

Machine Learning for Urban Heat Island (UHI) Analysis: Predicting Land Surface Temperature (LST) in Urban Environments



Ghazaleh Tanoori ^{a,*}, Ali Soltani ^b, Atoosa Modiri ^a

^a Department of Urbanism, Central Tehran Branch, Islamic Azad University, Tehran, Iran

^b Flinders Health and Medical Research Institute (FHMRI), Flinders University, Bedford Park, 5042, Australia

ARTICLE INFO

Keywords:

Urban Heat Island
Land surface temperature
Configuration metrics
Machine learning algorithms
Prediction
Deep Neural Network

ABSTRACT

This study investigates how urban configuration influences the distribution of heat, known as the Urban Heat Island (UHI) effect, in Shiraz, Iran. Several Machine Learning algorithms are employed to analyze Land Surface Temperature (LST) data across various land cover types, including built-up, soil, and vegetation. The analysis reveals that Deep Neural Networks (DNNs) and Extreme Gradient Boosting (XGB) models excel at predicting LST, outperforming other methods. These results highlight the significant impact of land use on LST patterns within the metropolitan regions. Furthermore, the study assesses the influence of specific configuration metrics within each land cover category. This allows researchers to pinpoint which urban morphology features most significantly affect LST. These insights can inform targeted interventions and management strategies implemented to mitigate heat and improve thermal comfort in specific areas of Shiraz.

1. Introduction

Temperature increase in urban environments is known as the urban heat island (UHI) effect. In recent decades, urbanization and the dramatic increase in urban populations have exacerbated the UHI effect (Yin et al., 2023). The uneven expansion and construction of the city will lead to an uneven temperature growth rate in the corresponding area (Jauregui et al., 1992). Land surface temperature (LST) has been commonly used to characterize the relationship between UHI intensity and land use conditions (Guo et al., 2020).

UHI is mostly caused by Land Use/Land Cover Change (LULCC) (Deilami et al., 2018). Changes in land use, such as types of land use, will result in extensive differences in regional climatic temperatures (Cao et al., 2015; Allan et al., 2022). Indeed, rapid urban expansion often leads to the shrinkage, disconnection, and fragmentation of green spaces, which can have negative consequences for their cooling capacity and overall environmental benefits (Li et al., 2021; Azizi et al., 2022).

The spatial distribution characteristics of cities, different land use types such as opaque land surface, vegetation, and water distribution, as well as the absorption of long-wave radiation in peri-urban areas, and the significant increase in the temperature difference of the underlying surface, all of which will lead to the formation of the heat island effect (Cheela et al., 2021; Yao et al., 2022). Meanwhile, most studies exploring the influence of city configurations on UHI have found that variables had relatively strong correlations with LST (Azhdari et al., 2018). One of the ways to understand landscape changes and their reactions is to be aware of the

* Corresponding author.

E-mail addresses: ghtanoori@gmail.com (G. Tanoori), ali.soltani@flinders.edu.au (A. Soltani).

dynamics of land use patterns. Because land use changes lead to changes in landscape patterns; hence, the analysis of land use changes can help to understand the changes in landscape patterns (Nagendra et al., 2004).

To measure landscape pattern, landscape metrics are utilized. The optimum size and landscape measurements vary according to the climatic zone (Zhao et al., 2017). These metrics aid in acquiring an awareness of the features of the urban landscape to manage urban settings in a sustainable manner (Magidi and Ahmed, 2019). Landscape composition and configuration can affect ecological processes independently and interactively. Thus, it is especially important to understand for each metric what aspect of landscape pattern is being quantified (Turner et al., 2001).

Several studies focused on examining the relationship between configuration metrics and LST (Athukorala and Murayama, 2020; Connors et al., 2013; Hou and Estoque, 2020; Kamarianakis et al., 2019; Liu et al., 2018), rather than specifically discussing the importance of each metric in increasing LST. Previous research explored the link between configuration metrics and surface temperature, but did not rank their importance.

Numerous studies have employed landscape pattern indices to explore the connections between urban configuration and LST (Li et al., 2017; Soltani and Sharifi, 2019). However, the relationship between landscape metrics and LST is acknowledged as complex and multifaceted (Yuan et al., 2024). While the existing literature often examines the correlation between landscape composition and LST through the lens of linear relationships (Gao et al., 2019), correlation coefficients (Zhou et al., 2022), or linear regression models (Zhao et al., 2017; Chun and Guldmann, 2014), these approaches may not fully capture the potential variations in how different landscape elements influence LST across density ranges (Guo et al., 2020).

Machine learning offers a powerful alternative for LST prediction by leveraging advanced computational techniques to analyze and predict LST patterns (Yin et al., 2023). These models utilize current and historical LST data alongside other relevant factors like climate patterns, land-use changes, and human activities to forecast future LST values (Rehman et al., 2022). By identifying patterns and relationships embedded within historical data, machine learning algorithms can make accurate predictions of future LST (Suthar et al., 2024). A diverse range of machine learning algorithms can be employed for LST prediction, including artificial neural networks (ANN) (Li et al., 2021), decision trees (DT) (Samardžić-Petrović et al., 2017), random forests (RF), support vector machines (SVM) (Karimi et al., 2019; Samardžić-Petrović et al., 2016), XGBoost regression (XGB) (Madaan et al., 2021; Mohammad et al., 2022), and AdaBoost (Chen et al., 2017). These models are trained on large datasets encompassing historical LST values and other relevant data like meteorological variables, land-use patterns, and topography. Once trained, the models can predict LST values for future periods or various locations (Suthar et al., 2024).

LST exhibits complex interactions with environmental factors. Studies have shown that XGB models are particularly well-suited for predicting average LST values. Khanifar and Khademalrasoul (2022) found XGB models to be more accurate in predicting warm season LST compared to cold seasons (Jabeur et al., 2021). This is likely due to the XGB algorithm's strength in handling complex relationships (Jabeur et al., 2021).

Another popular choice for LST prediction is RF due to its ability to identify non-linear relationships between variables and pinpoint the most influential factors (Shafizadeh-Moghadam et al., 2020). Suthar et al. (2024) utilized RF to model the link between LST and specific variables, while Li et al. (2021) demonstrated its effectiveness in revealing the impact of urban characteristics on LST.

SVM offers a supervised non-linear approach for LST prediction. This prevalent machine-learning method excels in various tasks like estimation, classification, and regression (Ge et al., 2018). SVM strikes a good balance between achieving high accuracy and avoiding overfitting the data (Ge et al., 2018). Studies by Samardžić-Petrović et al. (2017) suggest that SVM might be better at capturing land-use class changes compared to neural networks (NN) and DT.

The recent surge in computational power has fueled the rise of deep neural networks (DNNs) in LST prediction (Siqi et al., 2023). DNNs possess complex architectures that allow them to model highly intricate relationships. However, traditional linear models often struggle to capture the spatial variations in these relationships and fail to fully utilize the vast potential of remote sensing data (Bhandari et al., 2022). While DNNs excel at achieving high predictive accuracy (Azedou et al., 2023), they may generate less interpretable results for specific locations (Bhandari et al., 2022).

In contrast, DTs offer a simpler but powerful alternative. Their non-parametric nature makes them well-suited for analyzing diverse landscapes (Phiri et al., 2020). Additionally, DTs excel at identifying the most impactful variables influencing LST (Karimi et al., 2021). Samardžić-Petrović et al. (2017) highlight their advantage of easy interpretability compared to other algorithms like Neural Networks. Furthermore, DTs exhibit greater flexibility in handling high temporal frequency remote sensing data, as demonstrated by Liu et al. (2017) in their study of LST changes under urban renewal.

Moving beyond single algorithms, ensemble methods like AdaBoost have emerged with promising results. Pande et al. (2023) demonstrated AdaBoost's effectiveness in predicting climate variables and other datasets, while Li et al. (2022) found it, alongside XGBoost, to be highly accurate for identifying and extracting information about mangrove landscapes. Pande et al. (2024) further explored the potential of ensemble methods by comparing XGBoost, Bagging, and AdaBoost for LST prediction. This highlights the importance of model selection, as different algorithms offer a trade-off between flexibility for complex relationships and the interpretability of results.

Despite important efforts to collect and label training data using either conventional methods or crowdsourcing initiatives (Xiong et al., 2017), it is still challenging to generate a larger number of annotated samples required to successfully train deep learning algorithms (Gevaert and Belgiu, 2022). Collecting and labeling training data is indeed a crucial step in machine learning model development. While conventional methods and crowdsourcing initiatives have been widely used for this purpose, leveraging machine learning algorithms can further enhance results and improve accuracy. Employing six different machine learning algorithms including Support Vector Machine (SVM), DT, RF, AdaBoost, XGB, and Deep Neural Network (DNN) can lead to more robust and reliable outcomes.

This study aims to address several key research questions regarding the relationship between urban configuration, LST patterns, and the UHI effect in metropolitan Shiraz. Firstly, it seeks to understand how urban configuration contributes to the UHI effect. Secondly, the study investigates the temporal trends in LULC and urban configuration metrics from 2006 to 2021 within the city. Thirdly, it explores the predictive capabilities of machine learning algorithms in forecasting LST within different land cover types. Moreover, the study endeavors to identify the specific urban configuration metrics that exert the most significant influence on LST within each land cover category. Finally, it aims to elucidate the interplay between urban form, LST patterns, and the UHI effect in Shiraz. In this paper, the landscape composition at the class level were analyzed to identify the most critical land use categories driving LST increases. The ultimate goal was to pinpoint key factors contributing to rising LST and inform sustainable urban planning strategies. Five landscape metrics, including: percentage of landscape (PLAND), fractal dimension index (FRAC), landscape shape index (LSI), largest patch index (LPI), and edge density (ED) were selected, to assess urban configuration and investigate their association with LST variation (Azhdari et al., 2018; Yan et al., 2019; Yue et al., 2019; Zhang et al., 2020). These indicators can characterize the process of urban energy balance comprehensively (Han et al., 2023). Landscape metrics were calculated in FRAGSTATS 4.3. which is an independent piece of software that calculates several landscape measures for investigating landscape fragmentation (Singh et al., 2017).

2. Methodology

2.1. Study area

Several factors contribute to the formation of UHIs in major Iranian metropolises, including Shiraz. Inadequate provision of green spaces and the deterioration of historic gardens due to water scarcity and lack of maintenance reduce the cooling effect of vegetation. Plants absorb moisture, release water vapor, and provide shade, all of which help to regulate urban temperatures (Ghanbari et al., 2023). The widespread use of fossil fuels in Iranian cities, particularly for mass construction activities, heating and cooking, leads to increased greenhouse gas emissions, trapping heat in the urban environment. Poor fuel quality and outdated vehicle standards in Iranian cities contribute to air pollution with harmful pollutants like carbon monoxide and particulate matter (Soltani, 2017). These pollutants not only degrade air quality but also trap heat, exacerbating the UHI effect (Soltani and Sharifi, 2017). The layout of some Iranian cities, particularly those in windless regions, hinders effective heat dissipation, leading to a combination of poor air quality and high temperatures (Azhdari et al., 2018).

Shiraz, Iran's fifth-largest city by population, serves as a vital economic, social, and administrative center in the country's southern

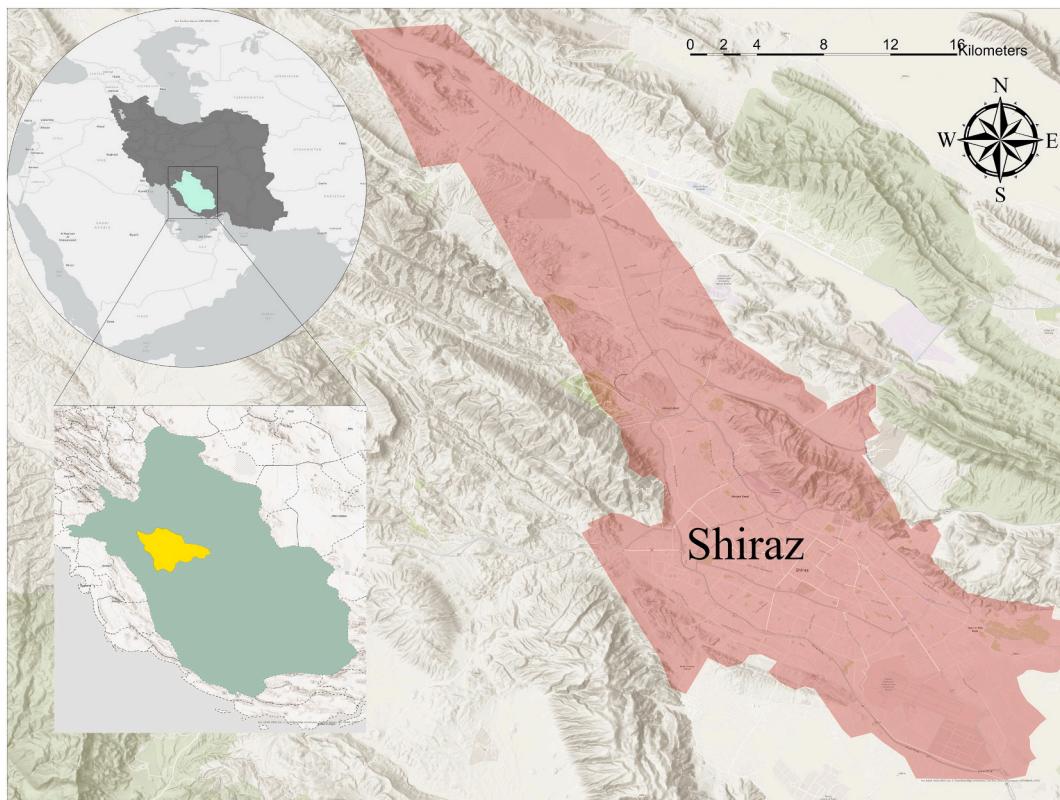


Fig. 1. Shiraz metropolitan region as the research area.

region (Fig. 1). The region's rapid urbanization is evident in its population growth rate of 2.04% and an area growth rate of 3.16% over the past three decades (Bagheri and Soltani, 2023). With a hot semi-arid climate classified by the Köppen system (Azhdari et al., 2018), Shiraz experiences hot summers with an average July high of 38.8 °C. All images used in this study were captured during July in different years. In Shiraz metropolitan region, climate changes and temperature increases have been observed, which can be attributed to the region's growing urbanization trend and expanding built-up areas. Additionally, the reduction of vegetation, particularly in historical infill gardens, has contributed to these changes.

2.2. Data collection and pre-processing

To examine how configuration affects LST, maps of the landscape (Lin et al., 2020) were required from LULC data between 2006 and 2021 and LST use of Landsat datasets in remote sensing. Satellite images were analysed using ArcGIS 10.2.1, ENVI 5.3.1, and finally Fragstats for landscape metrics. All coding practices for data manipulation and modelling were done in Python.

The land cover map of Shiraz included three categories; vegetation, soil, and built-up areas. Furthermore, urban morphology indices (Peng et al., 2022) were considered including five landscape metrics (PLAND, FRAC, LSI, LPI, and ED) to investigate their association with LST variation (Azhdari et al., 2018; Yan et al., 2019; Yue et al., 2019; Zhang et al., 2020; Chenary et al., 2023). These indicators can characterize the process of urban energy balance comprehensively (Han et al., 2023).

The accuracy and Kappa coefficient values were calculated for LULC maps. The accuracy was measured by randomly selecting 100 points from each land cover category and checking whether the classifications matched. The higher the match between the classification map and these images, the closer the accuracy was to 100%. The Kappa coefficient was calculated using the following equation;

$$\text{Overall accuracy} = (\text{No.of Corrected points}/\text{Total Number of Points})^*100 \quad (1)$$

The Kappa result shows an accuracy of 92% in the 2006 LULC map, 97% in 2010, 98% in 2014, 92% in 2018, and 97% in the 2021 LULC map of Shiraz produced from Landsat images.

The Kappa result shows an accuracy of 92% in the 2006 LULC map, 97% in 2010, 98% in 2014, 92% in 2018, and 97% in the 2021 LULC map of Shiraz produced from Landsat images. Mosaic maps from the LULC maps were generated and extracted configuration metrics maps and data using a moving window approach in Fragstats. The configuration metrics maps included PLAND, FRAC, LPI, LSI, and ED from LULC. To integrate the configuration maps with the corresponding LST map for each year, they were layered on top of each other. To ensure accuracy and prevent data overlap, the final output for each year contained both the configuration metric values and their respective geographic coordinates.

2.3. Retrieval of LST

Remotely sensed observations collected by the Landsat constellation of satellites are ideal for LST estimations to the thermal infrared radiometers carried by Landsat TM/ETM+/OLI/TIRS images provided by the USGS (<https://earthexplorer.usgs.gov/>). Pre-processing of data included radiometric calibration and FLAASH atmospheric correction, eliminating the influence of atmosphere and light on surface reflections (Wang et al., 2020), which make their data suitable for LST estimations, including for characterization of cooling. In this study, the LST estimations were derived from Band 10 Thermal infrared (TIR1), resampled from 100 m to 30 m, 10.60–11.19 μm of Landsat-8 observations (Goldblatt et al., 2021; Tanoori et al., 2024). Initially, four four-year intervals between 2006 and 2021 were selected. However, to identify the weather background to study UHI, the RS image for each year was sampled in July (Du et al., 2019). Therefore used for LST retrieval, a thermal channel (30 m resolution) is widely applied to retrieve LST data.

2.4. Machine learning algorithm

To predict LST, six different machine-learning algorithms were explored (Fig. 2). For training these models, the feature vector consisted of eight features including five configuration metrics, year, and geographic coordinates. All machine learning methods train with the dataset of size N*8, where N is the number of samples in the dataset and eight is the number of features. Geographic coordinates were used to clarify configuration metrics data for each patch in each year.

a. Preprocessing:

Data preprocessing is a crucial step to ensure the quality and suitability of a dataset for machine learning analysis. In this study, data preprocessing was performed to prepare the dataset for analysis. As the dataset did not contain any missing values, the focus was on normalization using the Min-Max technique. The purpose of normalization is to scale the features within a specific range, usually [0,1], to avoid any bias or dominance of certain features due to their original scale. To achieve this, the following normalization formula was applied to each feature in the dataset:

$$x_{\text{normalized}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (2)$$

Here, x represents a feature in the dataset, while x_{\min} and x_{\max} denote the minimum and maximum values of that feature, respectively, across the original dataset.

To assess the quality of the dataset, a five-number summary were computed, which included; min, max, median, first quartile (Q1), and third quartile (Q3) values for each feature vector.

Box plots for each feature vector, were presented demonstrating that the majority of the data points fall within the box

(interquartile range), indicating a relatively consistent distribution. (See Fig. 3.)

It were also identified and addressed some outliers that were beyond the whiskers in features. By calculating the interquartile range (IQR). The IQR is calculated as a difference between the third and first quartiles (Q3-Q1).

2.4.1. Support vector machine (SVM)

SVM is a supervised learning algorithm (Li et al., 2022) that can be used for both classification and regression tasks. When solving binary classification problems, SVMs formulate them as convex optimization problems and aim to find the maximum margin hyperplane that separates the classes while correctly classifying as many training points as possible (Awad and Khanna, 2015). The decision boundary is defined by drawing a hyperplane. The optimal hyperplane is obtained by solving a quadratic optimization problem:

$$f(x) = w^T x + b \quad (3)$$

Here w represents the weight vector, b is bias and x represents the multi-dimensional feature vector.

To generalize SVM for regression tasks, Support Vector Regression (SVR) introduces an ϵ -insensitive region around the function, known as the ϵ -tube. This tube reformulates the optimization problem to find the tube that best approximates the continuous-valued function while balancing model complexity and prediction error.

2.4.2. Decision Tree (DT)

DTs, as machine learning techniques, learn the relationships between predictor variables and target variables (Karimi et al., 2021) and are constructed by recursively partitioning a dataset and fitting a simple model to each partition (Loh, 2011). In the case of regression tasks, DTs are used to predict dependent variables that take continuous or ordered discrete values. The prediction error is typically measured by the squared difference between the observed and predicted values.

The construction of a DT involves selecting the best features and creating decision nodes to split the data based on those features. Each internal node of the tree represents a decision based on a selected feature, and each leaf node represents a prediction value. The decision process follows a hierarchical structure, where each level of the tree represents a different feature and its corresponding splitting criterion.

The advantage of DTs is their interpretability, as the resulting model can be easily visualized and understood. Additionally, DTs can handle both numerical and categorical features, making them flexible for various types of datasets.

2.4.3. Random Forest (RF)

RF is a popular ensemble learning method that combines multiple DTs to make predictions (Liaw and Wiener, 2002). Each DT in the RF independently predicts the class or value for observation, and the final prediction is determined by a majority vote (for classification) or averaging (for regression) of the individual tree predictions.

The process of building a RF involves creating a collection of DTs, where each tree is trained on a different subset of the data and a random subset of features. This randomness injects diversity into the ensemble, reducing overfitting and improving generalization performance. The prediction of RF classification can be formulated by the following equation (Li et al., 2021).

$$P_d(i|k) = \frac{1}{M} \sum_{n=1}^M I(h_n(x) = d) \quad (4)$$

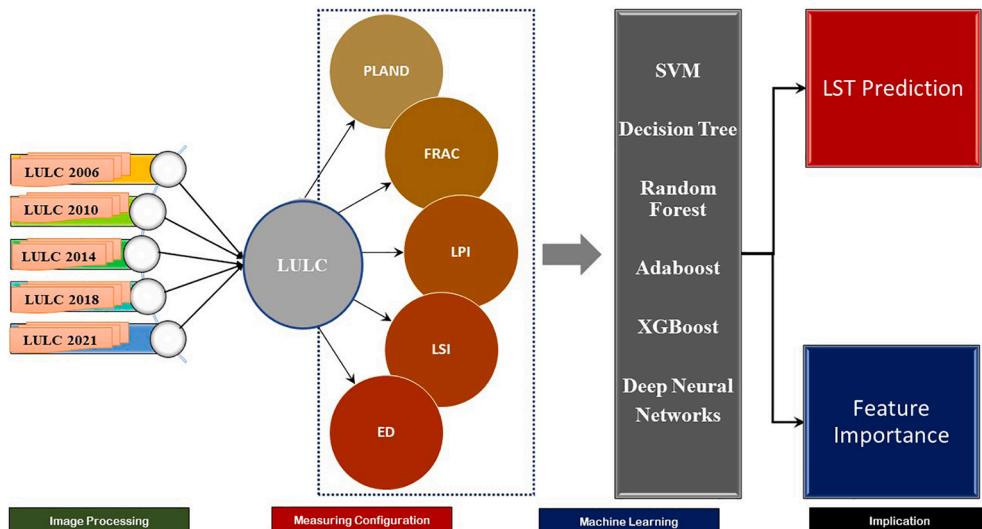


Fig. 3. Box plots of input parameters.

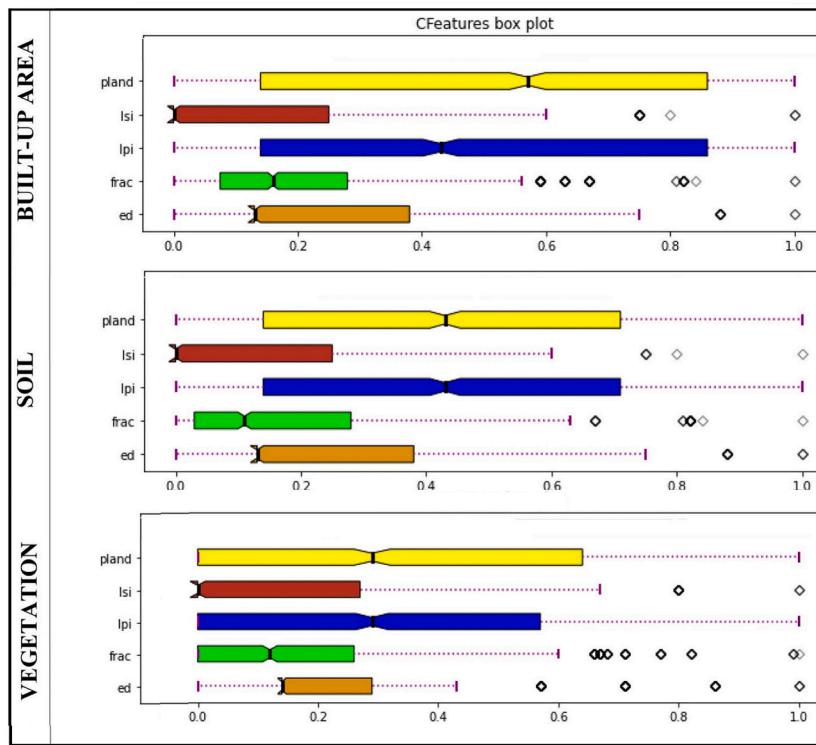


Fig. 2. The adopted framework for evaluating the importance of urban configuration metrics and predicting LST.

2.4.4. Adaptive Boosting (AdaBoost)

AdaBoost is an ensemble learning method that combines multiple weak learners to create a strong learner (Li et al., 2022). AdaBoost aims to improve the performance of weak learners by iterative training and combining them (Ying et al., 2013). The algorithm assigns different weights to each training sample throughout the iterations. It focuses on giving more weight to the misclassified samples from previous iterations, effectively “punishing” them, and placing emphasis on getting these samples correctly classified in subsequent iterations. AdaBoost has shown good generalization performance and can handle both classification and regression tasks.

2.4.5. Extreme Gradient Boosting (XGB)

XGB is an advanced gradient-boosting algorithm that has gained significant popularity in recent years. The gradient boosting algorithm is an ensemble learning method that predicts the target variable by iteratively combining the predictions from multiple weak models, typically DTs (Bentéjac et al., 2021). It works by minimizing a loss function through the gradient of the true values and the predicted values. In each iteration, a new weak model is trained to correct the errors made by the previous models, with a focus on reducing the loss function. XGB is an enhanced version of the gradient boosting algorithm that incorporates several additional features and optimizations to improve performance and accuracy. It is designed to handle both regression and classification tasks effectively.

2.4.6. Deep Neural Network (DNN)

Neurons are the building blocks of a DNN. They receive input signals, apply an activation function to the weighted sum of the inputs, and produce an output (Da Silva et al., 2017). Neurons are organized into layers. The input layer receives the initial input data, the output layer produces the final prediction, and the intermediate layers are referred to as hidden layers. Training a DNN involves iteratively adjusting the weights based on a loss function and a chosen optimization algorithm, such as backpropagation, to minimize the difference between the network's predicted output and the true values.

2.5. Validation of machine learning methods

For evaluation of different regression methods for temperature prediction, the k-fold cross-validation was employed. K-fold cross-validation is a widely used algorithm that divides the dataset into k folds, where each fold serves as a test set while the remaining folds are used for training. In this study, k = 5 was chosen for the cross-validation process. The k-fold cross-validation procedure involves the following steps:

Dataset Split: The dataset is divided into k equal-sized folds, ensuring that the distribution of samples is representative across the folds.

Iterative Training and Evaluation: For each iteration, one-fold is held out as a test set, and the remaining k-1 folds are used for training the model. The model is then evaluated using appropriate metrics.

To assess the performance of the models, four metrics were considered: Root Mean Squared Error (RMSE), Concordance Index (CI), Average Absolute Percent Relative Error (AAPRE %), and R-squared.

2.5.1. RMSE

RMSE was used to check the accuracy of the prediction results and measure the average difference between the predicted and actual values, indicating the model's overall accuracy (Suthar et al., 2024). Lower RMSE values indicate better performance.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - F(X_i))^2} \quad (5)$$

Where y_i is the true (Suthar et al., 2024) temperature value for instance with feature vector X_i , and $\hat{y}_i = F(X_i)$ is the predicted value of temperature and N is the number of test samples.

2.5.2. CI

Concordance Index, also known as the C-index, is a metric commonly used in survival (Alabdallah et al., 2024) analysis or ranking tasks (Alabdallah et al., 2024). It assesses the ability of the model to correctly order or rank the predicted values compared to the true values. A higher CI value indicates better predictive performance.

$$CI = \frac{1}{Z} \sum_i \sum_{y_i > y_j} h(\hat{y}_i - \hat{y}_j) \quad (6)$$

where \hat{y}_i is the predicted value for the larger temperature y_i , \hat{y}_j is the predicted value for the smaller temperature y_j , Z is a normalization constant, and $h(x)$ is the Heaviside step function:

$$h(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0.5 & \text{if } x = 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (7)$$

2.5.3. AAPRE

The Average Absolute Percent Relative Error (AAPRE %) is a statistical metric used to measure the accuracy of a prediction or estimation model. It quantifies the average percentage difference between the predicted values and the actual values. The formula for AAPRE % is:

$$AAPRE = \left(\left(\frac{1}{N} \right) \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_i} \right) * 100 \quad (8)$$

A lower AAPRE % indicates better accuracy, while a higher AAPRE % indicates poorer accuracy.

2.5.4. R-squared

The Coefficient of Determination, or R-squared, is a statistical measure that indicates the proportion of the variance in the dependent variable that can be explained by the independent variable(s) in a regression model. It is used as a measure of the goodness of fit of a regression model. R-squared is calculated as follows:

$$R^2 = 1 - \left(\frac{SSR}{SST} \right) \quad (9)$$

Where SSR and SST are calculated as follows:

$$SSR = \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (10)$$

$$SST = \sum_{i=1}^N (y_i - y_{avg})^2 \quad (11)$$

Where y_{avg} is the average of actual values. In the best case, the modeled values exactly match the observed values, which results in $SSR = 0$ and $R^2 = 1$.

2.6. Feature importance

This study utilized the Shapley Additive explanations (SHAP) method to measure the importance of landscape metrics when impacting LST (Yuan et al., 2024). SHAP is a powerful technique based on Game Theory that provides insights into the contribution of

each feature in the prediction of temperature. The SHAP algorithm operates by assigning a numerical value to each feature, representing its importance or impact on the prediction. These values are derived from the concept of Shapley values, which were originally developed to allocate the value of a cooperative game among its players. The metrics were arranged based on their feature importance, showcasing the contribution of metrics to LST and the overall relationship (Wu et al., 2020).

3. Result and discussion

3.1. LST temporal pattern changes in Shiraz

This study leveraged high-resolution remote sensing data to analyze temperature changes in Shiraz, focusing on LST variations as depicted in Fig. 4. To ensure accurate LST retrieval and land type interpretation, the August imagery was utilized, following best practices outlined by Huang et al. (2022). LST values were categorized into six groups, ranging from below 30 °C to above 50 °C, to investigate their spatial distribution across Shiraz. A concerning trend of rising LST was observed across all categories, with the average, minimum, and maximum LST all exhibiting a persistent increase over the study period. Notably, the minimum LST value in 2021 exceeded the maximum LST recorded in 2006, highlighting the importance of examining spatio-temporal changes within the Shiraz metropolitan area.

The analysis revealed higher LST values near the metropolitan boundaries, while gardens (vegetation) and built-up areas displayed cooler temperatures. This suggests a potential mitigating effect of both vegetation and urban infrastructure on overall region's LST. Consistent with global research (Athukorala and Murayama, 2020; Guo et al., 2019), areas with minimal vegetation exhibited the highest LST, while densely vegetated areas showed the coolest temperatures.

As illustrated in Fig. 5, LULC types significantly impact UHI patterns. Variations in land-use characteristics lead to disparities in LST balance between urban and non-urban regions. Similar to Azhdari et al. (2018), the findings suggest that the urban core maintains a consistently cooler surface temperature throughout the year compared to its surroundings. However, this disparity becomes more pronounced during the summer months, aligning with previous research. The presence of built-up areas appears to contribute to a reduction in overall region's surface temperature, suggesting that urban morphology plays a role in shaping UHI patterns.

3.2. LULCC and landscape changes between 2006 and 2021

Previous research has shown that urban expansion and regeneration, through landscape transformation and LULCC, are key drivers of urban thermal environment deterioration. Studying the detailed history of LULCC in a specific area can shed light on the spatial extent and severity of these changes (Xiao and Weng, 2007). The trend of changes in the soil areas in the period is decreasing (from 42.59% to 38.18%), and on the other hand, the trend of changes in the built-up areas is increasing (from 38% to 45.45%), this increase was taken place mainly by removing the natural vegetation cover (Basu and Das, 2023). This phenomenon results from the simultaneous increase in impermeable surfaces and reduction in vegetation cover within urban settings (Wang and Akbari, 2016; Tanoori et al., 2024).

The configuration metrics in the class from 2006 to 2021 were investigated to better distinguish landscape changes during study time. Fig. 6 shows the LSI trend changes and an increase means an increase in inner edge levels and a decrease in patch accumulation due to more irregular and complicated geometry (Ding et al., 2022). In all land cover types, it is noticed an increase from 2006 to 2014 and then a little decrease. ED equals the sum of the lengths of all edge segments involving the corresponding patch type, divided by the

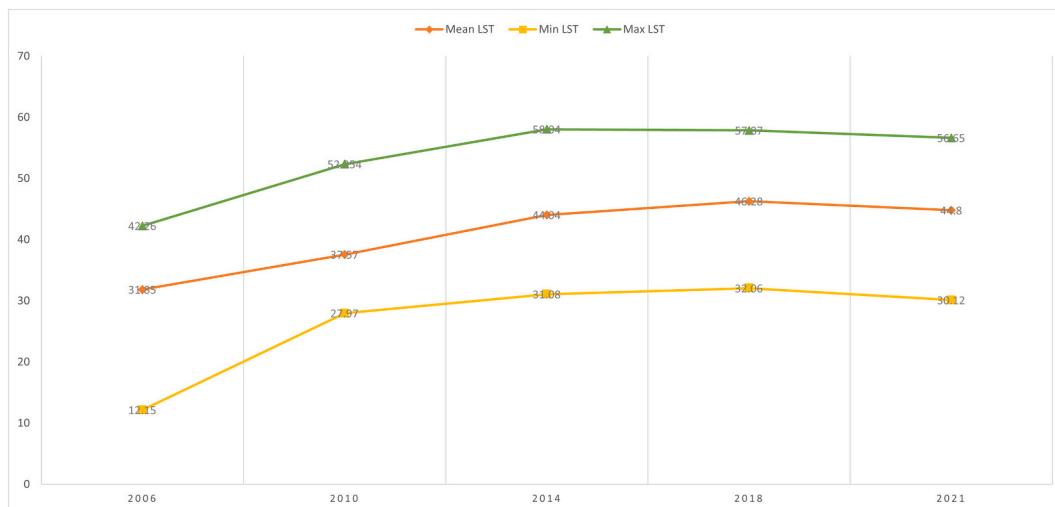


Fig. 4. Shiraz LST changes during 2006–2021.

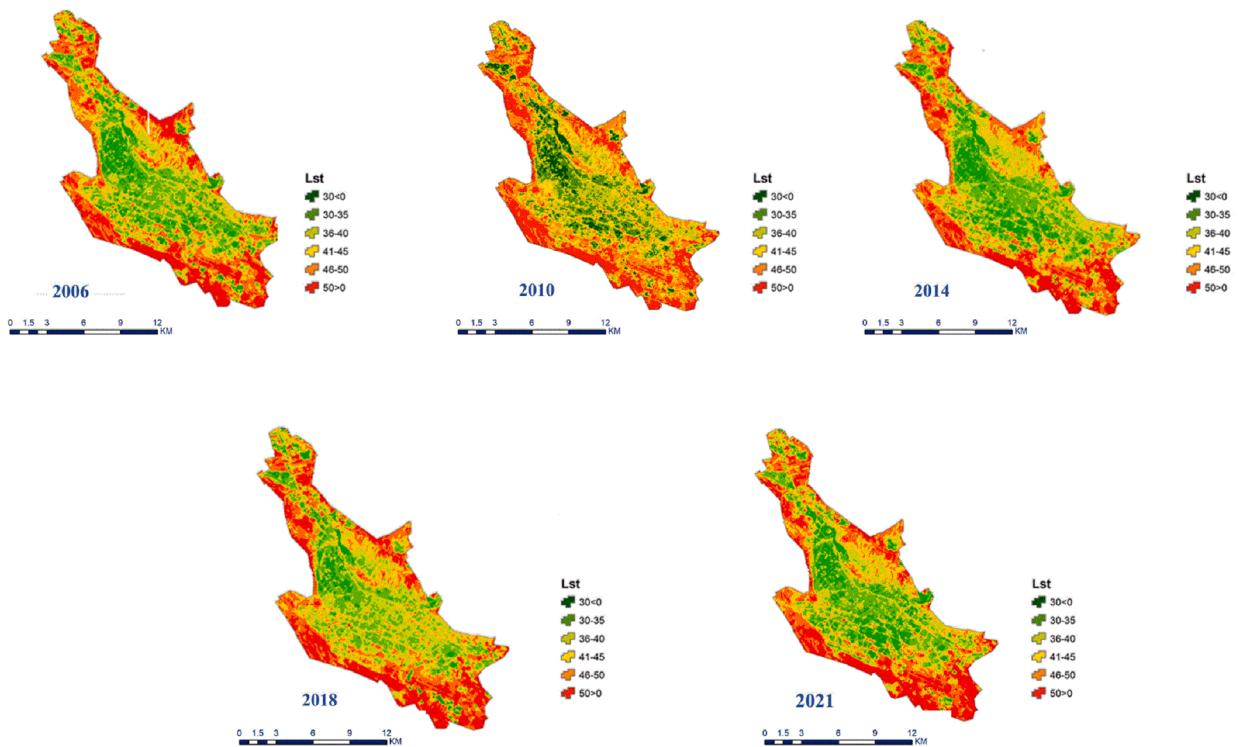


Fig. 5. LST of shiraz, (a); 2006, (b); 2010, (c); 2014, (d); 2018, (e); 2021.

total landscape area (Yan et al., 2019). The results showed that the connection and continuity of the landscape have decreased. The LPI metric of the built-up area increased from 2006 to 2021 which indicated that the largest patch had become larger (Wang et al., 2020) unlike other land cover types. The FRAC results showed decreasing complexity and fragmentation for built-up areas and an increasing complexity and fragmentation of vegetation and soil areas. Contrasting trends in the PLAND metric, the built-up areas, and vegetation experienced an increase, indicating an expansion, while soil areas witnessed a decrease in patch sizes.

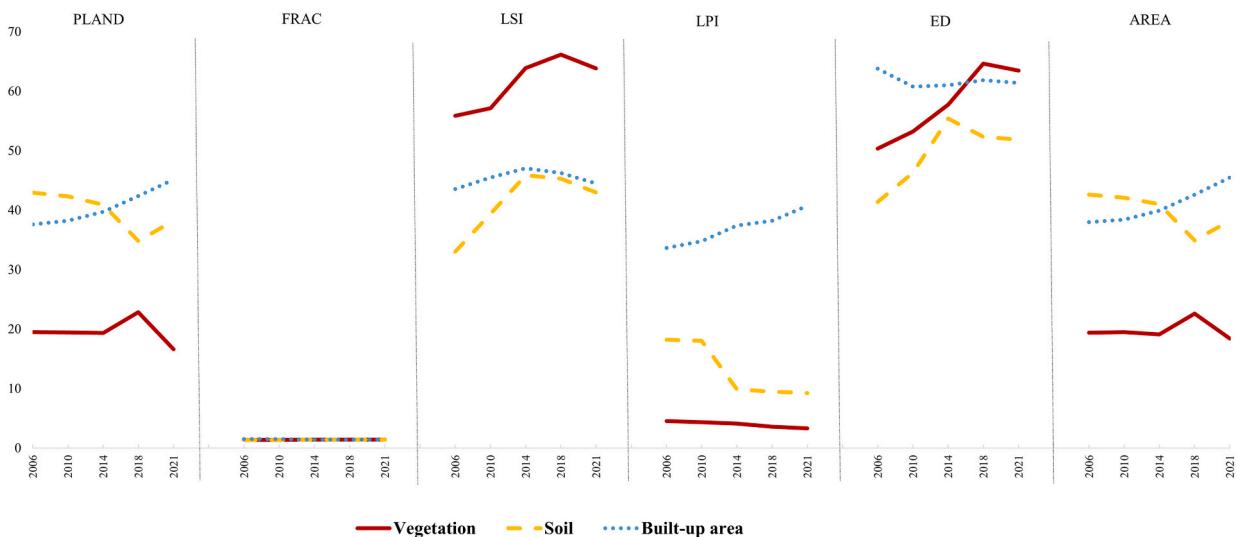


Fig. 6. LULCC and Configuration metrics changes in Shiraz between 2006 and 2021.

3.3. Machine learning prediction results

3.3.1. Hyperparameter tuning

Choosing the right hyper-parameter configuration for machine learning models is crucial for achieving optimal performance (Soltani and Lee, 2024). This task requires a deep understanding of machine learning algorithms and the appropriate techniques for hyper-parameter optimization. While there are automatic optimization methods available, they each have their strengths and limitations, and their effectiveness can vary depending on the type of problem being tackled (Yang and Shami, 2020).

A Random Search approach was employed to explore the hyperparameter space and identify optimal or near-optimal configurations for each algorithm. By utilizing Random Search, it is aimed to find the best hyperparameter settings for each algorithm, thereby maximizing their performance in predicting temperature from the given LULC features and geographic coordinates. The parameters of each of the machine algorithms and their explore values of parameters are described in Table 1.

For DTs, the complexity of trees is controlled through pruning techniques and by limiting the tree depth. This helps prevent the trees from becoming overly complex and reduces the likelihood of overfitting. Additionally, for algorithms such as XGB, RF, and AdaBoost that utilize DTs as base learners, similar measures are taken to control tree complexity and limit depth. For deep learning models, which are prone to overfitting, L2 regularization is employed. This regularization technique helps prevent the model from overemphasizing specific patterns in the training data, promoting better generalization.

3.3.2. Evaluation of the accuracy of machine learning algorithms

In this section, the results of the analysis using various machine learning algorithms for predicting LST in built-up areas, soil, and vegetation are presented. To assess the accuracy of prediction results, four different metrics were used to evaluate the performance of each regression algorithm across different land cover types. The machine learning models were implemented in two different scenarios. The first scenario involved applying 5-fold cross-validation. This involved splitting the dataset into five equal folds, with each fold being used as the test set once, and the remaining folds being used as the training set. This process was repeated five times to obtain a more robust evaluation of the models. The resulting RMSE, AAPRE, R², and CI values were calculated as the averages of the five different test sets, presented in Table 2.

Specifically, the RMSE metric is employed with 5-fold cross-validation. A low RMSE value is indicative of good performance of the prediction approach (Suthar et al., 2024). The results showed that the DNN algorithm performed the best overall, with the lowest error observed in built-up areas and soil. However, in vegetation areas, XGB performed the best, followed by DNN. These two algorithms consistently illustrated low RMSE values across all applications, indicating their effectiveness in accurately estimating LST and being well-suited for LST prediction in the study. The high performance of DNN and XGB can be attributed to their specific capabilities in capturing and modeling complex relationships between the input features, which include various configuration metrics and LST. In terms of AAPRE, it appears that the DNN is the most accurate algorithm for estimating or predicting built-up areas, followed by XGB. In soil results, DNN had the lowest AAPRE of 2.174%, indicating high accuracy in predicting LST values compared to other machine learning algorithms. XGB had the second-lowest AAPRE of 8.63%, suggesting good accuracy. Additionally, DNN showed the highest accuracy in predicting LST, in vegetation and soil configuration metrics, compared to the other algorithms.

R² is, quantifies the model's precision (Suthar et al., 2024), such as built-up, soil, and vegetation areas. The results of the study showed that XGB had the highest precision for built-up areas, with a score of 0.912, followed by DNN, RF, and AdaBoost. In soil areas, XGB was found to have the highest precision, with a score of 0.932, followed by AdaBoost, RF, and DNN, respectively. The precision levels for vegetation were close, with XGB and DNN both showing high levels of precision, with scores of 0.8974 and 0.8968, respectively.

CI is a metric for assessing the predictive performance of various algorithms. A higher CI value indicates a better ability of the model to correctly order or rank the predicted values compared to the true values. The study results reveal that XGB is the most effective

Table 1
Algorithms Parameters.

Algorithm		Values explored
SVM	Kernel ='rbf'	Kernel = {'rbf', 'linear'}
DTree	depth of trees = 8	depth of trees = {4,8,10,16,32}
RF	Number of estimators =100 depth of trees = 10	depth of trees = {4,8,10,16,32} Number of estimators = {100,200...,1000}
ADABOOST	Number of estimators =100 depth of trees = 8	depth of trees = {4,8,10,16,32} Number of estimators = {100,200...,1000}
XGB	Number of estimators =1000 depth of trees = 4	depth of trees = {4,8,10,16,32} Number of estimators = {100,200...,1000}
DNN	Number of layers =5, number units of layers = [128,64,32,16,1], Learning-rate = 0.0005, epoch = 1500, Batch-size =256, l2-regularization	Number of layers = {1,2,3,4,5,6,7} Num of units = {16,32,64,128,256} Learning-rate = {0.0005,0.001,0.005,0.01} Batch-size = {64,128,256}

Table 2

Evaluation of the Accuracy of Machine Learning Algorithms for Scenario One.

Metrics	BUILD UP				SOIL				VEGETATION			
	RMSE	AAPRE%	R ²	CI	RMSE	AAPRE%	R ²	CI	RMSE	AAPRE%	R ²	CI
SVM	2.8506				2.5026				3.1560			
	±0.103	9.194	0.693	0.9094	±0.109	10.371	0.83	0.8928	±0.110	8.755	0.574	0.8777
DTREE	2.4605				2.3413				2.7483			
	±0.047	9.702	0.792	0.8841	±0.10	10.519	0.853	0.9005	±0.065	9.036	0.723	0.8575
RF	2.1431				2.1018				2.2843			
	±0.057	9.505	0.875	0.9065	±0.122	10.338	0.915	0.9189	±0.098	8.872	0.82	0.8882
ADABOOST	2.1627				2.1117				2.3036			
	±0.053	9.471	0.873	0.9059	±0.130	10.323	0.916	0.917	±0.099	8.816	0.818	0.885
XGB	1.9069				2.0388				1.9949			
	±0.049	7.78	0.912	0.9146	±0.079	8.53	0.932	0.9192	±0.089	7.184	0.876	0.8974
DNN	1.8401				1.99							
	±0.082	2.444	0.886	0.9137	±0.096	2.074	0.908	0.9186	2.0565 ± 0.073	2.696	0.796	0.8968

algorithm regarding concordance ability in built-up areas, followed by DNN. In soil configuration metrics XGB outperforms other algorithms in ordering or ranking the predicted values compared to the true values, followed by RF, DNN, and AdaBoost. Similarly, DNN and XGB are the most suitable algorithms in vegetation, indicating their better predictive performance in this category. It is noteworthy that XGB consistently achieves the highest CI value 0.9146 in the built-up areas, 0.9192 in soil areas, and 0.9192 in vegetation areas indicating its superior ability to accurately rank and order the predicted temperature values compared to the actual values.

Based on the second scenario, where all samples of 2006–2018 were used as the training set and LST in 2021 was predicted, the results were compiled and presented in Table 3. XGB was found to perform the best with the lowest error in all LULC metrics. The study suggests that the DNN algorithm is the most accurate for predicting LST in built-up areas and soil. On the other hand, the XGB algorithm was found to be the most effective in predicting LST by configuration metrics in vegetation. Similar to the previous section, the CI results showed that XGB performed the best in terms of predictive performance. Overall, the study highlights XGB as the best model for quantifying precision in all LULC metrics.

3.4. LST prediction from machine learning algorithms

The intensity of LST is directly related to the rate of urbanization, land use patterns, and building density (Ahmed et al., 2013). LST is related to the patterns of LULCC, for example, the composition of the built-up area, vegetation, and water bodies (Mustafa et al., 2020). The findings showed that XGB and DNN were the most effective algorithms for predicting LST in both scenarios. The combination of low RMSE and high CI values for XGB highlights its effectiveness as a robust and reliable algorithm for LST prediction in various applications. The powerful modeling capabilities of XGB contribute to its superior performance, making it a valuable tool in LST predictions.

In addition, the results were visualised to show the effectiveness of each model. In this visualisation, the focus was on scenario one, to predict the LST of a test set in this section. The results indicate that the algorithms used were capable of accurately simulating and predicting LST values for future epochs. The observed and predicted values closely matched (Maduako et al., 2016). These visualizations provide a clear representation of the predicted LST values for different land use categories using the DNN and XGB models. The temperature ranges depicted in the graphs highlight the variability and potential impact of land use on LST patterns in Shiraz. The supplementary section contains figures that illustrate the results of LST prediction by various algorithms, these figures are presented for further comparison and analysis (Appendix A).

3.5. Uncertainty assessment

Standard prediction methods employed by machine learning algorithms provide a single point estimate for each new instance,

Table 3

Evaluation of the accuracy of machine learning algorithms for scenario two.

Metrics	BUILD UP				SOIL				VEGETATION			
	RMSE	AAPRE%	CI	R ²	RMSE	AAPRE%	CI	R ²	RMSE	AAPRE%	CI	R ²
SVM	4.184	8.155	0.659	0.522	3.422	6.087	0.788	0.65	3.954	7.167	0.632	0.455
DTREE	2.785	7.091	0.79	0.639	2.305	4.995	0.808	0.758	2.658	6.587	0.812	0.761
RF	2.432	6.759	0.838	0.812	2.122	4.79	0.864	0.867	2.491	6.444	0.813	0.837
ADABOOST	2.479	6.615	0.818	0.79	2.151	4.725	0.835	0.85	2.63	6.371	0.779	0.775
XGB	2.237	6.447	0.893	0.856	2.1	4.718	0.876	0.88	2.377	6.144	0.878	0.886
DNN	2.468	5.698	0.836	0.793	2.277	4.606	0.871	0.862	2.623	6.367	0.87	0.751

lacking information about the dispersion of observations around the predicted value and the confidence level of the prediction (Meinshausen and Ridgeway, 2006). Therefore, XGB Quantile Regression, Quantile Regression Forests, and Deep Learning-based models were used, which are capable of estimating conditional distributions.

To evaluate these models, the concentration of probabilities, also known as prediction intervals, were estimated. The sharpness metric quantifies the tightness of the predictive densities, where a narrower interval length indicates greater confidence in the predictions. The procedure for calculating sharpness is as follows:

1. Choose the prediction interval for evaluation.
2. Iterate through the predictions of the mean and standard deviation for each observation in the test set.
3. Calculate the boundaries of the interval for each observation.
4. Determine the length of the interval.
5. Calculate the average length across all intervals.

For XGB and Random Forest models, the procedure differs slightly due to predictions for multiple quantiles. To calculate the interval width corresponding to 60%, two quantiles (20% and 80%) were considered to compute the interval between these two using the difference in predictions. The sharpness at intervals of 60%, 70%, 80%, and 90% were calculated and compared the three models using this metric. The results are depicted in Fig. 7, illustrating the sharpness at different intervals for the built-up areas configuration, as well as for soil and vegetation.

As the interval coverage increases from 60% to 90%, the average prediction interval of the RF model expands at a much higher rate compared to the other algorithms in all three configurations. This indicates that the RF model may not provide optimal results for high-

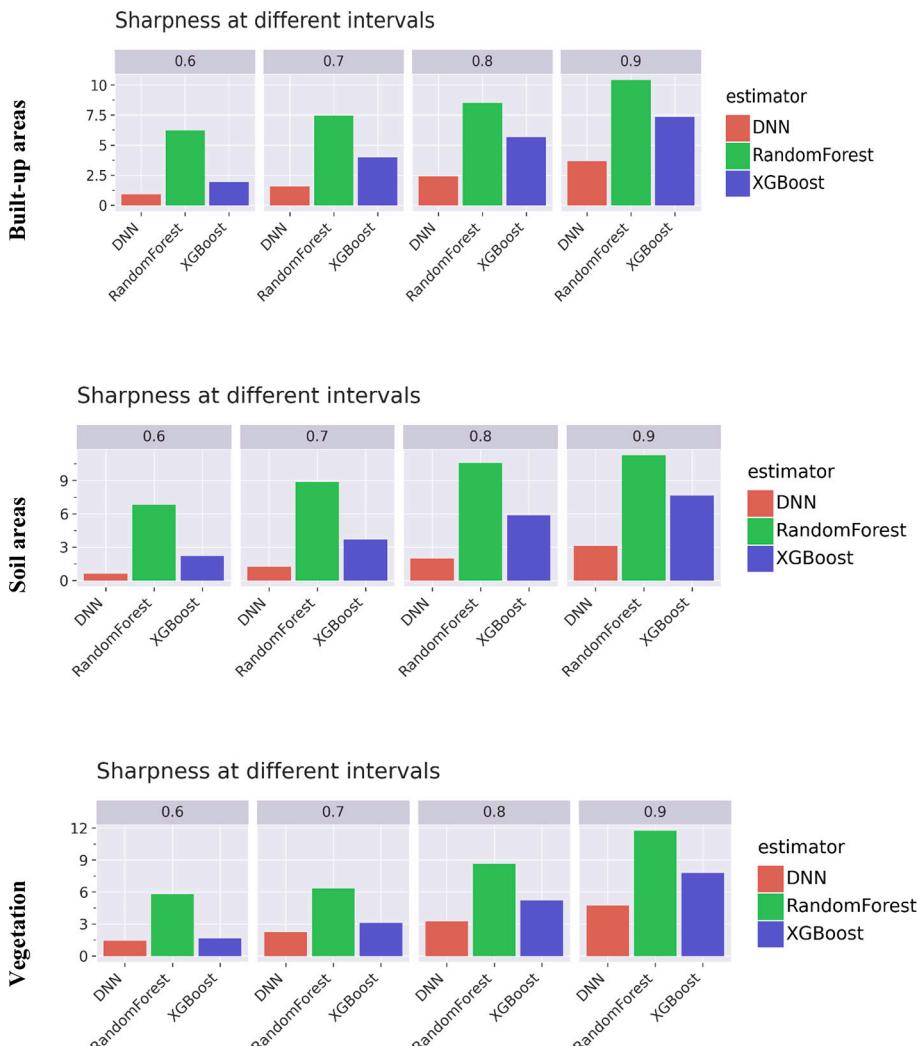


Fig. 7. The result of sharpness at different intervals.

interval coverage. On the other hand, the DNN demonstrates the highest confidence in the predictive densities among all three configurations.

Additionally, the proportion of real observations was assessed that fall outside the predictive boundaries of a given interval. Fig. 8 presents the number of missing observations within the interval. As the interval increases, the proportion of missing observations decreases. In this evaluation, XGB consistently yielded the best results across all three configurations (built-up, soil, vegetation).

Overall, the analysis indicates that XGB and deep learning models outperform the other algorithms in terms of sharpness, RMSE, CI, R², and AAPRE, while capturing the true values across all configurations. Deep learning demonstrates superior performance in terms of interval length, while XGB shows the highest confidence in the number of missing observations within the interval.

3.6. Feature importance of configuration metrics

In this section, the importance of each configuration metric was evaluated using the SHAP algorithm in combination with various algorithms. This approach helps to determine the significance of different metrics in predicting LST for specific land use categories. In the analysis of the factors influencing changes in temperature for various applications, it is sought to assess the effect of each factor, including (FRAC, LPI, LSI, PLAND, ED, and year) (Figs. 9 to 11).

The summary plot combines configuration metrics importance with LST effects. Each point on the summary plot is a Shapley value for a feature and an instance. The position on the y-axis is determined by the year and configuration metrics and on the x-axis LST by the Shapley value. The color represents the value of the feature from low to high. Overlapping points are jittered in the y-axis direction, therefore, there is a sense of the distribution of the Shapley values per feature. The features are ordered according to their importance.

A. Built-up area

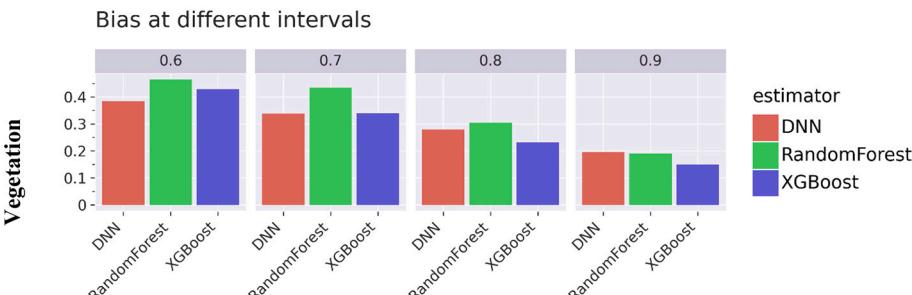
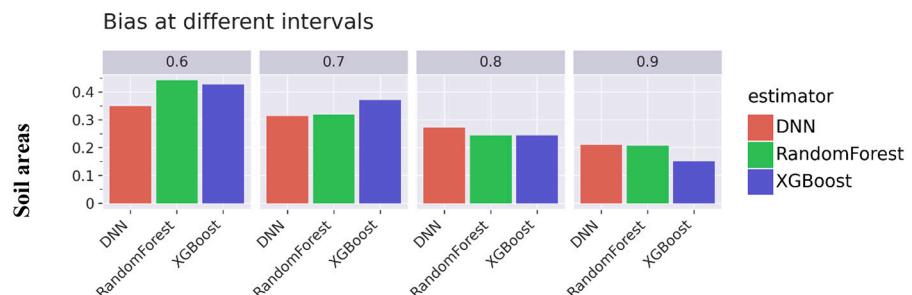
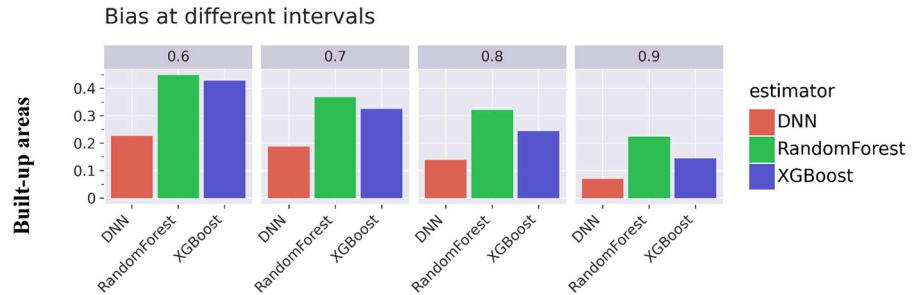


Fig. 8. the proportion of missing observation.

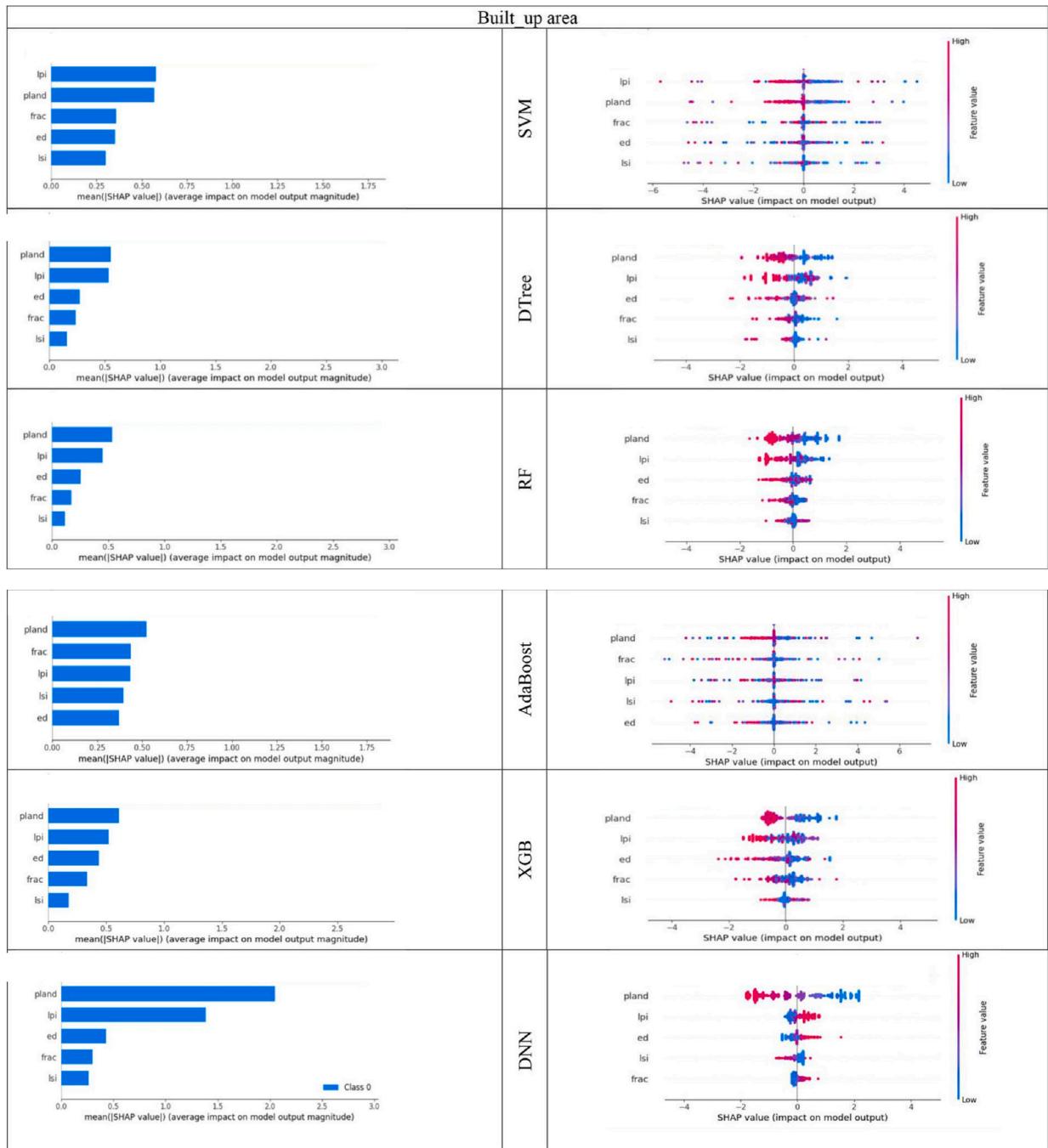


Fig. 9. The importance of configuration metrics in built-up area land use through machine learning algorithms and summary plots in a Shapley value.

When specifically considering the average SHAP values for the SVM regression model, the analysis revealed that the LPI and PLAND were the most important factors influencing LST changes in the built-up area. Following LPI and PLAND, the FRAC, ED, and LSI were ranked next in terms of importance. The summary plot, which combines the importance of configuration metrics with their effects on LST, indicated that an increase in LPI led to a decrease in LST. Similarly, an increase in PLAND also resulted in a decrease in LST.

The DTree model analysis revealed that PLAND and LPI had significant effects on LST for the built-up zone. These two metrics were identified as important factors influencing LST changes in built-up areas. Following PLAND and LPI, the analysis ranked ED, FRAC, and LSI in the next positions in terms of importance. The summary plot, which presents the combined information on the importance of configuration metrics and their effects on LST, indicated that an increase in PLAND and LPI led to a decrease in LST.

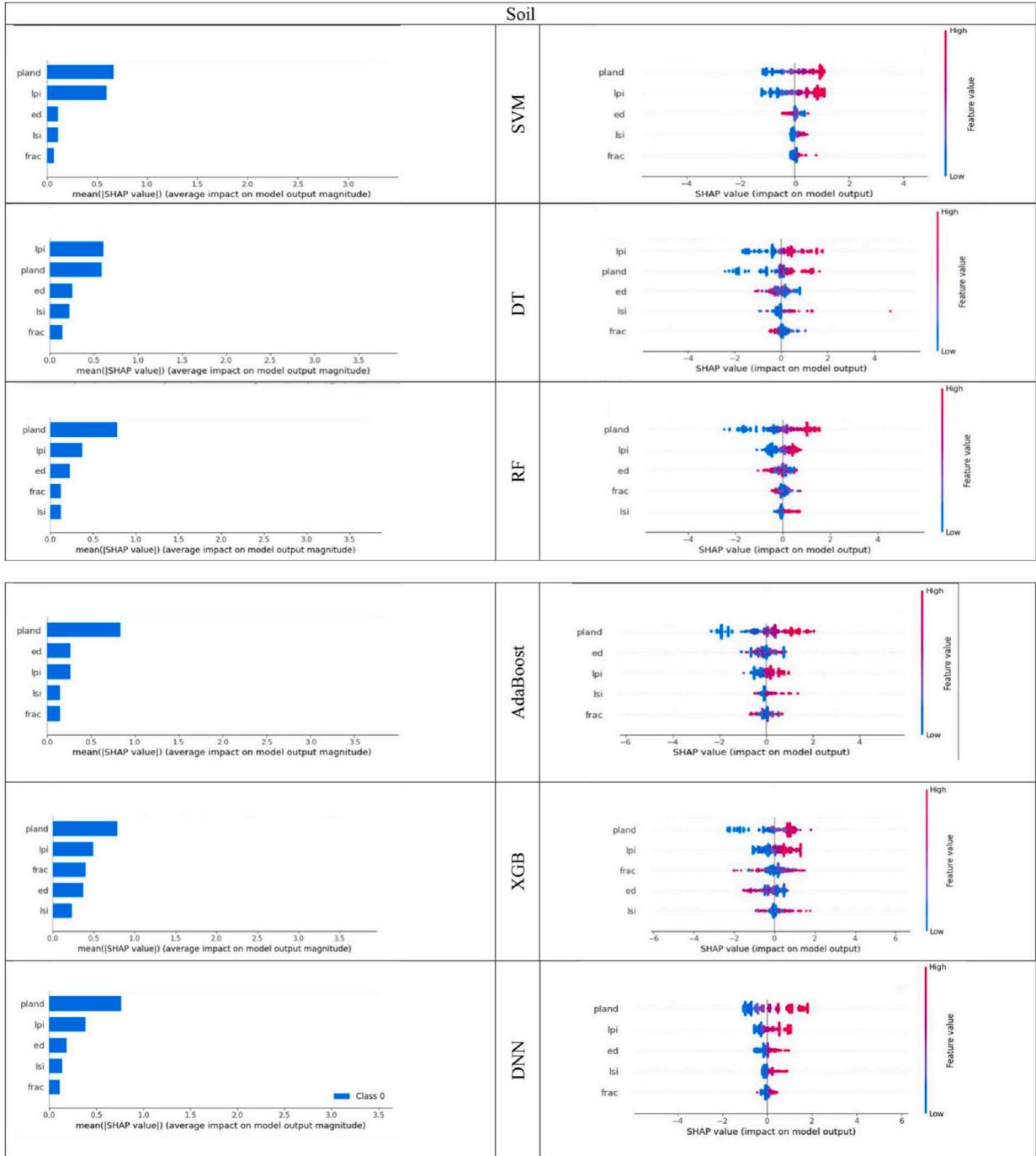


Fig. 10. The importance of configuration metrics in soil land use through machine learning algorithms and summary plots in a Shapley value.

The RF analysis confirmed that PLAND and LPI had significant effects on LST. These two metrics were identified as important factors influencing LST changes in built-up areas. Following PLAND and LPI, the analysis ranked ED, FRAC, and LSI in the next positions in terms of importance. The summary plot, which combines the information on the importance of configuration metrics and their effects on LST, indicated that an increase in PLAND and LPI resulted in a decrease in LST. The importance of configuration metrics obtained from the RF model aligns with the results from the DTs model.

The AdaBoost analysis revealed that PLAND and FRAC had much more significant effects on LST. These two metrics were identified as important factors influencing LST changes in built-up areas. Following the analysis ranked LP, LSI, and ED in the next positions in terms of importance. The summary plot, which presents the combined information on the importance of configuration metrics and their

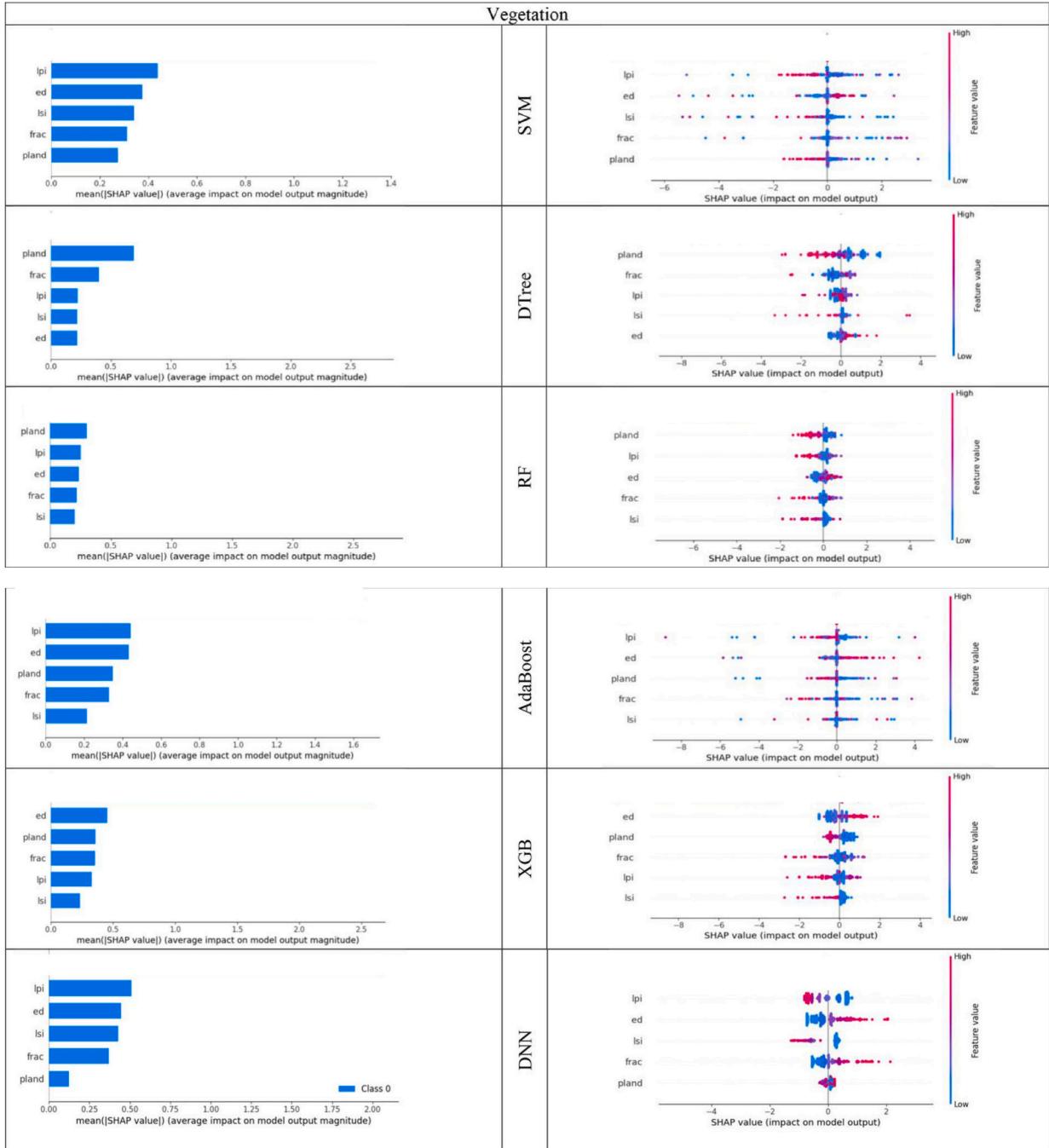


Fig. 11. The importance of configuration metrics in vegetation land use through machine learning algorithms and summary plots in a Shapley value.

effects on LST, indicated that an increase in PLAND and FRAC led to a decrease in LST.

The XGB analysis demonstrated that PLAND and LPI had particularly strong and significant effects on LST. These two metrics emerged as pivotal factors influencing LST changes in built-up areas. Moreover, the analysis ranked ED, FRAC, and LSI in subsequent positions of importance. The summary plot, which effectively combines the information on the significance of configuration metrics and their impact on LST, corroborated the results obtained from previous regression models. It reaffirmed that an increase in PLAND and LPI corresponded to a notable decrease in LST.

The DNN model demonstrated that PLAND and LPI had particularly strong and significant effects on LST. These two metrics emerged as pivotal factors influencing LST changes in built-up areas. Moreover, the analysis ranked ED, FRAC, and LSI in subsequent

positions of importance. The summary plot, which effectively combines the information on the significance of configuration metrics and their impact on LST, corroborated the results obtained from previous regression models. It reaffirmed that an increase in PLAND and LPI corresponded to a significant decrease in LST ([Figs. 9](#)).

B. SOIL

The first regression model was SVM, the result presented in soil land use, PLAND, and LPI were the most important factors influencing LST changes. Following PLAND, and LPI the ED, LSI, AND FRAC were ranked next in terms of importance. The summary plot indicated an increase in PLAND and LPI led to an increase in LST.

The DT model revealed that for the soil category, LPI, and PLAND were the most significant factors that led to changes in LST. After PLAND and LPI, ED, LSI, and FRAC were ranked as next in importance. The summary plot suggested that an increase in PLAND and LPI increased LST.

The analysis of RF model confirmed that PLAND and LPI had significant effects on LST. These two metrics were identified as important factors influencing LST changes. Following PLAND and LPI, the analysis ranked ED, FRAC, and LSI in the next positions in terms of importance. The summary plot, which combines the information on the importance of configuration metrics and their effects on LST, indicated that an increase in PLAND and LPI increased LST.

The AdaBoost analysis revealed that PLAND and ED had much more significant effects on LST. Following the analysis ranked LPI, LSI, and FRAC in the next positions in terms of importance. The summary plot, which presents the combined information on the importance of configuration metrics and their effects on LST, indicated that an increase in PLAND and ED led to an increase in LST.

The XGB model demonstrated that PLAND and LPI had particularly strong and significant effects on LST. These two metrics emerged as pivotal factors influencing LST changes in soil land use. Moreover, the analysis ranked ED, FRAC, and LSI in subsequent positions of importance. It reaffirmed that an increase in PLAND and LPI corresponded to a notable increase in LST.

The analysis of DNN model demonstrated that PLAND and LPI had particularly strong and significant effects on LST. These two metrics emerged as pivotal factors influencing LST changes in soil land use. Moreover, the analysis ranked ED, LSI, and FRAC in subsequent positions of importance. The summary plot, which effectively combines the information on the significance of configuration metrics and their impact on LST, corroborated the results obtained from previous regression models. It reaffirmed that an increase in PLAND and LPI corresponded to a notable increase in LST ([Fig. 10](#)).

C. VEGETATION

According to the SVM model, the average SHAP values highlighted the significance of two key factors, namely the LPI and ED, in influencing LST variations in the vegetation category. Following LPI and ED, LSI, FRAC, and PLAND were ranked next. The summary plot, which combines the importance of configuration metrics with their effects on LST, provided further insights. It indicated that an increase in LPI corresponded to a decrease in LST.

As per the analysis, the DT model indicated that changes in LST for vegetation were primarily influenced by PLAND, and FRAC. Following these factors, LPI, LSI, and ED were ranked as the next most important factors. The summary plot supported the results obtained suggesting that an increase in PLAND and FRAC led to a decrease in LST.

The RF model for vegetation confirmed that PLAND and LPI had significant effects on LST. These two metrics were identified as important factors influencing LST changes. Following PLAND and LPI, the analysis ranked ED, FRAC, and LSI in the next positions in terms of importance. The summary plot, which combines the information on the importance of configuration metrics and their effects on LST, indicated that an increase in PLAND and LPI decreases LST.

The AdaBoost analysis revealed that LPI and ED had much more significant effects on LST. These two metrics were identified as important factors influencing LST changes in vegetation. Following the analysis ranked PLAND, FRAC, and LSI in the next positions in terms of importance. The summary plot, which presents the combined information on the importance of configuration metrics and their effects on LST, indicated that an increase in LPI led to a decrease in LST and an increase in ED led to an increase in LST.

The XGB analysis demonstrated that ED and PLAND had particularly strong and significant effects on LST. These two metrics emerged as pivotal factors influencing LST changes in vegetation land use. Moreover, the analysis ranked FRAC, LPI, and LSI in subsequent positions of importance. It reaffirmed that an increase in ED corresponded to a notable increase in LST and an increase in PLAND led to a decrease in LST.

The DNN analysis demonstrated that LPI and ED had particularly strong and significant effects on LST. These two metrics emerged as pivotal factors influencing LST changes in vegetation land use. Moreover, the analysis ranked LSI, FRAC, and PLAND in subsequent positions of importance. The summary plot, which effectively combines the information on the significance of configuration metrics and their impact on LST, corroborated the results obtained from previous models. It reaffirmed that an increase in LPI corresponded to a decrease in LST and an increase in ED led to an increase in LST ([Fig. 11](#)).

3.7. Select the most important configuration metrics in each land use

In this part, it is described which configuration metrics had the most important role in each land use. These findings indicate that different configuration metrics play a varying role in influencing LST within different land use categories.

Analyzing the significance of configuration metrics in understanding LST variations across different land use types provides valuable insights. It is important to note that the importance of configuration metrics varied for each land use, and this ranking was

obtained through the utilization of various machine learning algorithms.

First, it is made decisions related to the built-up area land use. The average rank of configuration metrics that influenced LST revealed the following order: PLAND, LPI, ED, FRAC, and LSI. This means that the PLAND metric had the highest importance, followed by the LPI, ED, FRAC, and LSI. These results provide valuable insights into the factors driving LST variations in built-up areas. The high rank of PLAND suggests that the proportion of built-up areas has a substantial influence on LST. Additionally, the presence of larger and more connected patches of built-up areas, as indicated by LPI, contributes to lower LST. On the other hand, higher edge density and increased fractal dimension of the built-up areas may lead to higher LST. The Landscape Shape Index also plays a role, although it is ranked lower in terms of importance in most machine learning algorithms.

Next, the soil category was examined to assess the impact of configuration metrics on LST. The PLAND metric had the greatest importance in influencing LST, followed by LPI, ED, LSI, and FRAC. These findings shed light on the factors driving LST variations in soil area. The high rank of PLAND suggests that the proportion of soil areas within the landscape significantly affects LST. Furthermore, larger and more connected patches of soil areas, as indicated, contribute to lower LST. Conversely, higher edge density in the soil areas may lead to higher LST. The roles of landscape shape complexity and fractal dimension are also noteworthy, although they rank lower in importance.

The average rank of the configuration metrics influencing LST in vegetation category revealed distinct results compared to built-up and soil zones. Different algorithms yielded varying rankings of the importance of these metrics. In this case, utilizing the average rank of metrics provides a more reliable assessment, as it takes into account the collective influence of the metrics and because the first 3 places have close value. The average rank of the configuration metrics for vegetation land use resulted in the following order: LPI, ED, PLAND, FRAC, and LSI. This suggests that LPI had the highest importance in influencing LST, ED, PLAND, FRAC, and LSI. These findings highlight the varying importance of configuration metrics in driving LST variations in the vegetation land use category. The high rank of LPI indicates that larger and more connected patches of vegetation have a significant impact on reducing LST. The ED metric suggests that higher edge density, which indicates fragmented vegetation patches, is associated with increased LST. Furthermore, PLAND, albeit at a slightly lower rank, indicates that the proportion of vegetation within the landscape affects LST. FRAC and LSI also play a role, although they are ranked lower in importance.

Table 4 highlights the order of importance of configuration metrics across the land use categories. In both built-up and soil areas, PLAND, LPI, and ED were identified as the three most influential metrics in determining LST. This suggests that the proportion of the land covered by specific land use, the size and connectivity of the patches, and the edge density play crucial roles in influencing LST in both built-up and soil areas.

On the other hand, for vegetation category, the order of importance differs slightly. In this case, LPI was identified as the most important metric, followed by ED and then PLAND. This indicates that, in vegetation areas, the size and connectivity of patches and the edge density have a greater influence on LST compared to the proportion of vegetation in the landscape. Understanding the unique importance of these configuration metrics in each land use category can guide targeted interventions and management strategies to mitigate temperature effects and enhance thermal comfort.

4. Conclusion

The impact of different land covers on UHI has been extensively researched and documented in UHI literature. However, the degree of impact can vary depending on the local climate (Guo et al., 2020; Sharifi and Soltani, 2017). Previous studies have demonstrated that the transformation of the landscape and the LULCC resulting from urban expansion and regeneration are the primary factors responsible for the degradation of the urban thermal environment.

To investigate the impact of urban configuration on LST, landscape metrics (PLAND, LSI, LPI, ED, and FRAC) were appropriately selected. The findings indicate that the metropolitan region boundaries tend to exhibit higher LST values and that different UHI patterns were discernible based on LULC types. Specifically, vegetation and built-up spots were found to exhaust lower LST values.

The performance of six machine-learning algorithms was evaluated and compared using k-fold cross-validation. To ensure that the models were able to generalize, they were trained using configuration metrics data from 2006 to 2018 and tested with unseen data from 2021.

The study utilized four different metrics, namely RMSE, AAPRE, R², and CI, to evaluate the accuracy of the prediction results. The findings revealed that the XGB model exhibited the highest level of predictive performance compared to the other models tested, with DNN being the next best performing model. Additionally, the study performed an estimation of the length of the prediction interval and the number of missing observations within the interval, in order to evaluate the confidence of machine learning methods in predicting LST. The findings revealed that both XGB and DNN models exhibited a high level of confidence in predicting LST.

Table 4
Order of Importance of Configuration Metrics.

LAND LULC	Rank1	Rank2	Rank3	Rank4	Rank5
Built-up Area	PLAND	LPI	ED	FRAC	LSI
Soil	PLAND	LPI	ED	LSI	FRAC
Vegetation	LPI	ED	PLAND	FRAC	LSI

Additionally, the importance of different configuration metrics for each land use category were evaluated. In the built-up and soil zones, metrics such as PLAND, LPI, and ED emerged as the three most influential metrics in determining LST. This suggests that the proportion of land covered by specific land use activities, the size and connectivity of patches, and the edge density significantly influence LST in both built-up and soil areas. However, for vegetation land use, the order of importance differed slightly, with LPI identified as the most important metric, followed by ED and then PLAND. This indicates that in vegetation areas, the size and connectivity of patches and the edge density have a greater influence on LST compared to the proportion of vegetation in the landscape. Understanding the unique importance of these configuration metrics in each land use category can guide targeted interventions and management strategies to mitigate temperature effects and enhance thermal comfort in specific areas. Overall, this study underscores the significance of machine learning algorithms, particularly DNN and XGB, in predicting LST and comprehending the UHI effect in Shiraz. These models can serve as valuable tools for temperature predictions in urban environments, supporting urban planning, climate mitigation strategies, and decision-making processes aimed at improving the urban microclimate.

Building on this study, future research could explore long-term urban configuration changes and their impact on LST. Social factors influencing land use decisions, the role of building materials, and microclimates within specific areas are potential areas of investigation. Additionally, cost-effective interventions, the influence of vegetation types, and citizen science programs for LST data collection could be valuable avenues for further research. Finally, translating these findings into specific policy recommendations for urban planning in Shiraz could significantly contribute to creating a cooler and more comfortable city.

CRediT authorship contribution statement

Ghazaleh Tanoori: Conceptualization, Formal analysis, Methodology, Software, Writing – original draft, Writing – review & editing. **Ali Soltani:** Methodology, Visualisation, Formal analysis, Writing – review & editing. **Atoosa Modiri:** Resources, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Data availability

Data will be made available on request.

Acknowledgment

The authors would like to express their sincere gratitude to Dr. Betsabeh Tanoori for her invaluable contributions and support throughout the completion of this research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.uclim.2024.101962>.

References

- Ahmed, B., Kamruzzaman, M., Zhu, X., Rahman, M.S., Choi, K., 2013. Simulating land cover changes and their impacts on land surface temperature in Dhaka, Bangladesh. *Remote Sens* 5 (11), 5969–5998. <https://doi.org/10.3390/rs5115969>.
- Alabdullah, A., Ohlsson, M., Pashami, S., Rögnvaldsson, T., 2024. The concordance index decomposition: a measure for a deeper understanding of survival prediction models. *Artif. Intell. Med.* 148, 102781.
- Allan, A., Soltani, A., Abdi, M.H., Zarei, M., 2022. Driving forces behind land use and land cover change: a systematic and bibliometric review. *Land* 11 (8), 1222. <https://doi.org/10.3390/land11081222>.
- Athukorala, D., Murayama, Y., 2020. Spatial variation of land use/cover composition and impact on surface urban heat island in a tropical sub-Saharan city of Accra, Ghana. *Sustainability* 12 (19), 7953.
- Awad, M., Khanna, R., 2015. Support Vector Regression. Efficient Learning Machines. Apress, Berkeley, CA. In.
- Azedou, A., Amine, A., Kisekka, I., Lahssini, S., Bouziani, Y., Moukrim, S., 2023. Enhancing Land Cover/Land Use (LCLU) classification through a comparative analysis of hyperparameters optimization approaches for deep neural network (DNN). *Ecol. Inform.* 78, 102333.
- Azhdari, A., Soltani, A., Alidadi, M., 2018. Urban morphology and landscape structure effect on land surface temperature: evidence from shiraz, a semi-arid city. *Sustain. Cities Soc.* 41, 853–864. <https://doi.org/10.1016/j.scs.2018.06.034>.
- Aziziz, P., Soltani, A., Bagheri, F., Sharifi, S., Mikaeili, M., 2022. An integrated modelling approach to urban growth and land use/cover change. *Land* 11 (10), 1715. <https://doi.org/10.3390/land11101715>.
- Bagheri, B., Soltani, A., 2023. The spatio-temporal dynamics of urban growth and population in metropolitan regions of Iran. *Habitat Int.* 136, 102797. <https://doi.org/10.1016/j.habitatint.2023.102797>.
- Basu, T., Das, A., 2023. Urbanization induced degradation of urban green space and its association to the land surface temperature in a medium-class city in India. *Sustain. Cities Soc.* 90, 104373.
- Bentéjac, C., Csörgő, A., Martínez-Muñoz, G., 2021. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* 54, 1937–1967.

- Bhandari, R., Maithani, S., Karnataka, H., 2022. A deep neural network-based approach for studying the relationship between land surface temperature and surface air temperature. *J. Indian Soc. Remote Sens.* 50 (3), 563–568.
- Cao, Q., Yu, D., Georgescu, M., Han, Z., Wu, J., 2015. Impacts of land use and land cover change on regional climate: a case study in the agro-pastoral transitional zone of China. *Environ. Res. Lett.* 10 (12), 124025.
- Cheela, V.S., John, M., Biswas, W., Sarker, P., 2021. Combating urban heat island effect—a review of reflective pavements and tree shading strategies. *Buildings* 11 (3), 93.
- Chen, Y., Dou, P., Yang, X., 2017. Improving land use/cover classification with a multiple classifier system using AdaBoost integration technique. *Remote Sens.* 9 (10), 1055. <https://doi.org/10.3390/rs9101055>.
- Chenay, K., Soltani, A., Sharifi, A., 2023. Street network patterns for mitigating urban heat islands in arid climates. *Int. J. Digital Earth* 16 (1), 3145–3161.
- Chun, B., Guldmann, J.-M., 2014. Spatial statistical analysis and simulation of the urban heat island in high-density central cities. *Landscape and Urban Planning* 125, 76–88.
- Connors, J.P., Galletti, C.S., Chow, W.T., 2013. Landscape configuration and urban heat island effects: assessing the relationship between landscape characteristics and land surface temperature in Phoenix, Arizona. *Landsc. Ecol.* 28, 271–283.
- Da Silva, I.N., Hernane Spatti, D., Andrade Flauzino, R., Liboni, L.H.B., dos Reis Alves, S.F., da Silva, I.N., Hernane Spatti, D., Andrade Flauzino, R., Liboni, L.H.B., dos Reis Alves, S.F., 2017. Artificial Neural Network Architectures and Training Processes. Springer.
- Deilami, K., Kamruzzaman, M., Liu, Y., 2018. Urban heat island effect: a systematic review of spatio-temporal factors, data, methods, and mitigation measures. *Int. J. Appl. Earth Obs. Geoinf.* 67, 30–42. <https://doi.org/10.1016/j.jag.2017.12.009>.
- Ding, Q., Pan, T., Lin, T., Zhang, C., 2022. Urban land-cover changes in major cities in China from 1990 to 2015. *Int. J. Environ. Res. Public Health* 19 (23), 16079.
- Du, H., Ai, J., Cai, Y., Jiang, H., Liu, P., 2019. Combined effects of the surface urban heat island with landscape composition and configuration based on remote sensing: a case study of Shanghai, China. *Sustainability* 11 (10), 2890. <https://doi.org/10.3390/su11102890>.
- Gao, Z., Hou, Y., Chen, W., 2019. Enhanced sensitivity of the urban heat island effect to summer temperatures induced by urban expansion. *Environ. Res. Lett.* 14 (9), 094005.
- Ge, J., Meng, B., Liang, T., Feng, Q., Gao, J., Yang, S., Huang, X., Xie, H., 2018. Modeling alpine grassland cover based on MODIS data and support vector machine regression in the headwater region of the Huanghe River, China. *Remote Sens. Environ.* 218, 162–173.
- Gevaert, C.M., Belgiu, M., 2022. Assessing the generalization capability of deep learning networks for aerial image classification using landscape metrics. *Int. J. Appl. Earth Obs. Geoinf.* 114, 103054. <https://doi.org/10.1016/j.jag.2022.103054>.
- Ghanbari, R., Heidarimozaffar, M., Soltani, A., Arefi, H., 2023. Land surface temperature analysis in densely populated zones from the perspective of spectral indices and urban morphology. *Int. J. Environ. Sci. Technol.* 20 (3), 2883–2902. <https://doi.org/10.1007/s13762-022-04725-4>.
- Goldblatt, R., Addas, A., Crull, D., Maghrabi, A., Levin, G.G., Rubinyi, S., 2021. Remotely sensed derived land surface temperature (LST) as a proxy for air temperature and thermal comfort at a small geographical scale. *Land* 10 (4), 410. <https://doi.org/10.3390/land10040410>.
- Guo, A., Yang, J., Xiao, X., Xia, J., Jin, C., Li, X., 2020. Influences of urban spatial form on urban heat island effects at the community level in China. *Sustain. Cities Soc.* 53, 101972.
- Han, D., An, H., Cai, H., Wang, F., Xu, X., Qiao, Z., Jia, K., Sun, Z., An, Y., 2023. How do 2D/3D urban landscapes impact diurnal land surface temperature: insights from block scale and machine learning algorithms. *Sustain. Cities Soc.* 99, 104933. <https://doi.org/10.1016/j.scs.2023.104933>.
- Hou, H., Estoque, R.C., 2020. Detecting cooling effect of landscape from composition and configuration: An urban heat island study on Hangzhou. *Urban For. Urban Green.* 53, 126719.
- Huang, R., Yang, M., Lin, G., Ma, X., Wang, X., Huang, Q., Zhang, T., 2022. Cooling effect of green space and water on urban Heat Island and the perception of residents: a case study of Xi'an City. *Int. J. Environ. Res. Public Health* 19 (22), 14880. <https://doi.org/10.3390/ijerph192214880>.
- Jabeur, S.B., Mefteh-Wali, S., Viviani, J.-L., 2021. Forecasting gold price with the XGBoost algorithm and SHAP interaction values. *Annals of Operations Research* 1–21.
- Jauregui, E., Godinez, L., Cruz, F., 1992. Aspects of heat-island development in Guadalajara, Mexico. *Atmospheric Environment. Part B: Urban Atmosphere* 26 (3), 391–396.
- Kamarianakis, Y., Li, X., Turner, B., Brazel, A.J., 2019. On the effects of landscape configuration on summer diurnal temperatures in urban residential areas: application in Phoenix, AZ. *Front. Earth Sci.* 13, 445–463.
- Karimi, F., Sultana, S., Babakan, A.S., Suthaharan, S., 2019. An enhanced support vector machine model for urban expansion prediction. *Computers. Environ. Urban Syst.* 75, 61–75.
- Karimi, F., Sultana, S., Babakan, A.S., Suthaharan, S., 2021. Urban expansion modeling using an enhanced decision tree algorithm. *GeoInformatica* 25, 715–731. <https://doi.org/10.1007/s10707-019-00377-8>.
- Khanifar, J., Khademalrasoul, A., 2022. Modeling of land surface temperature–multiscale curvatures relationship using XGBoost algorithm (Case study: Southwestern Iran). *Int. J. Environ. Sci. Technol.* 19 (12), 11763–11774.
- Li, W., Cao, Q., Lang, K., Wu, J., 2017. Linking potential heat source and sink to urban heat island: Heterogeneous effects of landscape pattern on land surface temperature. *Sci. Total Environ.* 586, 457–465.
- Li, H., Li, Y., Wang, T., Wang, Z., Gao, M., Shen, H., 2021. Quantifying 3D building form effects on urban land surface temperature and modeling seasonal correlation patterns. *Build. Environ.* 204, 108132.
- Li, Y., Wen, H., Wang, F., 2022. Analysis of the evolution of mangrove landscape patterns and their drivers in Hainan Island from 2000 to 2020. *Sustainability* 15 (1), 759. <https://doi.org/10.3390/su15010759>.
- Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. *R News* 2 (3), 18–22.
- Lin, Y., Wang, Z., Jim, C.Y., Li, J., Deng, J., Liu, J., 2020. Water as an urban heat sink: blue infrastructure alleviates urban heat island effect in mega-city agglomeration. *J. Clean. Prod.* 262, 121411. <https://doi.org/10.1016/j.jclepro.2020.121411>.
- Liu, Y., Peng, J., Wang, Y., 2017. Diversification of land surface temperature change under urban landscape renewal: a case study in the main city of Shenzhen, China. *Remote Sens.* 9 (9), 919.
- Liu, Y., Peng, J., Wang, Y., 2018. Efficiency of landscape metrics characterizing urban land surface temperature. *Landsc. Urban Plan.* 180, 36–53.
- Loh, W.Y., 2011. Classification and regression trees. *Wiley Interdisciplinary Rev.: Data Mining and Knowledge Discov* 1 (1), 14–23. <https://doi.org/10.1002/widm.8>.
- Madaan, M., Kumar, A., Keshri, C., Jain, R., Nagrath, P., 2021. Loan default prediction using decision trees and random forest: a comparative study. *IOP Conference Series: Materials Science and Engineering*.
- Maduako, I., Yun, Z., Patrick, B., 2016. Simulation and prediction of land surface temperature (LST) dynamics within Ikrom City in Nigeria using artificial neural network (ANN). *J. Remote Sens. & GIS* 5 (1), 1–7. <https://doi.org/10.14355/ijrs.2016.06.010>.
- Magidi, J., Ahmed, F., 2019. Assessing urban sprawl using remote sensing and landscape metrics: a case study of City of Tshwane, South Africa (1984–2015). *Egypt. J. Remote Sens. Space Sci.* 22 (3), 335–346.
- Meinshausen, N., Ridgeway, G., 2006. Quantile regression forests. *J. Mach. Learn. Res.* 7 (6).
- Mohammad, P., Goswami, A., Chauhan, S., Nayak, S., 2022. Machine learning algorithm based prediction of land use land cover and land surface temperature changes to characterize the surface urban heat island phenomena over Ahmedabad city, India. *Urban Climate* 42, 101116.
- Mustafa, E.K., Co, Y., Liu, G., Kaloop, M.R., Beshr, A.A., Zarzoura, F., Sadek, M., 2020. Study for predicting land surface temperature (LST) using landsat data: a comparison of four algorithms. *Adv. Civil Engineer.* 2020, 1–16.
- Nagendra, H., Munroe, D.K., Southworth, J., 2004. From Pattern to Process: Landscape Fragmentation and the Analysis of Land Use/Land Cover Change, 101. Elsevier, pp. 111–115. <https://doi.org/10.1016/j.agee.2003.09.003>.
- Pande, C.B., Moharir, K.N., Varade, A.M., Abdo, H.G., Mulla, S., Yaseen, Z.M., 2023. Intertwined impacts of urbanization and land cover change on urban climate and agriculture in Aurangabad city (MS), India using google earth engine platform. *J. Cleaner Product.* 422, 138541. <https://doi.org/10.1016/j.jclepro.2023.138541>.

- Phiri, D., Simwanda, M., Nyirenda, V., Murayama, Y., Ranagalage, M., 2020. Decision tree algorithms for developing rulesets for object-based land cover classification. *ISPRS Int. J. Geo-Information* 9 (5), 329.
- Rehman, A., Qin, J., Pervez, A., Khan, M.S., Ullah, S., Ahmad, K., Rehman, N.U., 2022. Land-use/land cover changes contribute to land surface temperature: a case study of the Upper Indus Basin of Pakistan. *Sustainability* 14 (2), 934. <https://doi.org/10.3390/su14020934>.
- Samardžić-Petrović, M., Kovacević, M., Bajat, B., Dragičević, S., 2017. Machine learning techniques for modelling short term land-use change. *ISPRS Int. J. Geo-Information* 6 (12), 387.
- Sharifi, E., Soltani, A., 2017. Patterns of urban heat island effect in Adelaide: a mobile traverse experiment. *Modern Appl. Sci.* 11 (4), 80–90. <https://doi.org/10.5539/mas.v11n4p80>.
- Singh, P., Kikon, N., Verma, P., 2017. Impact of land use change and urbanization on urban heat island in Lucknow city, Central India. A remote sensing based estimate. *Sustain. Cities Soc.* 32, 100–114.
- Siqi, J., Yuhong, W., Ling, C., Xiaowen, B., 2023. A novel approach to estimating urban land surface temperature by the combination of geographically weighted regression and deep neural network models. *Urban Climate* 47, 101390.
- Soltani, A., 2017. Social and urban form determinants of vehicle ownership; evidence from a developing country. *Transportation Research Part A: Policy and Practice* 96, 90–100. <https://doi.org/10.1016/j.tra.2016.12.010>.
- Soltani, A., Lee, C.L., 2024. The non-linear dynamics of South Australian regional housing markets: a machine learning approach. *Appl. Geograph.* 166 <https://doi.org/10.1016/j.apgeog.2024.103248>.
- Soltani, A., Sharifi, E., 2017. Daily variation of urban heat island effect and its correlations to urban greenery: a case study of Adelaide. *Front. Architect. Res.* 6 (4), 529–538.
- Soltani, A., Sharifi, E., 2019. Understanding and analysing the urban heat island (UHI) effect in micro-scale. *Int. J. Soc. Ecol. Sustain. Develop. (IJSESD)* 10 (2), 14–28. <https://doi.org/10.4018/IJSESD.2019040102>.
- Suthar, G., Kaul, N., Khandelwal, S., Singh, S., 2024. Predicting land surface temperature and examining its relationship with air pollution and urban parameters in Bengaluru: a machine learning approach. *Urban Clim.* 53, 101830.
- Tanoori, G., Soltani, A., Modiri, A., 2024. Predicting Urban Land Use and Mitigating Land Surface Temperature: Exploring the Role of Urban Configuration with Convolutional Neural Networks, Urban Development and Planning. American Society of Civil Engineering, under publication.
- Turner, M.G., Gardner, R.H., O'Neill, R. V., & O'Neill, R. V., 2001. *Landscape Ecology in Theory and Practice*, vol. 401. Springer.
- Wang, Y., Akbari, H., 2016. Analysis of urban heat island phenomenon and mitigation solutions evaluation for Montreal. *Sustain. Cities Soc.* 26, 438–446.
- Wang, L., Wang, S., Zhou, Y., Zhu, J., Zhang, J., Hou, Y., Liu, W., 2020. Landscape pattern variation, protection measures, and land use/land cover changes in drinking water source protection areas: a case study in Danjiangkou reservoir, China. *Global Ecol. and Conserv.* 21, e00827.
- Wu, W., Li, C., Liu, M., Hu, Y., Xiu, C., 2020. Change of impervious surface area and its impacts on urban landscape: an example of Shenyang between 2010 and 2017. *Ecosyst. Health Sustain.* 6 (1), 1767511. <https://doi.org/10.1080/20964129.2020.1767511>.
- Xiao, H., Weng, Q., 2007. The impact of land use and land cover changes on land surface temperature in a karst area of China. *J. Environ. Manag.* 85 (1), 245–257.
- Xiong, J., Thenkabail, P.S., Gumma, M.K., Teluguntla, P., Poehnelt, J., Congalton, R.G., Yadav, K., Thau, D., 2017. Automated cropland mapping of continental Africa using Google earth engine cloud computing. *ISPRS J. Photogramm. Remote Sens.* 126, 225–244.
- Yan, J., Zhou, W., Jenerette, G.D., 2019. Testing an energy exchange and microclimate cooling hypothesis for the effect of vegetation configuration on urban heat. *Agric. For. Meteorol.* 279, 107666.
- Yang, L., Shami, A., 2020. On hyperparameter optimization of machine learning algorithms: theory and practice. *Neurocomputing* 415, 295–316.
- Yao, L., Sun, S., Song, C., Wang, Y., Xu, Y., 2022. Recognizing surface urban heat 'island' effect and its urbanization association in terms of intensity, footprint, and capacity: a case study with multi-dimensional analysis in northern China. *J. Clean. Prod.* 372, 133720. <https://doi.org/10.1016/j.jclepro.2022.133720>.
- Yin, Z., Liu, Z., Liu, X., Zheng, W., Yin, L., 2023. Urban heat islands and their effects on thermal comfort in the US: New York and New Jersey. *Ecol. Indic.* 154, 110765.
- Ying, C., Qi-Guang, M., Jia-Chen, L., Lin, G., 2013. Advance and prospects of AdaBoost algorithm. *Acta Automat. Sin.* 39 (6), 745–758.
- Yuan, B., Zhou, L., Hu, F., Wei, C., 2024. Effects of 2D/3D urban morphology on land surface temperature: contribution, response, and interaction. *Urban Clim.* 53, 101791.
- Yue, W., Liu, X., Zhou, Y., Liu, Y., 2019. Impacts of urban configuration on urban heat island: An empirical study in China mega-cities. *Sci. Total Environ.* 671, 1036–1046. <https://doi.org/10.1016/j.scitotenv.2019.03.421>.
- Zhang, Q., Chen, C., Wang, J., Yang, D., Zhang, Y., Wang, Z., Gao, M., 2020. The spatial granularity effect, changing landscape patterns, and suitable landscape metrics in the three gorges reservoir area, 1995–2015. *Ecol. Indic.* 114, 106259. <https://doi.org/10.1016/j.ecolind.2020.106259>.
- Zhao, X., Cao, C., Ni, X., Chen, W., 2017. Retrieval and application of leaf area index over China using HJ-1 data. *Geomat. Nat. Haz. Risk* 8 (2), 478–495.
- Zhou, L., Yuan, B., Hu, F., Wei, C., Dang, X., Sun, D., 2022. Understanding the effects of 2D/3D urban morphology on land surface temperature based on local climate zones. *Build. Environ.* 208, 108578.