

Article

GeoAI-Enabled Ensemble Modeling to Assess Land Use and Atmospheric Pollutant Impacts on Land Surface Temperature in the US Southwest

Bijoy Mitra  and Guiming Zhang * 

Department of Geography & the Environment, College of Natural Sciences and Mathematics,
University of Denver, Denver, CO 80208, USA; bijoy.mitra@du.edu

* Correspondence: guiming.zhang@du.edu; Tel.: +1-303-871-7908

Highlights

What are the main findings?

- The impact of biophysics dominates summer LST predictions, while atmospheric pollutants become the primary drivers of winter LST.
- Hyperparameter-tuned CatBoost and Extra Trees achieved high accuracy for spatial LST prediction.

What are the implications of the main findings?

- Regional LST hotspots were more driven by topography, and high urban intensity showed inverse thermal impacts during winter.
- Tropospheric ozone shows a negative association with surface temperature in the US Southwest based on model-attributed importance.

Abstract

The US Southwest is one of the driest and hottest regions, with a recent upsurge in land surface temperature (LST). Further, with land-use changes and global warming, anthropogenic pollution also significantly contributes to the rise in surface temperatures. While the impact of pollution on LST has been studied only in specific urban regions, insights from a broader, more diverse topography remain limited. This research incorporates LST with land cover parameters (NDBI, MNDWI, NDBSI, SAVI, WET), surface albedo, air pollutants (NO₂, SO₂, O₃, CO), aerosol particles, urban nighttime light, and digital elevation model to evaluate the non-linear spatial dependence of these variables for the summer (from June to August 2025) and winter (from December 2024 to February 2025) seasons in the US southwest. All multi-resolution inputs were harmonized by projecting to WGS84 and applying a ~11 km fishnet sampling grid commensurate with the coarsest-resolution dataset (Sentinel-5P), ensuring each sample captures a unique pixel value across all layers. AutoML was applied to benchmark learning algorithms, and we found that CatBoost, Extra Trees, LightGBM, HistGradientBoosting, and Random Forest were among the optimal models for predicting LST. After tuning these models using Bayesian optimization, we achieved a mean R² of 0.86 during summer and 0.84 during winter. After developing the hyperparameter-optimized model, explainable AI, e.g., SHAP, was employed to understand the complex nonlinear dynamics and top contributing features. Landcover variables had a more dominant impact on the spatial distribution of summer LST, while winter LST was more influenced by pollutant parameters. Partial Dependency Plot and Accumulated Local Effect were further incorporated to examine the marginal effects of the top-contributing features on spatial LST prediction. By extending the study area to the entire US Southwest, this study effectively



Academic Editors: Jochem Verrelst
and Jorge Vicent

Received: 25 January 2026

Revised: 26 February 2026

Accepted: 27 February 2026

Published: 1 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

captures urban–rural contrasts, climate- and land-cover–dependent pollutant responses, and regional climatic influences. It presents explicit spatial dependencies among LST, pollutants, land cover, topography, and nighttime activity that will aid future researchers and policymakers in effectively developing sustainable thermal planning for urban activities.

Keywords: LST; AutoML; bayesian optimization; explainable AI; Sentinel 5p

1. Introduction

Land surface temperature (LST) is an essential indicator for investigating hydrodynamic conditions, environmental changes, and the Earth’s energy balance. It directly drives the interchange of long-wave radiation and turbulent heat fluxes at the surface–atmosphere interface [1]. Numerous Earth science studies have shown the broad application of LST data. This includes tracking how global warming affects lakes and cryosphere melt [2], measuring the impact of urban heat islands [3], and comprehending the spatial patterns of heatwaves [4]. LST has also been broadly studied in agriculture for insect outbreaks and vector-borne diseases [5]. Since LST has been widely used in climatology and epidemiological studies and as a sustainability indicator, it has been recognized as one of the high-priority parameters of the International Geosphere-Biosphere Program (IGBP) [6].

Over the past few decades, LST has also risen dramatically, along with global atmospheric temperatures. For instance, the global mean LST has increased at the rate of 0.26–0.34 °C per decade between 2001 and 2020 [7]. The trend, however, is inconsistent across socioeconomic and climatological settings, as Wang et al. (2022) [7] reported that the Arctic, in particular, warmed at a rate 2.5–2.8 times the global average. The LST rose by 0.02 °C each year in North America from 2002 to 2018 [8]. Yan et al. (2020) [8] further stated the annual LST trend anomaly in the freezing northern extent (0.12 °C), the west coast from 20°N to 40°N (0.07 °C), and the tropics south of 20°N (0.04 °C). Moreover, Vose et al. (2017) [9] reported that the annual mean surface temperature increased across 95% of the contiguous United States.

In the US, the western region experienced one of the largest shifts in LST over the last century. The mean annual temperature in the southwestern US between 1986 and 2016 was 1.61 °F higher than the baseline average from 1901 to 1960 [9]. Furthermore, throughout the corresponding periods, the coldest single day of the year warmed by 3.99 °F, while the warmest single day warmed by 0.5 °F, with the highest shifts across the US. This region was further categorized as a catastrophic drought commencing at the turn of the century and anomalously high air temperatures in recent decades [10]. Based on integrated climate models (e.g., CMIP5), the difference between the average for late-century (2071–2100) and the average for near-present (1976–2005) in the southwest US is expected to rise between 4.9 and 8.65 °F under RCP 4.5 and RCP 8.5 [11,12]. In such thermal escalations, researchers often report land-use land-cover (LULC) changes as a principal factor.

Anthropogenic LULC changes modify biophysical surface properties and intensify climate change, land degradation, biodiversity loss, and the degradation of ecosystem services [13]. Moreover, converting natural vegetation to agricultural land alters surface roughness, albedo, leaf conductance, and other properties, thereby affecting energy exchanges between the land surface and the atmosphere [14]. The US Southwest is vastly dominated by shrubland and desert scrub, with significant areas of arid grassland and bare soil. However, over the last century, roughly 14% of southwest forests have experienced substantial mortality and structural alterations due to fire and insect outbreaks [15]. Further, shrubs and other woody species have invaded semiarid grasslands at unprece-

dented rates across the Southwest due to elevated CO₂, grazing, and fire suppression, transforming around 50 million hectares of semiarid Southwestern grasslands into woody shrublands [16,17]. Duman et al. (2021) [18] further demonstrated that shrub expansion and conifer mortality in the Southwestern US may raise daytime surface temperature by 1–2 °C. On the contrary, irrigated agriculture in arid land covers can enhance humidity and reduce LST, potentially influencing energy flux during the dry summer months [19].

Notwithstanding, LST is significantly impacted by impervious surfaces and associated anthropogenic activities. Impervious surfaces, particularly anthropogenic built-up areas, usually absorb more incoming shortwave solar radiation and slow nighttime cooling, thereby trapping heat and forming a heat island [20]. Further, cities are rapidly urbanizing, increasing anthropogenic fossil-dependent activities, open burning of garbage on streets or dump sites, automotive emissions, and industrial operations. Such activities are precursors to lethal pollutants, e.g., NO₂, SO₂, O₃, and particulate matter, which have a high potential to influence energy fluxes in a geographic proximity. Several studies demonstrated anomalous relations between pollutants and LST. For instance, Lai & Cheng (2009) [21] found a negative correlation between O₃ and LST in the Taiwan Strait, whereas ref. [22] found a positive correlation between these two variables in the Yangtze River Delta. However, LST had a positive correlation with NO₂, SO₂, and aerosol across several geographies [21–23]. In Bengaluru, Suthar et al. (2024) [24] mentioned seasonal LST correlations; e.g., SO₂ had a significant correlation during summer, while NO₂, CO, and O₃ were more significant during winter. While studies examining the relationship between pollution parameters and LST are evident, they have been limited to specific urban areas and often fail to account for the strong heterogeneity in climatic and topographic conditions.

To determine the exchange of energy between the land surface and the atmosphere, it is essential to sample land surface temperature both spatially and temporally, accounting for the effects of urbanization, atmospheric stability, and land surface characteristics. Even though several studies have incorporated time-series correlations between land cover, socioeconomic factors, or pollutant concentrations with LST, the integrated nonlinearity of all the variables in spatial dimensions remains largely understudied. Several recent studies applied machine learning (ML) models to capture LST dynamics, as they effectively capture nonlinear patterns in high-dimensional spatial data that are often ignored by traditional statistical models. For instance, Suthar et al. (2024) [24] demonstrated high model accuracy ($R^2 > 0.8$) for random forest (RF), artificial neural networks (ANN), support vector regression (SVR), and multiple linear regression (MLR) while predicting LST from pollutant parameters. Other ML models, e.g., Deep Neural Networks (DNNs) and Extreme Gradient Boosting (XGB), were found to be efficient at predicting LST from built-up areas, soil, and vegetation [25]. Badugu et al. (2023) [26] further mentioned the impact of seasonal aerosol, vegetation, and water indexes, road density, and elevation on LST using the Long Short-Term Memory (LSTM) method in urban Tiruchirappalli, India. While the studies incorporated advanced ML methodologies, they often trained the LST on independent categories, e.g., land cover, climatic, or pollutant parameters, and failed to explain the localized impact of individual features.

Here, Geospatial AI (GeoAI) may serve as an ideal learning strategy for understanding how the spatial distribution of land covers and anthropogenic contaminants interacts with LST. GeoAI accelerates the extraction and understanding of complex data, enabling us to resolve planetary challenges and patterns in rapidly growing datasets [27]. Earlier surface temperature studies [1,8,24,26,28], however, focused primarily on a smaller geographical extent, limiting the literature on the underlying climatological distinctness of LST dynamics. GeoAI can further enable us in this circumstance by effectively monitoring and analyzing events, assets, and entities to enable faster response and proactive decision-making.

Moreover, geospatial automation can boost productivity through its intelligent, scalable mapping workflow for remotely sensed data [27].

The existing literature indicates a significant gap in demonstrating the spatial heterogeneity of LST and its major driving components across a broader geographic extent. Therefore, this study aims to integrate land cover parameters (e.g., SAVI, NDBI, NDBSI, MNDWI, WET, and Albedo), atmospheric pollutants (e.g., NO₂, SO₂, O₃, CO, and AI) column concentrations, elevation profile, and nighttime light to evaluate their spatial nonlinearity with surface temperature in the southwestern US. An AutoML pipeline was developed and tested on two datasets from the summer and winter seasons to identify optimal learning models that capture seasonal fluctuations in LST. To improve model learning and lessen overfitting, the hyperparameters of the top five models from the AutoML results were tuned using Bayesian optimization. Lastly, explainable AI (XAI) models, such as SHAP, PDP, and ALE, were used to determine the independent feature importance, localized impacts, direction, and correlation with LST values. To the best of the authors' knowledge, the present paper is the first study to (1) apply an AutoML-driven ensemble benchmarking with Bayesian optimization for LST prediction across the diverse topography of the US Southwest, (2) integrate multi-sensor land cover, atmospheric pollutants, and socioeconomic variables at a regional scale beyond single urban areas, and (3) employ XAI methods including SHAP, PDP, and ALE to explicitly quantify and interpret the nonlinear contributions of individual predictors to seasonal LST dynamics. This research will explicitly elucidate the importance of the selected features and evaluate the performance of machine learning models for geospatial land surface temperature prediction.

2. Materials and Methods

2.1. Study Area

Although there are no official boundaries, the US Southwest is a geographic region of the US that includes New Mexico and Arizona, as well as significant portions of adjacent states, e.g., California, Nevada, Utah, Colorado, and Texas [29]. For the purpose of this study, the US Southwest is operationally defined as the complete state polygons of California, Nevada, Arizona, Utah, Colorado, New Mexico, and Texas, with coastal zones excluded using the NLCD land cover mask [30]. It stretches from the southern border between the US and Mexico to the southern regions of Colorado, Utah, and Nevada (39° N latitude). The region spans multiple Köppen climate classifications, predominantly hot desert and cold semi-arid, with Mediterranean. The region also includes humid continental climates along the California coast and at higher elevations in Colorado [31,32]. The region also experiences persistent elevated atmospheric pressure, rain-shadow effects, and seasonality in jet stream flow [31,33]. The elevation profile of the US Southwest is varied. Colorado and New Mexico have higher elevations (over 1500 m), while the southernmost states have lower elevations (under 20 m) (Figure 1). The most common types of land cover in the study area are shrubland, barren or sparse vegetation, cropland, and urban or built-up areas. Forests are mostly found at higher elevations. The study area spans several EPA Level III ecoregions, including the Mojave Basin, Sonoran Desert, Colorado Plateaus, and Southern Rockies [30,32,34]. Major landscapes include the Basin and Range Province, the Colorado Plateau, the Sonoran and Mojave Deserts, and mountains such as the Rocky Mountains and the Sierra Nevada [34]. To understand the heterogeneity of the LST-influencing features, we incorporated all land areas from the seven states of California, Nevada, Arizona, Utah, Colorado, New Mexico, and Texas.

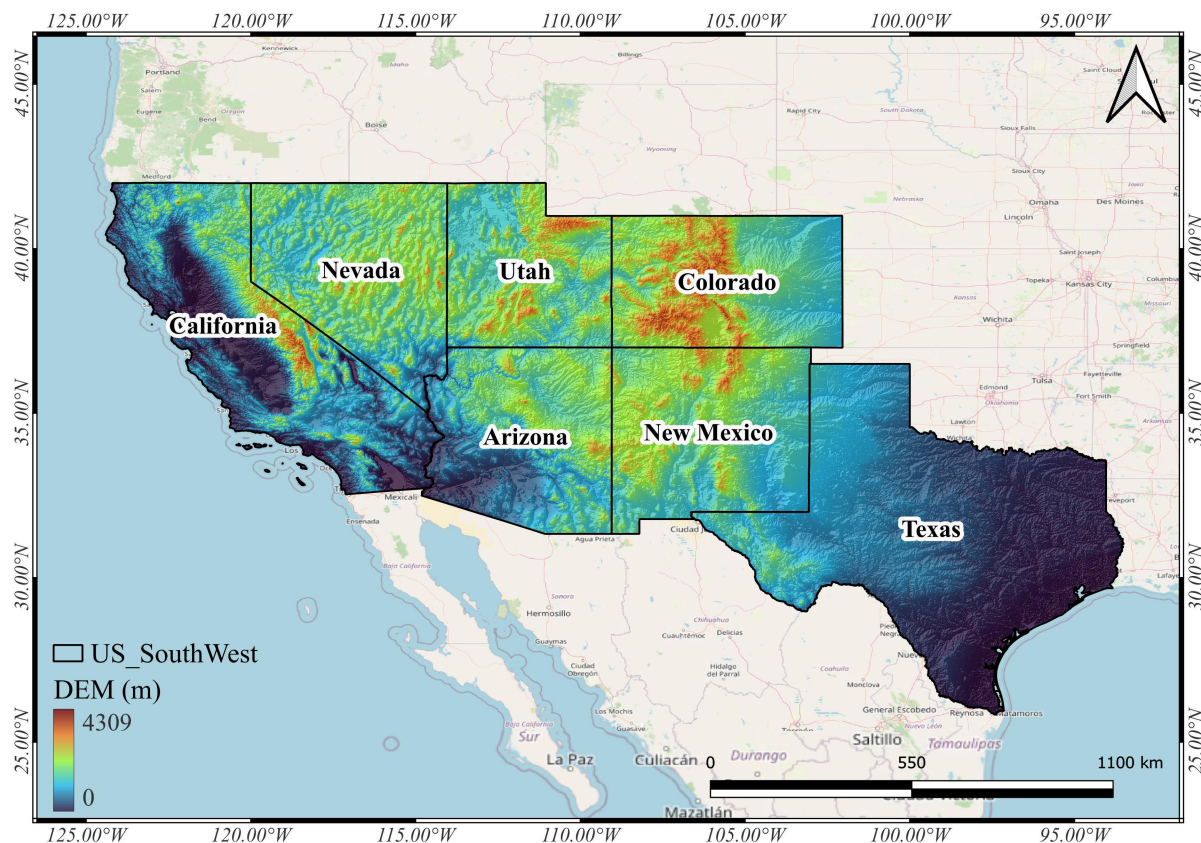


Figure 1. Study area and its digital elevation profile.

2.2. Data

This study incorporated multiple independent variables, including land cover, pollution, topographic, and socioeconomic features, to understand their combined impact on seasonal LST. This comprises NDBI, MNDWI, NDBSI, SAVI, WET, Albedo, NO₂, SO₂, O₃, CO, AI, NTL, and DEM from different satellite instruments for the summer (June, July, and August) and winter (December, January, and February), 2025, using Google Earth Engine.

Land cover variables, e.g., NDBI, MNDWI, NDBSI, SAVI, WET, and Albedo, and the thermal dataset, LST, were collected from Landsat-9 Level 2, Collection 2, Tier 1. This dataset provides atmospherically adjusted surface reflectance and LST from Landsat 9 OLI/TIRS sensors [35]. Collection 2 adds angle coefficients to the data to calculate per-pixel solar and viewing-geometry characteristics [36]. Further, it provides useful per-pixel masks for cloud, shadow, cirrus cloud, and snow using the CFmask algorithm [36]. Consequently, the collection reduces atmospheric noise and potential outliers per pixel, enabling researchers to assess land cover characteristics with greater precision. For this study, 869 images from the summer period and 786 from the winter period were analyzed. Using the QA_PIXEL band, cloud masking was carried out at the pixel level. Bitwise operations were used to mask bits 3 (cloud shadow), 4 (cloud), and 5 (snow). Without further gap-filling, the clean pixel collection was then composited using a mean across the specified seasonal frame.

Sentinel-5P is one of the top-priority and prominent satellite instruments for assessing atmospheric column concentrations of pollution [37–40]. The satellite carries the most advanced TROPOMI (TROPOspheric Monitoring Instrument), which measures UV-visible (270–500 nm), NIR (675–775 nm), and SWIR (2305–2385 nm) spectral bands, enabling more accurate capture of air pollutants. The individual ground pixels are 7.2 km in the longitudinal path and 3.6 km in the perpendicular direction [37,38]. The entire swath diameter is approximately 2600 km, with the exception of short stretches among orbits

that are roughly 0.5° wide at the equator. This allows the instrument to collect data at a very large temporal (daily) scale. For this study, we incorporated level-3 Sentinel-5p data for NO_2 , SO_2 , O_3 , CO , and Aerosol Index (AI) to comprehend their spatial and seasonal impacts on LST.

The elevation profile (DEM) was used as a topographic dataset to understand the influence of landscape characteristics on LST. NASADEM is a near-global elevation dataset with a 30 m spatial resolution, developed by reconstructing Shuttle Radar Topography Mission (SRTM) radar data and integrating it with revised ASTER GDEM elevations. SRTM voids are decreased using sophisticated interferometric decoding methods and greater vertical control of each SRTM image swath using ICESat elevations [41]. We incorporated DEM from the NASADEM instrument to improve the model's efficiency, yielding a more precise elevation profile with lower error pixels. Further, socioeconomic variability is dynamic and often clustered in a specific geographic extent. Proxy datasets, e.g., nighttime light radiance, are ideal to track such socioeconomic changes over the period [39,40]. This will help us to determine whether specific urban clusters exhibit higher LST values than other regions. We incorporated the Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band (DNB) instrument [42,43] to track major urban areas and determine whether their distinct nighttime light has specific impacts on LST.

A list of the selected variables, their corresponding satellite instrument, spatial resolution, and the defined study period is presented in Table 1. To address any temporal mismatches arising from the use of multi-source satellite datasets, all variables were temporally harmonized by calculating mean values over specific seasonal periods (summer and winter of 2025). All satellite-derived outputs were temporally aligned using the same seasonal frames to guarantee consistency between datasets.

Table 1. Description of the dataset.

Dataset	Abbreviations	Unit	Satellite Instrument	Spatial Resolution	Study Period
Land Surface Temperature	LST	°C	Landsat-9	30 m	Summer (June, July, and August 2025) and Winter (December 2024, January, and February 2025)
Normalized Difference Built-up Index	NDBI				
Modified Normalized Difference Water Index	MNDWI	-			
Normalized Difference Built-up and Soil Index	NDBSI				
Soil-Adjusted Vegetation Index	SAVI				
Tasseled Cap Wetness Component	WET				
Surface Albedo	Albedo				
Nitrogen Dioxide	NO ₂	μmol m ⁻²	Sentinel 5p-OFFL L3 Product	11.2 km	
Sulfur Dioxide	SO ₂	μmol m ⁻²			
Ozone	O ₃	mmol m ⁻²			
Carbon Monoxide	CO	mmol m ⁻²			
Aerosol Index	AI	-			
Urban Nighttime Light Radiance	NTL	nW sr ⁻¹ cm ⁻²	VIIRS	463 m	
Digital Elevation Model	DEM	m	NASA-DEM	30 m	

2.3. Sampling Technique

To align the multiple spatial resolutions across Landsat/DEM, VIIRS, and Sentinel-5P, a fishnet sampling method was used. It ensures that no sample points were collected from the same coarse-resolution pixel. This specifically avoids the repetition of identical pollutant values across multiple fine-resolution samples, eliminating false sample-size inflation and reducing spatial dependence bias in training and validation. We incorporated 16,381 and 15,272 samples into the summer and winter models, respectively. Be-

fore sampling, all raster datasets were projected to the same coordinate reference system (WGS84). The fishnet sampling grid was established at about 11 km intervals, commensurate with the coarsest-resolution dataset (Sentinel-5P), ensuring that each sample point records a unique pollutant value. Pixel values were retrieved at each sample point directly from each raster layer using the extraction tool, with no spatial aggregation or resampling. To assess comparability across spatial supports, zonal statistics of Landsat LST were computed within each $\sim 11 \text{ km} \times 11 \text{ km}$ grid cell (consistent with Sentinel-5P resolution). The Intraclass Correlation Coefficient between point-sampled and cell-mean LST was 0.83 (summer) and 0.90 (winter), confirming minimal distortion from scale mismatch (Supplementary Tables S1 and S2) [44]. No buffer zone was applied between blocks, as the sampling interval inherently ensures each point captures a unique coarse-resolution pixel, providing natural spatial separation between neighboring blocks.

The model samples for both seasons were divided into three distinct sets, e.g., training, validation, and testing, as a 64-16-20% ratio using stratified spatial splitting (Figure 2). The samples were separated into bins by creating unique spatial groups. The study area was divided into ~ 200 spatial grid blocks along X and Y coordinates, with samples stratified into five LST quantile groups to ensure balanced distribution across training, validation, and test sets. Spatial blocking was implemented using a 14×14 fishnet grid (number of bins = $\sqrt{200} \approx 14$), yielding 196 candidate cells. After removing empty cells with no sample points, 126 populated spatial blocks remained, which were split into 75 training, 25 validation, and 26 test blocks, respectively. This assures statistical independence of data and removes the spatial dependency bias caused by resolution mismatches. Moreover, such spatial splitting enabled us to better model learning and reduce data leakages within the training sets. Model evaluation metrics, e.g., R^2 , RMSE, and MAE, were computed for each set to assess model accuracy, potential overfitting or underfitting, and prediction error ratios. Model validation and testing were conducted on geographically independent samples collected within the same fishnet grid, maintaining a consistent spatial framework across all datasets.

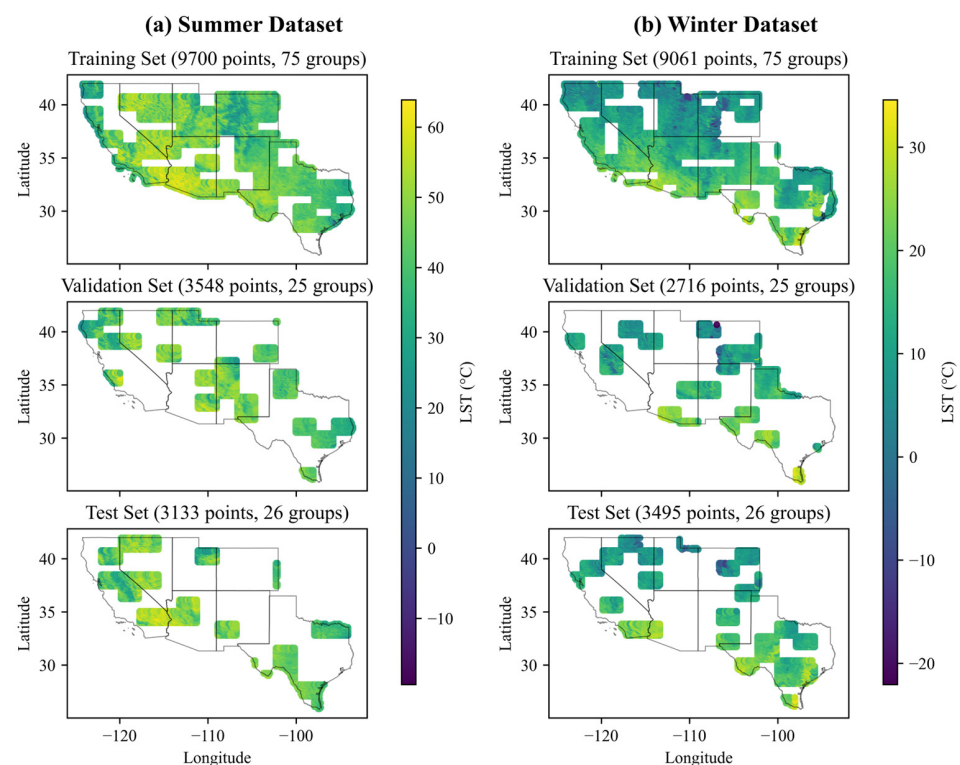


Figure 2. Samples for each of the sets for (a) summer and (b) winter.

2.4. Automated Machine Learning Pipeline

Automated machine learning, or AutoML, streamlines the machine learning pipeline by automating model selection, hyperparameter search, and performance benchmarking, significantly reducing manual effort while still benefiting from expert-guided problem formulation and data preprocessing. AutoML may streamline manual processes and allow domain experts to build and maintain ML pipelines without ML or statistical knowledge [45]. It uses ML learning tool customization and supervision to self-adapt to difficulties [46]. A comprehensive AutoML system may create an intuitive end-to-end ML pipeline system by dynamically combining different methodologies. Thus, for this study, optimal models for predicting LST were identified using PyCaret's AutoML based on multilayer raster extractions of socioeconomic, environmental, and topographic factors. PyCaret is an open-source end-to-end machine learning and model management platform that streamlines machine learning operations [47]. A total of 29 models were analyzed with predefined hyperparameters to comprehend the predictive power of major baseline models.

2.5. Hyperparameter Tuning from Bayesian Optimization

After the AutoML pipeline, we found that CatBoost, Extra Trees, LightGBM, HistGradientBoosting, and Random Forest were the top-performing models for both seasons in predicting LST from the selected independent variables. We further optimized the hyperparameters of these models with Bayesian optimization to reduce overfitting and minimize prediction errors [48]. Bayesian optimization generates a preliminary distribution that most accurately characterizes the function being optimized. The probability increases with more observations, and the approach becomes more confident about which areas of the hyperparameter space are being examined [48]. Table 2 presents the selected hyperparameters, their boundaries, and the observed optimal values for each season in this study. Bayesian hyperparameter optimization was performed using BayesSearchCV with negative mean squared error as the objective metric over 30 iterations per model, applying 3-fold cross-validation on the training set only and using the validation set solely for early stopping; spatial/group structure was preserved throughout by ensuring hyperparameter search did not cross block boundaries. It is critical to emphasize that Bayesian optimization was employed solely as a hyperparameter refining strategy. The final models are deterministic and generate predictions for each input, rather than probabilistic outputs or uncertainty estimates.

Table 2. Hyperparameters searching boundary and best hyperparameters for the selected top models.

Models	Hyperparameters	Parameters Boundary	Optimal Parameters	
			Summer	Winter
CatBoosting	bagging_temperature	0.0, 1.0	1	0.7458
	border_count	32, 255	255	152
	depth	4, 12	7	8
	iterations	500, 2000	2000	1588
	l2_leaf_reg	1, 10	1.0387	10
	leaf_estimation_iterations	1, 10	10	8
	learning_rate	0.01, 0.3	0.0327	0.0279
	random_strength	0.1, 2.0	0.7649	0.9806
Extra Trees	bootstrap	True, False	False	False
	max_depth	5, 50	50	50
	max_features	0.1, 1.0	1	1
	min_samples_leaf	1, 10	1	1
	min_samples_split	2, 20	2	2
	n_estimators	100, 500	500	500

Table 2. *Cont.*

Models	Hyperparameters	Parameters Boundary	Optimal Parameters	
			Summer	Winter
LightGBM	colsample_bytree	0.5, 1.0	0.7182	0.9116
	learning_rate	0.01, 0.3	0.0420	0.01
	max_depth	3, 20	18	16
	min_child_samples	5, 100	5	5
	n_estimators	100, 1000	1000	814
	num_leaves	20, 200	20	200
	reg_alpha	0.001, 10.0	0.0011	0.0028
	reg_lambda	0.001, 10.0	0.001	0.3127
	subsample	0.5, 1.0	0.7061	0.5
HistGradientBoosting	l2_regularization	0.0, 10.0	8.4823	10
	learning_rate	0.01, 0.3	0.0433	0.0421
	max_bins	100, 255	177	100
	max_depth	3, 15	13	13
	max_iter	100, 500	480	500
	max_leaf_nodes	15, 100	78	100
	min_samples_leaf	10, 50	28	10
Random Forest	bootstrap	True, False	True	True
	max_depth	5, 50	47	50
	max_features	0.1, 1.0	0.7695	0.7989
	min_samples_leaf	1, 10	1	1
	min_samples_split	2, 20	4	14
	n_estimators	50, 250	249	119

2.6. Feature Importance and Direction Analysis

After developing hyperparameter-optimized ensemble models, we finally comprehended the individual feature importance using popular explainable AI (XAI) methods, e.g., SHAP, PDP, and ALE. SHAP (Shapley Additive exPlanations) is a method for assessing the impact of each feature on a machine learning model's prediction [49]. All feature combinations and the predictions they yield are examined to calculate a feature's Shapley value in a machine learning model. The ideal Shapley values found in game theory serve as the foundation for SHAP. The value of the feature is then determined by averaging the difference in the prediction results across all potential feature combinations that comprise it [49]. SHAP values in a model particularly define the features' association and direction with the dependent features and how they interact with the other confounding variables.

PDP (Partial Dependence Plot) and ALE (Accumulated Local Effects) are two model-agnostic XAI approaches that visualize how a feature affects a model's predictions [50,51]. PDP helps to understand the relationship between a feature and the target variable by showing the feature's marginal influence on the expected outcome. On the other hand, ALE focuses more on differences in predictions than on the mean. While ALE plots are a more efficient and impartial alternative to PDP, both are effective tools for examining how factors affect model predictions within localized features [50,51]. While previous studies have developed predictive models, the global and local nonlinearity of LST has not been widely studied. We used SHAP to explain the contributions of individual predictions by assigning a feature to each prediction, while PDP/ALE depicted the average marginal impact of a feature on the model's output across the dataset. A generalized methodological workflow is presented in Figure 3.

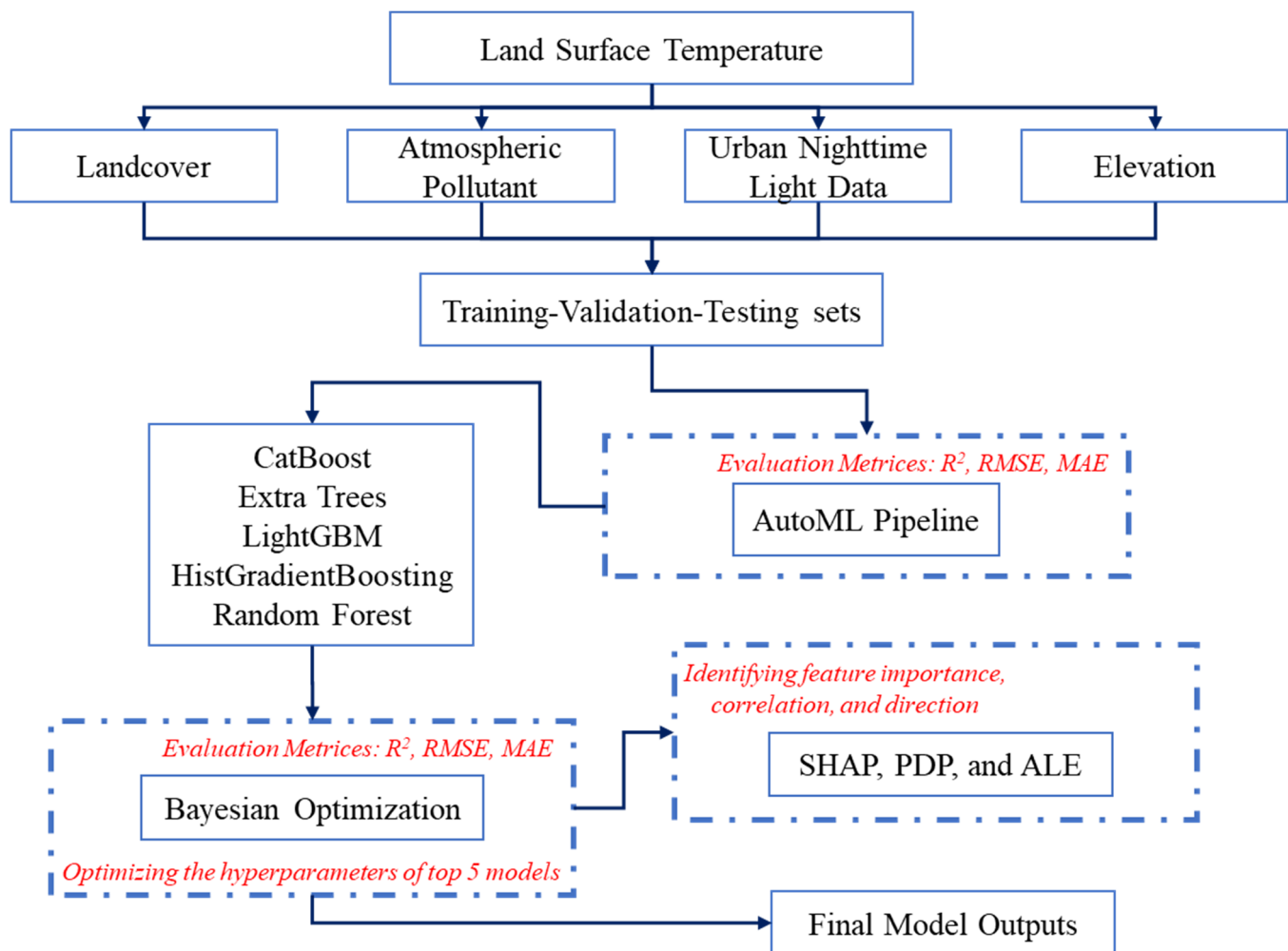


Figure 3. Methodological flowchart.

3. Results

3.1. Spatiotemporal Distribution of the Selected Thermal, Landcover, and Pollution Parameters

The mean spatiotemporal distribution of LST for summer and winter 2025 shows a clear distinction, with elevated temperatures at higher latitudes (37–42°N) during summer and little variation in mean temperature at lower latitudes (28–32°N) (Figure 4).

The mean summer temperature across the study area was 44.02 ± 8.26 °C, with states such as Nevada, Arizona, Utah, and New Mexico reporting LST > 42 °C. In California, the Great Basin and Sierra Escarpment exhibit clear, heterogeneous temperature variation relative to the surroundings, making it a unique temperature region with an LST ranging from 45 to 60 °C. On the contrary, a clear negative correlation between LST and elevation is evident, as LST in the Sierra Nevada and the Rockies drops by 18–25 °C. Moreover, the coasts of California and Texas also had a similar low mean LST (ranging from 20 to 35 °C) during summer. During winter, the mean LST drops to 13.99 ± 6.89 °C, and the highest variations were observed for Utah and Colorado, where the LST ranges between 0 and 10 °C.

In contrast, Arizona and New Mexico had mean LSTs ranging from 20 to 35 °C during winter. The variation in LST was also relatively low along the California and Texas coasts (20–30 °C).

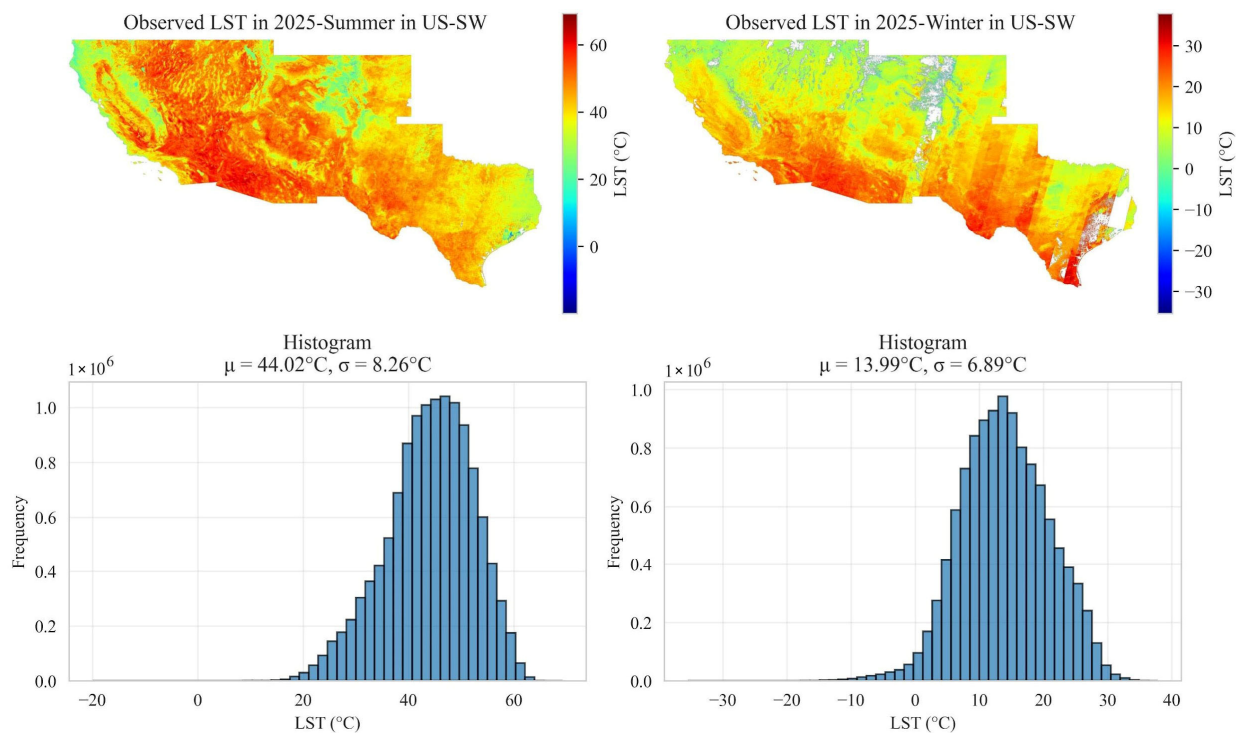


Figure 4. Spatiotemporal patterns of land surface temperature across the US Southwest.

The selected land cover features, e.g., SAVI, NDBSI, NDBI, MNDWI, WET, and Albedo at US-south, consisted of sharp longitudinal gradients and high seasonal patterns (Figure 5). One of the significant changes was found for SAVI, WET, and albedo owing to phenological cycles and moisture availability in the region. The SAVI in regions such as East Texas, Northern California, and parts of central Colorado ranged from 0.35 to 0.4 and decreased significantly during winter. This pattern was also evident for the WET and the Albedo index. Albedo was, however, almost constant (0.18–0.23) in eastern Colorado and New Mexico and northern Texas, where steppe landforms were dominant. These regions also showed very little seasonal contrast in the NDBSI and MNDWI indexes. MNDWI and WET had similar spatial patterns at the mountain ice caps. The surface moisture index (WET) increased slightly (ranging from -0.3 to -0.2 in summer to from -0.2 to -0.1) on the high Colorado Plateau and decreased in East Texas (from almost 0.05 during summer to -0.1 in winter).

The built-up index-NDBI (Figure 5) presents some proxy outcomes from SAVI, NDBSI, and WET. While major urban areas in the US Southwest were distinguishable for the summer NDBI, massive vegetation loss and constant bare soil landforms caused “masking failure” for the winter NDBI. Therefore, the urban structure was assessed using night-time light (NTL) from the VIIRS instrument (Figure 6). The spatiotemporal distribution of NTL clearly distinguishes the urban centers, with no major shifts or urban expansions during 2–25. NTL further demonstrates a higher mean nighttime light radiance during summer ($1.69 \pm 9.8 \text{ nW sr}^{-1} \text{ cm}^{-2}$) than during winter ($1.49 \pm 10.17 \text{ nW sr}^{-1} \text{ cm}^{-2}$). Such an increase in NTL during summer can be attributed to increased human activity, such as prolonged nighttime outdoor activities, greater tourism, and longer operating hours of commercial businesses [52,53]. Urban centers in California, Texas, Nevada, Colorado, and Arizona, where $\text{NTL} > 350 \text{ nW sr}^{-1} \text{ cm}^{-2}$, witnessed a mean $\text{LST} > 45^\circ\text{C}$ during summer 2025.

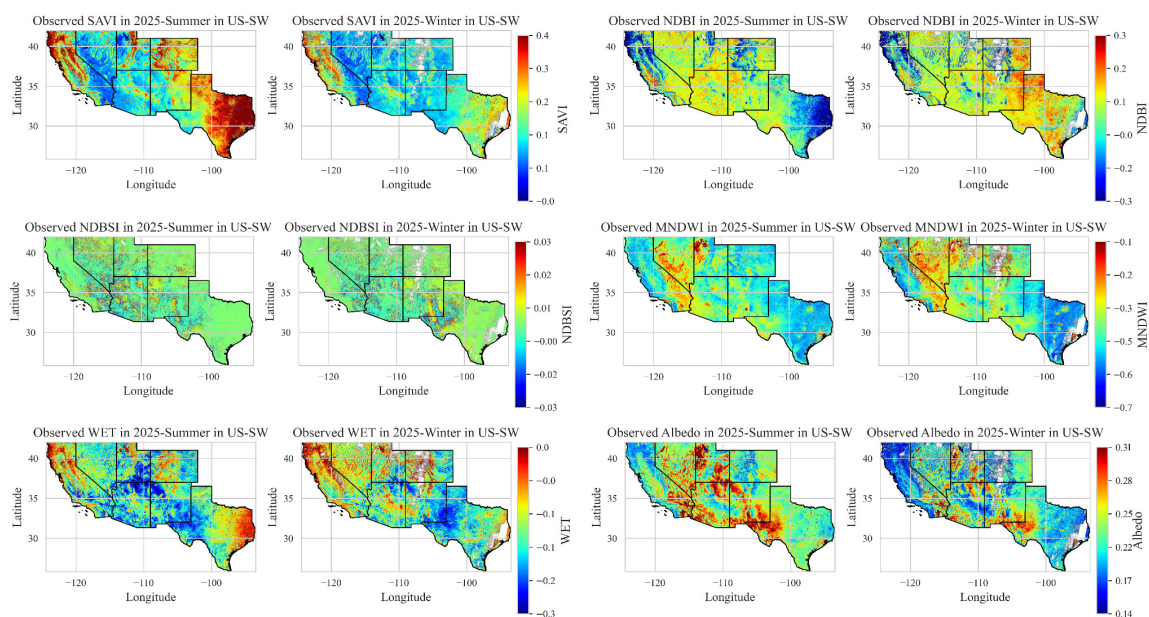


Figure 5. Regional evolution of selected land-cover indicators across the US Southwest.

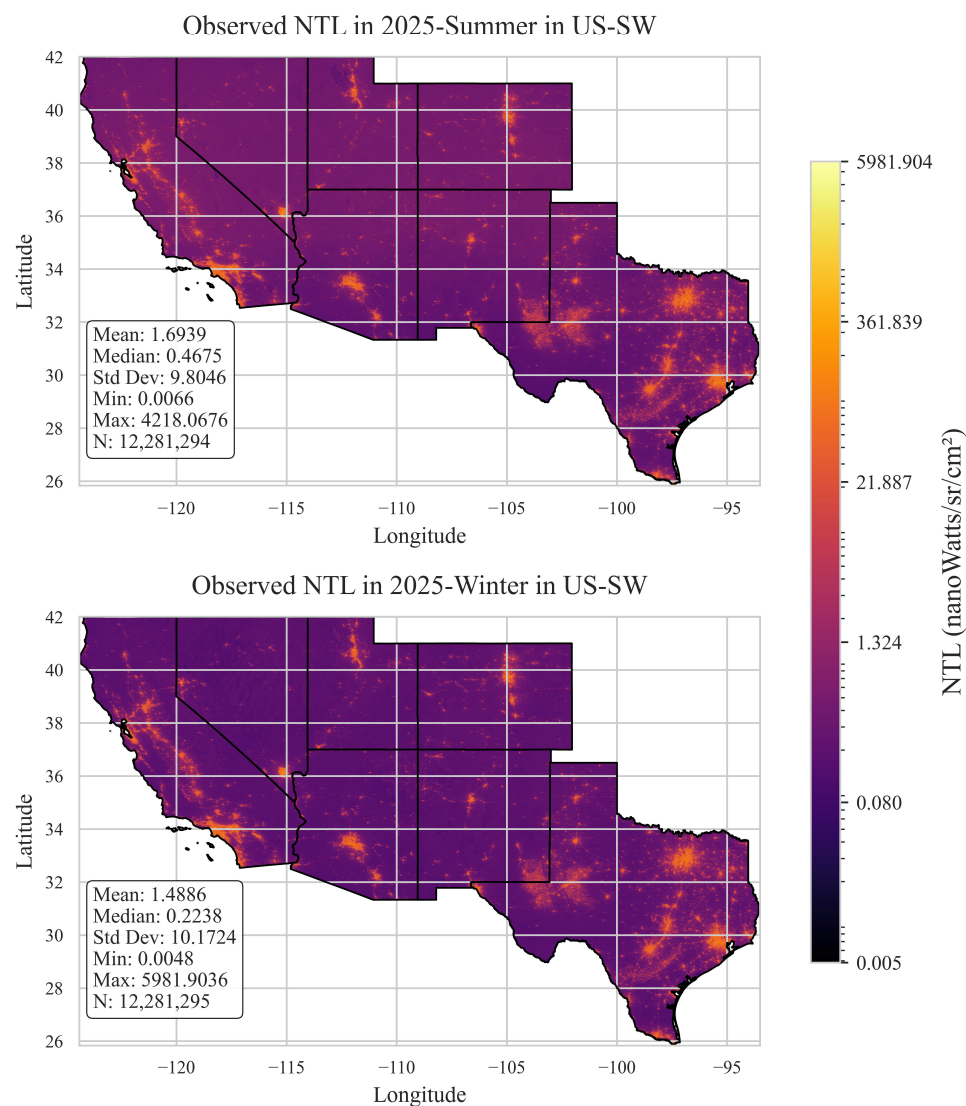


Figure 6. Spatial and temporal dynamics of urban nighttime light radiance in the US Southwest.

Figure 7 presents the vertically integrated column density of the selected pollution parameters, e.g., NO_2 , SO_2 , O_3 , CO , and AI. Seasonal and geographical variations are significant with high pollution concentrations near the urban centers. NO_2 was relatively high ($\text{NO}_2 > 75 \mu\text{mol m}^{-2}$) during both seasons near the urban centers from the NTL image (Figure 6), while the column concentration reduced slightly ($\text{NO}_2 < 42 \mu\text{mol m}^{-2}$) at the northern latitudes during winter. The SO_2 column showed the most drastic change throughout the seasons: during summer, the SO_2 mean ranged from 45 to $75 \mu\text{mol m}^{-2}$ across the US Southwest, but it skyrocketed to $>160 \mu\text{mol m}^{-2}$ during winter. Similar seasonal changes were also evident in the O_3 column, with summer levels ranging from 135 to 140 mmol m^{-2} across the study area. On the contrary, during winter, the mean O_3 column was very high ($>145 \text{ mmol m}^{-2}$) at the upper latitude and plummeted ($<130 \text{ mmol m}^{-2}$) at the lower latitudes.

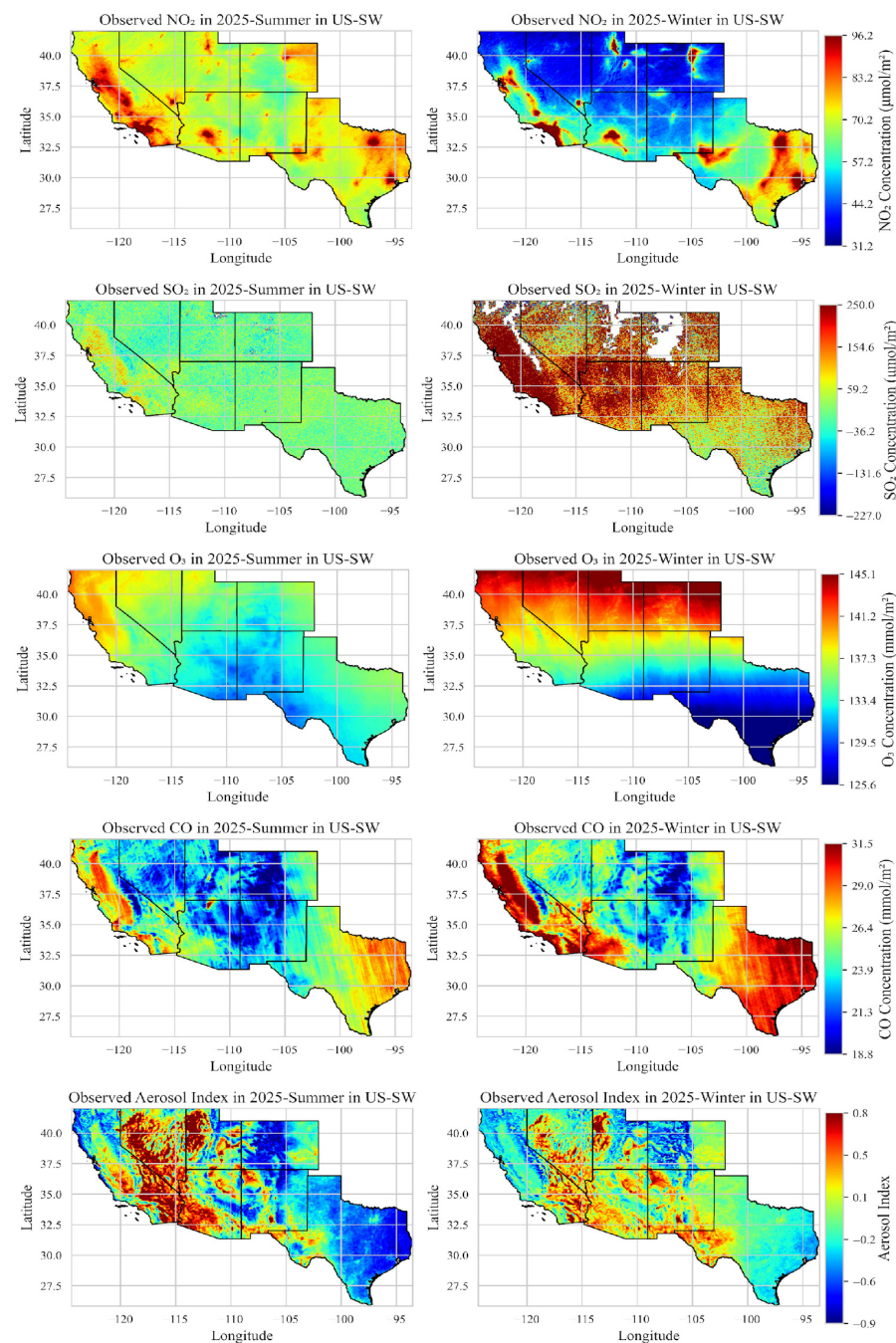


Figure 7. Spatiotemporal variability of key atmospheric pollutant concentrations over the US Southwest.

The CO column was high ($>30 \text{ mmol m}^{-2}$) on the California coast and in the Central Valley in both seasons and further elevated in eastern Texas during winter. It also increased slightly (from 21 mmol m^{-2} to 26 mmol m^{-2}) in parts of Nevada and Utah. During summer, the aerosol particles were comparatively high ($\text{AI} > 0.75$) in the east of the Sierra Nevada, some parts near the Rockies, and in southern Arizona. On the contrary, it was very low (ranging between -0.6 and -0.8) in eastern Texas. However, during winter, the aerosol level plummeted in the high AI zones of summer and was relatively balanced (ranging from -0.2 to 0.1) across the study area.

3.2. Benchmarking the Hyperparameter-Tuned Ensemble Model

An AutoML pipeline was developed to comprehend the model performance of LST learning from the selected variables for both seasons (Table 3). The coefficient of determination for summer was highest for the CatBoost model ($R^2 = 0.875$), while in winter, the Extra Trees had the highest R^2 (0.848). For summer, CatBoost was followed by Extra Trees, LightGBM, and Hist Gradient Boosting with RMSE and MAE ranging between $2.9\text{--}3.1 \text{ }^\circ\text{C}$ and $2.1\text{--}2.16 \text{ }^\circ\text{C}$, respectively. On the contrary, in winter, the model performance slightly deteriorates compared to summer, with top models, e.g., CatBoost, LightGBM, and Random Forest, showing an MAE of $2.12\text{--}2.18 \text{ }^\circ\text{C}$.

Table 3. Output from the AutoML pipeline on the test set.

Model	Summer			Winter		
	R^2	RMSE	MAE	R^2	RMSE	MAE
AdaBoost	0.738	4.227	3.296	0.684	3.874	3.043
Bagging Regressor	0.840	3.311	2.347	0.817	2.944	2.183
Bayesian Ridge	0.760	4.046	2.949	0.616	4.268	3.275
CatBoost	0.875	2.928	2.080	0.834	2.803	2.097
Decision Tree	0.710	4.452	3.236	0.686	3.863	2.841
Dummy Regressor	0.000	8.265	6.548	0.000	6.889	5.596
Elastic Net	0.665	4.782	3.503	0.553	4.603	3.610
Extra Tree (Single)	0.708	4.469	3.246	0.628	4.201	3.049
Extra Trees	0.866	3.021	2.136	0.848	2.687	1.971
Gaussian Process	-0.010	8.306	5.180	0.112	6.491	4.633
Gradient Boosting	0.841	3.296	2.393	0.768	3.315	2.501
Hist Gradient Boosting	0.860	3.096	2.193	0.819	2.928	2.191
Huber Regressor	0.756	4.083	2.930	0.613	4.286	3.270
K-Nearest Neighbors	0.828	3.425	2.461	0.777	3.250	2.380
Lasso Regression	0.671	4.743	3.504	0.556	4.589	3.589
LightGBM	0.860	3.097	2.177	0.822	2.903	2.177
Linear Regression	0.760	4.046	2.949	0.616	4.268	3.275
Linear SVR	0.756	4.086	2.930	0.609	4.310	3.280
MLP Regressor	0.858	3.115	2.244	0.777	3.257	2.436
Nu-SVR	0.821	3.497	2.428	0.720	3.644	2.677
Orthogonal Matching Pursuit	0.466	6.037	4.587	0.374	5.450	4.220
Passive Aggressive	0.634	5.002	3.671	0.394	5.365	3.901
Random Forest	0.854	3.153	2.225	0.832	2.823	2.074
RANSAC Regressor	-8.729	25.779	5.428	-36.834	42.374	6.755
Ridge Regression	0.760	4.046	2.949	0.616	4.268	3.275
SGD Regressor	0.760	4.051	2.940	-2.520	12.925	3.648
Support Vector Regression	0.824	3.472	2.403	0.723	3.627	2.656
Theil-Sen Regressor	0.686	4.630	3.105	-1.432	10.744	4.462
XGBoost	0.858	3.112	2.220	0.809	3.014	2.260

In the AutoML output, hyperparameters for CatBoost, Extra Trees, LightGBM, Hist-GradientBoosting, and Random Forest were tuned using Bayesian Optimization to improve model performance and generalization (Table 4). The model performance for each set (training-validation-testing) shows a high model efficiency with a high R^2 ($>83\%$), with a low RMSE ($3.1\text{ }^{\circ}\text{C}$) and MAE ($2.19\text{ }^{\circ}\text{C}$).

Table 4. Model output from Bayesian optimization-based tuned models.

Models	Dataset	Summer			Winter		
		R^2	RMSE	MAE	R^2	RMSE	MAE
CatBoost	Train	0.9695	1.436	1.1107	0.9526	1.4954	1.1371
	Validation	0.893	2.7282	2.0513	0.8402	2.7214	2.0091
	Test	0.8769	2.8992	2.0624	0.8438	2.7228	2.0231
Extra Trees	Train	0.927	2.13535	1.6273	0.90355	2.04635	1.5188
	Validation	0.8845	2.8347	2.1439	0.8545	2.5973	1.9005
	Test	0.8669	3.0156	2.13	0.8489	2.6777	1.962
LightGBM	Train	0.958	1.684	1.3091	0.9804	0.9612	0.7432
	Validation	0.8852	2.8265	2.1239	0.8471	2.6618	1.9696
	Test	0.8675	3.0085	2.1294	0.8404	2.7519	2.0354
HistGradientBoosting	Train	0.9689	1.4497	1.0941	0.9727	1.1346	0.8566
	Validation	0.8855	2.8224	2.1283	0.839	2.7318	2.0171
	Test	0.8649	3.0376	2.1219	0.8362	2.7878	2.0528
Random Forest	Train	0.9798	1.1692	0.8518	0.9449	1.6115	1.2106
	Validation	0.8765	2.9315	2.1891	0.8392	2.7301	2.0178
	Test	0.8585	3.109	2.1893	0.8323	2.8214	2.0784

After tuning the hyperparameters, CatBoost's model performance slightly increased. The R^2 was still the highest (0.876) in the test set, along with a low RMSE ($2.8\text{ }^{\circ}\text{C}$) and MAE ($2.06\text{ }^{\circ}\text{C}$) in summer. On the contrary, for winter, the Extra Tree regressor had the highest model performance with an R^2 of 0.848, followed by CatBoost ($R^2 = 0.844$) and LightGBM ($R^2 = 0.840$). Between the selected top models, LightGBM and HistGradientBoosting benefited most from Bayesian optimization, with RMSE reductions of 2.8% and 1.9% in summer and 5.2% and 4.8% in winter, respectively. Random Forest had the lowest model performance among the hyperparameter-tuned models during the selected study period. While the R^2 for the summer models was higher than for the winter models, the RMSE and MAE were slightly lower in winter, reflecting lower variability. Notwithstanding, Extra Trees and Random Forest showed minimal gains, suggesting they were already near-optimal at baseline. Moran's I computed on block-level mean residuals showed no significant spatial autocorrelation in the training and test sets and significant ($p < 0.05$) but very low (Moran's I < 0.2) on the validation set for both seasons, indicating that the spatial blocking strategy effectively minimized leakage at the chosen sampling scale (Supplementary Table S3).

The scatterplot matrix (Figure 8) shows the degree of scatter for the selected ensemble models before and after hyperparameter tuning for both seasons. For summer (Figure 8a), several values are consistently underpredicted by all models, but this underprediction reduced slightly after the model hyperparameters were tuned. For winter, on the other hand, values below $0\text{ }^{\circ}\text{C}$ were closest to the linear fit line. While Bayesian optimization reduced proximity, there were some under- and over-predictions, more than in the summer models.

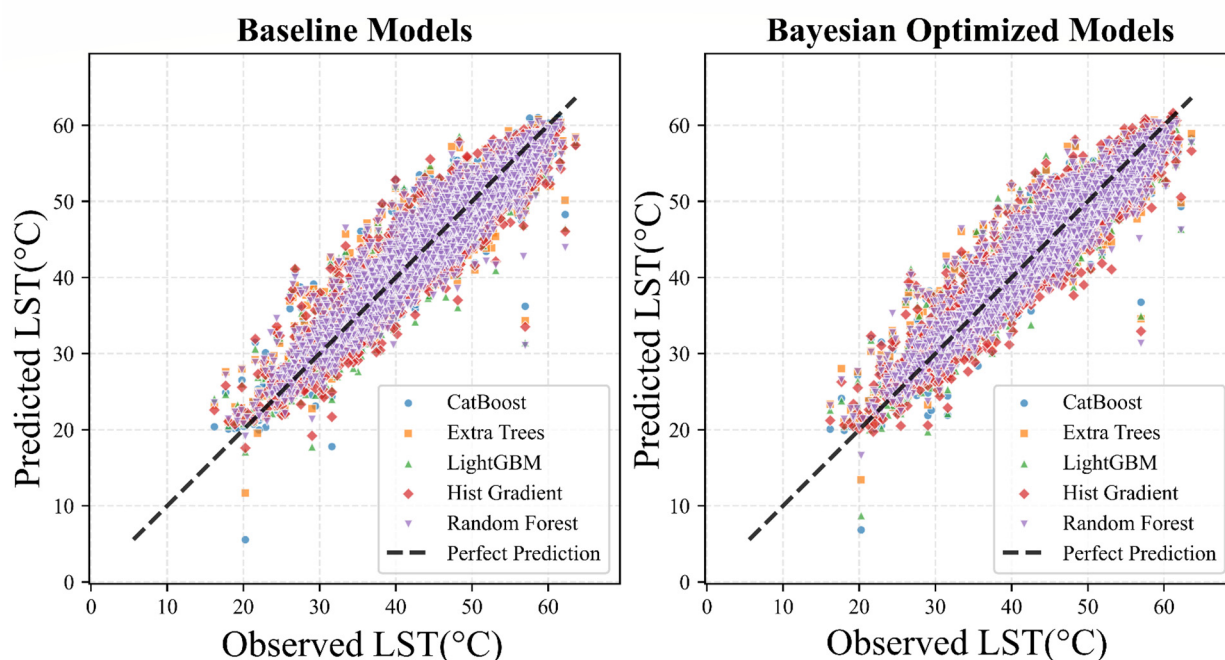
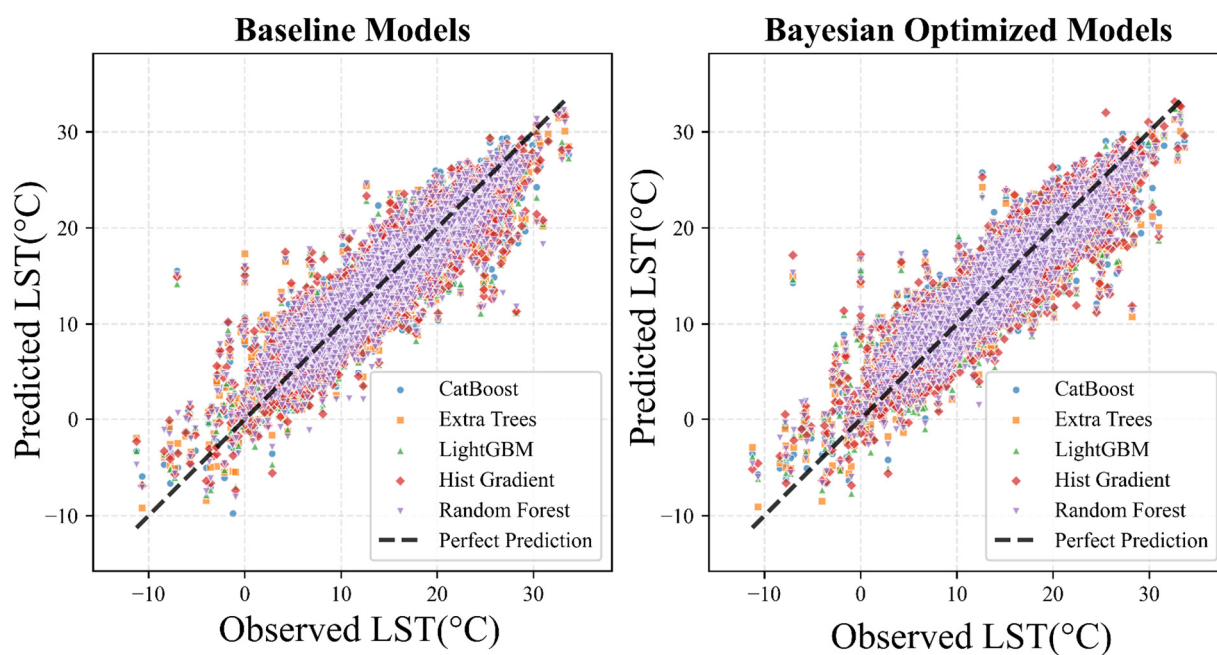
(a) Scatter Plots for prediction in Summer**(b) Scatter Plots for prediction in Winter**

Figure 8. Scatterplot matrix on the test set before and after hyperparameter tuning for the (a) summer and (b) winter.

3.3. Identifying the Relative Feature Importance

In the SHAP analysis, a clear distinction of top-contributing features was significant between both seasonal models (Figures 9 and 10). SAVI, DEM, NDBI, and WET were the top features to predict summer LST, while winter LST prediction was dominated by O_3 , Aerosol, DEM, and WET.

For the summer LST prediction, SAVI had the highest SHAP importance for CatBoost, Extra Trees, and LGB, followed by DEM (Figure 9). Both SAVI and DEM's high feature values were associated with negative SHAP values. Other top features, e.g., Aerosol and

NDBI's high feature values, were associated with positive SHAP values, indicating positive significance in the models. Here, SO_2 , NTL, and NO_2 constantly had the lowest feature importance for all the tuned models, while NTL and NO_2 had a slight positive correlation for LST prediction.

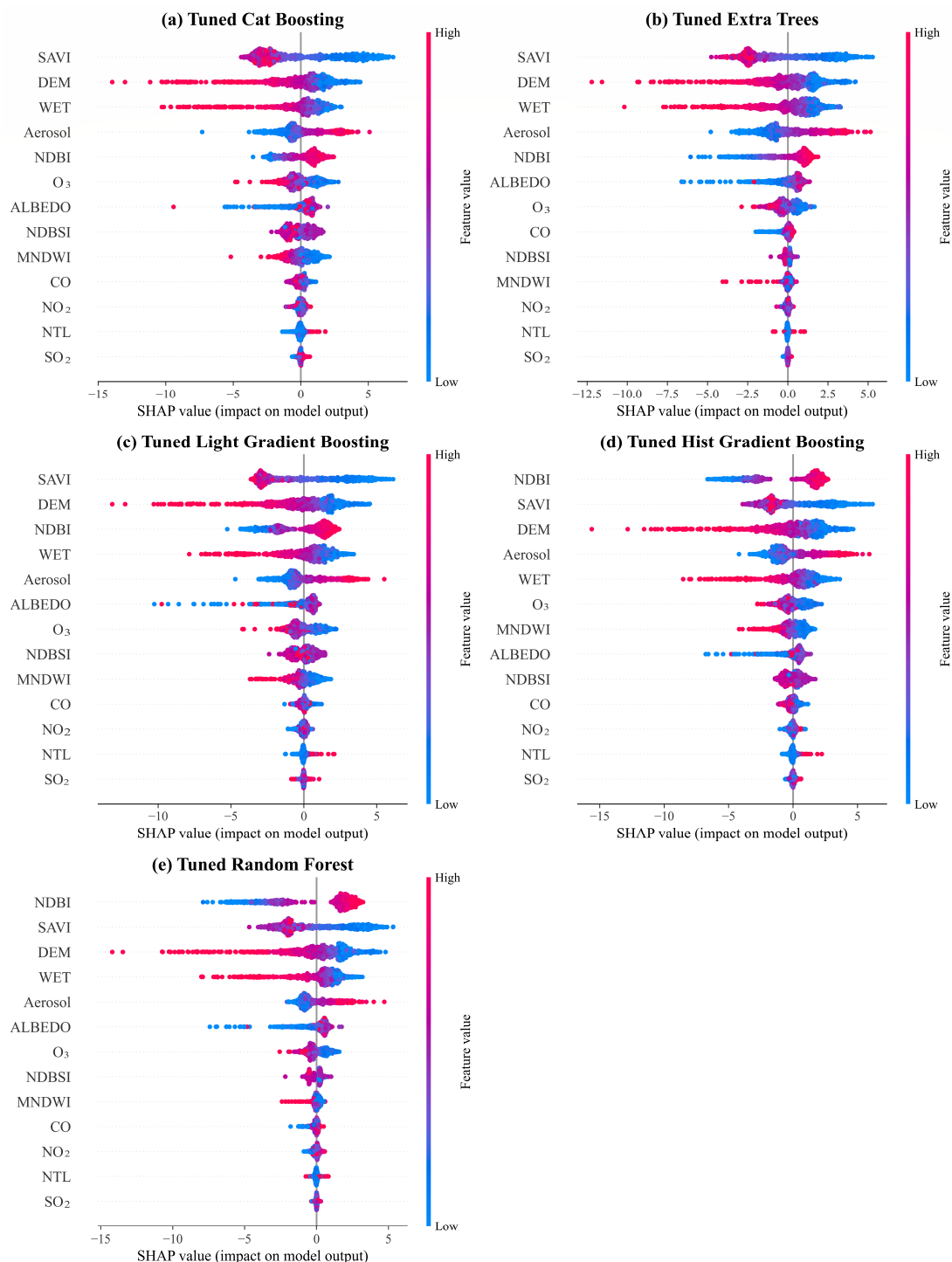


Figure 9. SHAP analysis plot of the features after training in the Bayesian optimized models for the summer dataset.

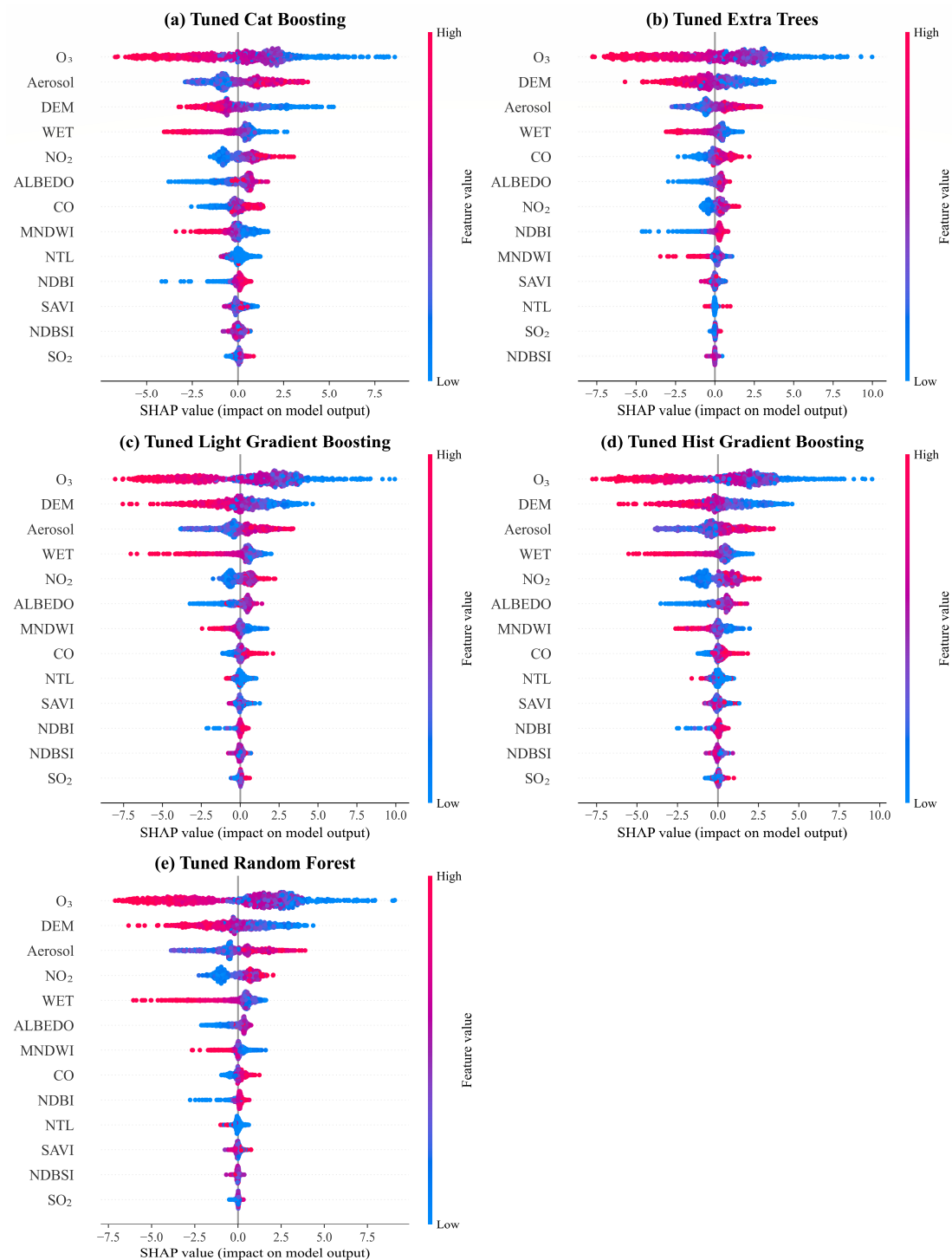


Figure 10. SHAP analysis plot of the features after training in the Bayesian optimized models for the winter dataset.

On the other hand, for the winter LST model, O₃ and DEM were the top two features across all models, except CatBoost. O₃ and DEM's high feature values were associated with negative SHAP. Other pollution parameters, e.g., aerosol, had high feature importance, while NO₂ and CO had moderate feature importance and showed a positive association with LST prediction. While SO₂ again had the lowest feature importance, SHAP values for SAVI and NDBI were low for the winter LST prediction. For winter, NTL's feature increased a bit, but the high NTL values were associated with negative SHAP, an inverse relation from the summer LST model.

Figures 11 and 12 present the PDP and ALE plots, which demonstrate the individual influence of the top five features on the model's average prediction. For the summer LST prediction, it is evident that DEM had the only steep trend (negative), indicating that elevation consistently influences LST across its entire observed range (Figure 9). Other features, e.g., SAVI, WET, MNDWI, and O₃, also had a downward trend, indicating a negative correlation with LST. For NDBI and Aerosol, LST inclines at from 0.0 to 0.2 and from -0.5 to 1.5 , respectively, indicating that high anthropogenic built-up and particulate materials are associated with high LST values. Albedo had the only unique relationship: Albedo < 0.2 had a positive impact on LST, while Albedo > 0.2 had a negative impact on LST.

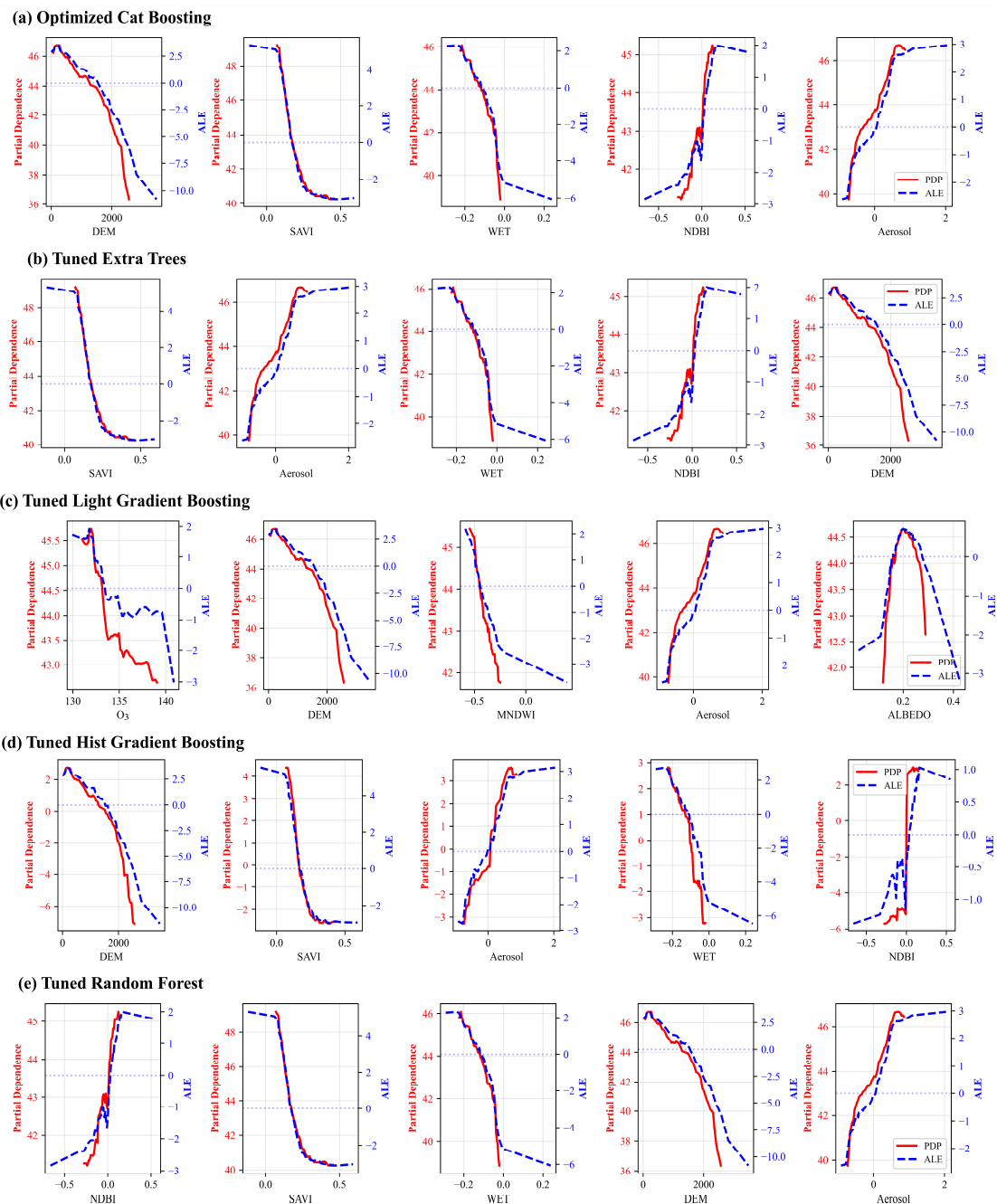


Figure 11. PDP–ALE evaluation of global and localized effects of the five dominant features identified by the tuned models on the summer dataset.

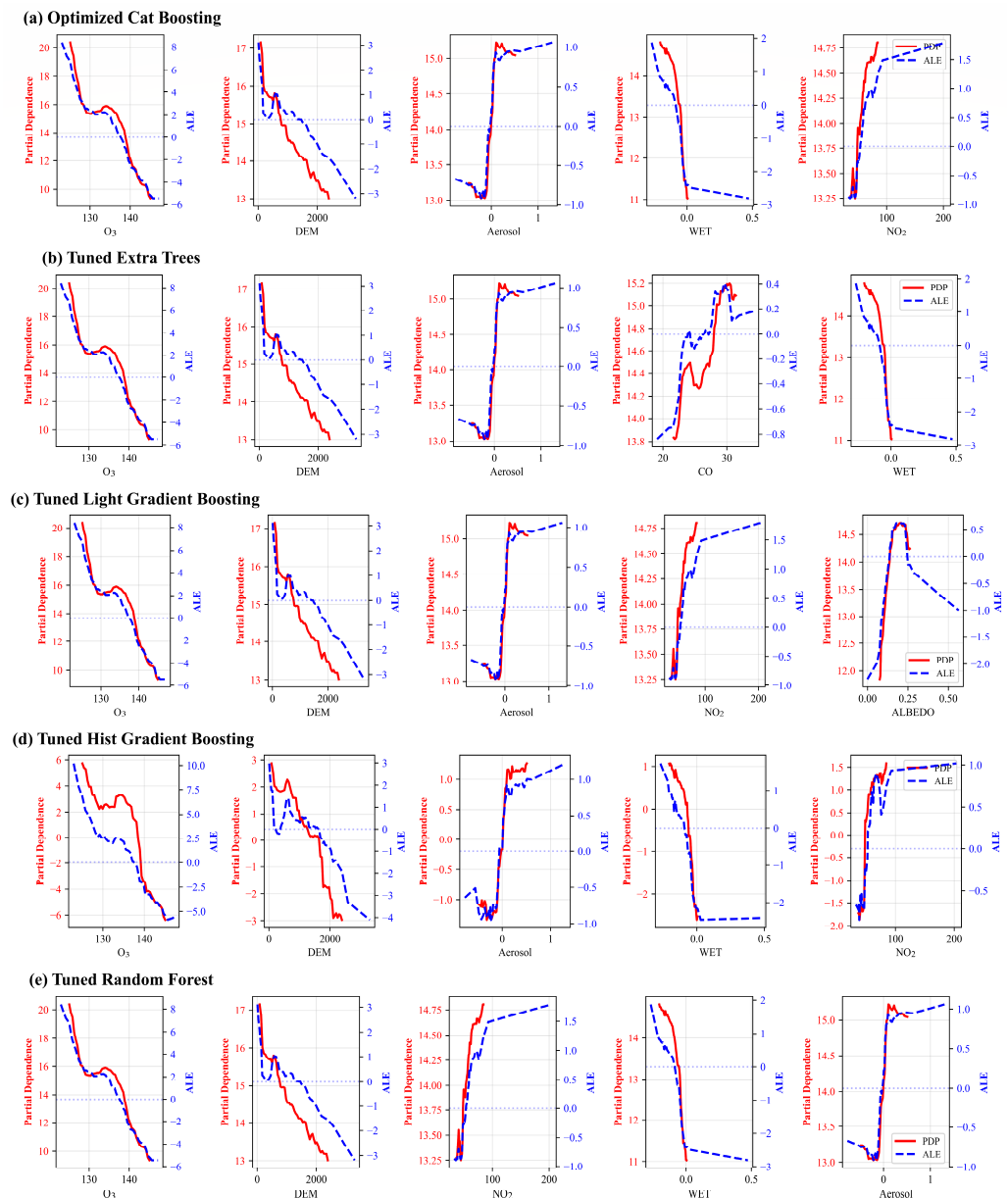


Figure 12. PDP–ALE evaluation of global and localized effects of the five dominant features identified by the tuned models on the winter dataset.

The winter LST prediction was more dominated by the elevation and pollution parameters than by the land cover variables. The PDP and ALE line shows that O_3 and DEM were consistently the most significant features, followed by NO_2 , aerosol, and CO. O_3 and DEM showed a steep decline, with a flattening zone at 15–16 °C. On the other hand, the highest positive slope was observed for NO_2 (within the range of 100 $\mu\text{mol m}^{-2}$) and then flattened. Albedo also showed an inverse trigger point, similar to the summer LST prediction, at an albedo value of 0.25. CO, however, showed a general upward trend, whereas the PDP and ALE both plummeted for CO values of 25 and 32 mmol m^{-2} . Aerosol showed a positive trend and was indifferent to summer LST prediction.

4. Discussion

While the US Southwest is among the hottest regions in the US, recent anthropogenic and climatological changes are further elevating surface temperatures there. From our study, it is evident that the majority of the study area experiences elevated LST (35–55 °C

during summer and 5–25 °C during winter), with significant seasonal fluctuations. Several spatial and temporal studies on the US Southwest or parts of it, e.g., the Mojave & Sonoran Deserts or urban areas, have demonstrated increases in extreme heatwaves and urban heat island effects [8,28,54,55]. Yan et al. (2020) [8] identified the US West Coast as one of the major LST anomaly areas and found a significant positive correlation between LST and El Niño–Southern Oscillation. It was further validated by strong warm air circulation from the Atlantic, which caused major heatwaves in 2023 [55]. From our study, the high mean LST (>40 °C) during summer is evident, indicating the strong climatological impacts of the tropical Atlantic or Gulf of Mexico as one of several factors (Figure 4).

Recent studies further demonstrated pronounced LULC changes in the region, with a shift towards industrial agriculture practices or urban growth. Humans in this region began developing large-scale agriculture and impervious infrastructure, potentially altering natural vegetation structure and adding additional pressure on resources [56–58]. We found NDBI or NTL as one of the dominant factors in LST increase (Figures 5 and 6). Surface temperature cooling agents, e.g., MNDWI or WET, exerted a strong negative influence (Figures 6, 11 and 12). A reduction of water bodies can significantly deteriorate the overall climatic and energy balance. Moreover, researchers found that shrub and woody species have been replaced by semi-arid grasslands as a response to changes in carbon flux and grazing [55,56,58]. Precursor pollutants, e.g., NO₂ and CO, and aerosol particles were more clustered around high LST hotspots and have a positive correlation with LST (Figures 5, 10 and 11). This indicates how anthropogenic actions are strongly associated with LST dynamics in the southwestern US.

The application of spatial LST prediction [1,8,24,26,28] is not novel, yet studies often struggle to explain LST distributions with higher accuracy and identify their key predictors. Consequently, we developed a robust GeoAI pipeline to benchmark LST predictive models and define the critical non-linearity of the selected variables. From the AutoML, we identified several models that predict LST with high precision ($R^2 > 80\%$). Among those models, decision tree-based ensemble models showed the best baseline performance, followed by support vector, k-nearest neighbors, Gaussian, and linear models. Earlier studies demonstrated the capabilities of SVR, Gaussian, and linear models [24,26], but in our case, these models achieved moderate to very low accuracies ($R^2 < 0.7$). Our systematic spatial stratification and modeling strategy further indicate high fluctuations in the model evaluation statistics across the training-validation-testing sets, suggesting their limitations for predicting LST over large areas or across distinct categories of confounding features. We further improved the five top-performing AutoML models by tuning their hyperparameters using Bayesian optimization. This optimization technique employs a surrogate model and acquisition function to explore the hyperparameter space through a systematic iterative process [48]. After tuning, Bayesian optimization reduced prediction errors by refining hyperparameter configurations, improving model robustness, and generalizing across spatially independent test locations. Although a moderate divergence between training and test R^2 (~0.06–0.12) is observed, comparable or larger train-test gaps have been reported in ensemble geospatial machine learning studies without indicating overfitting [59–61]. More critically, the negligible difference between validation and test R^2 across all five models (<0.01) demonstrates consistent generalization to spatially independent locations, suggesting that the observed train-test divergence reflects the inherent complexity and spatial heterogeneity of the data rather than model overfitting.

One of our study objectives was to fill the gap in explaining the nature of individual features to predict LST in the US Southwest. While our study does not explicitly examine the long-term temporal effect on LST, we found significant phenological changes across seasons and their effect on LST dynamics. Here, SHAP values may be scattered across asso-

ciated predictors; the uniformity of feature rankings across five distinct ensemble models increases the dependability of these attributions. From Section 3.3, it is, however, clear that summer LST dynamics were more strongly influenced by land cover features, whereas pollutant parameters were more significant for predicting winter LST. This contrast highlights the distinctiveness of energy flux [62] in which pollutant parameters show statistical associations consistent with greenhouse-type mechanisms, though direct causal attribution is limited by the absence of meteorological controls and potential confounding factors [3,23]. During the summer, the mean SAVI and WET indices were high, controlling the LST cooling effect and also having a significant negative impact on low LST (Figures 9–12). However, during winter, SAVI and WET weaken, and control over LST is lost. NDBI had one of the top positive impacts during summer, but during winter, its impact was replaced by NTL owing to an anomalous spatial distribution. This conclusion defines the significance of NTL for LST prediction as a proxy for socioeconomic activities in urban areas. While SAVI, NDBI, and WET are naturally associated as land cover indices, their different SHAP directionality (negative for SAVI/WET and positive for NDBI) suggests meaningful individual contributions rather than random importance splitting under collinearity.

The impact of pollution parameters remains insufficiently explored, with only a limited number of studies focused on specific urban areas [3,21–23]. In this study, prominent pollutant parameters were incorporated, and XAI was used to identify the critical values for each pollutant. NO₂, CO, and aerosols consistently exhibited positive influences, with NO₂ serving as a low-significance predictor in the summer model and a primary contributor in the winter models [22,23]. This transition may be attributed to reduced incoming shortwave radiation and elevated outgoing longwave radiation during winter, which are frequently retained by tropospheric NO₂, though direct causal inference is constrained by the absence of meteorological controls and potential confounding factors. Conversely, O₃ exerted a negative influence on both seasonal models. SHAP, PDP, and ALE corroborated these findings, showing a consistent decline in LST with increasing O₃ concentrations. ALE plots were used as a condition on the observed data distribution, making them more resilient to feature collinearity and delivering credible marginal impact estimates for correlated predictors. In this climatic region, O₃ shows a statistically negative association with LST, possibly confounded by transport patterns, elevation, and seasonal meteorology [3,23]. An additional explanation is that the majority of tropospheric O₃ in the southwestern US is transported over considerable distances by air masses, leading the model to interpret O₃ dynamics as those of a negative contributor [63]. It should be underscored that SHAP-based attributions are based on model-learned statistical connections rather than direct causal pathways; therefore, assessments of individual pollutant impacts must be assessed within this statistical framework.

5. Conclusions

Although LST in the US Southwest is widely studied as a response to climatic changes, we demonstrated the impact of land cover indices, socioeconomic variability, topographic characteristics, and anthropogenic pollution using GeoAI. The spatiotemporal distribution of LST is strongly influenced by seasonal variations and is linked to physical and anthropogenic variability. Significant temperature hotspots, e.g., the Great Basin, Sierra Escarpment, Mojave, and Sonoran Deserts, form a unique cluster as constant high LST hotspots, while urban hotspots, e.g., Phoenix or Las Vegas, and other coastal and low-latitude zones from California and Texas have high seasonal fluctuations driven by land cover changes and pollution concentration. AutoML revealed that the variables are efficient in capturing the spatial LST pattern and effectively explain their underlying nonlinearity. Applying Bayesian optimization hyperparameter tuning helped us achieve $R^2 > 0.84$, and

the RMSE and MAE scores were reduced by 2–6.5% compared to baseline models. This will further assist future researchers to benchmark optimal models and hyperparameters for effective decision-making.

XAI, while a comparatively prominent algorithm for understanding a feature's characteristics in machine learning, has not been adopted to capture the nonlinearity of variables across the entire US Southwest. From SHAP, we found that summer LST is more strongly influenced by land cover variables, e.g., SAVI, WET, and NDBI, while winter LST is more strongly influenced by atmospheric pollutants, e.g., O₃, aerosols, and NO₂. Ozone, usually characterized as a positive factor for LST increase, was found to have a strong negative association with spatial LST dynamics. Further, the localized impacts of the individual features were explained by PDP and ALE, demonstrating the critical values of each parameter at which LST starts to escalate sharply. An albedo index of 0.22–0.25 is the critical point where LST starts to decrease substantially.

We applied open-source satellite datasets from GEE. Satellite instruments are popular for large-scale spatiotemporal studies, but using multi-satellite data is often limited by differences in spatial and temporal scales. For example, Landsat-9 provides data every 14 days, but the VIIRS instrument used in this study offers data at 30-day intervals. Meanwhile, Sentinel 5P collected pollution data at a spatial resolution of about 11.2 km—coarser than the other two satellites. Although we carefully collected samples from each raster, combining satellite data with consistent spatial and temporal scales would better support learning complex nonlinear relationships. In the winter images, significant vegetation loss and bare soil led to masking failures, resulting in inadequate NDBI images. Further, we used NTL as a socioeconomic indicator for urban activities; a more precise NDBI raster would have provided the model with additional information. Even though important radiative and land surface variables are included in this study, the inclusion of direct surface energy balance components like soil moisture, wind speed, air temperature, humidity, cloud cover, and latent/sensible heat fluxes would further improve the robustness of LST prediction. The relationships identified between LST and pollutants represent statistical associations and should not be interpreted as direct radiative forcing mechanisms. Establishing causal links would require additional meteorological controls and boundary layer context beyond the scope of this study. Future research should incorporate this variable to capture its underlying impact on spatiotemporal LST dynamics.

However, this study streamlines the integration of remote sensing spectral indices with GeoAI learning to predict LST across large-scale areas with significant geographic and climatic differences. Our findings can inform decision-making on environmental management and the mitigation of the urban heat island effect. As a direction for future research, we recommend enhancing temporal variability to capture climatological influences, along with adequate in situ measurements.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs18050746/s1>, Table S1: Within-cell LST Homogeneity and Scale Representativeness Diagnostics. Table S2: Spearman Correlation Preservation Analysis—Point-Sampled vs. Cell-Mean LST. Table S3: Spatial Autocorrelation Statistics of the model residuals from CatBoosting.

Author Contributions: B.M.: conceptualization, methodology, formal analysis, writing—original draft preparation, G.Z.: conceptualization, writing—review and editing, supervision, project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The satellite data from Landsat, Sentinel5-p, VIIRS, and NASADEM are openly accessible from Google Earth Engine. The codes for this study are available from Bijoy Mitra upon reasonable request.

Acknowledgments: The authors acknowledge the support received from the Department of Geography & the Environment, University of Denver, US, to conduct this research.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Li, Z.L.; Tang, B.H.; Wu, H.; Ren, H.; Yan, G.; Wan, Z.; Trigo, I.F.; Sobrino, J.A. Satellite-Derived Land Surface Temperature: Current Status and Perspectives. *Remote Sens. Environ.* **2013**, *131*, 14–37. [[CrossRef](#)]
- Hall, D.K.; Comiso, J.C.; Digirolamo, N.E.; Shuman, C.A.; Key, J.R.; Koenig, L.S. A Satellite-Derived Climate-Quality Data Record of the Clear-Sky Surface Temperature of the Greenland Ice Sheet. *J. Clim.* **2012**, *25*, 4785–4798. [[CrossRef](#)]
- Kim, S.W.; Brown, R.D. Urban Heat Island (UHI) Intensity and Magnitude Estimations: A Systematic Literature Review. *Sci. Total Environ.* **2021**, *779*, 146389. [[CrossRef](#)]
- Campbell, S.; Remenyi, T.A.; White, C.J.; Johnston, F.H. Heatwave and Health Impact Research: A Global Review. *Health Place* **2018**, *53*, 210–218. [[CrossRef](#)]
- Neteler, M.; Roiz, D.; Rocchini, D.; Castellani, C.; Rizzoli, A. Terra and Aqua Satellites Track Tiger Mosquito Invasion: Modelling the Potential Distribution of *Aedes Albopictus* in North-Eastern Italy. *Int. J. Health Geogr.* **2011**, *10*, 49. [[CrossRef](#)]
- Townshend, J.R.; Justice, C.O.; Skole, D.; Malingreau, J.P.; Cihlar, J.; Teillet, P.; Sadowski, F.; Ruttenberg, S. The 1 Km Resolution Global Data Set: Needs of the International Geosphere Biosphere Programme†. *Int. J. Remote Sens.* **1994**, *15*, 3417–3441. [[CrossRef](#)]
- Wang, Y.R.; Hessen, D.O.; Samset, B.H.; Stordal, F. Evaluating Global and Regional Land Warming Trends in the Past Decades with Both MODIS and ERA5-Land Land Surface Temperature Data. *Remote Sens. Environ.* **2022**, *280*, 113181. [[CrossRef](#)]
- Yan, Y.; Mao, K.; Shi, J.; Piao, S.; Shen, X.; Dozier, J.; Liu, Y.; Ren, H.; Bao, Q. Driving Forces of Land Surface Temperature Anomalous Changes in North America in 2002–2018. *Sci. Rep.* **2020**, *10*, 6931. [[CrossRef](#)]
- Vose, R.S.; Easterling, D.R.; Kunkel, K.E.; LeGrande, A.N.; Wehner, M.F. Temperature Changes in the United States. In *Climate Science Special Report: Fourth National Climate Assessment*; US Global Change Research Program: Washington, DC, USA, 2017; Volume 1.
- Williams, A.P.; Allen, C.D.; Macalady, A.K.; Griffin, D.; Woodhouse, C.A.; Meko, D.M.; Swetnam, T.W.; Rauscher, S.A.; Seager, R.; Grissino-Mayer, H.D.; et al. Temperature as a Potent Driver of Regional Forest Drought Stress and Tree Mortality. *Nat. Clim. Chang.* **2012**, *3*, 292–297. [[CrossRef](#)]
- Sillmann, J.; Kharin, V.V.; Zwiers, F.W.; Zhang, X.; Bronaugh, D. Climate Extremes Indices in the CMIP5 Multimodel Ensemble: Part 2. Future Climate Projections. *J. Geophys. Res. Atmos.* **2013**, *118*, 2473–2493. [[CrossRef](#)]
- Fischer, E.M.; Beyerle, U.; Knutti, R. Robust Spatially Aggregated Projections of Climate Extremes. *Nat. Clim. Chang.* **2013**, *3*, 1033–1038. [[CrossRef](#)]
- Pielke, R.A.; Pitman, A.; Niyogi, D.; Mahmood, R.; McAlpine, C.; Hossain, F.; Goldewijk, K.K.; Nair, U.; Betts, R.; Fall, S.; et al. Land Use/Land Cover Changes and Climate: Modeling Analysis and Observational Evidence. *Wiley Interdiscip. Rev. Clim. Change* **2011**, *2*, 828–850. [[CrossRef](#)]
- Pielke, R.A.; Marland, G.; Betts, R.A.; Chase, T.N.; Eastman, J.L.; Niles, J.O.; Niyogi, D.D.S.; Running, S.W. The Influence of Land-Use Change and Landscape Dynamics on the Climate System: Relevance to Climate-Change Policy beyond the Radiative Effect of Greenhouse Gases. *Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci.* **2002**, *360*, 1705–1719. [[CrossRef](#)]
- Clifford, M.J.; Cobb, N.S.; Buenemann, M. Long-Term Tree Cover Dynamics in a Pinyon-Juniper Woodland: Climate-Change-Type Drought Resets Successional Clock. *Ecosystems* **2011**, *14*, 949–962. [[CrossRef](#)]
- Van Auken, O.W. Shrub Invasions of North American Semiarid Grasslands. *Annu. Rev. Ecol. Syst.* **2000**, *31*, 197–215. [[CrossRef](#)]
- Biederman, J.A.; Scott, R.L.; Bell, T.W.; Bowling, D.R.; Dore, S.; Garatuza-Payan, J.; Kolb, T.E.; Krishnan, P.; Krofcheck, D.J.; Litvak, M.E.; et al. CO₂ Exchange and Evapotranspiration across Dryland Ecosystems of Southwestern North America. *Glob. Chang. Biol.* **2017**, *23*, 4204–4221. [[CrossRef](#)] [[PubMed](#)]
- Duman, T.; Huang, C.W.; Litvak, M.E. Recent Land Cover Changes in the Southwestern US Lead to an Increase in Surface Temperature. *Agric. For. Meteorol.* **2021**, *297*, 108246. [[CrossRef](#)]
- Kueppers, L.M.; Snyder, M.A.; Sloan, L.C.; Cayan, D.; Jin, J.; Kanamaru, H.; Kanamitsu, M.; Miller, N.L.; Tyree, M.; Du, H.; et al. Seasonal Temperature Responses to Land-Use Change in the Western United States. *Glob. Planet. Change* **2008**, *60*, 250–264. [[CrossRef](#)]
- Lafortezza, R.; Carrus, G.; Sanesi, G.; Davies, C. Benefits and Well-Being Perceived by People Visiting Green Spaces in Periods of Heat Stress. *Urban For. Urban Green.* **2009**, *8*, 97–108. [[CrossRef](#)]

21. Lai, L.W.; Cheng, W.L. Air Quality Influenced by Urban Heat Island Coupled with Synoptic Weather Patterns. *Sci. Total Environ.* **2009**, *407*, 2724–2733. [\[CrossRef\]](#)
22. Wang, Y.; Du, H.; Xu, Y.; Lu, D.; Wang, X.; Guo, Z. Temporal and Spatial Variation Relationship and Influence Factors on Surface Urban Heat Island and Ozone Pollution in the Yangtze River Delta, China. *Sci. Total Environ.* **2018**, *631–632*, 921–933. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Ngarambe, J.; Joen, S.J.; Han, C.H.; Yun, G.Y. Exploring the Relationship between Particulate Matter, CO, SO₂, NO₂, O₃ and Urban Heat Island in Seoul, Korea. *J. Hazard. Mater.* **2021**, *403*, 123615. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Suthar, G.; Kaul, N.; Khandelwal, S.; Singh, S. Predicting Land Surface Temperature and Examining Its Relationship with Air Pollution and Urban Parameters in Bengaluru: A Machine Learning Approach. *Urban Clim.* **2024**, *53*, 101830. [\[CrossRef\]](#)
25. Tanoori, G.; Soltani, A.; Modiri, A. Machine Learning for Urban Heat Island (UHI) Analysis: Predicting Land Surface Temperature (LST) in Urban Environments. *Urban Clim.* **2024**, *55*, 101962. [\[CrossRef\]](#)
26. Badugu, A.; Arunab, K.S.; Mathew, A. Predicting Land Surface Temperature Using Data-Driven Approaches for Urban Heat Island Studies: A Comparative Analysis of Correlation with Environmental Parameters. *Model. Earth Syst. Environ.* **2023**, *10*, 1043–1076. [\[CrossRef\]](#)
27. Li, W. Artificial Intelligence in Earth Science: A GeoAI Perspective. *J. Geophys. Res. Mach. Learn. Comput.* **2025**, *2*, e2025JH000691. [\[CrossRef\]](#)
28. Nowicki, S.A.; Inman, R.D.; Esque, T.C.; Nussear, K.E.; Edwards, C.S. Spatially Consistent High-Resolution Land Surface Temperature Mosaics for Thermophysical Mapping of the Mojave Desert. *Sensors* **2019**, *19*, 2669. [\[CrossRef\]](#)
29. U.S. Department of Commerce. *Census Regions and Divisions of the United States*; U.S. Department of Commerce: Washington, DC, USA, 2010.
30. Levick, L.; Fonseca, J.; Goodrich, D.; Hernandez, M.; Semmens, D.; Stromberg, J.; Leidy, R.; Scianni, M.; Guertin, D.P.; Tluczek, M. *The Ecological and Hydrological Significance of Ephemeral and Intermittent Streams in the Arid and Semi-Arid American Southwest*. US Environmental Protection Agency and USDA/ARS Southwest Watershed Research Center; Office of Research and Development: Washington, DC, USA, 2008.
31. Beck, H.E.; Zimmermann, N.E.; McVicar, T.R.; Vergopolan, N.; Berg, A.; Wood, E.F. Present and Future Köppen-Geiger Climate Classification Maps at 1-Km Resolution. *Sci. Data* **2018**, *5*, 180214. [\[CrossRef\]](#)
32. Omernik, J.M.; Griffith, G.E. Ecoregions of the Conterminous United States: Evolution of a Hierarchical Spatial Framework. *Environ. Manag.* **2014**, *54*, 1249–1266. [\[CrossRef\]](#)
33. Blodgett, L. Climatology of the United States and of the Temperate Latitudes of the North American Continent. In *Meteorology in Nineteenth-Century Society*; Routledge: London, UK, 2025; pp. 187–199.
34. Dewitz, J. National Land Cover Database (NLCD) 2019 Products (Ver. 3.0, February 2024). *US Geol. Surv. (USGS) Data Release* **2021**, 624. [\[CrossRef\]](#)
35. EROS. *Landsat 8–9 Operational Land Imager/Thermal Infrared Sensor Level-2, Collection 2 [Dataset]*; US Geological Survey: Reston, VA, USA, 2020.
36. Masek, J.G.; Wulder, M.A.; Markham, B.; McCorkel, J.; Crawford, C.J.; Storey, J.; Jenstrom, D.T. Landsat 9: Empowering Open Science and Applications through Continuity. *Remote Sens. Environ.* **2020**, *248*, 111968. [\[CrossRef\]](#)
37. Irizar, J.; Melf, M.; Bartsch, P.; Koehler, J.; Weiss, S.; Greinacher, R.; Erdmann, M.; Kirschner, V.; Perez Albinana, A.; Martin, D. Sentinel-5/UVNS. In *Proceedings of the International Conference on Space Optics—ICSO 2018*; SPIE: Chania, Greece, 12 July 2019; Volume 11180, pp. 41–58.
38. Compernelle, S.; Argyrouli, A.; Lutz, R.; Snee, M.; Lambert, J.C.; Mari Fjæraa, A.; Hubert, D.; Keppens, A.; Loyola, D.; O'Connor, E.; et al. Validation of the Sentinel-5 Precursor TROPOMI Cloud Data with Cloudnet, Aura OMI O₂-O₂, MODIS, and Suomi-NPP VIIRS. *Atmos. Meas. Tech.* **2021**, *14*, 2451–2476. [\[CrossRef\]](#)
39. Mitra, B.; Hridoy, A.-E.E.; Mahmud, K.; Uddin, M.S.; Talha, A.; Das, N.; Nath, S.K.; Shafiullah, M.; Rahman, S.M.; Rahman, M.M. Exploring Spatial and Temporal Dynamics of Red Sea Air Quality through Multivariate Analysis, Trajectories, and Satellite Observations. *Remote Sens.* **2024**, *16*, 381. [\[CrossRef\]](#)
40. Mahmud, K.; Mitra, B.; Uddin, M.S.; Hridoy, A.-E.E.; Aina, Y.A.; Abubakar, I.R.; Rahman, S.M.; Tan, M.L.; Rahman, M.M. Temporal Assessment of Air Quality in Major Cities in Nigeria Using Satellite Data. *Atmos. Environ. X* **2023**, *20*, 100227. [\[CrossRef\]](#)
41. NASA JPL. *NASADEM Merged DEM Global 1 Arc Second V001 [Data Set]*. NASA EOSDIS Land Processes DAAC; NASA: Greenbelt, MD, USA, 2020. [\[CrossRef\]](#)
42. Elvidge, C.D.; Baugh, K.; Zhizhin, M.; Hsu, F.C.; Ghosh, T. VIIRS Night-Time Lights. *Int. J. Remote Sens.* **2017**, *38*, 5860–5879. [\[CrossRef\]](#)
43. Elvidge, C.D.; Zhizhin, M.N.; Baugh, K.; Zhizhin, M.; Hsu, F.C. Why VIIRS Data Are Superior to DMSP for Mapping Nighttime Lights. *Proc. Asia-Pac. Adv. Netw.* **2013**, *35*, 62–69. [\[CrossRef\]](#)
44. Cicchetti, D.V. Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. *Psychol. Assess.* **1994**, *6*, 284–290. [\[CrossRef\]](#)

45. Zöller, M.A.; Huber, M.F. Benchmark and Survey of Automated Machine Learning Frameworks. *J. Artif. Intell. Res.* **2021**, *70*, 409–472. [[CrossRef](#)]
46. Yao, Q.; Wang, M.; Chen, Y.; Dai, W.; Li, Y.-F.; Tu, W.-W.; Yang, Q.; Yu, Y. Taking Human out of Learning Applications: A Survey on Automated Machine Learning. *arXiv* **2018**, arXiv:1810.13306.
47. Ali, M. PyCaret: An Open Source, Low-Code Machine Learning Library in Python. *PyCaret Version* **2020**, *2*. Available online: <https://github.com/pycaret/pycaret> (accessed on 26 February 2026).
48. Frazier, P.I. A Tutorial on Bayesian Optimization. *arXiv* **2018**, arXiv:1807.02811. [[CrossRef](#)]
49. Shapley, L.S. Stochastic Games. *Proc. Natl. Acad. Sci. USA* **1953**, *39*, 1095–1100. [[CrossRef](#)]
50. Evans, L.C. *Partial Differential Equations*; American Mathematical Society: Providence, Rhode Island, USA, 2022; Volume 19.
51. 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.* **2016**, *82*, 1059–1086. [[CrossRef](#)]
52. Levin, N. The Impact of Seasonal Changes on Observed Nighttime Brightness from 2014 to 2015 Monthly VIIRS DNB Composites. *Remote Sens. Environ.* **2017**, *193*, 150–164. [[CrossRef](#)]
53. Mokhtari, Z.; Bergantino, A.S.; Intini, M.; Elia, M.; Buongiorno, A.; Giannico, V.; Sanesi, G.; Laforteza, R. Nighttime Light Extent and Intensity Explain the Dynamics of Human Activity in Coastal Zones. *Sci. Rep.* **2025**, *15*, 1663. [[CrossRef](#)]
54. Brazel, A. June Temperature Trends in the Southwest Deserts of the USA (1950–2018) and Implications for Our Urban Areas. *Atmosphere* **2019**, *10*, 800. [[CrossRef](#)]
55. Lopez, H.; Lee, S.K.; West, R.; Kim, D.; Jia, L. The Longest-Lasting 2023 Western North American Heat Wave Was Fueled by the Record-Warm Atlantic Ocean. *Nat. Commun.* **2025**, *16*, 6544. [[CrossRef](#)]
56. Howey, T.; North, L.; Kerry, R. Land Use and Land Cover Change and Potential Implications for Water Levels of the Great Salt Lake. *Environments* **2025**, *12*, 381. [[CrossRef](#)]
57. Nedd, R.; Anandhi, A. Land Use Changes in the Southeastern United States: Quantitative Changes, Drivers, and Expected Environmental Impacts. *Land* **2022**, *11*, 2246. [[CrossRef](#)]
58. Li, X.; Tian, H.; Lu, C.; Pan, S. Four-Century History of Land Transformation by Humans in the United States (1630–2020): Annual and 1gkm Grid Data for the HIStory of LAND Changes (HISLAND-US). *Earth Syst. Sci. Data* **2023**, *15*, 1005–1035. [[CrossRef](#)]
59. Liu, L. An Ensemble Framework for Explainable Geospatial Machine Learning Models. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *132*, 104036. [[CrossRef](#)]
60. Hoang, N.D.; Tran, V.D.; Huynh, T.C. From Data to Insights: Modeling Urban Land Surface Temperature Using Geospatial Analysis and Interpretable Machine Learning. *Sensors* **2025**, *25*, 1169. [[CrossRef](#)]
61. Das, P.K.; Mukherjee, I.; Prasad, P.; Pushkar, S. Downscaling MODIS Land Surface Temperature to 90 m Using Random Forest Regression to Assess Transferability. *PeerJ Comput. Sci.* **2025**, *11*, e3246. [[CrossRef](#)]
62. Liou, Y.-A.; Le, M.S.; Chien, H. Normalized Difference Latent Heat Index for Remote Sensing of Land Surface Energy Fluxes. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 1423–1433. [[CrossRef](#)]
63. Parrish, D.D.; Faloona, I.C.; Derwent, R.G. Maximum Ozone Concentrations in the Southwestern US and Texas: Implications of the Growing Predominance of the Background Contribution. *Atmos. Chem. Phys.* **2025**, *25*, 263–289. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.