Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Land Surface Temperature Responses to Land-Cover Change: Machine Learning Insights from Semi-Arid Peshawar, Pakistan

Sami Ur Rahman
samiurrahman418@gmail.com
Fujian Normal University
Abdul Samad
Fujian Normal University
Jinming Sha
Fujian Normal University
Xiaomei Li
Fujian Normal University
Attaur Rahman
Shaheed Benazir Bhutto University
Waqar Akhtar
Fujian Normal University
Zeeshan Ali
Fujian Normal University

Research Article

Keywords: Land surface temperature, Land-use/land-cover change, Surface urban heat island, Google Earth Engine, Tree-ensemble regression

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11663077/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Multi-epoch assessments of land surface temperature (LST) often associate thermal differences with land-use/land-cover (LULC) change without separating compositional effects from epoch-wide variations in meteorological conditions, sensors, and retrieval processing. This study examines LULC-associated summer (June–August) LST patterns in Peshawar District (≈ 1,259 km²), a semi-arid basin in northwestern Pakistan, for 2001, 2014, and 2024. Landsat Collection 2 Level-2 imagery was processed in Google Earth Engine, and LULC was classified using gradient tree boosting, achieving overall accuracies of 93–97% and kappa coefficients of 0.90–0.96. Random Forest (RF), XGBoost, and LightGBM were employed to model LST using five spectral indices and elevation. Between 2001 and 2024, forest cover declined from 308 to 87 km², while built-up land expanded from 287 to 445 km². District-mean composite LST increased by 10.3°C, although comparable increases across land-cover classes, including forest (+ 8.2°C) and water (+ 6.1°C), indicate substantial epoch-wide differences. Compositional decomposition attributed less than 1°C of the district-mean difference to changing class proportions, leaving the remaining variation unresolved. Relative thermal contrasts also changed, with built-up land becoming warmer than agricultural and forest areas, particularly during 2001–2014. The three models showed similar performance under random pixel splitting (r^2 ≈ 0.78–0.85; RMSE ≈ 1.7–2.1°C). Predictor importance showed partial agreement but requires reassessment because of a Landsat 8 bare soil index calculation error and predictor redundancy. These findings highlight the importance of compositional decomposition and relative class contrasts when interpreting multi-epoch LST patterns.

1. Introduction

Urban expansion converts vegetated and agricultural land into impervious surfaces. This land-use change affects the surface energy balance, reduces evapotranspiration, and increases radiant heat storage, leading to local temperature increases commonly known as the surface urban heat island (SUHI) effect. Urban heat islands interact with heatwaves, amplifying urban overheating, and affect air quality, energy use, and human health (Santamouris, 2020; Zhou et al., 2019). In low-latitude developing cities, where population growth and unplanned expansion often occur with minimal green space planning, the thermal impact of land cover change can be significant. In arid and semi-arid regions, however, the sign and magnitude of the daytime SUHI depend strongly on the surroundings: where the rural background is sparsely vegetated desert, urban areas with irrigated vegetation can be cooler than their surroundings, whereas where the background is densely vegetated, urban surfaces are usually warmer (Imhoff et al., 2010). Daytime LST in such settings is governed largely by surface energy partitioning the fraction of net radiation consumed by evapotranspiration, which depends on vegetation cover and soil moisture (including irrigation), versus the fraction converted to sensible and stored heat by dry soils and impervious materials (Oke, 1982). For scale, a global analysis of 419 large cities reported a mean annual daytime SUHI intensity of 1.5 ± 1.2°C (Peng et al., 2012), which provides an order-of-magnitude benchmark for the temperature contrast that urban land cover produces relative to its surroundings.

Satellite remote sensing offers consistent, spatially detailed observations for monitoring urban surface temperatures. Thermal data from Landsat and MODIS missions have enabled the detection of detailed land surface temperature (LST) variations across different environments (Li et al., 2017; Weng et al., 2014). However, understanding the causes of these variations, such as vegetation, built-up density, soil moisture, and topography, requires robust multivariate modeling frameworks capable of capturing nonlinear relationships. Because linear parametric models may not capture threshold and interaction effects among these drivers, machine learning (ML) methods are increasingly used in remote-sensing applications such as soil-moisture estimation (Lamichhane et al., 2025).

Among ML algorithms, ensemble-tree frameworks such as Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM) are widely used. RF builds many decorrelated decision trees, reducing variance relative to single trees(Breiman, 2001). XGBoost and LightGBM instead use gradient boosting, adding new trees to address residuals from earlier trees (Chen & Guestrin, 2016; Ke et al., 2017a). Recent studies report higher accuracy than linear or geographically weighted regression for soil-moisture prediction (Ali, 2025) and LST downscaling (Su et al., 2025), and tree ensembles have reached $R^2 > 0.8$ for LST estimation (Rengma & Yadav, 2024). Comparing these algorithms is scientifically informative only if it goes beyond ranking. If structurally different learners (bagging versus boosting; level-wise versus leaf-wise tree growth) reach similar held-out accuracy and similar predictor rankings, the LST–predictor association is less likely to be an artefact of one algorithm; divergence indicates sensitivity to model structure or to redundant predictors. Neither outcome establishes a physical mechanism, because importance measures describe how a fitted model uses correlated predictors, not how the land surface partitions energy (Strobl et al., 2007). Simultaneously, cloud computing platforms such as Google Earth Engine (GEE) provide access to consistently processed satellite archives together with server-side parallel processing (Gorelick et al., 2017a). Coupling GEE's data infrastructure with ML modeling environments such as R or Python enables researchers to combine efficient data preprocessing with advanced analytics.

Peshawar, the provincial capital of Khyber Pakhtunkhwa (KP) in north-western Pakistan, has expanded rapidly since the early 2000s, with built-up land expanding at the expense of vegetated land. The district lies in a semi-arid basin where much of the vegetated land depends on irrigation, so its surface temperature reflects irrigation-dependent evapotranspiration and soil moisture as well as urban expansion. Regional studies have documented substantial vegetation loss (Tabassum et al., 2025) and examined land transformation and temperature in the Peshawar Valley (Hussain et al., 2024), but we are not aware of multi-epoch LST analyses for Peshawar that quantify how much of an observed LST change is associated with changes in land-cover composition.

Three gaps motivate this study. *Gap 1*: multi-epoch LST studies of South Asian cities commonly report absolute LST differences between single-season images and attribute them to LULC change, without separating the contribution of changing class composition from epoch-wide shifts that affect all classes. *Gap 2*: relationships between LST and spectral or topographic predictors are usually evaluated with a single algorithm and random pixel-level validation, which can overstate predictive skill in spatially autocorrelated data (Ploton et al., 2020; Roberts et al., 2017) and leaves the robustness of predictor rankings untested. *Gap 3*: In semi-arid, irrigated basins, SUHI estimates are sensitive to the choice of rural reference surface (Imhoff et al., 2010) yet many studies report normalized anomalies without a physically defined reference.

Each gap is addressed by one objective. (O1) To map LULC for 2001, 2014 and 2024 with accuracy assessment, quantify area change, and decompose the inter-epoch difference in district-mean LST into a compositional (area-change) component and a within-class component (Gap 1). (O2) To compare the held-out performance and predictor rankings of RF, XGBoost and LightGBM on a held-out random test set, and to identify the spatially independent validation needed to test their transferability, as a test of whether LST–predictor associations are robust to model structure (Gap 2). (O3) To quantify SUHI as a temperature difference (°C) between built-up land and explicitly defined reference surfaces, alongside a dimensionless standardized LST anomaly used only for mapping (Gap 3).

We test the hypothesis (H1) that the LST difference associated with LULC transformation can be distinguished from the epoch-wide difference that affects all land-cover classes. Under the null hypothesis (H0), changes in class composition explain only a small fraction of the change in district-mean LST, and class-specific LST changes are of similar magnitude in converted and stable classes. Because only one summer composite is available per epoch, H1 is evaluated descriptively; the design cannot by itself attribute the within-class component to climatic variability, sensor differences or local intensification.

The contribution of the study is methodological and regional rather than algorithmic. For a data-scarce, semi-arid South Asian city, it (i) combines accuracy-assessed LULC maps for three epochs with a simple decomposition that shows how much of a multi-epoch LST difference can be associated with land-cover composition; (ii) evaluates three widely used tree ensembles with a common predictor set on a held-out random test set, treating cross-model agreement as a robustness check rather than as evidence of mechanism; and (iii) documents a GEE–R workflow that could potentially be adapted to other cities, although its transferability has not been tested here.

Once validated, such quantification could support spatially explicit planning in emerging cities, for example by identifying thermal hotspots and locating cooling green belts. Beyond local application, the workflow may be adapted to other data-limited cities, subject to local accuracy assessment and validation.

2. Materials and methods

2.1 Study area

The study concentrates on Peshawar District (33°44'–34°15' N, 71°22'–71°42' E) in northwestern Pakistan, covering approximately 1,259 km² within the Peshawar Valley. The district lies in a small valley bordered by the Khyber Hills to the west, the Bara River to the south and the Budhni River to the north (Development Bank, 2023). The climate is semi-arid subtropical, with average summer highs above 40°C, winter lows below 5°C, and annual precipitation of around 338 mm (Rehmani et al., 2018). Surface water is concentrated in the Kabul River system in the north of the district (Fig. 2), and much of the cropland is irrigated; summer evapotranspiration therefore varies with irrigation scheduling and canal flows as well as with land cover. Over the past two decades, Peshawar has experienced rapid urban expansion. Built-up expansion and vegetation loss have been documented over this period (Tabassum et al., 2025) and are expected to alter surface albedo and evapotranspiration. Figure 1 shows the location and boundary of the study area.

2.2 Data sources and processing environment

Multi-temporal optical and thermal imagery were obtained from Landsat 5 TM (2001) and Landsat 8 OLI/TIRS (2014, 2024) Collection 2 Level-2 surface reflectance (SR) and surface temperature (ST) products for the 2001, 2014, and 2024 epochs. All data were processed in Google Earth Engine (GEE), which provides atmospherically corrected surface reflectance and temperature bands, consistent radiometric calibration, and a large-scale computational framework (Gorelick et al., 2017a). An elevation model (Section 2.8) was incorporated to capture elevation variation across Peshawar's valley. All images were clipped to the region of interest (ROI) defined by the administrative boundary of Peshawar District (≈ 1,259 km²), and re-projected to EPSG:4326. Table 1 summarizes the data and sample sizes for each epoch. Table 2 documents the source, identifier, resolution, period, units, processing, and role of every dataset used.

Table 1
Data inventory and sample sizes by epoch.

Epoch	Sensor (Collection 2, Level-2)	Composite window	Regression samples (n)
2001	Landsat 5 TM (SR + ST_B6)	1 Jun–31 Aug 2001	22,843
2014	Landsat 8 OLI/TIRS (SR + ST_B10)	1 Jun–31 Aug 2014	22,843
2024	Landsat 8 OLI/TIRS (SR + ST_B10)	1 Jun–31 Aug 2024	22,843

Table 2
Data sources, specifications and processing used in this study.

Dataset / product	Variables or bands	Provider; identifier / version	Spatial resolution	Temporal coverage and acquisition	Units / scaling	Access	Quality control and preprocessing
Landsat 5 TM, Collection 2 Level-2 SR and ST (2001)	SR_B1–SR_B5, SR_B7 (blue, green, red, NIR, SWIR1, SWIR2); ST_B6; QA_PIXEL	USGS; GEE: LANDSAT/LT05/C02/T1_L2	30 m for surface reflectance; 120 m native thermal resolution	Summer 2001 (1 June–31 August); WRS-2 path/row 150/036	Surface reflectance: DN × 0.0000275 − 0.2. Surface temperature: DN × 0.00341802 + 149 K; converted to °C	GEE Data Catalog; USGS	Reflectance composites: QA_PIXEL bits 3 and 5 masked; surface-temperature composites: bits 0–4 masked; reflectance clamped to [0, 1]
Landsat 8 OLI/TIRS, Collection 2 Level-2 SR and ST (2014, 2024)	SR_B2–SR_B7; ST_B10; QA_PIXEL	USGS, accessed through GEE;	30 m; ST_B10 resampled from native 100 m	Medians of scenes 1 Jun–31 Aug 2014 and 2024; WRS-2 150/036;	As above	GEE Data Catalog; USGS	As above, but reflectance not clamped; BSI computed with the green band in place of red (Section 2.4)
Spectral indices NDVI, NDBI, NDWI, BSI, NDMI	Eqs. (1)–(5); SWIR = SWIR1; NDBI = − NDMI	Derived in GEE from the Landsat composites	30 m	One composite per epoch	Dimensionless (− 1 to 1)	Exported as GeoTIFF	Clipped to the district boundary
ASTER Global Emissivity Dataset v3, elevation band	Elevation	NASA JPL; NASA/ASTER_GED/AG100_003	100 m, resampled to 30 m	Static	m	GEE Data Catalog	Clipped to the district boundary
LULC reference samples	Class labels (built-up, agriculture, forest, bare land, water)	Digitized from high-resolution imagery and topographic maps	Pixels at 30 m; count per class	One sample set per epoch	Categorical	—	Random 80/20 pixel split without a fixed seed
LULC maps (derived)	Five-class maps	SMILE gradient tree boosting in GEE (Section 2.5)	30 m	2001, 2014, 2024	Categorical; areas in km² by pixel counting	—	Overall accuracy 93–97%, kappa 0.90–0.96 (random split)
Peshawar District boundary	Region of interest (≈ 1,259 km²)	Source	Vector	—	—	—	Used to clip all rasters; data reprojected to EPSG:4326
Regression samples	LST (response); NDVI, NDBI, NDWI, NDMI, BSI, elevation (predictors)	Extracted in GEE; analysed in R 4.4.1	30 m pixels; n = 22,843 per epoch	One set per epoch	°C; dimensionless; m	—	Missing values omitted; random 80/20 split (caret::createDataPartition, seed 123)

2.3 Cloud masking and composite generation

To reduce cloud contamination, composites were generated from all Collection 2 Level-2 scenes acquired between 1 June and 31 August of each target year over the district (WRS-2 Path/Row 150/036). Different quality masks were applied to the two composite types. For the reflectance composites used for LULC classification and spectral indices, pixels flagged by QA_PIXEL bit 3 (cloud) or bit 5 (snow) were masked; cloud shadow (bit 4), dilated cloud (bit 1) and cirrus (bit 2) were not masked. For the surface-temperature composites, pixels flagged by QA_PIXEL bits 0–4 (fill, dilated cloud, cirrus, cloud and cloud shadow) were masked. Unmasked cloud shadows may bias reflectance and index values in shaded areas (Section 5). A per-pixel median of the masked scenes was then computed separately for the reflectance and surface-temperature bands. Because a median composite combines acquisitions from different days, its LST is not a single-date observation; the scenes contributing to each composite are summarised in Table 1.

2.4 Spectral indices

Five indices representing vegetation (NDVI), built-up surfaces (NDBI), water (NDWI), bare soil (BSI) and moisture (NDMI) (Table 3; Eqs. (1)–(5)) were computed from the reflectance composites using standard formulas. For Landsat 5 TM, SR_B1–SR_B5 and SR_B7 correspond to blue, green, red, NIR, SWIR1

and SWIR2; for Landsat 8 OLI, the corresponding bands are SR_B2–SR_B7. Reflectance was scaled with the Collection 2 factors (×2.75 × 10⁻⁵, − 0.2) before index computation. This scaling was applied consistently across all three epochs; however, the resulting reflectance values were clamped to the valid [0, 1] range only in the 2001 (Landsat 5) script and not in the 2014/2024 (Landsat 8) scripts; this inconsistency is present in the results reported here. In the scripts that produced the reported results, the Bare Soil Index (BSI) for the Landsat 8 composites (2014 and 2024) used the green band (SR_B3) in place of the red band (SR_B4), whereas the 2001 (Landsat 5) BSI correctly used the TM red band (SR_B3). The 2014 and 2024 BSI layers, and the model and predictor-importance results that depend on them, are therefore affected by this error (Section 3.6); a corrected script has been prepared, but the analyses have not been regenerated. All indices were scaled and clipped to the ROI, then exported as 30-meter GeoTIFF files. Basic statistics, such as minimum, maximum, and mean, were computed using the reduceRegion() function. As defined in Eqs. (2) and (5), NDBI and NDMI use the same NIR and SWIR bands with opposite signs; if the same SWIR band (SWIR1) is used in both, NDBI = − NDMI exactly and the two indices carry identical information. SWIR1 (SR_B5 for Landsat 5; SR_B6 for Landsat 8) was used for both NDBI and NDMI in all three epochs (2001, 2014, 2024), so NDBI = − NDMI exactly throughout the study; no SWIR2 band was used for either index. This redundancy is taken into account when interpreting predictor importance (Section 3.6). For this reason NDMI was excluded from the XGBoost model whose importance values are reported in Section 3.6.

Table 3
Spectral indices used as predictors (Eqs. (1)–(5)). SWIR denotes SWIR1 (SR_B5 for Landsat 5 TM; SR_B6 for Landsat 8 OLI) in all cases.

$NDVI = \frac{(NIR - RED)}{(NIR + RED)}$	(Huang et al., 2021) Eq. (1)
$NDBI = \frac{(SWIR - NIR)}{(SWIR + NIR)}$	(Ramzan et al., 2022) Eq. (2)
$NDWI = \frac{(Green - NIR)}{(Green + NIR)}$	(Ashraf et al., 2021) Eq. (3)
$BSI = \frac{(SWIR1 + Red) - (NDBI + Blue)}{(SWIR1 + Red) + (NDBI + Blue)}$	(Nguyen et al., 2021) Eq. (4)
$NDMI = \frac{(NIR - SWIR)}{(NIR + SWIR)}$	(Delgado-Moreno & Gao, 2022) Eq. (5)

2.5 Land-use/land-cover classification and accuracy assessment

A five-class LULC scheme was implemented, including built-up areas, agriculture, forest, bare land, and water. Training samples were digitized with high-resolution imagery and verified against historical topographic maps. In Google Earth Engine (GEE), the SMILE Gradient Tree Boosting (GBT) classifier, an implementation of stochastic gradient boosting, was used with 100 trees, a learning rate of 0.1, a sampling rate of 0.7, and a maximum of 16 leaf nodes (Natekin & Knoll, 2013). Inputs comprised optical reflectance bands and five spectral indices. For accuracy assessment, samples were randomly split into 80% for training and 20% for validation without a fixed random seed, so the split cannot be reproduced exactly. Confusion matrices, overall accuracy (OA), producer's and user's accuracies, and Kappa coefficients (KC) were derived using errorMatrix(). If validation pixels were drawn by random splitting from the same digitized polygons as the training pixels, they are spatially dependent and the accuracies may be optimistic; a polygon-level split (all pixels of a polygon in the same set) is preferable. Class areas were obtained by pixel counting; error-adjusted area estimates with 95% confidence intervals (Olofsson et al., 2014) require a probability-based validation sample and were not computed.

2.6 Land surface temperature retrieval and quality control

LST was taken from the Collection 2 Level-2 surface-temperature band (ST_B6 for Landsat 5 TM; ST_B10 for Landsat 8 TIRS). This product is generated by the USGS with a single-channel algorithm that uses ASTER GED-based emissivity and reanalysis atmospheric profiles (Malakar et al., 2018; U.S. Geological Survey, 2021). Digital numbers were converted to degrees Celsius with the published scale factor and offset (Eq. (6)):

$$LST_{^\circ C} = 0.00341802 \times DN_{ST} + 149.0 - 273.15$$

6

ST_B6 (Landsat 5, 2001) and ST_B10 (Landsat 8, 2014 and 2024) were used directly; Eq. (6) is the only conversion applied, and no emissivity correction was applied beyond that embedded in the USGS Level-2 product. Quality control consisted of the QA_PIXEL mask described in Section 2.3; no ST_QA uncertainty screening and no statistical outlier trimming were applied, so the extremes reported in Section 3.2 are single-pixel values of the masked composites. Because TM band 6 (native 120 m) and TIRS band 10 (native 100 m) differ in spectral response and resolution, and both are resampled to 30 m in the Level-2 product, inter-sensor differences are a potential source of bias in the 2001 versus 2014/2024 comparison (Section 5). The resulting LST maps (2001, 2014, and 2024) were used to compute the mean, minimum, and maximum for each land-cover class.

2.7 Surface urban heat island (SUHI) and standardized LST anomaly

SUHI intensity was calculated as:

$$SUHI = Tu - Tr \qquad (7)$$

where T_u and T_r represent the mean LST of urban (built-up) and rural reference surfaces. Tu is the mean LST of built-up pixels and T_r is computed separately for three reference surfaces agriculture, forest and bare land because in a semi-arid basin the sign of SUHI depends on the reference surface (Imhoff et al.,

2010). SUHI from Eq. (7) is a temperature difference in °C. For mapping only, a dimensionless standardized LST anomaly (z_{LST}) was also computed for each pixel i, where the overbar and σ_{LST} denote the district mean and standard deviation for the same epoch:

$$SUHI_{norm} = \frac{LSTi - \overline{LST}}{\sigma\ LST}$$

Positive values identify warmer-than-average pixels. Outputs were visualized using an eight-stage blue-to-red temperature palette. Because Eq. (8) is a z-score, its district mean is zero and its standard deviation is one by construction. It expresses each pixel's position within that epoch's LST distribution in units of standard deviations, not in °C, and it cannot by itself measure the intensity or spatial extent of a heat island. In this manuscript, "SUHI" refers only to Eq. (7).

2.8 Topographic predictor

The elevation predictor in the models reported here was the elevation band of the ASTER Global Emissivity Dataset v3 (NASA/ASTER_GED/AG100_003; (Hulley et al., 2015), clipped to the ROI and resampled from 100 m to 30 m to match the Landsat-derived predictors. This band is an ancillary layer of an emissivity product rather than a dedicated topographic model; a revised script using the 30 m SRTM Global 1 Arc-Second DEM (USGS/SRTMGL1_003; absolute vertical accuracy ± 16 m at 90% confidence; (Farr et al., 2007) has been prepared but was not used for the results reported here. Slope and relief were not derived or included as additional predictors in this study; given that elevation alone showed measurable importance in the predictive models (Fig. 5), future work should evaluate whether slope- and aspect-derived variables, which more directly capture solar exposure and drainage effects relevant to land surface temperature in arid basin environments, improve model performance beyond elevation alone.

2.9 Machine-learning regression and validation

Pixel samples extracted from GEE were merged with the index and elevation values to form one predictor matrix per epoch, with n = 22,843 sample points per epoch (Table 1); the response was composite LST and the predictors were NDVI, NDBI, NDWI, NDMI, BSI and elevation. After omitting missing values, 80% of the records were assigned to training and 20% to testing with createDataPartition() in the R package caret, using a random seed of 123 set once at the start of the script. Because createDataPartition() stratifies on quantiles of the response but ignores location, neighbouring, spatially autocorrelated pixels can fall into both sets, which can inflate accuracy (Ploton et al., 2020; Roberts et al., 2017). Spatially blocked cross-validation (Valavi et al., 2019) was not performed, so the reported accuracies describe prediction within the sampled area rather than transfer to spatially independent areas (Section 5). Three ensemble models were trained (Table 4):

Table 4
Regression model configuration (fixed, untuned hyperparameters). Analyses were run in R 4.4.1.

Model	Key Parameters	Reference
Random Forest (RF)	200 trees, default mtry (⌊p/3⌋ = 2 for p = 6 predictors); R package randomForest	(Breiman, 2001)
Extreme Gradient Boost (XGBoost)	objective = "reg:squarederror", learning rate = 0.1, max depth = 6, nrounds = 100 ; R package xgboost	(Chen & Guestrin, 2016)
Light Gradient Boosting Machine (LGBM)	objective = "regression", num_leaves = 31, learning rate = 0.1, iterations = 100 ; R package lightgbm	(Ke et al., 2017b)

The LULC classifier (SMILE gradient tree boosting in Earth Engine, Section 2.5) and the LST regression models (R packages randomForest, xgboost and lightgbm) are separate implementations used at different stages; settings and results of one do not transfer to the other. Hyperparameters were not tuned; the settings are listed in Table 4. Training time depends on the implementation, number of trees and threading, and is reported (Section 3.5) but not used as an evaluation criterion.

Performance Metrics

$$RMSE = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

9

$$MAE = \frac{1}{n}\sum\nolimits_{i=1}^{n}|y_i - \hat{y}_i| \qquad (10)$$

MAPE was not used because LST in °C is on an interval scale with an arbitrary zero, so percentage errors depend on the temperature unit chosen.

$$R^2 = 1 - \frac{\sum\nolimits_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum\nolimits_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$$

11

$$r^2 = (\mathrm{cor}\,(y_i,\ \hat{y}_i))^2 \qquad (12)$$

Goodness of fit was computed as the squared Pearson correlation between observed and predicted LST (r^2, Eq. (12)). This quantity ignores systematic bias and scale errors and coincides with the coefficient of determination (Eq. (11)) only for least-squares fits evaluated on the data used for fitting (Kv¥lseth, 1985) ; on held-out data, Eq. (11) can be lower and even negative. The coefficient of determination (Eq. (11)) was not computed for the reported results, which are therefore given as r^2 together with RMSE and MAE. In Eqs. (9)–(12), y_i and $\hat{y}_i$ are observed and predicted LST of test pixel i, $\bar{y}$ is the mean observed LST and n is the number of test pixels.

Feature importance was taken from each model's built-in measure: IncNodePurity for RF and gain for XGBoost and LightGBM. These measures are on different scales, are computed on training data and, for IncNodePurity in particular, are biased toward continuous predictors with many candidate split points (Strobl et al., 2007); the importance values are therefore compared only by rank. NDMI was excluded from the XGBoost model used for importance because it duplicates NDBI (Section 2.4); in RF and LightGBM, NDBI and NDMI were scored separately, so their individual ranks are not interpretable. All computations were scripted in R.

2.10 Decomposition of inter-epoch LST differences

To evaluate H1, the difference in district-mean LST between two epochs, $\Delta\bar{T}$, was decomposed exactly into a compositional component and a within-class component:

$$\Delta\ \bar{T} = \sum_k \Delta\, a_k\, T_{k,2001} + \sum_k a_{k,2024}\, \Delta\, T_k$$

13

where a_k is the area fraction and T_k the mean LST of class k. The first term is the change expected if each class had kept its 2001 mean temperature (the effect of land-cover conversion alone); the second is the change within classes. Both orderings of the decomposition (evaluating the compositional term at 2001 or at 2024 class means) are reported. The decomposition is descriptive: within-class change can include LULC-related intensification (e.g., densification within built-up land) as well as meteorological, sensor and processing differences. In addition, contrasts between classes in the same epoch (e.g., built-up minus agriculture), which are insensitive to epoch-wide shifts, are used as LULC-related signals. Formal significance tests of inter-epoch differences were not applied to pixel-level LST because pixels are spatially autocorrelated and each epoch is represented by a single composite (n = 1 per epoch). Placing the epoch differences in climatological context requires reference data, such as ERA5-Land summer skin or 2 m air temperature anomalies for 2001, 2014 and 2024 relative to a multi-year baseline, and the LST of pixels whose class did not change across epochs; neither analysis has yet been performed, and both are identified here as necessary extensions for future work rather than completed components of the present study.

3. Results

3.1 Land-use/land-cover change and classification accuracy

Based on the study area's geographical scenario, the five basic land cover types were selected: Agricultural land, Barren land, Built-up areas, Forest, and Water bodies. The classification results derived from Landsat 5 (2001) and Landsat 8 (2014, 2024) show substantial change in surface composition between 2001 and 2024. Mapped forest decreased from 308 km² in 2001 to 118 km² in 2014 and 87 km² in 2024 (− 221 km², equivalent to 17.55 percentage points of the district area and a 72% relative decrease). Built-up surfaces expanded from 287 km² (2001) to 337 km² (2014) and 445 km² (2024) (+ 158 km²; +12.55 percentage points; +55% relative), while agriculture increased from 266 km² to 399 km² and 467 km² (+ 201 km²; +15.96 percentage points; +76% relative), consistent with conversion of former forest and bare land to cropland in Fig. 2, although class-to-class transitions were not quantified. Bare-land extent declined by 11.20 percentage points (≈ 141 km²) (389, 355 and 248 km² in 2001, 2014 and 2024), consistent with the spread of cultivation and housing, whereas mapped water was 9 km² in 2001 and 12 km² in 2024 but 50 km² in 2014; the 2014 water area is four to five times that of the other epochs and is interpreted with caution (Section 5). Changes are summarized in Fig. 2 and Table 5. Because much of the area mapped as forest in 2001 lies in the irrigated northern plain and is mapped as agriculture in 2014 and 2024 (Fig. 2), part of this transition may reflect spectral confusion between dense irrigated vegetation or orchards and forest rather than deforestation.

Classification validation gave overall accuracies of 94% (KC = 0.92) for 2001, 97% (KC = 0.96) for 2014, and 93% (KC = 0.90) for 2024. Producer's accuracy ranged from 0.71 to 1.00 and user's accuracy from 0.83 to 1.00. These accuracies may be optimistic if validation pixels came from the same polygons as training pixels (Section 2.5).

Table 5
LULC area in 2001, 2014 and 2024 and change between 2001 and 2024 (district area ≈ 1,259 km²).

Class	2001 (km²)	2014 (km²)	2024 (km²)	Change (km²)	Change (percentage points of district)	Relative change (%)
Forest	308	118	87	−221	−17.55	−71.8
Built-up	287	337	445	+ 158	+ 12.55	+ 55.1
Agriculture	266	399	467	+ 201	+ 15.96	+ 75.8
Bare land	389	355	248	−141	−11.20	−36.3
Water	9	50	12	+ 3	+ 0.24	33.3

3.2 Land surface temperature

The minimum, mean and maximum LST were higher in each later composite. In 2001, the minimum, mean, and maximum LST were 22.89, 34.46, and 46.13°C, respectively. In 2014, the corresponding values were 24.88, 41.44, and 57.3°C, respectively. In 2024, the LST minimum, mean, and maximum increased to 25.9, 44.73, and 59.3°C; that is, the district-mean composite LST was 10.27°C higher in 2024 than in 2001 (6.98°C between 2001 and 2014 and 3.29°C between 2014 and 2024). The minimum and maximum LST values increased concurrently. These are statistics of single median composites, one per epoch, and the differences between them are not trends in the climatological sense. The minimum and maximum are single-pixel extremes and are sensitive to residual cloud and mixed pixels; no outlier trimming was applied.

Assessment of the 2024 maximum (59.3°C). A pixel-level maximum of 59.3°C is not implausible a priori for dry bare soil or dark roofs under clear-sky summer conditions, because LST is a radiometric skin temperature that can exceed air temperature by many degrees; it is a single-pixel maximum, not an area average. However, Fig. 4 shows maxima of ≈ 57–59°C in 2024 for every class, including water and forest, which is physically implausible for open water and dense canopy. This points to mixed pixels at class boundaries (the thermal bands have a native resolution of 100–120 m), misclassification (e.g., dry channel beds or sandbars labelled as water) or residual cloud-edge artefacts. Before the maximum is interpreted, its location should be traced and checked against the QA_PIXEL and ST_QA flags, the contributing acquisition dates and high-resolution imagery, and robust upper percentiles (e.g., the 99th) should be reported alongside it. These checks were not performed in the present analysis; the 59.3°C value is therefore reported descriptively, and pixel-level provenance verification and percentile reporting are identified as necessary steps for future work rather than completed components of this study.

The standardized LST anomaly (z_{LST}, Eq. (8)) ranged from − 5.27 to 2.47 in 2001, from − 3.53 to 2.51 in 2014 and from − 3.12 to 2.45 in 2024 (dimensionless; extremes as shown in the Fig. 3 legends). Its mean is zero by construction, and the reported means (-2.1×10^{-5}, -1.9×10^{-6} and -1.6×10^{-5}) are numerical residuals with no physical meaning (Table 7). Results are mapped in Fig. 3. Consistency check. If zLST had been computed from the LST layers summarized above, the standard deviation implied by each extreme, σ = (LST − mean)/z, would be the same at the minimum and the maximum. For 2024 the two estimates agree (σ ≈ 6.0°C), but for 2001 (4.7 vs 2.2°C) and 2014 (6.3 vs 4.7°C) they do not. The spatial patterns of the LST and zLST panels of Fig. 3 also differ in places, whereas a z-score is a linear transformation that preserves spatial pattern. The 2001 and 2014 zLST layers therefore do not appear to have been derived from the LST layers shown, and they are not interpreted further.

3.3 LST by land-cover class and decomposition of the inter-epoch difference

Thermal variation among individual land cover types is pronounced (Fig. 4). In 2001, bare land and built-up areas already exhibited elevated mean LSTs (38.44 and 35.54°C, respectively), while forest and water surfaces remained cool (30.15°C and 29.47°C). In 2014, the class means were 45.40°C (bare land), 43.23°C (built-up), 38.89°C (agriculture), 37.67°C (water) and 34.72°C (forest). By 2024, built-up zones reached 46.61°C and bare land areas 48.53°C; forest also increased, to 38.38°C. The contrast between built-up and forest classes widened from 5.4°C in 2001 to 8.5°C in 2014 and was 8.2°C in 2024. Agriculture remained intermediate, showing a 42.3°C mean LST in 2024 (+ 9.7°C since 2001). In all three epochs bare land was the warmest class, followed by built-up land and agriculture; water was the coolest class in 2001 and 2024, and forest in 2014.

All five classes show increases of similar magnitude between 2001 and 2024: +11.1°C (built-up), + 10.1°C (bare land), + 9.7°C (agriculture), + 8.2°C (forest) and + 6.1°C (water) (Fig. 4C). Because classes that are not associated with urban conversion warmed nearly as much as built-up land, most of the inter-epoch difference cannot be explained by land-cover conversion. Applying Eq. (13) to the class areas and means (Tables 5 and 6) gives a compositional component of + 0.14°C (evaluated at 2001 class means) to + 0.52°C (at 2024 class means), compared with a district-mean difference of 10.27°C; the within-class component therefore accounts for about 95–99% of the difference. The area-weighted class means reproduce the reported district means to within 0.01°C (2001, 2014) and 0.02°C (2024), which supports the internal consistency of these statistics. By contrast, the contrasts between classes changed: built-up minus agriculture increased from ≈ 2.9°C (2001) to ≈ 4.3°C (2014 and 2024) and built-up minus forest from 5.4°C to 8.5°C and 8.2°C, whereas bare land remained warmer than built-up land (by 2.9, 2.2 and 1.9°C). Most of the change in these contrasts occurred between 2001 and 2014, the interval that also spans the change from Landsat 5 TM to Landsat 8 OLI/TIRS, so part of it may be sensor-related. These derived values are computed from rounded class statistics; a separate analysis of pixels whose class did not change was not performed.

Table 6
Mean summer-composite LST by class and decomposition of the 2001–2024 difference (class means for 2001, 2014 and 2024).

Class	Mean LST 2001 (°C)	Mean LST 2014 (°C)	Mean LST 2024 (°C)	ΔT (°C)	Δ area (percentage points)
Forest	30.15	34.72	38.38	+ 8.2	−17.55
Built-up	35.54	43.23	46.61	+ 11.1	+ 12.55
Agriculture	32.64	38.89	42.3	+ 9.7	+ 15.96
Bare land	38.44	45.40	48.53	+ 10.1	−11.20
Water	29.47	37.67	35.56	+ 6.1	≈+0.24
District mean (reported)	34.46	41.44	44.73	+ 10.27	–
Compositional component (Eq. 13, first term)				+ 0.14 to + 0.52	
Within-class component (Eq. 13, second term)				+ 9.74 to + 10.11	

Class means are from the zonal statistics underlying Fig. 4. The range of the compositional component reflects the two orderings of the decomposition (compositional term evaluated at 2001 vs. 2024 class mean LST); both components were computed from the rounded values shown.

3.4 Surface urban heat island and standardized LST anomaly

Table 7
reports the zLST statistics. Because zLST is standardized within each epoch, its mean is zero and its range mainly reflects the shape of the LST distribution; the near-zero means and the narrowing of the minimum (− 5.27 to − 3.12) cannot be interpreted as a broader spatial distribution of elevated temperatures or as a city-wide heat canopy. SUHI in °C (Eq. (7)), derived from the class means in Table 6, is ≈ 2.9°C (2001), ≈ 4.3°C (2014) and ≈ 4.3°C (2024) relative to agriculture, 5.4°C, 8.5°C and 8.2°C relative to forest, and negative (− 2.9°C, − 2.2°C and − 1.9°C) relative to bare land. The widening contrast with vegetated references between 2001 and 2014 is consistent with increasing differentiation between built-up and vegetated surfaces, although it coincides with the sensor change, whereas the negative contrast with bare land illustrates the reference-dependence of SUHI in a semi-arid setting. The spatial extent of positive SUHI (e.g., the area of built-up pixels exceeding the reference by a fixed threshold) has not been computed in the present analysis and is identified here as a necessary extension for future work.

Table 7. Statistics of the standardized LST anomaly z_{LST} (Eq. (8); dimensionless). The 2001 and 2014 values are not consistent with the LST layers (Section 3.2).

Periods	Minimum	Mean	Maximum
2001	-5.27	-0.000021	2.47
2014	-3.53	-0.0000019	2.51
2024	-3.12	-0.000016	2.45

3.5 Machine-learning model performance

Under a random 80/20 pixel split, the three models performed almost identically within each epoch (Table 8): r^2 (Eq. (12)) was 0.83 in 2001, 0.85 in 2014 and 0.78–0.79 in 2024, and RMSE was 1.66–1.67°C, 1.75–1.76°C and 2.05–2.09°C, respectively. Differences between models (≤ 0.01 in r^2 and ≤ 0.04°C in RMSE) are too small to be resolved from a single split, so these results do not show that any one algorithm is more accurate. Because neighbouring pixels could occur in both training and test sets, the values are also likely to overstate performance in spatially independent areas. Spatially independent validation and the coefficient of determination (Eq. (11)) were not computed, so these random-split metrics are reported descriptively. LightGBM had the lowest RMSE in every epoch, by 0.01–0.03°C.

Table 8
Model performance under a random 80/20 pixel split (n = 22,843 points per epoch; single split; R^2 computed as squared correlation, i.e., r^2 of Eq. (12)).

Year	Model	RMSE (°C)	MAE (°C)	r^2
2001	LGBM	1.66	1.31	0.83
	RF	1.67	1.31	0.83
	XGB	1.67	1.32	0.83
2014	LGBM	1.75	1.39	0.85
	RF	1.76	1.39	0.85
	XGB	1.75	1.39	0.85
2024	LGBM	2.05	1.59	0.79
	RF	2.09	1.61	0.78
	XGB	2.07	1.60	0.79

The lower accuracy in 2024 is consistent with, but does not demonstrate, greater sub-pixel heterogeneity of the 2024 surface. Model RMSE relative to a composite LST does not measure LST retrieval accuracy. Training time was 0.48–0.60 s for LightGBM, 4.1–11.1 s for XGBoost and 54–79 s for RF across the three epochs (about two orders of magnitude between LightGBM and RF), but it depends on software implementation, number of trees and threading and is not used to rank the models.

3.6 Predictor importance

Predictor rankings agreed only partly across models (Fig. 5). In 2001, BSI ranked first for XGBoost and LightGBM, whereas NDMI ranked first for RF, with BSI second; DEM ranked second or third in all three models. In 2014, NDMI ranked first for LightGBM and RF, whereas NDBI ranked first for XGBoost. In 2024, NDBI ranked first for XGBoost and LightGBM (gain ≈ 0.57 and 0.29), with NDMI a close second for LightGBM, whereas DEM ranked first for RF. NDMI does not appear in the XGBoost rankings because it was excluded from that model (Section 2.9); in the XGBoost model, NDBI accounted for 33.7%, 67.2% and 57.1% of the total gain in 2001, 2014 and 2024, and elevation for 17.4%, 22.5% and 25.9%. Considered jointly, the NDBI–NDMI pair, which contrasts NIR and SWIR reflectance and therefore responds both to surface moisture and to dry built-up and bare materials, was among the most important predictors in every model after 2001, BSI was dominant only in 2001 (39.8% of XGBoost gain in 2001 but 1.4% and 1.3% in 2014 and 2024, when BSI was computed with the erroneous band; Section 2.4), and DEM ranked in the top three for every model and epoch.

The shift from BSI in 2001 to the NDBI–NDMI pair in 2014 and 2024 coincides with the change from Landsat 5 TM to Landsat 8 OLI with land-cover change and, for 2014 and 2024, with the BSI band error (Section 2.4), so it is not interpreted as a change in the physical controls of LST. Interpreting it would require corrected BSI layers, inter-sensor harmonisation of reflectance and held-out permutation importance.

4. Discussion

4.1 What the multi-epoch LST differences can and cannot show

District-mean summer-composite LST was ≈ 10°C higher in 2024 than in 2001, during a period in which built-up land expanded and mapped forest declined. However, the decomposition in Section 3.3 shows that changes in class composition account for less than 1°C of this difference and that the forest and water classes warmed by 6–8°C. The ≈ 10°C figure is therefore not evidence that land-cover transformation, rather than background conditions, caused the warming; consistent with H0, it is mainly an epoch-wide difference whose sources this design cannot separate. Plausible contributors include differences in summer meteorology and soil moisture between the three composite periods, differences in the dates and number of scenes in each median composite, Landsat 5 TM versus Landsat 8 TIRS differences in spectral response, native resolution and overpass time, and differences in the atmospheric inputs to Level-2 processing. Increases of 6–11°C in the means of all classes are also much larger than typical daytime urban–rural contrasts (1.5 ± 1.2°C across 419 cities (Peng et al., 2012) which places the decomposition result in context. Studies of rapidly urbanizing South Asian cities, including Karachi (Afzal et al., 2025), Delhi (Sarkar, 2021), and Barishal (Hasan et al., 2025), have associated LST increases with forest loss and impervious-surface expansion. Our results suggest that such comparisons should report class-specific changes and compositional decompositions before attributing an increase to LULC change.

4.2 Physical interpretation in a semi-arid, irrigated basin

The LULC-related signal that the data do support the widening contrast between built-up land and vegetated references is consistent with the lower evapotranspiration and greater heat storage of built-up materials. Replacing vegetation with concrete and asphalt increases sensible heat flux and limits soil moisture exchange (Oke, 1982). Greater daytime heat storage is also expected to prolong nocturnal heat release, which cannot be observed with morning Landsat overpasses but contributes to persistent heat stress (Santamouris, 2020). In Peshawar, irrigated agriculture rather than natural vegetation forms most of the vegetated background. Its cooling depends on irrigation and soil moisture at the time of acquisition, so the contrast between built-up and agricultural land can vary between years independently of land-cover change. Dry bare soil was the warmest class in every epoch and was warmer than built-up land, as commonly observed where the rural background is sparsely vegetated (Imhoff et al., 2010).

DEM ranked among the top three predictors (here the ASTER GED elevation band; Section 2.8) for every model and epoch. Elevation in the district co-varies with the gradient from the irrigated Kabul River plain in the north to the drier, predominantly bare southern areas (Figs. 2–3), so its importance may reflect a spatial proxy for soil type, irrigation access and land cover rather than lapse-rate or ventilation effects. The direction and shape of the elevation–LST association have not been shown; partial-dependence or accumulated-local-effects plots evaluated on held-out data are needed.

The apparent prominence of NDMI in 2014 and of NDBI in 2024 should not be interpreted as a switch from moisture control to built-up control: because NDBI and NDMI are redundant as defined, their relative ranks depend on how each algorithm allocates importance between them (Section 3.6). Jointly, the pair shows that the NIR–SWIR contrast, which responds both to vegetation and soil water content and to dry built-up and bare materials, is a strong predictor of LST. Separating these effects requires predictors that are not algebraically related, such as fractional impervious cover or a SWIR2-based moisture index.

4.3 SUHI definition and reference surface

The maximum of zLST (≈ 2.5) is a z-score, not a temperature, and its near-zero mean is a property of standardization (Section 3.4), so zLST cannot show whether a heat-island footprint expanded. Expressed in °C relative to vegetated references, the built-up contrast increased between 2001 and 2024, mostly between 2001 and 2014. Whether the area of built-up land exceeding the reference also increased requires a threshold-based footprint analysis (Qiao et al., 2019) applied to recalculated SUHI.

Plot-scale experiments indicate that reflective coatings and grass remain markedly cooler at midday than bare soil or synthetic turf (Ahmed et al., 2025)

4.4 What the model comparison does and does not reveal

Under random pixel validation, all three ensembles explained about 80% of the variance of composite LST from six predictors (RMSE ≈ 1.7–2.1°C), which is in the range reported by other machine-learning LST studies (Rengma & Yadav, 2024; Su et al., 2025). Their near-identical accuracy is itself informative: structurally different learners extracted essentially the same predictive information, which is consistent with accuracy being limited by the predictor set and by unmodelled factors such as soil moisture at acquisition time and sub-pixel mixing rather than by the choice of algorithm, although this was not tested. For

the urban thermal environment, this means that spectral indices and elevation capture the dominant within-epoch contrasts between moist vegetated, dry bare and built-up surfaces, but they leave 15–22% ($1 - r^2$) of the within-epoch LST variance unexplained and, by construction, say nothing about the between-epoch shift, because each model is fitted to one epoch. Whether these conclusions hold in spatially independent areas remains to be shown, as spatial block cross-validation was not evaluated in the present study and is identified as an essential requirement for future spatial generalization.

LightGBM had the lowest RMSE in every epoch, but its advantage (≤ 0.01 in r^2 and ≤ 0.04°C in RMSE) was obtained from a single random split with fixed, untuned hyperparameters, and its much shorter training time reflects implementation choices (e.g., histogram-based split finding and multithreading in the lightgbm package versus 200 trees in randomForest) as well as algorithm design (Ke et al., 2017b). Predictive accuracy, computational cost, interpretability and operational suitability are therefore treated as separate criteria: on the present evidence, the three models are equivalent in accuracy, LightGBM was the fastest in this implementation, and none can be recommended as operationally superior.

4.5 Predictor importance: association, not mechanism

Cross-model agreement in predictor ranks was partial (Section 3.6). Where it occurs, it indicates rank-order agreement in how the fitted models use correlated predictors, not that the predictors represent distinct physical processes. RF IncNodePurity and boosting gain are computed on training data, are on different scales and favour continuous predictors with many candidate split points (Strobl et al., 2007). Held-out permutation importance, with redundant predictors grouped, would provide a common basis for comparison. The 2001-to-2024 shift in top-ranked predictors is consistent with the expansion of built-up land, but because it coincides with the change of sensor and is affected by predictor redundancy and the BSI band error, it is not interpreted here as evidence of a transition from natural to anthropogenic controls of LST.

The GEE–R workflow is a practical strength for data-limited cities: preprocessing uses openly available, consistently processed Level-2 products (Gorelick et al., 2017b), and model fitting uses open-source R packages. Its reproducibility depends on releasing the scripts, scene lists, sample locations and package versions, and its transferability to other cities would need to be demonstrated with local accuracy assessment.

4.6 Implications for urban planning

The persistent LST contrast between built-up land and vegetated or irrigated land indicates where surface cooling is currently provided. Remaining forest occupied ≈ 7% of the district in 2024 (87 km²), and vegetation along canals may provide local cooling, but these cool areas have not been mapped or quantified here. Evidence that urban trees reduce LST in European cities (Schwaab et al., 2021) suggests that extending tree cover could lower surface temperatures, but the magnitude in Peshawar's semi-arid setting, where water availability constrains greening, remains to be quantified. Once validated, the workflow could support municipal screening of thermal hotspots.

5. Limitations and future work

(1) *One composite per epoch.* Each epoch is represented by one June–August median composite. Interannual variability in summer meteorology, soil moisture and irrigation can change LST by several degrees in all classes and cannot be separated from LULC effects with this design. This is the most important limitation for the reported ≈ 10°C difference (Section 4.1). Using all clear-sky acquisitions from several summers around each epoch (e.g., 2000–2002, 2013–2015 and 2023–2025) and comparing them with ERA5-Land or station temperature anomalies would allow epoch-wide variability to be estimated.

(2) *Sensor and retrieval differences.* The 2001 epoch uses Landsat 5 TM (band 6, native 120 m) and the later epochs Landsat 8 TIRS (band 10, native 100 m). Differences in spectral response, calibration, overpass time and the atmospheric inputs to Level-2 processing could add a systematic offset to the 2001 versus 2014/2024 comparison; its magnitude for this scene has not been quantified. Reflectance-based indices are affected by TM–OLI band differences in the same way.

(3) *Outlier treatment and quality screening. No statistical trimming or ST_QA screening was applied, so single-pixel extremes (e.g., the 59.3°C maximum) may reflect mixed or cloud-edge pixels, and cloud shadows were not masked in the reflectance composites.*

(4) *Causal attribution.* The models are predictive. Predictive accuracy and importance rankings do not establish that LULC change caused the observed warming; attribution requires reference-area comparisons, climate covariates and designs such as comparing converted with stable pixels.

(5) *Scope and transferability.* The analysis covers one district and three summers; the transferability of the models and conclusions to other cities and seasons has not been tested.

Future work. Higher-resolution spectral inputs (Sentinel-2 MSI) and thermal observations at other times of day (ECOSTRESS) would improve the characterization of narrow green and water features and of diurnal behaviour. Coupling LST with population density and energy use would support assessment of heat exposure and thermal equity, and monthly or multi-year seasonal composites would clarify the temporal dynamics of SUHI.

6. Conclusions

This study combined accuracy-assessed LULC maps, summer Landsat LST composites and three tree-ensemble regression models to examine how land-cover change between 2001 and 2024 relates to surface temperature in semi-arid Peshawar. Three conclusions follow from the evidence.

First, land-cover composition changed substantially mainly through loss of mapped forest and expansion of built-up land and cropland but these changes account for only a small part of the difference in district-mean summer LST between epochs. Most of that difference is shared by all classes, and with one

composite per epoch it cannot be attributed to land-cover change, climatic variability or sensor differences. Multi-epoch LST studies in comparable settings should therefore report class-specific changes and compositional decompositions rather than attributing absolute LST differences to urbanization.

Second, the LULC-related signal that is insensitive to uniform epoch-wide shifts the contrast between built-up land and vegetated references widened, mainly between 2001 and 2014, whereas dry bare land remained warmer than built-up land. In semi-arid, irrigated basins, SUHI must therefore be defined against an explicit reference surface and expressed in °C.

Third, structurally different tree ensembles reached nearly identical accuracy under random validation and agreed only partly on predictor rankings, partly because two predictors were redundant. In this dataset model choice made little difference to accuracy, and no algorithm can be identified as superior on the present evidence.

The main contribution is a transparent framework for distinguishing compositional from epoch-wide LST change in a data-scarce city, together with an explicit account of the validation and reference data needed for attribution. Its principal limitations one composite per epoch, a sensor changes between epochs, processing errors in the BSI and elevation inputs, and the absence of spatially independent validation define the next steps: re-analysis with the corrected scripts, multi-year composites, climatological reference data and spatially blocked validation.

Declarations

Conflict of interest statement

The authors reported no potential conflict of interest.

Author contributions

Authors Contributions: Conceptualization, S.R.; methodology, A.S. and X.L.; software, A.S, S.R and .A.R,; validation, Z.A., J.S., and X.L.; formal analysis, S.R. and W.A.; investigation, Z.A..; resources, J.S.; data curation, A.S., S.R and A.R.; writing—original draft preparation, S.R.; writing—review and editing, J.S. and X.L.; visualization, A.R, supervision, J.S.; project administration, J.S. All authors have read and agreed to the published version of the manuscript.

Declarations of generative AI and AI-assisted technologies in the writing process:

During the preparation of this work the author(s) used DeepSeek-V3 in order to polish the paper’s language. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Funding:

This research received no external funding.

Author Contribution

Authors Contributions: Conceptualization, S.R.; methodology, A.S. and X.L.; software, A.S, S.R and .A.R,; validation, Z.A., J.S., and X.L.; formal analysis, S.R. and W.A.; investigation, Z.A..; resources, J.S.; data curation, A.S., S.R and A.R.; writing—original draft preparation, S.R.; writing—review and editing, J.S. and X.L.; visualization, A.R, supervision, J.S.; project administration, J.S. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

We want to express our sincere gratitude to Fujian Normal University's Faculty of the School of Geographical Science for providing the necessities to conduct this study. A special thanks to our research team for their efforts, which included gathering data and creating a proper timeline for the research.

Data Availability

Landsat Collection 2 Level-2 data are openly available from the USGS and through the Google Earth Engine data catalog (https://developers.google.com/earth-engine/datasets). The Earth Engine and R scripts used in this study are available from the corresponding author on reasonable request.

References

1. Afzal MT, Tripathi NK, Ninsawat S, Pal I (2025) Urban heatwave vulnerability assessment and ecological evaluation of land use/land cover in Karachi, Pakistan: A developing megacity perspective. Environ Sustain Indic 28:100965. https://doi.org/10.1016/j.indic.2025.100965

2. Ahmed SZ, Rosa D, La, Jayakumar S, Ahmed SZ, Rosa D, La, Jayakumar S (2025) Seasonal and Diurnal Dynamics of Urban Surfaces: Toward Nature-Supportive Strategies for SUHI Mitigation. Land 2025 14(12):2412. https://doi.org/10.3390/LAND14122412. *14*
3. Ali A (2025) Machine learning approaches for soil moisture prediction: enhancing agricultural water management with integrated data. Model Earth Syst Environ 2025 12:1(1):16. https://doi.org/10.1007/S40808-025-02607-5. *12*
4. Ashok A, Rani HP, Jayakumar KV (2021) Monitoring of dynamic wetland changes using NDVI and NDWI based landsat imagery. Remote Sens Applications: Soc Environ 23:100547. https://doi.org/10.1016/j.rsase.2021.100547
5. Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324
6. Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-Augu*, 785–794. https://doi.org/10.1145/2939672.2939785
7. Delgado-Moreno D, Gao Y (2022) Forest Degradation Estimation Through Trend Analysis of Annual Time Series NDVI, NDMI and NDFI (2010–2020) Using Landsat Images. Lecture Notes Geoinf Cartography 149–159. https://doi.org/10.1007/978-3-030-98096-2_11
8. Development Bank A (2023) *Regional: Establishing a Platform for Climate-Resilient and Low-Carbon Urban Development Peshawar District, Pakistan: Climate Change Assessment Prepared by Olga Chepelianskaia, Climate Finance Specialist (Urban Climate Investment Plans) For ADB and the G…*
9. Farr TG, Rosen PA, Caro E, Crippen R, Duren R, Hensley S, Kobrick M, Paller M, Rodriguez E, Roth L, Seal D, Shaffer S, Shimada J, Umland J, Werner M, Oskin M, Burbank D, Alsdorf DE (2007) The shuttle radar topography mission. Rev Geophys 45(2). https://doi.org/10.1029/2005RG000183
10. Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R (2017a) Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens Environ 202:18–27. https://doi.org/10.1016/j.rse.2017.06.031
11. Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R (2017b) Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens Environ 202:18–27. https://doi.org/10.1016/j.rse.2017.06.031
12. Hasan I, Goni O, Katha ZT, Rabby MI, Hossain S, Banik A, Hasan S, Rahman I (2025) Prediction modeling of land surface temperature in relation to land cover dynamics and health risk perception analysis in barishal city of Bangladesh. Sci Rep 15(1):30730. https://doi.org/10.1038/s41598-025-14868-2
13. Huang S, Tang L, Hupy JP, Wang Y, Shao G (2021) Correction to: Commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing (Journal of Forestry Research, (2021), 32, 1, (1–6), 10.1007/s11676-020-01155-1). In *Journal of Forestry Research*(Vol. 32, Issue 6, p. 2719). Northeast Forestry University. https://doi.org/10.1007/s11676-020-01176-w
14. Hulley GC, Hook SJ, Abbott E, Malakar N, Islam T, Abrams M (2015) The ASTER Global Emissivity Dataset (ASTER GED): Mapping Earth's emissivity at 100 meter spatial scale. Geophys Res Lett 42(19):7966–7976. https://doi.org/10.1002/2015GL065564
15. Hussain MF, Meng X, Shah SF, Hussain MA (2024) Integrating Spatiotemporal Analysis of Land Transformation and Urban Growth in Peshawar Valley and Its Implications on Temperature in Response to Climate Change. ISPRS Int J Geo-Information 13(7):239. https://doi.org/10.3390/ijgi13070239
16. Imhoff ML, Zhang P, Wolfe RE, Bounoua L (2010) Remote sensing of the urban heat island effect across biomes in the continental USA. Remote Sens Environ 114(3):504–513. https://doi.org/10.1016/j.rse.2009.10.008
17. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY (2017a) LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, *2017-Decem*, 3147–3155. https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html
18. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY (2017b) LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, *2017-Decem*, 3147–3155. https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html
19. Kv¥lseth TO (1985) Cautionary note about r2. Am Stat 39(4):279–285. https://doi.org/10.1080/00031305.1985.10479448
20. Lamichhane M, Mehan S, Mankin KR (2025) Soil Moisture Prediction Using Remote Sensing and Machine Learning Algorithms: A Review on Progress, Challenges, and Opportunities. In *Remote Sensing*(Vol. 17, Issue 14, p. 2397). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/rs17142397
21. Li X, Zhou Y, Asrar GR, Imhoff M, Li X (2017) The surface urban heat island response to urban expansion: A panel analysis for the conterminous United States. Sci Total Environ 605–606:426–435. https://doi.org/10.1016/j.scitotenv.2017.06.229
22. Malakar NK, Hulley GC, Hook SJ, Laraby K, Cook M, Schott JR (2018) An Operational Land Surface Temperature Product for Landsat Thermal Data: Methodology and Validation. IEEE Trans Geosci Remote Sens 56(10):5717–5735. https://doi.org/10.1109/TGRS.2018.2824828
23. Natekin A, Knoll A (2013) Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, *7*(DEC), 63623. https://doi.org/10.3389/fnbot.2013.00021
24. Nguyen CT, Chidthaisong A, Diem PK, Huo LZ (2021) A modified bare soil index to identify bare land features during agricultural fallow-period in southeast asia using landsat 8. Land 10(3):1–18. https://doi.org/10.3390/land10030231
25. Oke TR (1982) The energetic basis of the urban heat island. Q J R Meteorol Soc 108(455):1–24. https://doi.org/10.1002/qj.49710845502
26. Olofsson P, Foody GM, Herold M, Stehman SV, Woodcock CE, Wulder MA (2014) Good practices for estimating area and assessing accuracy of land change. In *Remote Sensing of Environment*(Vol. 148, pp. 42–57). Elsevier Inc. https://doi.org/10.1016/j.rse.2014.02.015
27. Peng S, Piao S, Ciais P, Friedlingstein P, Ottle C, Bréon FM, Nan H, Zhou L, Myneni RB (2012) Surface urban heat island across 419 global big cities. Environ Sci Technol 46(2):696–703. https://doi.org/10.1021/es2030438
28. Ploton P, Mortier F, Réjou-Méchain M, Barbier N, Picard N, Rossi V, Dormann C, Cornu G, Viennois G, Bayol N, Lyapustin A, Gourlet-Fleury S, Pélissier R (2020) Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nat Commun 11(1). https://doi.org/10.1038/s41467-020-18321-y

29. Qiao Z, Wu C, Zhao D, Xu X, Yang J, Feng L, Sun Z, Liu L (2019) Determining the boundary and probability of surface urban heat island footprint based on a logistic model. Remote Sens 11(11). https://doi.org/10.3390/rs11111368
30. Ramzan M, Saqib ZA, Hussain E, Khan JA, Nazir A, Dasti MYS, Ali S, Niazi NK (2022) Remote Sensing-Based Prediction of Temporal Changes in Land Surface Temperature and Land Use-Land Cover (LULC) in Urban Environments. Land 11(9):1610. https://doi.org/10.3390/land11091610
31. Rehmani MIA, Tajbar S, Begum B, Rafiq L, Dawood M, Darand M, Nasim IW (2018) Trends over Peshawar City. J Environ Agricultural Sci 14:23–31. https://www.researchgate.net/profile/Mia-Rehmani/publication/325131140_TRMM-Precipitation_Data_for_Estimating_Seasonal_and_Annual_Trends_over_Peshawar_City/links/5af9f8d50f7e9b3b0beff786/TRMM-Precipitation-Data-for-Estimating-Seasonal-and-Annual-Trends-ov TRMM-Precipitation Data for Estimating Seasonal and Annual
32. Rengma NS, Yadav M (2024) Machine learning-based estimation of land surface temperature variability over a large region: a temporally consistent approach using single-day satellite imagery. Environ Monit Assess 196(8):738. https://doi.org/10.1007/s10661-024-12856-z
33. Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillera-Arroita G, Hauenstein S, Lahoz-Monfort JJ, Schröder B, Thuiller W, Warton DI, Wintle BA, Hartig F, Dormann CF (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography, vol 40. Blackwell Publishing Ltd, pp 913–929. 8 https://doi.org/10.1111/ecog.02881
34. Santamouris M (2020) Recent progress on urban overheating and heat island research. Integrated assessment of the energy, environmental, vulnerability and health impact. Synergies with the global climate change. In *Energy and Buildings*(Vol. 207, p. 109482). Elsevier. https://doi.org/10.1016/j.enbuild.2019.109482
35. Schwaab J, Meier R, Mussetti G, Seneviratne S, Bürgi C, Davin EL (2021) The role of urban trees in reducing land surface temperatures in European cities. Nat Commun 12(1). https://doi.org/10.1038/s41467-021-26768-w
36. Strobl C, Boulesteix AL, Zeileis A, Hothorn T (2007) Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*. https://doi.org/10.1186/1471-2105-8-25
37. Su Q, Meng X, Sun L, Guo Z (2025) Quantification of MODIS Land Surface Temperature Downscaled by Machine Learning Algorithms. Remote Sens 17(14). https://doi.org/10.3390/rs17142350
38. Tabassum I, Sultan M, Ihsanullah, Khan J, Syed (2025) NaeemUdDin. *Land Use/Land Cover Change Dynamics in Rapidly Growing City: A Potential Threat to Sustainability: A Case Study of the Peshawar District of Pakistan*(pp. 391–407). Springer, Singapore. https://doi.org/10.1007/978-981-97-7455-5_22
39. U.S. Geological Survey (2021) Landsat Collection 2 Level-2 Science Products (Fact Sheet 2021–3055). *United Nations Geological Survey*, *November*, 1–2. https://www.usgs.gov/landsat-missions/landsat-collection-2-level-2-science-products
40. Valavi R, Elith J, Lahoz-Monfort JJ, Guillera-Arroita G (2019) blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. Methods Ecol Evol 10(2):225–232. https://doi.org/10.1111/2041-210X.13107;REQUESTEDJOURNAL:JOURNAL:2041210X;WEBSITE:WEBSITE:BESJOURNALS;WGROUP:STRING:PUBLICATION
41. Weng Q, Fu P, Gao F (2014) Generating daily land surface temperature at Landsat resolution by fusing Landsat and MODIS data. Remote Sens Environ 145:55–67. https://doi.org/10.1016/j.rse.2014.02.003
42. Zhou D, Xiao J, Bonafoni S, Berger C, Deilami K, Zhou Y, Frolking S, Yao R, Qiao Z, Sobrino JA (2019) Satellite remote sensing of surface urban heat islands: Progress, challenges, and perspectives. In *Remote Sensing*(Vol. 11, Issue 1). https://doi.org/10.3390/rs11010048

Figures

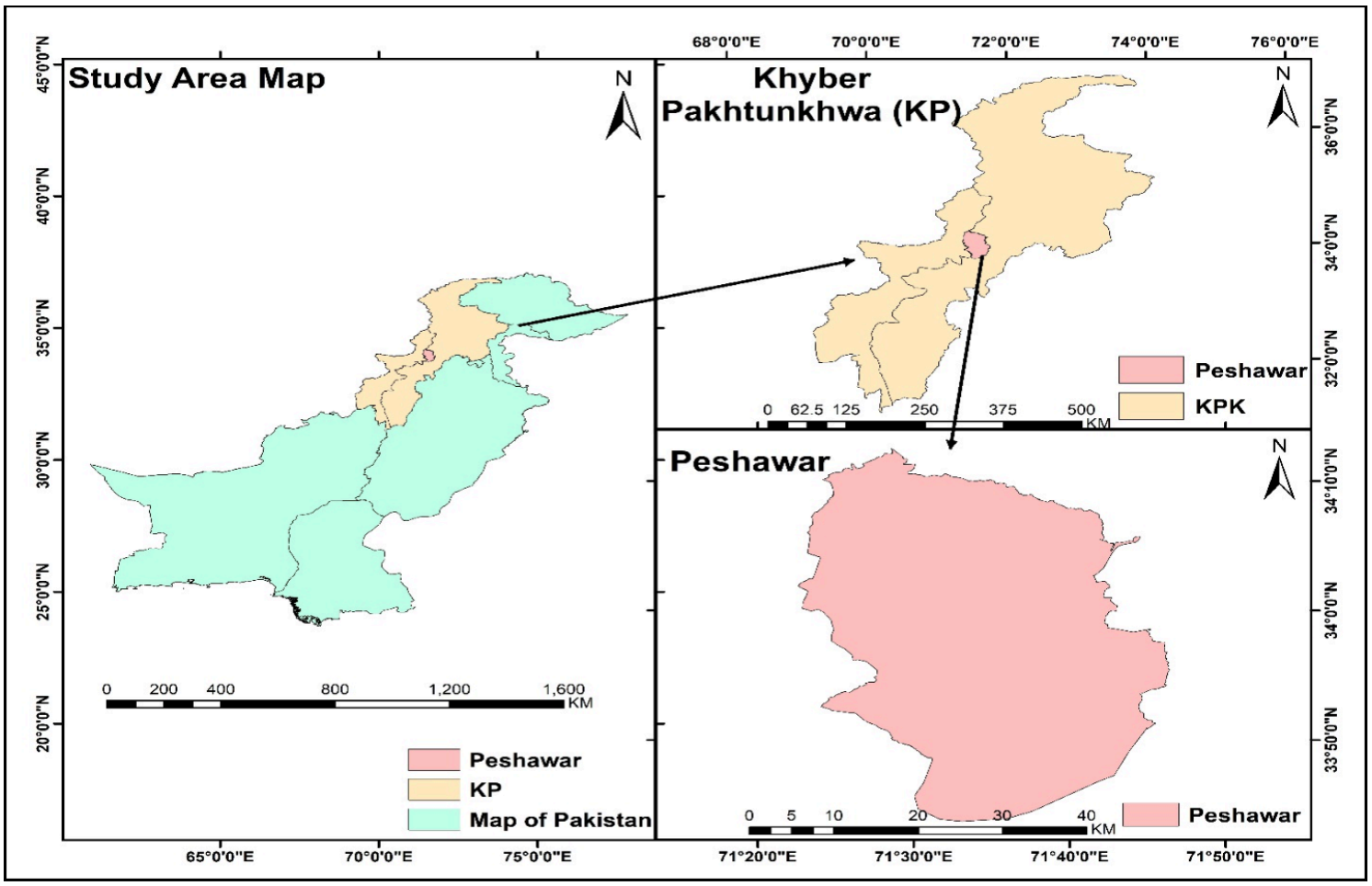


Figure 1

Location of Peshawar District in Pakistan and Khyber Pakhtunkhwa.

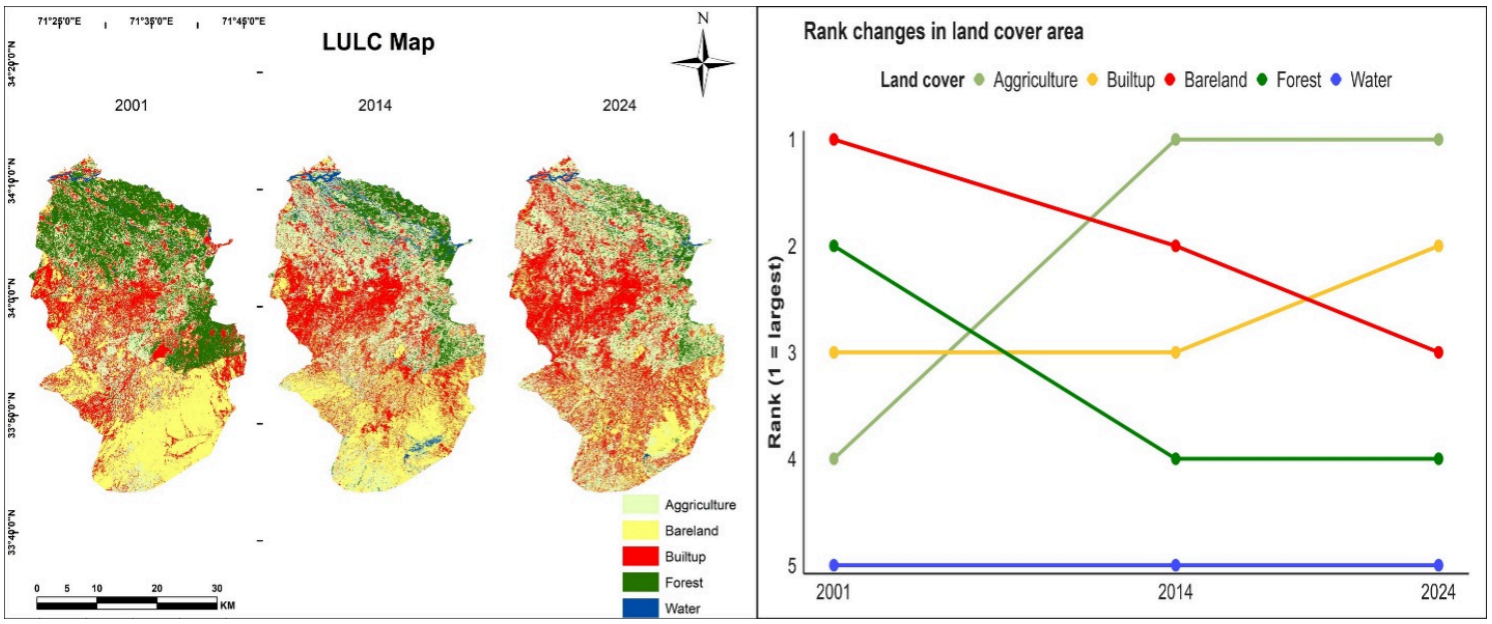


Figure 2

LULC maps of Peshawar District for 2001, 2014 and 2024 (left) and rank of class areas (right).

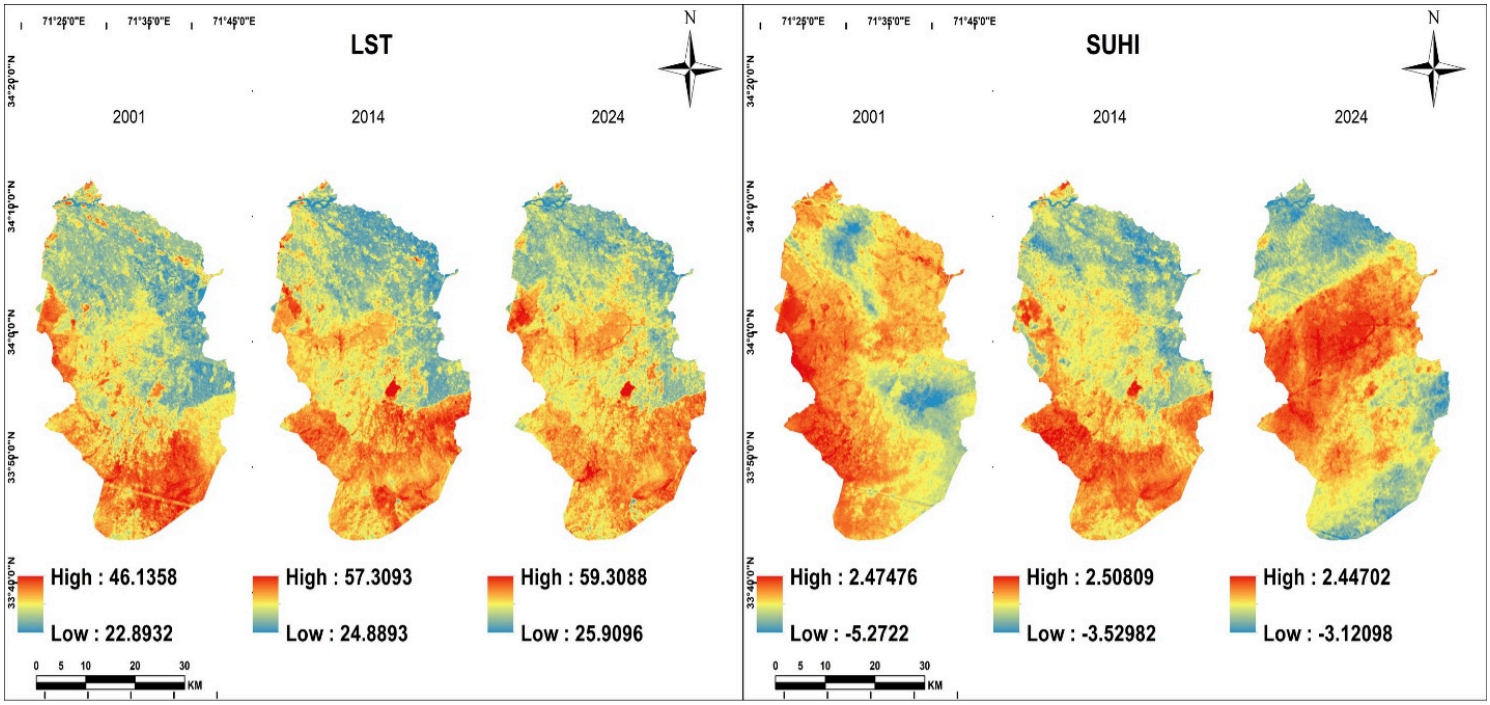


Figure 3

Summer-composite LST (°C; left) and standardized LST anomaly z_{LST} (dimensionless, Eq. (8); right; labelled "SUHI" in the panel title) for 2001, 2014 and 2024.

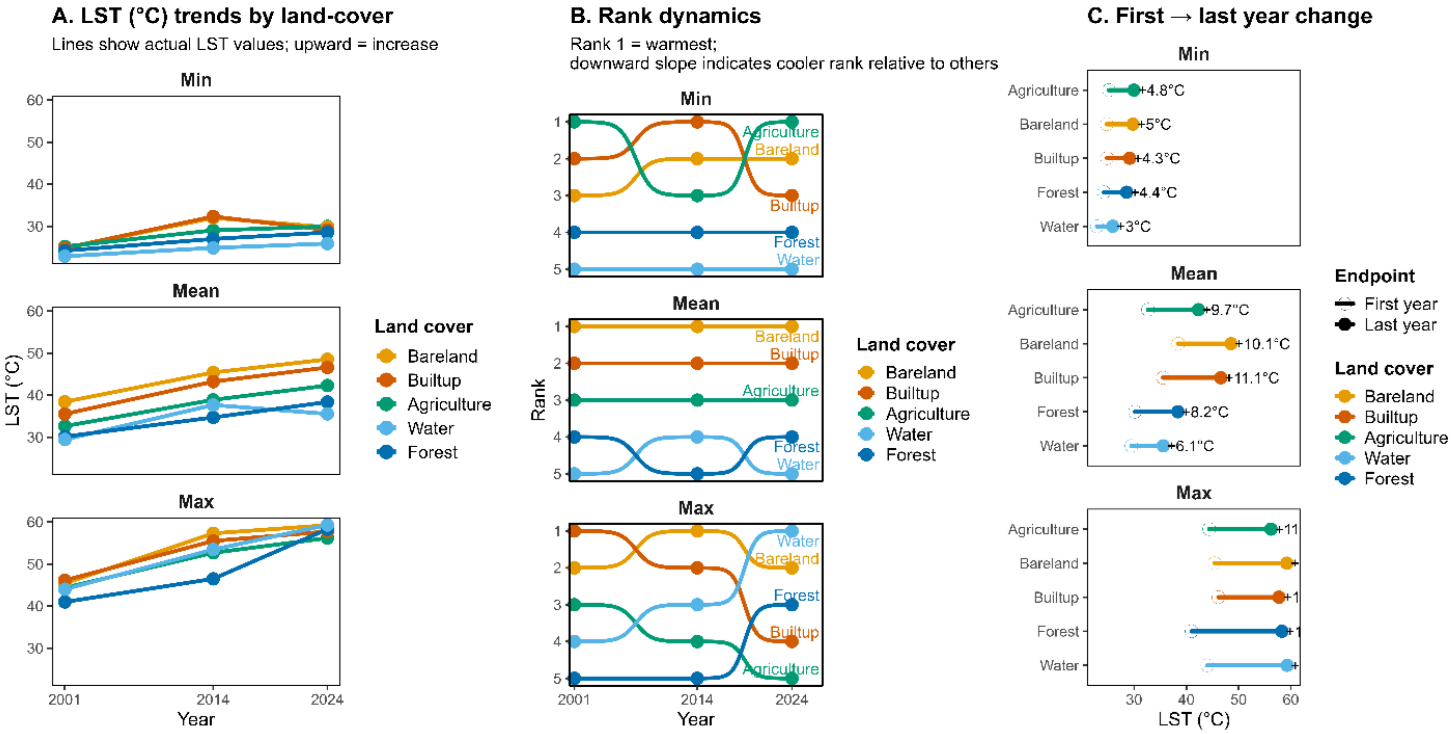


Figure 4

Minimum, mean and maximum summer-composite LST by land-cover class: (A) values by epoch, (B) rank of classes, (C) change between 2001 and 2024.

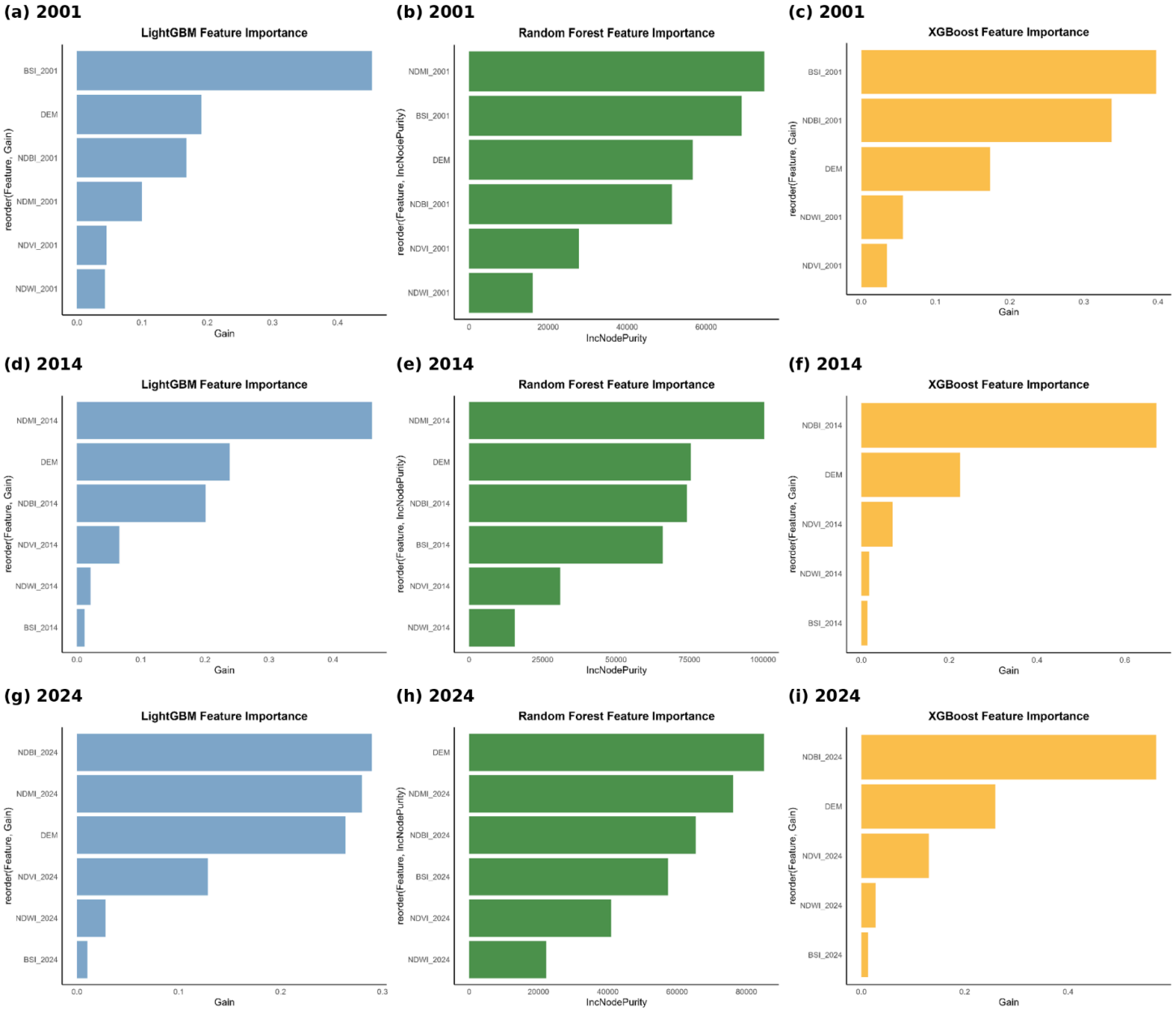


Figure 5

Predictor importance in 2001 (a–c), 2014 (d–f) and 2024 (g–i) for LightGBM (gain), Random Forest (IncNodePurity) and XGBoost (gain), computed on the training data; XGBoost excluded NDMI, and 2014/2024 BSI values are affected by the band error described in Section 2.4. Importance scales differ between models and are comparable only by rank.