestation for urban expansion and agricultural purposes. At the same time, the decrease in bare land indicates the conversion of previously unused areas for development and cultivation. These changes highlight the need for sustainable land management practices to balance economic growth with environmental conservation in the industrial region. These changes are not merely structural but have pronounced thermal implications, as represented by the LST maps (figure 3). The LULC class-wise LST patterns in figure 8 also clearly indicate that bare land and built-up areas consistently record the highest median LST, often exceeding 45 °C. Again, the class-wise widening LST ranges suggest increased surface heterogeneity.

5.2. Biophysical drivers and vegetation thresholds

Spectral indices further highlighted the biophysical control on LST. NDVI showed a negative correlation with LST. In contrast, NDBI, Albedo, and EBBI correlate positively with LST, suggesting that more vegetation tends to a cooler surface using mechanisms such as shading and evapotranspiration. Urban and bare areas are warmer, suggesting increased impermeable development is associated with higher surface temperatures. Similarly, EBBI relations indicate that areas dominated by exposed soil or construction materials experience elevated temperatures. These findings align with observations by Rahimi *et al* (2025) and Pearson's r was found to be equal to 0.65–0.68 with NDBI and $r = 0.63$ with EBBI, respectively, by Sarif and Gupta 2022. While albedo is typically linked negatively to LST, this can vary in semi-arid or industrial areas. Studies have observed that bright but barren surfaces (e.g., mine tailings, fly-ash dumps, dry salt flats, or rocky outcrops) can exhibit positive LST–albedo correlations (Haashemi *et al* 2016). This is often due to the prevalence of impervious surfaces that both reflect more sunlight (higher albedo) and retain heat, leading to higher surface temperatures (Singh *et al* 2023). These observations highlighted the broader trend that shows replacing the natural land cover with impervious surfaces is linked to increased local LST. Impervious materials possess high thermal inertia and absorb solar energy, contributing to the development of urban and industrial heat islands due to both their heat retention and the release of anthropogenic heat.

Recent studies indicate that the relationship between NDVI and LST is non-linear and can be better understood through segmented fits (Song *et al* 2025). Our piecewise analysis identifies a critical NDVI value of 0.32, beyond which the cooling effect becomes more pronounced. While many studies report cooling thresholds regarding canopy coverage or area, rather than NDVI. For instance, Zhou *et al* (2023), specified cooling thresholds and a breakpoint for urban green spaces. Rahimi *et al* (2025) also examined MODIS LST in conjunction with Sentinel-2 vegetation data using polynomial regression. They found that nighttime LST is most effectively reduced when the NDVI surpasses a critical density threshold of 0.4. Such a phenomenon aligns with the concept of biophysical cooling efficiency, indicating that denser vegetation dissipates heat more effectively. Thus, the threshold value 0.32 appears to signify the onset of functionally effective greenness.

5.3. Landscape fragmentation and spatial configuration

Our study highlights the importance of spatial configuration of land cover, quantified through a Composite Fragmentation Index (CFI), in influencing Land Surface Temperature (LST). In our industrial context, we observed that hotspots were especially pronounced in areas where greenery was sparse and highly fragmented by barren lands or infrastructure, creating a mosaic of hot and cool spots. This pattern reinforces that thermal stress arises from the quantity of green cover lost and the spatial distribution of the remaining vegetation. This suggests that the distribution of greenery and built areas affects cooling efficiency. Large, contiguous green patches provide better cooling effects than numerous small, isolated patches. The result supports that fragmented vegetation loses some microclimate regulation capacity due to edge effects and reduced transpiration (Kowe *et al* 2021). In urban areas, fragmented greenery amidst barren lands creates temperature hotspots. Our results align with landscape ecology studies showing that fragmented vegetation can lose some cooling power compared to continuous green spaces (Wu *et al* 2021).

To effectively mitigate thermal patterns, it is crucial to consider the amount of green cover and its connectivity. A cohesive green infrastructure (e.g., corridors, larger parks) yields cooler temperatures than fragmented layouts. This highlights the importance of enhancing vegetation quantity and spatial coherence in peri-urban areas to mitigate heat and reduce fragmentation, thereby promoting cooling and enhancing overall environmental health.

5.4. Machine learning model performance and interpretation

The XGBoost model achieved strong predictive accuracy for LST, with a cross-validated R^2 of 0.77 and an RMSE of 1.7 °C, slightly surpassing the Random Forest model (R^2 of 0.76, RMSE of 1.7 °C). This aligns with XGB's effectiveness in managing non-linear interactions. In a similar remote-sensing study by Li *et al* (2019) Tree-based ensembles have demonstrated improved predictive power compared to linear models. In another study, gradient-boosted trees, such as XGB, typically outperformed bagged trees, like RF, due to their error-correcting mechanisms. (Sun *et al* 2022). The cross-validation result also reported that the performance drops to R^2 (0.70) using the industrial area as a special zone, which involves a special block cross-validation process. The hold-out validation using 2024 data also yielded a decent R^2 (0.61) result, as shown in figure 13. This indicates that while the models learned genuine environment-LST relationships, predicting entirely new locations remains a challenge.

The SHAP values and sensitivity analysis showed that NDVI and NDBI were the dominant factors. Additionally, secondary factors, such as CFI, showed conditional effects, suggesting that spatial configuration has a significant impact on local temperatures, particularly at moderate levels of vegetation. The XGBoost model has a 90% confidence interval for land surface temperature (LST) of approximately $\pm$2.8 °C, indicating a 90% probability that the true LST lies within this range of the predicted value. Such interval estimates enhance decision-making robustness and are rarely reported in remote sensing LST models. The intervals were slightly narrower for XGBoost compared to RF, showing comparable reliability in uncertainty calibration. A study found a similar result, which confirms the significance of composition and configuration for analysing LST (Mao *et al* 2020). Our model's performance and interpretations align with peer-reviewed studies, reinforcing that data-driven LST models achieve high accuracy. Both land cover composition and configuration have a significant influence on surface temperature.

5.5. Ecological and urban planning implications of the findings

The findings of this study demonstrate that land surface temperature in industrial peri-urban landscapes is predominantly governed by the quantity and configuration of vegetation, as well as the extent of built-up and impervious surfaces. Over the study period, mean LST in built-up areas increased by approximately 2.67 °C, and LST in vegetation increased moderately by 0.52 °C. It underscores the inconsistent anthropogenic influence in urbanised and industrial areas. The overall increase in mean LST of 2.9 °C showed an increase in thermal variability, reflecting the spatial heterogeneity induced by rapid land cover transitions. Critically, piecewise regression analysis identified an NDVI threshold of 0.32 as a breaking point for thermal mitigation. Areas with an NDVI below this threshold experienced local temperatures up to 2.5 °C higher than those of adjacent, well-vegetated patches. These quantitative results underscore the necessity of maintaining adequate vegetation cover and minimising landscape fragmentation to counteract industrial heat effects. Given these findings, urban planners and policymakers should aim for an NDVI greater than 0.32 in industrial planning. This can be achieved through the use of green buffers, minimising vegetation fragmentation, and implementing strategic landscape connectivity requirements within and around industrial clusters. The statistical outputs from the machine learning models in this study support the integration of composite fragmentation for climate adaptation and targeted greening interventions in peri-urban and industrial growth areas.

5.6. Limitations and future work

While this study provides important insights, it is essential to acknowledge some limitations inherent in our research. The models indicate that the 90% conformal prediction interval for land surface temperature (LST) is approximately $\pm$2.8 °C. These highlights uncertainties arising from the use of Landsat-derived LST data at 30-meter resolution, which may obscure fine-scale heterogeneity and localised heating effects, particularly in industrial sub-clusters and densely built-up areas. Additionally, uncertainties related to atmospheric correction algorithms, assumptions about surface emissivity, and parameterisation of retrieval models may propagate through the analysis. Furthermore, the study does not account for important anthropogenic factors that influence air temperature and air pollution, as well as socio-economic variables such as population density, which may limit the predictive accuracy in complex urban environments and could lead to overestimation or underestimation of cooling effectiveness. Future research could address these challenges by utilising multi-sensor fusion and machine learning-based downscaling approaches to better capture sub-pixel temperature variations. Incorporating socio-economic data and anthropogenic heat flux would enhance our understanding of local microclimate drivers and allow for a more robust quantification of uncertainty propagation.

6. Conclusion

This study addressed the critical research problem of assessing the impact of the composition and configuration of the landscape on LST. The primary objective of this study was to develop a machine-learning model to analyse how landscape distribution impacts LST using biophysical and landscape matrices as affecting factors. Key findings indicate that LST has significantly increased in the study area from 2006 to 2024. The most considerable warming was observed in areas with expanding cropland and bare land, which replaced natural vegetation. Our findings showed that effective heat mitigation in industrial areas requires a dual approach: (i) ensuring that vegetation surpasses a critical threshold and (ii) maintaining connected green networks to enhance cooling. A critical threshold at NDVI = 0.32 indicates that increasing healthy vegetation can significantly reduce temperatures, suggesting a target for urban planning. The Composite Fragmentation Index (CFI) showed that the arrangement of vegetation is crucial, as fragmented green spaces lose cooling capacity compared to larger, continuous areas. The ensemble models explain 77% of the variance with an RMSE of 1.7. The XGBoost model consistently outperformed the Random Forest model by a small margin and demonstrated robust generalisation across temporal and spatial validation splits. Conformal prediction intervals (± 3 °C) reinforced the model's reliability, providing invaluable uncertainty bounds for decision-makers. The implications of these findings are profound. They underline the need for policymakers to integrate the measures into economic planning and development strategies. Incorporating vegetation threshold and fragmentation factors could help mitigate LST to some extent. Moreover, the study contributes to the existing literature by highlighting the critical aspect of landscape configuration. This research acknowledges limitations, such as considering factors like air pollution and socioeconomic status, which could help. Additionally, the image resolution may not effectively capture micro-hotspots or nocturnal temperature dynamics. Future research could benefit from integrating more driving factors and higher-resolution images. In conclusion, our study emphasises that effective heat mitigation in industrialising areas requires a two-pronged strategy. This strategy involves increasing vegetation to critical levels, ensuring that green spaces exceed threshold densities, and optimising vegetation configuration by maintaining cohesive green networks and managing built surfaces.

Declaration

The authors declare that all statements regarding funding, competing interests, and publication status have been provided below as required by the journal.

Funding

The authors did not receive support from any organisation for the submitted work.

Competing interests

All authors certify that they have no affiliations with or involvement in any organisation or entity with any financial or non-financial interest in the subject matter or materials discussed in this manuscript.

Dual publication

The results/data/figures in this manuscript have not been published elsewhere, nor are they under consideration (from you or one of your Contributing Authors) by another publisher.

Third-party material

The authors own all the material, and/or no permissions are required.

Data availability statement

Due to the large size of the geospatial and model-generated datasets and server storage limitations, the complete dataset cannot be hosted publicly. Processed subsets are available upon reasonable request. The data that support the findings of this study are available upon reasonable request from the authors.

Author contributions

Sheela
Conceptualization (lead), Data curation (lead), Formal analysis (lead), Methodology (lead), Supervision (equal), Validation (lead), Visualization (lead), Writing – original draft (lead), Writing – review & editing (equal)

Medha Jha 0000-0001-8532-0633
Supervision (equal)

Anurag Ohri 0000-0002-6963-6109
Supervision (equal)

References

Bala R, Prasad R and Pratap Yadav V 2019 Disaggregation of MODIS land surface temperature in urban areas using improved thermal sharpening techniques *Adv. Space Res.* **64** 591–602
Bao T L G, Li X M, Zhang J, Zhang Y J and Tian S Z 2016 Assessing the distribution of urban green spaces and its anisotropic cooling distance on urban heat Island pattern in Baotou, China *ISPRS International Journal Of Geo-Information* **5** 1–13
Dai Z, Guldmann J M and Hu Y 2018 Spatial regression models of park and land-use impacts on the urban heat island in central Beijing *Sci. Total Environ.* **626** 1136–1147
De Montis A, Serra V, Ganciu A and Ledda A 2020 Assessing landscape fragmentation: a composite indicator *Sustainability* **12** 1–25
Debnath M, Poddar I, Saha A, Islam N and Roy R 2025 Analysis of land surface temperature and its correlation with urban, water, vegetation, and surface indices in the Siliguri Urban Agglomeration area of the Himalayan foothill region *Geomatics, Natural Hazards and Risk* **16** 2470393
Guha S and Govil H 2020 Land surface temperature and normalized difference vegetation index relationship: a seasonal study on a tropical city *SN Applied Sciences* **2** 1–14
Guha S and Govil H 2022 Seasonal impact on the relationship between land surface temperature and normalized difference vegetation index in an urban landscape *Geocarto Int.* **37** 2252–72
Guha S, Govil H, Gill N and Dey A 2021 A long-term seasonal analysis on the relationship between LST and NDBI using Landsat data *Quat. Int.* **575** 249–58
Guo L, Liu R, Men C, Wang Q, Miao Y and Zhang Y 2019 Quantifying and simulating landscape composition and pattern impacts on land surface temperature: a decadal study of the rapidly urbanizing city of Beijing, China *Sci. Total Environ.* **654** 430–40
Haashemi S, Weng Q H, Darvishi A and Alavipanah S K 2016 Seasonal variations of the surface urban heat Island in a Semi-Arid City *Remote Sens.* **8** 1–17
Jiang J and Tian G 2010 Analysis of the impact of Land use/Land cover change on Land Surface Temperature with Remote Sensing *Procedia Environ. Sci.* **2** 571–5
Jiang W, Wang Y and Zhang M 2025 Exploring the industrial heat island effects and key influencing factors in the guangzhou–foshan metropolitan area *Sustainability* **17** 1–21
Kedron P J, Frazier A E, Ovando-Montejo G A and Wang J 2018 Surface metrics for landscape ecology: a comparison of landscape models across ecoregions and scales *Landscape Ecology* **33** 1489–504
Kotharkar R, Bagade A and Ramesh A 2019 Assessing urban drivers of canopy layer urban heat island: A numerical modeling approach *Landscape and Urban Planning* **190** 1–12
Kowe P, Mutanga O, Odindi J and Dube T 2021 Effect of landscape pattern and spatial configuration of vegetation patches on urban warming and cooling in Harare metropolitan city, Zimbabwe *Gisci. Remote Sens.* **58** 261–80
Kulsum U and Moniruzzaman M 2022 Exploring the relationship of climate change and land-use dynamics with satellite-derived surface indices and temperature in greater Dhaka, Bangladesh *J. Earth Syst. Sci.* **131** 117
Li B, Zhang L, Yan Q and Xue Y J 2014 Application of piecewise linear regression in the detection of vegetation greenness trends on the Tibetan Plateau *Int. J. Remote Sens.* **35** 1526–39
Li Z, Chi D, Chen T, Xia G and Wu Q 2019 Application of penalized linear regression and ensemble methods for drought forecasting in Northeast China *Meteorol. Atmos. Phys.* **132** 113–30
Liu Y, Peng J and Wang Y 2018 Efficiency of landscape metrics characterizing urban land surface temperature *Landscape and Urban Planning* **180** 36–53
Lu X-Y Y, Chen X, Zhao X-L L, Lv D-J J and Zhang Y 2021 Assessing the impact of land surface temperature on urban net primary productivity increment based on geographically weighted regression model *Sci. Rep.* **11** 22282
Macarof P and Statescu F 2017 Comparasion of NDBI and NDVI as indicators of surface urban heat Island effect in landsat 8 imagery: A case study of Iasi *Present Environment and Sustainable Development* **11** 141–50
Maimaitiyiming M, Ghulam A, Tiyip T, Pla F, Latorre-Carmona P, Halik Ü, Sawut M and Caetano M 2014 Effects of green space spatial pattern on land surface temperature: Implications for sustainable urban planning and climate change adaptation *ISPRS J. Photogramm. Remote Sens.* **89** 59–66

Mallick J, Rahman A and Singh C K 2013 Modeling urban heat islands in heterogeneous land surface and its correlation with impervious surface area by using night-time ASTER satellite data in highly urbanizing city, Delhi-India *Adv. Space Res.* **52** 639–55
Mao C, Xie M and Fu M 2020 Thermal response to patch characteristics and configurations of industrial and mining land in a Chinese mining city *Ecol. Indic.* **112** 106075
Masoudi M, Tan P Y and Fadaei M 2021 The effects of land use on spatial pattern of urban green spaces and their cooling ability *Urban Climate* **35** 1–17
Mcgarigal K and Marks B J 1995 *FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structure* 351 United States Deparment of Agriculture Pacific Nrothwest Research Station 122
Miniandi N D, Jamal M H, Muhammad M K I, Sharrar L and Shahid S 2025 Machine learning in modeling urban heat Islands: A data-driven approach for Kuala Lumpur *Earth Syst. Environ.* **9** 1037–60
Motlagh Z K, Lotfi A, Pourmanafi S, Ahmadizadeh S and Soffianian A 2020 Spatial modeling of land-use change in a rapidly urbanizing landscape in central Iran: integration of remote sensing, CA-Markov, and landscape metrics *Environ. Monit. Assess.* **192** 1–19
Mustafa A and Szydlowski M 2020 The impact of spatiotemporal changes in land development (1984-2019) on the increase in the runoff coefficient in Erbil, Kurdistan region of Iraq *Remote Sens.* **12** 1–21
Pandey A, Mondal A, Guha S, Upadhyay P K and Singh D 2023 A long-term analysis of the dependency of land surface temperature on land surface indexes *Papers in Applied Geography* **9** 279–94
Pandeya A, Mondalb A and Guhac S 2022 A seasonal investigation on land surface temperature and spectral indices in Imphal City, India *Journal of Landscape Ecology* **15** 1–18
Rahimi D, Kii M, Shimadera H and Causone F 2025 Exploring seasonal and diurnal variations of the thermal environment in metropolitan scale analysis based on remote sensing data *Remote Sens.* **17** 1–25
Rahimi E, Barghjelveh S and Dong P 2021 Quantifying how urban landscape heterogeneity affects land surface temperature at multiple scales *J. Ecol. Environ.* **45** 1–13
Rahimi E, Dong P and Jung C 2025 Global NDVI-LST correlation: Temporal and spatial patterns from 2000 to 2024 *Environments - MDPI* **12** 1–14
Ramaiah M, Avtar R and Rahman M M 2020 Land cover influences on LST in two proposed smart cities of India: comparative analysis using spectral indices *Land* **9** 292
Sahoo S *et al* 2022 Analysis of decadal land use changes and its impacts on Urban Heat Island (UHI) using remote sensing-based approach: a smart city perspective *Sustainability.* **14** 11892
Sarif M and Gupta R 2022 Evaluation of seasonal ecological vulnerability using LULC and thermal state dynamics using Landsat and MODIS data: a case study of Prayagraj City, India *Environ. Sci. Pollut. Res.* **29** 77502–35
Sharma M, Bangotra P, Gautam A S and Gautam S 2021 Sensitivity of normalized difference vegetation index (NDVI) to land surface temperature, soil moisture and precipitation over district Gautam Buddh Nagar, UP, India *Stochastic Environmental Research and Risk Assessment* **36** 1779–89
Singh P, Verma P, Chaudhuri A, Singh V and Rai P 2023 Evaluating the relationship between Urban Heat Island and temporal change in land use, NDVI and NDBI: a case study of Bhopal city, India *Int. J. Environ. Sci. Technol.* **21** 3061–72
Singh V K, Mohan M and Bhati S 2023 Industrial heat island mitigation in Angul-Talcher region of India: evaluation using modified WRF-single urban canopy model *Sci. Total Environ.* **858** 159949
Song Y, Song X and Shao G 2020 Effects of green space patterns on urban thermal environment at multiple spatial-temp *Sustainability* **12** 1–18
Song Y, Xu H, Liu T, Xu J and Song X 2025 Linking spatiotemporal variations in urban land surface temperature to land use and land cover: a case study in Hangzhou City, China *Ecol. Indic.* **173** 113336
Sultana S and Satyanarayana A N V V 2019 Impact of urbanisation on urban heat island intensity during summer and winter over Indian metropolitan cities *Environ. Monit. Assess.* **191** 789
Sun X, Zhang H, Wang J, Shi C, Hua D and Li J 2022 Ensemble streamflow forecasting based on variational mode decomposition and long short-term memory *Sci. Rep.* **12** 1–19
Sun Z, Li Z and Zhong J 2022 Analysis of the impact of landscape patterns on urban heat Islands: A case study of Chengdu, China *Int. J. Environ. Res. Public Health* **19** 1–15
Toms J and Lesperance M 2003 Piecewise regression: a tool for identifying ecological thresholds *Ecology* **84** 2034–41
Wu C, Li J, Wang C, Song C, Haase D, Breuste J and Finka M 2021 Estimating the cooling effect of pocket green space in high density urban areas in Shanghai, China *Front. Environ. Sci.* **9** 657969
Wu Q, Li Z Y, Yang C B, Li H Q, Gong L W and Guo F X 2022 On the scale effect of relationship identification between land surface temperature and 3D landscape pattern: The application of random forest *Remote Sens.* **14** 1–24
Zawadzka J E, Harris J A and Corstanje R 2021 The importance of spatial configuration of neighbouring land cover for explanation of surface temperature of individual patches in urban landscapes *Landscape Ecology* **36** 3117–36
Zhang Y, Wang Y and Ding N 2022 Spatial effects of landscape patterns of urban patches with different vegetation fractions on urban thermal environment *Remote Sens.* **14** 1–20
Zhao H, Wang H, Li Y, Yao N, Chen X and Liu F 2020 Monitoring monthly soil moisture conditions in China with temperature vegetation dryness indexes based on an enhanced vegetation index and normalized difference vegetation index *Theor. Appl. Climatol.* **143** 159–76
Zheng B, Myint S W and Fan C 2014 Spatial configuration of anthropogenic land cover impacts on urban warming *Landscape and Urban Planning* **130** 104–11
Zhou W, Yu W, Zhang Z, Cao W and Wu T 2023 How can urban green spaces be planned to mitigate urban heat island effect under different climatic backgrounds? A threshold-based perspective *Sci. Total Environ.* **890** 164422
Zhou X and Wang Y 2011 Dynamics of land surface temperature in response to land-use/cover change *Geogr. Res.* **49** 23–36