

Explainable Machine Learning for Spatiotemporal Prediction of Urban Heat Dynamics: Insights from Chicago

S M Redwan Kabir

`smredwan.kabir@wmich.edu`

Western Michigan University

Laiyin Zhu

Western Michigan University

Farhana Kabir Zisha

Western Michigan University

Article

Keywords:

Posted Date: October 1st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10722684/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Explainable Machine Learning for Spatiotemporal Prediction of Urban Heat Dynamics: Insights from Chicago

S M Redwan Kabir, Laiyin Zhu & Farhana Kabir Zisha

Abstract

Urbanization is intensifying Land Surface Temperature (LST) dynamics and Urban Heat Island (UHI) effects in rapidly expanding cities. Yet, the intricate interactions among land cover, carbon storage (CS), and socioeconomic factors remain inadequately understood at detailed spatial and temporal scales. Here we synthesize remote sensing, GIS, and census tract data with ensemble machine learning (ML) techniques to model the spatiotemporal variability of LST in Chicago from 2000 to 2022. Using Landsat and MODIS datasets, we apply a quantile mapping approach to reduce inter-sensor bias and align temperature distributions. Land-cover and CS assessments are derived from the National Land Cover Database (NLCD) and the Integrated Valuation of Ecosystem Services and Trade-Offs (InVEST) model. Long-term temperature records reveal a statistically significant warming trend over 52 years ($P = 0.03$), with the frequency of extreme temperature days increasing from 5 in 2000 to 28 in 2022. Adjusted LST maps show a 5–6 °C rise in average surface temperature over two decades, with the most pronounced warming concentrated in industrial and high-density developed areas. Extreme Gradient Boosting (XGBoost) outperformed Random Forest (RF), achieving higher predictive performance ($R^2 = 0.55$ vs. 0.49) and marginally better classification of temperature extremes (Area under the Curve, AUC = 0.85 vs. 0.84). SHAP analysis identifies vegetation cover and carbon stock density as the strongest cooling factors, whereas built-up intensity and population density are associated with higher LST. Partial dependence analysis reveals nonlinear predictor–response relationships, including a diminishing marginal cooling effect of vegetation at high NDVI values. Total carbon storage across Chicago’s census tracts declined by 0.38% over the study period, with the spatial concordance between carbon depletion and thermal amplification providing evidence that carbon conservation serves dual climate mitigation and heat reduction objectives. These results demonstrate the utility of interpretable machine learning for understanding urban thermal dynamics and provide a scalable, transferable framework for climate-informed urban planning and heat mitigation.

Introduction

Urbanization and climate change jointly escalate environmental and thermal risks in cities worldwide. Urban areas, which cover approximately 3% of Earth’s land surface, account for 60–80% of global energy consumption and nearly 75% of carbon emissions¹. Rapid urban growth alters land use, reduces vegetation cover, and increases anthropogenic heat emissions, thereby raising land surface temperature (LST) and intensifying the Urban Heat Island (UHI) effect. Changes in temperature, precipitation, sea levels, and hazard frequency threaten urban populations and ecosystems alike. Vegetation helps mitigate these impacts by regulating the water cycle, maintaining soil moisture, cooling surfaces through evapotranspiration, and sequestering carbon². However,

urban expansion and land conversion accelerate carbon emissions while simultaneously depleting urban carbon sinks³. Understanding the interplay between these processes is essential for developing effective climate adaptation strategies.

LST is a key indicator of urban climate change and environmental degradation, reflecting the thermal response of land surfaces to urbanization and land use and land cover (LULC) change. In densely urbanized areas, buildings and impervious surfaces absorb and retain heat, reduce natural ventilation, and alter local energy and radiation balances, thereby intensifying the UHI effect⁴. A substantial body of literature has documented strong links between LST and urban morphology, vegetation loss, population density, and built-up intensity^{5–7}. The loss of vegetation and forests further diminishes carbon sequestration capacity, increasing atmospheric CO₂ concentrations and reinforcing warming trends⁸. Thus, a comprehensive understanding of LST variability—and the factors that drive it—is vital for urban planning and climate adaptation, particularly in cities experiencing rapid transformation.

Chicago represents a compelling case for studying urbanization, LULC change, carbon dynamics, and thermal variability. Since the post-industrial era, the city and its surrounding metropolitan area have undergone substantial expansion and landscape transformation. From 1955 to 1992 alone, approximately 25% of agricultural land in the broader region was converted to urban use⁹. Prior studies have documented significant urbanization-driven changes in LST patterns across Chicago^{10–13}. While green spaces contribute to carbon sequestration and local cooling, industrialization and dense development continue to amplify thermal stress and emissions. The city's proximity to Lake Michigan adds a further dimension of complexity, as the lake–land temperature gradient influences local circulation patterns and moderates surface temperatures in adjacent areas.

Machine learning approaches have emerged as powerful tools for modeling complex environmental processes and predicting LST variability. Traditional statistical models frequently fail to capture the nonlinear relationships and spatial dependencies inherent in georeferenced environmental data^{14,15}. Machine Learning (ML) methods—including Random Forest (RF), Extreme Gradient Boosting (XGBoost), support vector regression (SVR), artificial neural networks (ANN), and K-nearest neighbor (K-NN)—have demonstrated improved predictive performance by capturing nonlinear interactions among climatic, environmental, and socioeconomic variables^{16–22}. These methods are recognized as essential for enhancing urban climate prediction and environmental management at operational scales. Explainability frameworks, particularly SHAP (SHapley Additive exPlanations) analysis, further enable researchers to decompose model predictions and identify the specific contributions of individual predictors, thereby bridging the gap between predictive accuracy and physical understanding of key mechanisms²³.

Despite these advances, few studies have comprehensively integrated remote sensing, carbon assessment, socioeconomic indicators, and explainable machine learning within a unified analytical framework. Many existing investigations overlook seasonal variability, nonlinear interactions, or the robust validation of LST estimates across multiple sensors. Limited research has examined how vegetation, carbon storage, urban density, and population simultaneously influence long-term urban thermal dynamics at the census-tract scale. Furthermore, the integration of long-term temperature

trend analysis with machine learning–based prediction remains underexplored. Explainable ML techniques can reveal the functional form of predictor–response relationships. Despite their potential to guide targeted intervention strategies within urban thermal studies, these techniques have not been extensively investigated in this domain.

This study addresses these gaps by combining remote sensing, GIS, census tract data, and ensemble machine learning to analyze LST variability in Chicago from 2000 to 2022. The study is designed to: 1) characterize long-term temperature trends and spatiotemporal LST patterns across the metropolitan area; 2) estimate carbon stock changes attributable to urban growth and assess their thermal implications; 3) develop and compare Random Forest and XGBoost models to predict LST variability using environmental and socioeconomic predictors; and 4) elucidate the nonlinear relationships among vegetation cover, carbon stock, urban density, population, and LST through SHAP and partial dependence analysis. By bridging trend analysis and predictive modeling, this study provides transferable insights into urban heat dynamics and establishes a geospatial framework applicable to climate-resilient planning in metropolitan regions globally.

Results

Long-term temperature characteristics and LST trend

Before developing predictive models, we characterized long-term temperature trends and LST patterns in Chicago to establish the climatological context for the machine-learning analysis. By analyzing daily maximum temperature and 5-year moving average temperature (Fig. 1a) we see a recurring pattern of peaks and troughs, with high-temperature plateaus occurring in the early 1990s, early 2000s, and early 2020s (approximately 15.75–15.80 °C), and cooler interludes dropping to around 14.3 °C. This cyclical behavior suggests the influence of multi-decadal climate oscillations superimposed on the underlying warming trajectory. Since 2020, however, the 5-year moving average has risen sharply from 14.9 °C to 15.6 °C, a rate of increase that exceeds the natural variability observed in earlier decades. As a result, a significant long-term warming trend was observed from the NOAA station at Chicago O'Hare, with annual mean temperature increasing by 0.165°C decade⁻¹ ($p < 0.005$; linear regression), equivalent to approximately 0.86°C over 52 years period of time 1970–2022^{24,25} (Fig 1a). Monthly temperature distributions show distinct seasonal cycles, with summer peaks (June–August) averaging approximately 27.90 °C and winter troughs (December–February) near 0.87 °C. Winter months display wider temperature ranges, indicating greater inter-annual variability, while summer ranges are comparatively narrow, suggesting that summer warming is more persistent and less buffered by natural climatic variability (Fig. 1b).

A critical finding from the trend analysis is the marked escalation in the frequency of extreme temperature days over the study period. Using a 5-day moving average threshold of 30 °C to identify extreme heat events (Fig. 1a), we observed that the number of extreme temperature days rose from merely 5 in 2000 to 28 in 2022, with intermediate values of 27 (2007), 18 (2013), and 23 (2016) in Fig. 2. Importantly, this increase was not strictly monotonic: extreme heat days dropped to 18 in 2013 before rising again, likely reflecting inter-annual climate variability associated with phenomena such as the Pacific Decadal Oscillation and regional precipitation anomalies. The spatial analysis of LST across five study periods (2000, 2007, 2013, 2016, and 2022) suggests that this warming is not

spatially uniform: the adjusted LST temperature range expanded from 26–35 °C in 2000 to 25–43 °C in 2022 (Fig. 2).

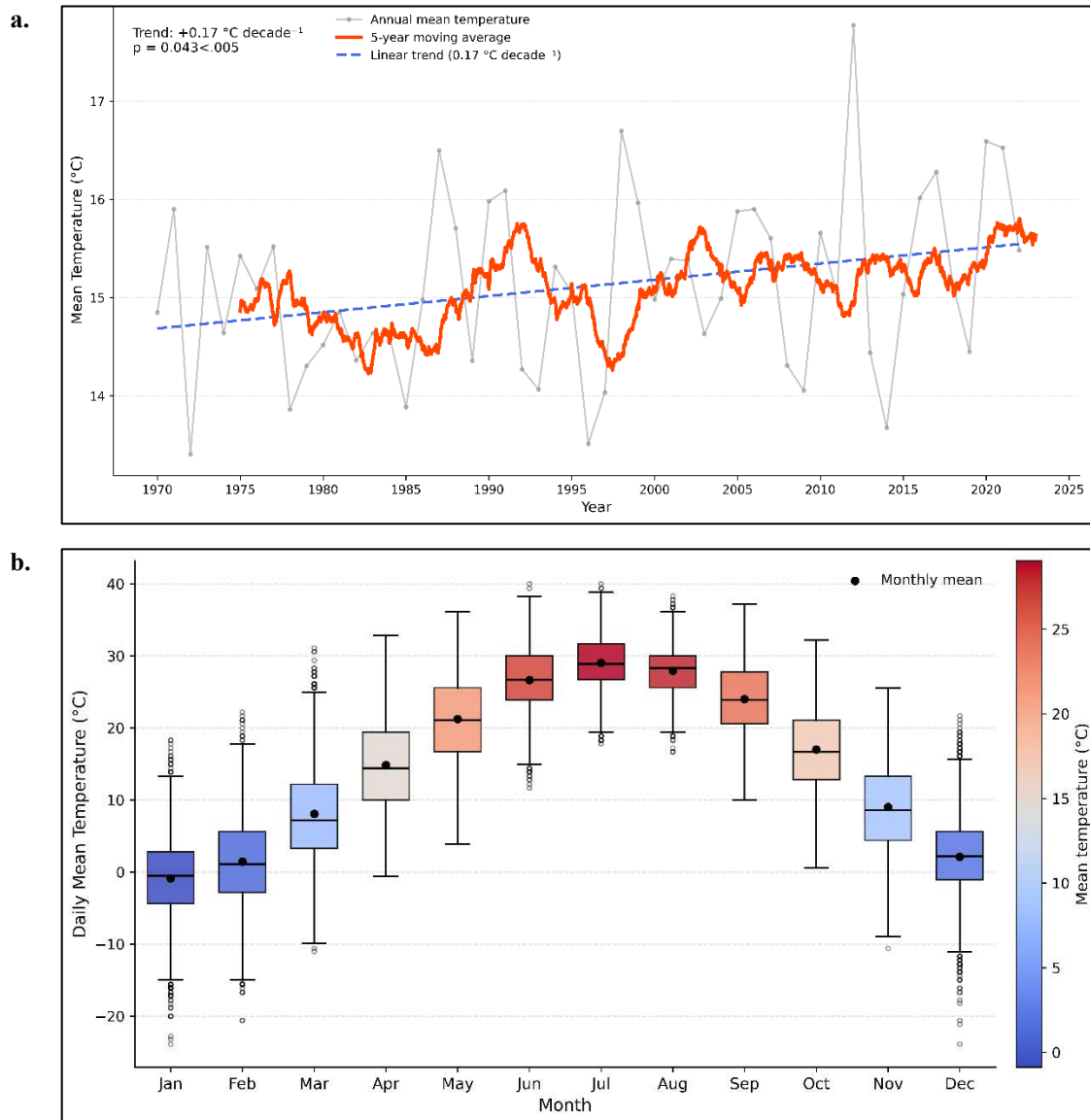


Fig. 1 | Long-Term Seasonal Trend and variability of daily temperature at Chicago O'Hare International Airport from Weather Station Data (1970–2022). **a.** Long-term temperature variability and warming trend at Chicago O'Hare International Airport, 1970–2022. The semi-transparent gray line shows annual mean air temperatures, while the orange line represents the 5-year moving average. The blue dashed line represents the linear trend estimated from annual mean temperatures. **b.** Seasonal Variability of daily mean temperature are represented using boxplots, showing the median, interquartile range, and temperature extremes across the annual cycle. The color gradient from blue to red indicates the transition from cooler winter months to warmer summer months.

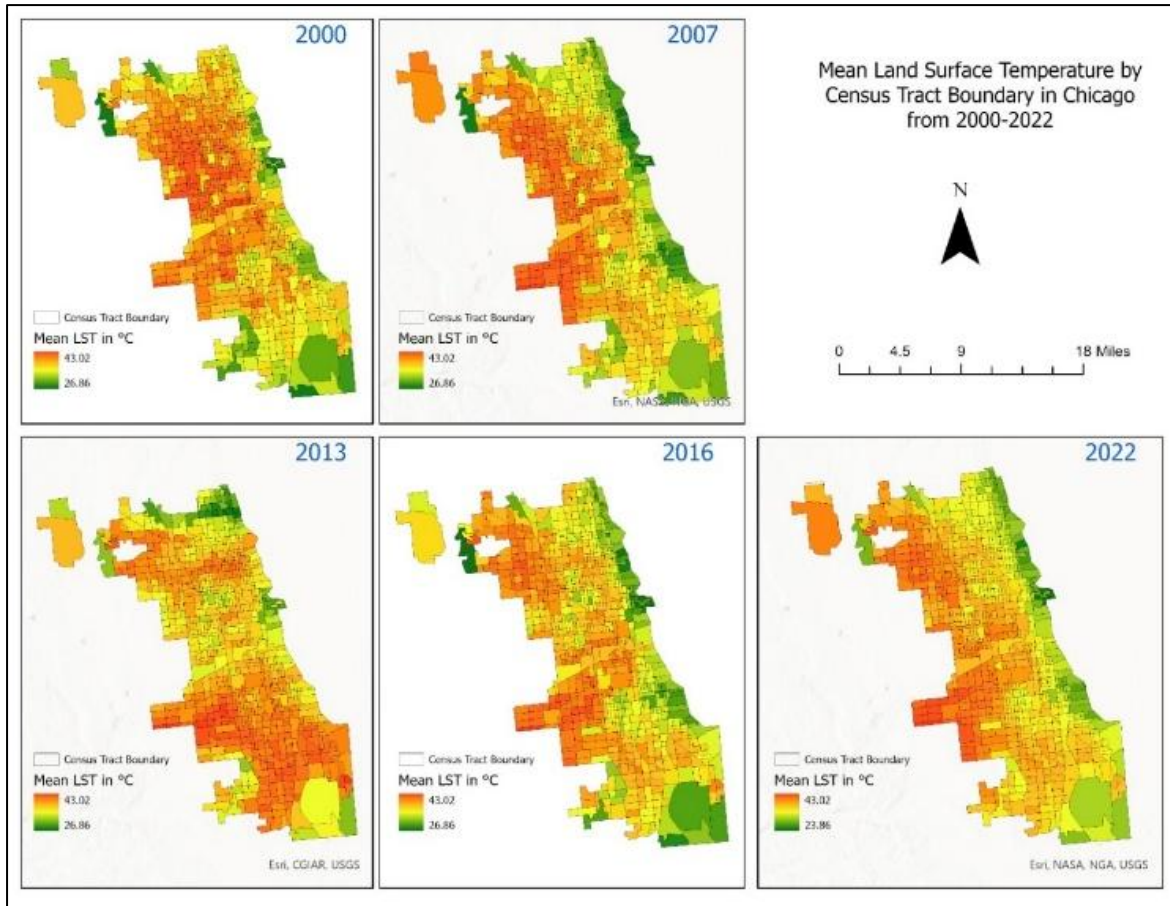


Fig. 2 | Land surface temperature dynamics across Chicago, 2000–2022 and Climatic context and seasonal thermal patterns. Five panels showing Chicago census tract LST for study period. Color scale 26.86–43.02°C (green=cool, red=hot). Clear spatial pattern: western/southern tracts are consistently hotter; lakeshore/northeastern tracts cooler

Over the 22-year study period (2000–2022), mean land surface temperature (LST) increased substantially from 34.1°C to 39.3°C, representing a net rise of 5.2°C (2.3°C per decade). The extent of extreme heat expanded dramatically, with census tracts exceeding 35°C increasing from 12.2% (98 tracts) to 98.9% (792 tracts). Most of these high-temperature areas (87.6%; 694 tracts) emerged after 2000, while only 98 tracts experienced persistent extreme heat throughout the study period. Warming patterns varied considerably across neighborhoods, with average warming of 0.34°C per decade; however, nearly one-third of tracts showed slight cooling, likely associated with localized vegetation increases. The mean LST in the ten hottest census tracts increased from approximately 35.3–35.4 °C in 2000 to 43.0 °C in 2022, representing a temperature differential of roughly 7–8 °C (Fig. 2). These spatially derived LST estimates provide the foundational dataset for the development, training, and evaluation of the subsequent machine learning models.

Model performance comparison

Following the characterization of temperature trends and LST patterns, we developed and compared RF and XGBoost models to predict census tract-level LST using a suite of environmental and land cover type predictors across all five time points (2000, 2007, 2013, 2016, and 2022). We pooled the

data to maximize sample size and capture temporally robust relationships between urban land cover and thermal emissions. Firstly, we split the pooled dataset into an 80% training set and a 20% independent holdout test set. We then performed 5-fold cross-validation on the training set to assess model stability before final evaluation on the unseen test set. We documented the detailed fold-level results and residual plots in Appendix Table S1 and Fig. S1–S2.

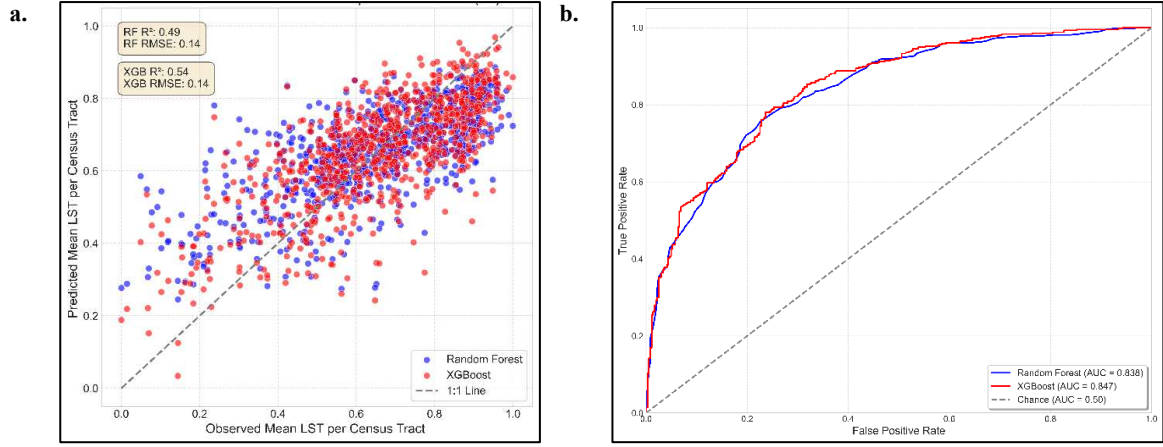


Fig. 3 | Predictive performance on independent test data. a Observed versus predicted standardized LST per census tract for the held-out 20% test set. The 1:1 line indicates perfect prediction. **b** ROC curves for classification of temperature extremes. AUC = 1.0 represents perfect discrimination; 0.5 indicates random performance.

Both models exhibited low root mean squared error (RMSE) values of approximately 0.14 in standardized units on the test set, demonstrating comparable precision at the level of individual predictions. XGBoost, however, achieved a higher coefficient of determination ($R^2 = 0.55$) compared to RF ($R^2 = 0.49$), indicating an enhanced capacity to capture spatial variability in LST across diverse urban environments. Observed-versus-predicted plots (Fig. 3a) suggests that XGBoost’s predictions were more tightly clustered around the 1:1 line, exhibiting lower variance and better calibration. The consistency between cross-validation and test-set performance—XGBoost marginally outperformed RF in both settings—suggests that the observed predictive accuracy was not dependent on a single random data partition. We evaluated the regression and classification tasks independently: RMSE and R^2 were derived from the continuous LST predictions on the held-out test set. The Receiver Operating Characteristic (ROC) and Area under the curve (AUC) assessment were computed from a separate binary classification of temperature extremes (above vs. below the training-set mean LST), ensuring that each evaluation metric reflected its intended predictive task. In evaluating the probabilistic classification of temperature extremes using ROC analysis, XGBoost demonstrated a marginal advantage, with an Area Under the ROC Curve (AUC) of 0.85 compared to RF’s 0.84 (Fig. 3b). Although the difference is modest, XGBoost consistently outperformed RF at various classification thresholds, showing slightly better sensitivity in detecting high-temperature events across the 2000–2022 study period.

Feature importance and SHAP-based driver analysis

We generated SHAP summary plots (Fig. 4) to decompose the model predictions and quantify the magnitude and direction of each predictor's influence on LST in the ML model. A key finding is the strong inter-model consensus between RF and XGBoost on the primary drivers of LST: the normalized difference vegetation index (NDVI) and carbon stock density (CSD) emerged as the two most influential predictors, both exhibiting strong negative correlations with LST. Higher NDVI values, indicative of denser and healthier vegetation canopies, were consistently associated with substantially lower surface temperatures. The strength and consistency of the NDVI–LST relationship across both models and all study periods reinforce the conclusion that vegetation is the most effective natural mitigation factor for urban heat (Fig. 4). Carbon stock density exhibited a clear negative correlation with LST in both models, extending the conventional urban heat paradigm by demonstrating that an ecosystem-service indicator integrating aboveground biomass, belowground biomass, soil organic carbon, and dead organic matter provides additional and complementary explanatory power for urban thermal variability⁸.

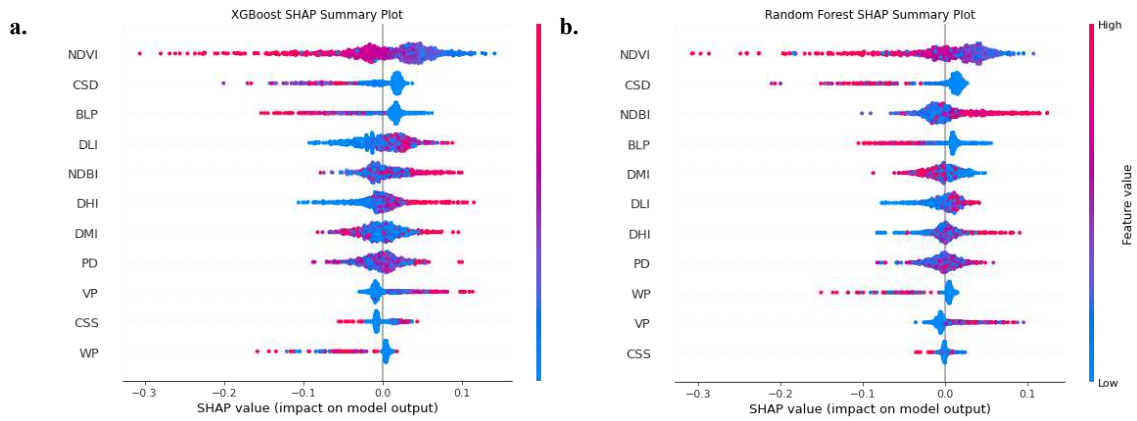


Fig. 4 | SHAP summary plots for feature importance. a XGBoost b Random Forest. Each point represents one census tract; color indicates standardized feature value (red = +1 SD, blue = -1 SD). Features ordered by mean absolute SHAP value. Standardization enables direct comparison across variables of differing original scales.

Variables related to developed land cover exhibited consistently positive correlations with LST. The percentage of high-intensity developed areas (DHI) showed consistently positive SHAP values, confirming that impervious surfaces and densely constructed environments increase heat retention. The normalized difference built-up index (NDBI), a spectral indicator commonly used to characterize the intensity and spatial heterogeneity of built-up surfaces and urban land-use patterns, provided additional insight into variations in urban form. SHAP analysis revealed a value-dependent effect: high NDBI values (indicating dense built-up surfaces) were associated with positive SHAP contributions to LST, whereas low NDBI values (indicating minimal built-up cover) corresponded to negative SHAP values, reflecting a relative cooling influence. Population density (PD) showed a moderate positive association with LST, though its influence was consistently ranked below that of the built-environment and vegetation variables. Notably, the percentage of barren land (BLP) demonstrated a counterintuitive negative correlation with LST in both models, likely attributable to the extensive presence of vegetated open spaces or transitional landscapes classified as barren land

within the Chicago metropolitan area. Water bodies (WP) showed the expected negative effect on LST, highlighting the evaporative cooling capacity of blue infrastructure, a finding particularly relevant given Chicago's proximity to Lake Michigan.

Partial dependence analysis of predictor–LST relationships

We also generated partial dependence plots (PDPs; Fig. 5) for both models to explain the functional relationships between individual predictors and LST, holding all other variables at their mean values. The PDP analysis revealed several important nonlinearities that the SHAP summary plots alone could not fully resolve. NDVI exhibited a predominantly negative relationship with LST in both models, but the functional form differed: RF captured a slightly linear negative slope, whereas XGBoost modeled a distinctly nonlinear pattern characterized by a steep cooling response at low-to-moderate NDVI values (approximately 0.1–0.3) that plateaus at higher NDVI values. This suggests that the marginal cooling benefit of additional vegetation is greatest in sparsely vegetated areas and diminishes as canopy cover approaches saturation, a finding consistent with the ecological concept of diminishing returns in ecosystem service provision²⁶. The XGBoost model's ability to capture this nonlinearity likely contributes to its superior predictive performance. A positive relationship in XGBoost (more vegetation associated with higher LST), which appears counterintuitive but likely reflects the fact that vegetated areas in this dataset coexist with urban development; the PDP isolates the marginal effect while holding other variables constant, meaning this relationship emerges after accounting for the cooling effect of NDVI already captured in the model.

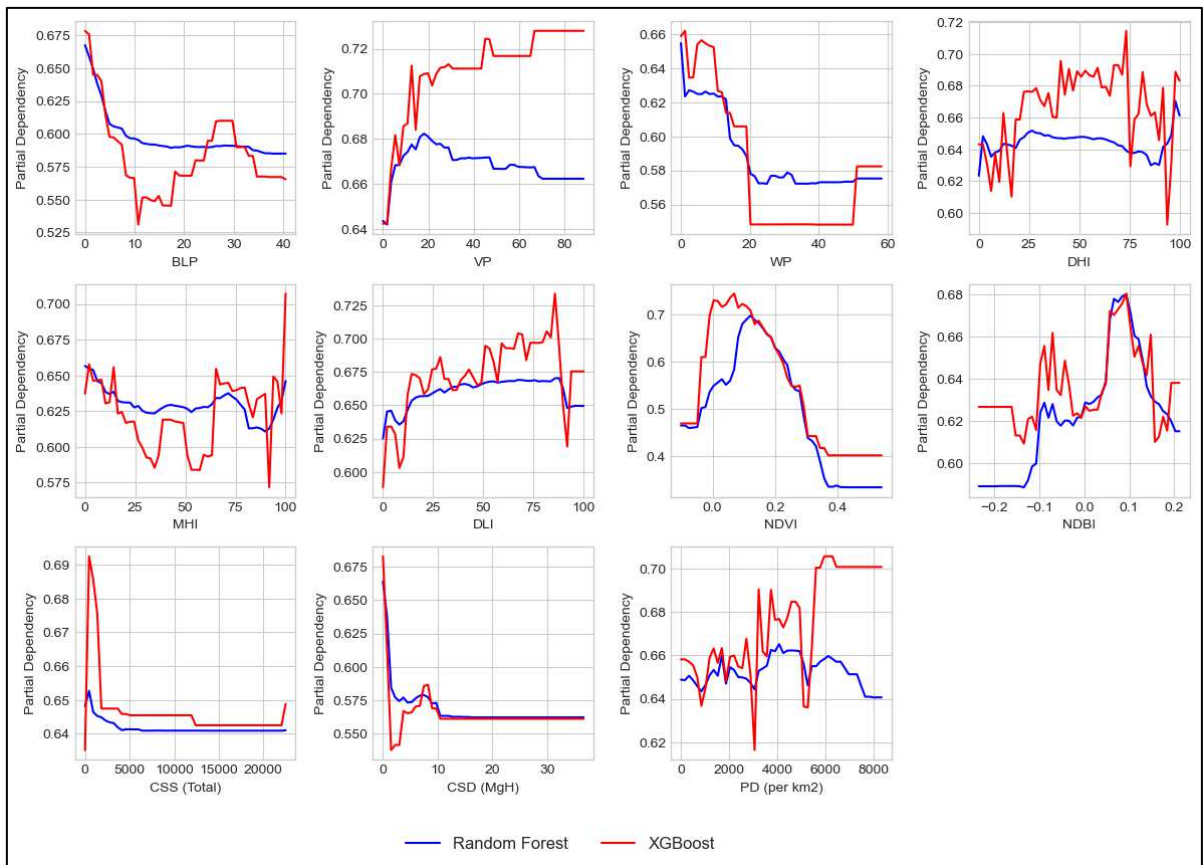


Fig. 5 | Partial dependence relationships between land surface temperature and predictor variables. Partial dependence plots illustrate the marginal effects of 11 environmental, built-environment and

demographic predictors on standardized mean land surface temperature. The plots highlight nonlinear responses and differences in the estimated predictor–temperature relationships between the two machine-learning approaches.

Among the developed area variables, Developed High Intensity (DHI) Percentage exhibited a more pronounced positive relationship with LST in RF compared to XGBoost, whereas Developed Medium Intensity Percentage (MHI) showed a clearer positive relationship in XGBoost. Developed Low Intensity (DLI) Percentage demonstrated a similar positive relationship in both models, although with a slightly steeper slope in XGBoost. NDBI exhibited a consistently positive relationship in both models, with XGBoost showing a slightly steeper slope. This aligns with the well-established urban heat island effect: impervious surfaces, reduced vegetation cover increase thermal absorption, and decrease evaporative cooling.

Waterbody Percentage (WP) showed a negative relationship in both models, with XGBoost capturing a sharper decline in LST as water cover increases — consistent with water's high heat capacity and evaporative cooling. Carbon Stock Density (CSD) displayed a clearly nonlinear relationship in XGBoost, with the strongest negative effect on LST occurring at intermediate-to-high carbon stock densities, whereas RF showed a less defined pattern. This suggests that carbon-rich ecosystems (likely forested or wetland areas) provide nonlinear cooling benefits. Population Density (PD) demonstrated a positive relationship with LST in both models, with RF exhibiting a slightly steeper slope — reflecting the concentration of anthropogenic heat sources, impervious surfaces, and reduced vegetation in densely populated areas. Barren Land Percentage (BL) showed a positive relationship in both models (higher barren land associated with higher LST), which differs from the negative SHAP contribution for this variable. This divergence is consistent with the distinct interpretative frameworks of the two methods: PDPs estimate average marginal effects while holding other variables at their means, whereas SHAP values quantify local feature contributions accounting for the joint distribution of predictors and interactions. The positive PDP relationship suggests that, after controlling for other variables, barren land itself acts as a warming surface (low albedo, no evaporative cooling), while the negative SHAP contribution may reflect barren land's correlation with cooling variables like waterbody percentage or NDVI in the joint feature space²⁷.

Carbon stock dynamics and their thermal implications

Census tract–level carbon stock density across Chicago for five study periods (2000, 2007, 2013, 2016, 2022), derived from the InVEST Carbon Storage and Sequestration model. Although aggregate carbon stock declined by only 0.38% over the two decades (from 103,337 to 102,940 Mg), tract-level analysis revealed a spatially heterogeneous and predominantly loss-dominated pattern (Fig. 6a). Carbon stock density decreased in 28.6% of census tracts ($n = 229$), compared with increases in only 18.2% ($n = 146$), resulting in a negatively skewed distribution of changes among non-stationary tracts (mean = -0.061 ; median = -0.014 Mg C ha⁻¹). Notably, the majority of the observed change was concentrated in just 4.2% of tracts that experienced changes exceeding ± 0.5 Mg C ha⁻¹. This pattern suggests that localized land-use transitions and vegetation changes may have played a disproportionate role in shaping urban carbon storage trajectories²⁸ (Fig. 6b). The temporal trajectory was non-monotonic: total stock increased from 2000 to 2013, decreased in 2016, and recovered

partially by 2022. These fluctuations likely reflect the interplay of land use policy changes, the timing of urban development projects, natural disturbances, and the succession dynamics of urban vegetation communities²⁹. Spatially, the northern and western sections of the city consistently maintained the highest carbon storage, with Census Tract 7608 recording the maximum values (ranging from approximately 21,421 Mg in 2000 to 22,390 Mg in 2022) and Census Tract 8388 ranking second (12,344–13,800 Mg). These high-carbon tracts likely correspond to large urban parks, forest preserves, or riparian corridors. Conversely, the central business district and heavily developed areas exhibited the lowest carbon storage⁹.

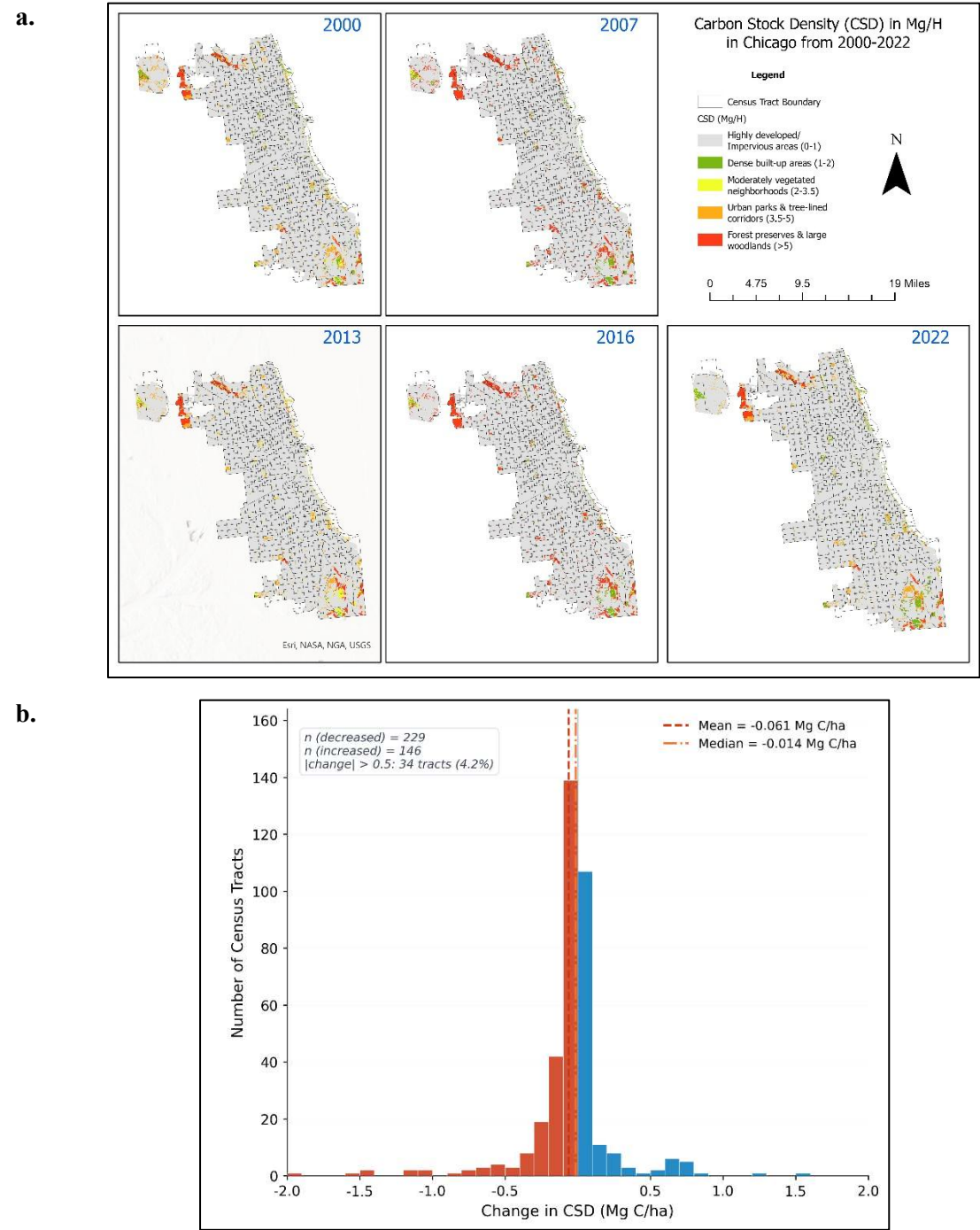


Fig. 6 | Census tract–based carbon stock maps (2000–2022) derived from the InVEST model. a. Carbon stock density (CSD) dynamics at the census-tract level, 2000-2022. Distribution of tracts across five CSD

categories where higher CSD tracts are in Red. Over 87% of tracts consistently fall below 1.0 Mg C/ha, rendering temporal differences **b**. Histogram of Non-Zero Change Tracts (n=375) in terms of CS only

Census tracts that maintained or increased their carbon storage over the study period generally exhibited moderate LST increases, while those experiencing the greatest carbon loss tended to show the most pronounced warming. For example, tracts in the western industrial corridor, which experienced significant vegetation loss and a 22% reduction in green cover, recorded the highest LST increases (up to 6–8 °C). This spatial concordance between carbon depletion and thermal amplification reinforces the SHAP-derived finding that CSD is a primary cooling factor, and it suggests that carbon conservation and restoration can serve dual objectives: climate mitigation through carbon sequestration and local heat reduction through ecosystem cooling services³⁰.

Disproportionate thermal impacts of industrial land use

Census tracts intersecting industrial zone boundaries (n = 227, representing 28% of the total tracts) exhibited a mean LST of 40.33 °C, significantly higher than the 38.88 °C mean in non-industrial tracts (Fig. 7). The spatial pattern of temperature change from 2000 to 2022 revealed a pronounced gradient, with the largest increases exceeding 6 °C, primarily within or adjacent to industrial zones. The mean temperature difference in industrial-intersecting tracts was 6.00 °C, compared with 4.87 °C in non-industrial tracts. Land cover analysis revealed a 22% reduction in vegetation from 2000 to 2022, coinciding with the expansion of high- and medium-intensity developed zones. During this period, the average LST increased from 32 °C to 38 °C, marking a 20% rise, with the most significant temperature increases observed in high-, medium-, and low-intensity developed regions. These results highlight the disproportionate role of industrial land use in driving localized warming and intensifying the UHI effect.

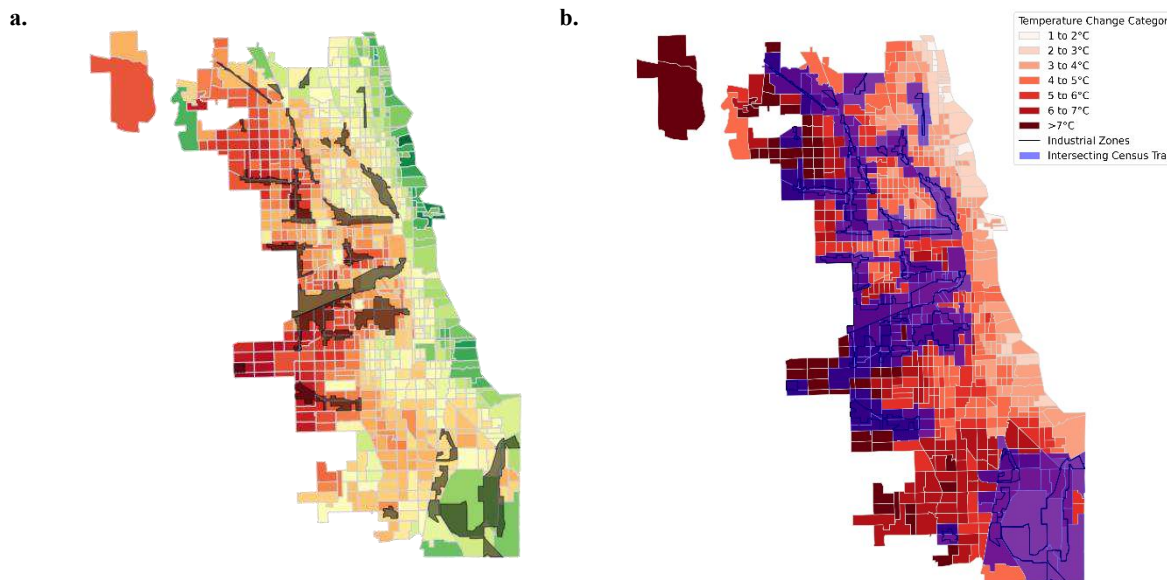


Fig. 7 | Spatial pattern of temperature change from 2000 to 2022 across Chicago census tracts. a mean Highest Temperature Census Tracts with Industrial Zone Overlay. b Census Tracts with highest increase in temperature change over the years including industrial zone overlay. Industrial zones and adjacent areas exhibit the largest warming (>6 °C), highlighting the disproportionate thermal burden in industrial corridors.

Discussion

This study demonstrates that machine-learning methods, particularly when combined with explainable frameworks; offer an effective approach for predicting and understanding urban thermal variation at fine spatial scales. XGBoost consistently outperformed Random Forest—a result we attribute to its sequential boosting and built-in regularization, which reduce overfitting and better capture nonlinear interactions among environmental and socioeconomic variables³¹. Both models achieved comparable RMSE values, yet XGBoost's higher R^2 and AUC suggesting reliable spatial LST predictions. These performance levels match or exceed those reported in analogous urban thermal studies³², suggesting the framework transfers readily to other cities. Notably, XGBoost captured physically meaningful nonlinearities—such as the diminishing cooling return of vegetation at high NDVI and the threshold response of carbon density to LST—that RF approximated as linear.

A central finding is the strong agreement between both models on the dominant LST drivers: NDVI and carbon stock density consistently ranked as the two most influential predictors. This inter-model consensus rules out algorithm-specific artifacts and reinforces confidence in these environmental controls. The emergence of CSD as a primary cooling factor is particularly noteworthy. Conventional urban heat studies typically rely on vegetation indices and impervious-surface metrics alone^{5,6}; CSD, by integrating aboveground and belowground biomass, soil organic carbon, and dead organic matter into one predictor, captures the cumulative thermal effect of overall ecosystem health—giving planners a more holistic metric for heat-mitigation decisions.

The long-term temperature record contextualizes these machine-learning findings. A statistically significant 52-year warming trend ($P = 0.03$), together with the rise in extreme heat days from 5 to 28 (albeit non-monotonic, with 7 days in 2013), places Chicago within the broader pattern of urban thermal amplification. Crucially, this warming is far from spatially uniform. Industrial and high-density tracts warmed by up to 7–8 °C over two decades, whereas less-developed eastern tracts changed little—a disparity invisible in citywide averages. Analyzing heat at census-tract resolution, as done here, is therefore essential for identifying the localized thermal inequities that disproportionately burden vulnerable communities.

Built-up indicators—NDBI and all development-intensity classes—showed positive LST associations, confirming that urban warming extends well beyond dense commercial cores into medium- and low-density neighborhoods where cumulative impervious cover meaningfully contributes to regional heat³³. Industrial zones stood out as the most intense hotspots, exceeding non-industrial tracts by over 1 °C on average, with some areas warming more than 6 °C. This concentration reflects the convergence of extensive impervious surfaces, manufacturing-related anthropogenic heat, and sparse vegetation. The infrastructure implications are serious: extreme heat amplifies power-outage risk³⁴, and industrial neighborhoods—often home to disadvantaged communities—face compounding exposures from heat, air pollution, and limited cooling access^{35,36}.

Water bodies showed moderate cooling in both models, underscoring the thermal value of blue infrastructure³⁷. Chicago's proximity to Lake Michigan suggests that preserving and expanding water features and riparian corridors could meaningfully temper nearby temperatures. Population density,

though positively associated with LST, was less influential than vegetation, carbon stock, or built-environment variables—implying that landscape configuration matters more than demographic density for heat mitigation, and that physical urban design interventions are likely to outperform population-level strategies.

Three implications follow. First, methodologically, combining SHAP (observation-level importance and direction) with PDPs (average functional form) yields richer insight than either alone, as illustrated by the divergent barren-land results²⁷—a point reinforced by recent SHAP applications in environmental ML³⁸. Second, carbon stock density is a powerful and previously underused predictor of urban heat; it extends the conventional NDVI–impervious-surface paradigm by capturing ecosystem health in a single, integrative metric. Third, for planning practice, the strongest cooling levers are interventions that jointly maximize vegetation cover and carbon storage—particularly large-canopy native tree planting and preservation of mature woodlands.

We should be considering key limitations when interpreting these results. First, census-tract aggregation may mask sub-tract thermal heterogeneity as satellite-derived LST remains subject to sensor-related uncertainties hence; future work could leverage parcel-level or street-level thermal data to capture finer-scale heterogeneity. Second, the InVEST carbon estimates rely on generalized lookup tables rather than field measurements. Third, the partial dependence analysis assumes predictor independence; instead accumulated local effects (ALE) plots could provide more reliable interpretations in the presence of strong predictor interaction³⁹. Despite these limitations, the consistency of key findings across SHAP, PDP, trend analysis, and two independent model architectures lends confidence to the robustness of the reported relationships.

Methods

Data sources and satellite image processing

This study combined multiple geospatial, remote sensing, and socioeconomic data sources. Landsat images were obtained and preprocessed using Google Earth Engine (GEE) and visualized with Folium within a boundary covering the Chicago metropolitan area. Image collections were filtered by date range and cloud cover (maximum 10%), and study years were selected at the intersection of peak summer temperature extremes and cloud-free Landsat availability, as determined by the trend analysis. A 5-day moving average temperature of 30 °C during summer served as the threshold for standardizing image dates. Given Landsat’s 16-day revisit cycle, acquisition dates were adjusted to the nearest cloud-free pass within ± 7 days of the identified temperature extremes. Daily MODIS LST data (MOD11A1) were retrieved for the corresponding dates to enable inter-sensor calibration via quantile mapping. Land cover data were retrieved from the NLCD for the study years (2001, 2006, 2011, 2016, and 2019, corresponding approximately to 2000, 2007, 2013, 2016, and 2022) through the Multi-Resolution Land Characteristics (MRLC) Consortium. NLCD classifications were grouped into six categories: vegetation, water bodies, barren land, and three levels of development intensity (low, medium, high). Census tract boundaries and population data for Cook County, Illinois, were obtained from the U.S. Census Bureau for the 2000, 2010, and 2020 decennial censuses. Population

density (persons/km²) was calculated at the tract level and combined with LST and carbon stock estimates for integrated analysis.

LST extraction and quantile mapping

LST was estimated from the thermal bands of Landsat 5 TM and Landsat 8 TIRS imagery for the study periods. Spectral radiance ($L\lambda$) was first calculated from the digital number (DN) values using sensor-specific equations: $L\lambda = LMIN + ((LMAX - LMIN) / (QCALMAX - QCALMIN)) \times QCAL$ for Landsat 5 TM, and $L\lambda = (ML \times DL) + AL$ for Landsat 8 TIRS, where ML and AL are band-specific rescaling factors from the metadata. The at-satellite brightness temperature (TB) was then derived as $TB = K2 / \ln(K1 / L\lambda + 1)$, where K1 and K2 are sensor-specific calibration constants. Finally, LST was obtained from TB using the equation $LST = TB / [1 + ((\lambda \times TB / \rho) \times \ln \epsilon)] - 273.15$, where λ is the wavelength of emitted radiance, ρ is a physical constant, and ϵ is emissivity (typically 0.97–0.99). The entire process was executed in ArcGIS Model Builder for consistency across study periods^{40–44}. We developed a geostatistical quantile-mapping tool in Python to adjust Landsat LST values using MODIS as the reference distribution. This approach has been previously validated for bias correction in climate applications^{45,46}.

Land cover and carbon stock datasets

NLCD land cover rasters were converted to GeoTIFF format and resampled to match the spatial resolution of the LST data. Carbon stock density (CSD) was computed using the InVEST Carbon Storage and Sequestration model (version 3.11.0)⁴⁷. Carbon pool parameters—including aboveground biomass, belowground biomass, dead organic matter, and soil organic carbon—were assembled from global biomass carbon density datasets and SoilGrids 2.0 stock density maps⁴⁸. Mean carbon values were assigned to each NLCD land cover class using ArcGIS Zonal Statistics as Table. These carbon pool data and reclassified land-cover rasters were input into the InVEST model, which generated total current carbon storage (Mg C per pixel) as a GeoTIFF file. Carbon stock density (Mg C ha⁻¹) was then calculated at the census tract level for each study period.

Historical temperature trend analysis

Historical temperature trends were characterized using daily maximum temperature records from the NOAA National Centers for Environmental Information (NCEI) for the Chicago O'Hare International Airport station (1970–2022)⁴⁹. A 5-year moving average was applied to smooth interannual variability and reveal underlying trends. Linear regression was used to assess the statistical significance of long-term warming, and extreme temperature days were identified using a 5-day moving average threshold of 30 °C. This long-term record provides essential climatological context for the subsequent machine-learning analysis of LST patterns.

Machine learning framework for multi-criteria LST prediction

Two ML models—Random Forest and Extreme Gradient Boosting (XGBoost)—were trained on the dataset^{50,51}. The models were trained to predict census tract level standardized mean LST (dependent variable Y) using 11 explanatory variables representing land cover composition, vegetation condition, urban development, ecosystem carbon storage, and population distribution from five time points: 2000, 2007, 2013, 2016, and 2022. Specifically, the predictor set included the proportional

coverage of major National Land Cover Database (NLCD) land cover classes (barren land, vegetation, water bodies, and low-, medium-, and high-intensity developed land), the Normalized Difference Vegetation Index (NDVI), the Normalized Difference Built-up Index (NDBI), total carbon stock (CSS), carbon stock density (CSD), and population density (PD). Rather than training separate models for each year, we pooled all observations into a single dataset to maximize sample size and enable the models to learn generalizable relationships between land cover characteristics and LST that persist across time. The pooled dataset therefore contains repeated spatial observations (census tracts) at multiple temporal snapshots, with the year of observation included as an implicit stratifying factor in the data structure. We calculated land-cover percentages at the census-tract level for each study year using a custom Python script in ArcGIS Pro, and all data were combined into year-specific master tables. All variables were standardized using z-score normalization prior to model training.

To ensure an unbiased assessment of generalization, we first split the full dataset into a training set (80%) and an independent hold-out test set (20%) using a fixed random seed (`random_state = 50`). We performed the split before any model fitting, ensuring that the test set remained completely unseen during training and hyperparameter selection. We then performed 5-fold cross-validation exclusively on the training set to evaluate model stability and guard against overfitting. In each fold, the training data were divided into five equal partitions; the model was trained on four partitions and validated on the remaining one, rotating until each partition served as validation once. After cross-validation, we retrained each model on the entire training set and evaluated their performance on the held-out 20% test set. This two-stage procedure—cross-validation for internal validation followed by a single independent test-set evaluation on pooled multi-year data—follows best practices in predictive modeling and ensures that our reported test metrics reflect true out-of-sample performance across the full study period (2000–2022). We evaluated the performance using RMSE and R^2 on the test set. Separately, a binary classification task was defined by converting the target variable to an above/below training-set-mean threshold, and ROC curves with AUC values were computed to assess classification of temperature extremes. SHAP analysis was applied to interpret feature importance and directionality in both models²³. Lastly, we generated partial dependence plots to characterize the marginal functional relationships between each predictor and LST, providing complementary interpretability to the observation-level SHAP analysis.

Data availability

The datasets analyzed and generated during the current study are available at <https://doi.org/10.5281/zenodo.21970121>

References

1. UN News. UN relief chief appeals for concerted action to tackle deadly heatwave threat | UN News. *UN News: Global perspective Human stories* <https://news.un.org/en/story/2022/10/1129407> (2022).
2. Aljerf, L. Biodiversity is key for More Variety for Better Society. *Biodivers. Int. J.* **1**, 5–7 (2017).
3. Sarathchandra, C. *et al.* Impact of land use and land cover changes on carbon storage in

- rubber dominated tropical Xishuangbanna, South West China. *Ecosyst. Heal. Sustain.* **7**, (2021).
4. Bader, D. A. *et al.* Urban Climate Science. in *Climate Change and Cities* 27–60 (Cambridge University Press, 2018). doi:10.1017/9781316563878.009.
 5. Amani-Beni, M., Chen, Y., Vasileva, M., Zhang, B. & Xie, G. di. Quantitative-spatial relationships between air and surface temperature, a proxy for microclimate studies in fine-scale intra-urban areas? *Sustain. Cities Soc.* **77**, 103584 (2022).
 6. Fu, P. & Weng, Q. A time series analysis of urbanization induced land use and land cover change and its impact on land surface temperature with Landsat imagery. *Remote Sens. Environ.* **175**, 205–214 (2016).
 7. Weng, Q., Firozjaei, M. K., Sedighi, A., Kiavarz, M. & Alavipanah, S. K. Statistical analysis of surface urban heat island intensity variations: A case study of Babol city, Iran. *GIScience Remote Sens.* **56**, 576–604 (2019).
 8. Houghton, R. A. *et al.* Carbon emissions from land use and land-cover change. *Biogeosciences* **9**, 5125–5142 (2012).
 9. Mills, E. & Simmons, C. *Growing Populations, Changing Landscapes. Growing Populations, Changing Landscapes* (National Academies Press, Washington, D.C., 2001). doi:10.17226/10144.
 10. Alfraihat, R., Mulugeta, G. & Gala, T. S. Ecological Evaluation of Urban Heat Island in Chicago City, USA. *J. Atmos. Pollution, Vol. 4, 2016, Pages 23-29* **4**, 23–29 (2016).
 11. Coseo, P. & Larsen, L. Cooling the Heat Island in Compact Urban Environments: The Effectiveness of Chicago’s Green Alley Program. *Procedia Eng.* **118**, 691–710 (2015).
 12. Cosgrove, A. & Berkelhammer, M. Downwind footprint of an urban heat island on air and lake temperatures. *npj Clim. Atmos. Sci.* **2018 11** **1**, 1–10 (2018).
 13. Liu, H. & Weng, Q. An examination of the effect of landscape pattern, land surface temperature, and socioeconomic conditions on WNV dissemination in Chicago. *Environ. Monit. Assess.* **159**, 143–161 (2009).
 14. Jetz, W., Rahbek, C. & Lichstein, J. W. Local and global approaches to spatial data analysis in ecology. *Glob. Ecol. Biogeogr.* **14**, 97–98 (2005).
 15. Li, S., Zhao, Z., Miaomiao, X. & Wang, Y. Investigating spatial non-stationary and scale-dependent relationships between urban surface temperature and environmental factors using geographically weighted regression. *Environ. Model. Softw.* **25**, 1789–1800 (2010).
 16. AlDousari, A. E. *et al.* Modelling the impacts of land use/land cover changing pattern on urban thermal characteristics in Kuwait. *Sustain. Cities Soc.* **86**, 104107 (2022).
 17. Edan, M. H., Maarouf, R. M. & Hasson, J. Predicting the impacts of land use/land cover change on land surface temperature using remote sensing approach in Al Kut, Iraq. *Phys. Chem. Earth, Parts A/B/C* **123**, 103012 (2021).
 18. Lee, J. & Kim, J. Evaluation of Spatial and Temporal Performance of Deep Learning Models for Travel Demand Forecasting: Application to Bike-Sharing Demand Forecasting. *J. Adv. Transp.* **2022**, 5934670 (2022).
 19. Mansourmoghaddam, M. *et al.* Modeling and Estimating the Land Surface Temperature (LST) Using Remote Sensing and Machine Learning (Case Study: Yazd, Iran). *Remote Sens.* **2024, Vol. 16, Page 454** **16**, 454 (2024).
 20. Valipour, M. & Guzmán, S. M. Identification of the Meteorological Variables Influencing Evapotranspiration Variability Over Florida. *Environ. Model. Assess.* **27**, 645–663 (2022).

21. Tian, L. *et al.* Prediction of Land Surface Temperature Considering Future Land Use Change Effects under Climate Change Scenarios in Nanjing City, China. *Remote Sens.* **15**, 2914 (2023).
22. Ahmad, A. *et al.* Carbon Emissions With Forest Cover Change and Wood Harvest in the Dry Temperate Region of Pakistan Between 1908 and 2015. *Front. Environ. Sci.* **10**, (2022).
23. Lundberg, S. M., Allen, P. G. & Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *Adv. Neural Inf. Process. Syst.* **30**, (2017).
24. Calvin, K. *et al.* IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland. (2023) doi:10.59327/IPCC/AR6-9789291691647.
25. Oke, T. R. The energetic basis of the urban heat island. *Q. J. R. Meteorol. Soc.* **108**, 1–24 (1982).
26. Mohd Tajuddin Abdullah, P. Fas. & <https://akademisains.academia.edu/MohdTajuddinAbdullahPhDFASc>. The value of estuarine and coastal ecosystem services. at https://www.academia.edu/3359823/The_value_of_estuarine_and_coastal_ecosystem_services.
27. Lundberg, S. M. *et al.* From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020 21 2**, 56–67 (2020).
28. Nowak, D. J., Greenfield, E. J., Hoehn, R. E. & Lapoint, E. Carbon storage and sequestration by trees in urban and community areas of the United States. *Environ. Pollut.* **178**, 229–236 (2013).
29. Mallick, J., Almesfer, M. K., Alsubih, M., Ahmed, M. & Ben Kahla, N. Estimating Carbon Stocks and Sequestration With Their Valuation Under a Changing Land Use Scenario: A Multi-Temporal Research in Abha City, Saudi Arabia. *Front. Ecol. Evol.* **10**, 905799 (2022).
30. Bowler, D. E., Buyung-Ali, L., Knight, T. M. & Pullin, A. S. Urban greening to cool towns and cities: A systematic review of the empirical evidence. *Landsc. Urban Plan.* **97**, 147–155 (2010).
31. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* **13-17-August-2016**, 785–794 (2016).
32. Zhao, L., Lee, X., Smith, R. B. & Oleson, K. Strong contributions of local background climate to urban heat islands. *Nature* **511**, 216–219 (2014).
33. Arnfield, A. J. Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island. *Int. J. Climatol.* **23**, 1–26 (2003).
34. Kabir, S. M. R., Rahman, M., Zisha, F. K. & Meng, L. Quantifying Event-Based Heatwave-Induced Power Outage Risk: A Multi-Year Spatiotemporal Analysis in Texas. *Sustain.* **2026, Vol. 18, Page 6205 18**, 6205 (2026).
35. Heavyside, C., Macintyre, H. & Vardoulakis, S. The Urban Heat Island: Implications for Health in a Changing Environment. *Curr. Environ. Heal. reports* **4**, 296–305 (2017).
36. Jesdale, B. M., Morello-Frosch, R. & Cushing, L. The racial/ ethnic distribution of heat risk-related land cover in relation to residential segregation. *Environ. Health Perspect.* **121**, 811–817 (2013).
37. Vancutsem, C., Ceccato, P., Dinku, T. & Connor, S. J. Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. *Remote*

- Sens. Environ.* **114**, 449–465 (2010).
38. Evans, C., Zhu, L., Baker, K. & Meng, L. Unveiling multiscale drivers of wind speed in Michigan using machine learning. *npj Clim. Atmos. Sci.* **2025** *81* **8**, 286- (2025).
 39. Apley, D. W. & Zhu, J. Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **82**, 1059–1086 (2020).
 40. Pal, S. & Ziaul, S. Detection of land use and land cover change and land surface temperature in English Bazar urban centre. *Egypt. J. Remote Sens. Sp. Sci.* **20**, 125–145 (2017).
 41. Shataee, S. Non-parametric forest attributes estimation using Lidar and TM data. in *The Asian Conference on Remote Sensing, Taiwan* 3–7 (2011).
 42. Shatnawi, N. & Abu Qdais, H. Mapping urban land surface temperature using remote sensing techniques and artificial neural network modelling. *Int. J. Remote Sens.* **40**, 3968–3983 (2019).
 43. Uehara, K., Murakami, S., Oikawa, S. & Wakamatsu, S. Wind tunnel experiments on how thermal stratification affects flow in and above urban street canyons. *Atmos. Environ.* **34**, 1553–1562 (2000).
 44. Zeng, Z. *et al.* A reversal in global terrestrial stilling and its implications for wind energy production. *Nat. Clim. Chang.* **9**, 979–985 (2019).
 45. Bahari, N. I. S., Muharam, F. M., Zulkafli, Z., Mazlan, N. & Husin, N. A. Modified Linear Scaling and Quantile Mapping Mean Bias Correction of MODIS Land Surface Temperature for Surface Air Temperature Estimation for the Lowland Areas of Peninsular Malaysia. *Remote Sens.* **2021**, Vol. 13, Page 2589 **13**, 2589 (2021).
 46. Qian, W. & Chang, H. H. Projecting Health Impacts of Future Temperature: A Comparison of Quantile-Mapping Bias-Correction Methods. *Int. J. Environ. Res. Public Health* **18**, 1–12 (2021).
 47. Kohestani, N., Rastgar, S., Heydari, G., Jouibary, S. S. & Amirnejad, H. Spatiotemporal modeling of the value of carbon sequestration under changing land use/land cover using InVEST model: a case study of Nour-rud Watershed, Northern Iran. *Environ. Dev. Sustain. A Multidiscip. Approach to Theory Pract. Sustain. Dev.* **26**, 14477–14505 (2024).
 48. Yang, H. *et al.* Global increase in biomass carbon stock dominated by growth of northern young forests over past decade. *Nat. Geosci.* **16**, 886–892 (2023).
 49. Rajput, M. & Augenbroe, G. Impact of meso-scale weather uncertainty on building performance simulation outcomes. *Build. Environ.* **263**, 111835 (2024).
 50. Sameh, S., Zarzoura, F. & El-Mewafi, M. Automated Mapping of Urban Heat Island to Predict Land Surface Temperature and Land use/cover Change Using Machine Learning Algorithms: Mansoura City. *Int. J. Geoinformatics* **18**, 47–67 (2022).
 51. Samb, R., Jena, A., Manavvi, S., Roy, U. K. & Yadav, B. Machine learning driven land surface temperature prediction and urban heat risk assessment in the Gambia. *Discov. Cities* **2026** *31* **3**, 49- (2026).

Acknowledgements and Funding Information

S.R.K. acknowledges the support from the Harrison Endowment, funded by the School of Environment, Geography, and Sustainability at Western Michigan University. L.Z. and F.K.Z. acknowledge the support from the Milton E. and Ruth M. Scherer Fellowship.

Authors and Affiliations

S M Redwan Kabir, Laiyin Zhu & Farhana Kabir Zisha; School of Environment, Geography, and Sustainability, Western Michigan University, Kalamazoo, MI, USA

Corresponding author

Correspondence to S M Redwan Kabir; Email: smredwan.kabir@wmich.edu

Ethics declarations

Competing interests: The authors declare no competing interests.

Author Contributions Statement

Conceptualization: S.R.K. and L.Z.; Supervision: L.Z.; Methodology: S.R.K. and L.Z.; Data curation, S.R.K.; Software: S.R.K.; Formal analysis: S.R.K. and F.K.Z.; Visualization: S.R.K. ; Writing—original draft preparation: S.R.K ; Writing—review and editing: L.Z. and F.K.Z.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementMaterials.docx](#)