TẠP CHÍ KHOA HỌC VÀ CÔNG NGHỆ ĐẠI HỌC DUY TÂN
DTU Journal of Science and Technology 5(72) (2025) 24-31

ĐẠI HỌC DUY TÂN
DUYTAN UNIVERSITY

# Machine learning-based geospatial modeling of urban land surface temperature using topographical features and urban morphology

## Mô hình hóa không gian địa lý dựa trên máy học về nhiệt độ bề mặt đô thị sử dụng các đặc điểm địa hình và hình thái đô thị

Hoang Nhat Duc[a,b*], Nguyen Quoc Lam[b], Pham Quang Nhat[b]
Hoàng Nhật Đức[a,b*], Nguyễn Quốc Lâm[b], Phạm Quang Nhật[b]

[a]*Institute of Research and Development, Duy Tan University, Da Nang, 550000, Viet Nam*
[a]*Viện Nghiên cứu và Phát triển Công nghệ cao, Đại học Duy Tân, Đà Nẵng, Việt Nam*
[b]*Faculty of Civil Engineering, School of Engineering and Technology, Duy Tan University, Da Nang, 550000, Viet Nam*
[b]*Khoa Xây dựng, Trường Công nghệ và Kỹ thuật, Đại học Duy Tân, Đà Nẵng, Việt Nam*

*(Date of receiving article: 27/03/2025, date of completion of review: 19/08/2025, date of acceptance for posting: 05/09/2025)*

## Abstract

Land surface temperature (LST) is a crucial parameter for assessing the urban thermal comfort, especially in the context of climate change and urban expansion. This study presents a machine learning-based method for geospatial modeling of urban LST. The topographical and urban morphological features are employed as explanatory variables. LST data retrieved from the thermal band of Landsat-8 imaginary is used as the dependent variable, which is modeled by the extreme gradient boosting machine regressor (XGBoost). The urban center of Hue city is selected as the study area to apply the proposed framework. Google Earth Engine (GEE) platform is employed to retrieve the topographical data, including the elevation, slope, and aspect. A land use/ land cover (LULC) map for the study area is constructed via a Random Forest model and the spectral bands of the Sentinel-2 imaginary. The LULC data classification is performed in GEE platform. Experimental results point out that XGBoost can model the spatial variation of urban LST with a mean absolute percentage error of 4.21% and a coefficient of determination of 0.72. Among the explanatory variables, the built-up density strongly correlates with LST. The factors of green space density and waterbody density have apparent negative correlations with the predicted output, demonstrating their cooling effects in the study area. The findings in this study provide more insights into the spatial distribution of LST in Hue City, helping planners in urban planning and mitigating the negative effects of urban heat island phenomenon.

*Keywords:* urban land surface temperature; topographical features; urban morphology; remote sensing; XGBoost.

## Tóm tắt

Nhiệt độ bề mặt đất là một thông số quan trọng để đánh giá nhiệt độ đô thị, đặc biệt là trong bối cảnh biến đổi khí hậu và quá trình mở rộng của các đô thị. Bài báo của chúng tôi trình bày một phương pháp dựa trên máy học để lập mô hình không gian địa lý của nhiệt độ bề mặt đất đô thị. Các đặc điểm hình thái địa hình và đô thị được sử dụng làm biến số ảnh hưởng. Dữ liệu nhiệt độ bề mặt đất thu được từ dải nhiệt của vệ tinh Landsat-8 thu thập được sử dụng làm biến số được mô hình hóa bằng phương pháp XGBoost. Thành phố Huế được chọn làm khu vực nghiên cứu để áp dụng phương pháp đề xuất. Nền tảng Google Earth Engine (GEE) được sử dụng để thu thập dữ liệu địa hình, bao gồm độ cao, độ dốc, và hướng. Bản đồ sử dụng lớp phủ đất cho khu vực nghiên cứu được xây dựng thông qua mô hình Rừng ngẫu nhiên và các

*Corresponding author:* Hoang Nhat Duc
*Email:* hoangnhatduc@duytan.edu.vn

dải quang phổ của vệ tinh Sentinel-2. Phân loại dữ liệu lớp phủ đất được thực hiện trên nền tảng GEE. Kết quả tính toán chỉ ra rằng XGBoost có thể mô phỏng sự biến động của nhiệt độ bề mặt đô thị với sai số phần trăm tuyệt đối trung bình là 4,21% và hệ số xác định là 0,72. Trong số các biến ảnh hưởng, mật độ xây dựng có mối tương quan mạnh nhất với nhiệt độ bề mặt đô thị. Các yếu tố về mật độ không gian xanh và mật độ khối nước có mối tương quan âm với biến được mô phỏng; điều này chỉ ra tác dụng làm giảm nhiệt độ bề mặt của chúng trong khu vực nghiên cứu. Những phát hiện trong nghiên cứu này cung cấp thêm thông tin chi tiết về sự phân bố không gian của nhiệt độ bề mặt đô thị tại thành phố Huế, giúp các nhà hoạch định trong quy hoạch đô thị và giảm thiểu tác động tiêu cực của hiện tượng đảo nhiệt đô thị.

*Từ khóa:* nhiệt độ bề mặt đất đô thị; đặc điểm địa hình; hình thái đô thị; viễn thám; XGBoost.

## 1. Introduction

The fast pace of urban expansion and global warming has urged urban planners to investigate spatial variations in urban land surface temperature (LST) and their drivers [1]. As urban areas expand, natural landscapes, such as green spaces, are increasingly replaced by impervious surfaces such as asphalt roads, concrete structures, and built infrastructure. These materials exhibit low albedo and high thermal mass, absorbing a substantial proportion of solar radiation. This leads to prolonged heat retention, with considerable difference in temperature between impervious surfaces and vegetated areas.

The loss of green spaces inevitably intensifies thermal stress; it is because urban parks and tree canopies normally provide cooling through evapotranspiration. However, many cities have seen vegetative cover decline during rapid expansion phases. As pointed out in [2], the proportion of vegetation cover in urban areas has significantly reduced, from a median of 47% (in 2000) to a median of 42% (in 2015). The changes in LULC essentially alter the thermal properties and dynamics of urban centers, making the urban core significantly warmer than the surrounding rural areas. Therefore, understanding these spatial relationships between LST variation and its governing factors through GIS-based modeling is crucial for formulating targeted mitigation strategies, including construction density regulation, development of green infrastructure, and promotion of environmental friendly materials [3].

The combined effect of urban topography and morphology has been shown to impose significant impacts on the variation of LST. This complex relationship involves various factors that dictate how urban areas absorb, retain, and dissipate heat. The primary drivers of LST variation include built-up density and green space density. Moreover, topographic features (e.g., elevation, slope, and aspect) also significantly affect the spatial distribution of LST. To understand the thermal behavior of a city, it is required to carry out the analysis of these factors to reveal their functional relationships with the urban LST.

Machine learning has been successfully employed for geospatial modeling of the urban LST. This method provides robust and capable methods to analyze complex spatial patterns and relationships in GIS datasets. By leveraging multi-layered data, machine learning algorithms can handle non-linear interactions between explanatory variables and predict LST with good accuracy. Advanced machine learning models pave the way for the development of data-driven approaches that can be applied across different urban settings. These models are highly useful for policymakers and urban planners in the tasks of designing targeted interventions for UHI mitigation.

In recent years, Google Earth Engine (GEE) provides a powerful cloud-based platform designed for large-scale geospatial analysis. GEE offers unparalleled access to a vast source of satellite imagery and geospatial datasets. GEE's scalability and ease of use allow researchers to efficiently model and predict urban LST. This capability is particularly valuable for urban planning and sustainable development. Data of LST and its influencing factors can be easily extracted from various datasets in GEE, including Landsat 8 imaginary, Sentinel-2 imaginary, and Shuttle Radar Topography Mission (SRTM).

This study relies on machine learning, remote sensing data, and geospatial data analysis to model the spatial variation of urban LST in Hue City, Vietnam. The GEE platform serves as a critical tool for retrieving and processing relevant remote sensing data. Random forest classifier is employed to construct the LULC map. Based on this map, the features of bare land density, built-up density, green space density, waterbody density, and Shannon Entropy are computed to characterize the urban morphology in the study area. These variables provide crucial insights into the spatial distribution and complexity of urban structures in Hue City, which are crucial for understanding how different land cover types influence LST.

## 2. Research methods and materials

### *2.1. The study area and remote sensing datasets*

Hue City, located in the Central Coast region of Vietnam, serves as the capital of Thua Thien Hue province. Hue is widely recognized as a world cultural heritage city due to its rich historical background and vibrant cultural traditions. The region is home to the Complex of Hue Monuments, which was the first site in Vietnam to be recognized by UNESCO as a World Cultural Heritage. In recent years, Hue City has faced numerous challenges related to climate change, experiencing frequent heatwaves that have intensified extreme heat and drought conditions. Accordingly, the urban center of Hue City is selected as the study area of the current study.

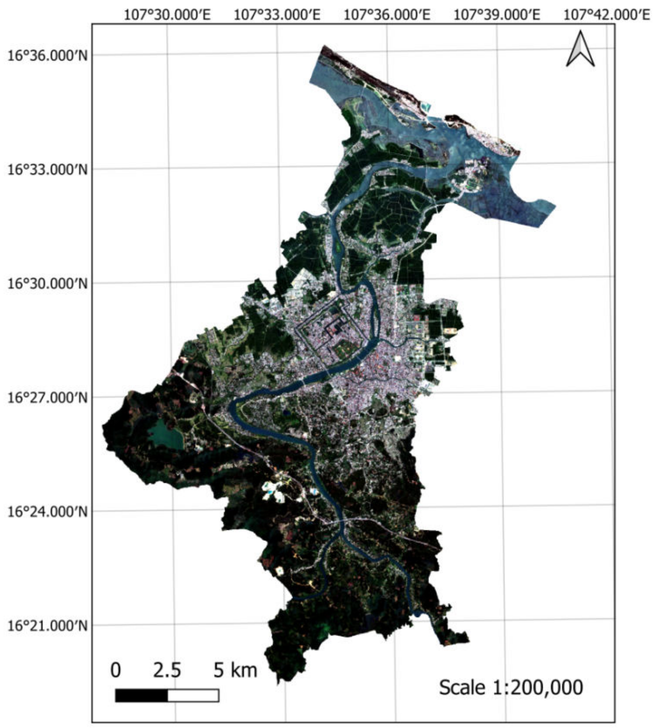


Figure 1. The study area shown as a true-color composite from Sentinel-2 bands (Dataset provider: European Union/ESA/Copernicus)

In the study area (refer to Figure 1), the combination of high population density and intensified urban LST causes various negative effects due to the UHI phenomenon. The dense infrastructure and economic activities collectively escalate heat retention. This fact results in prolonged periods of extreme heat. Therefore, it is essential to investigate the spatial variations in urban heat stress within Hue City with respect to urban morphology and topographical features. This analysis can offer valuable insights into the local temperature patterns and significantly assist urban planning strategies. To support the geospatial modeling of urban LST in Hue, this study has collected remote sensing data from Landsat 8, Sentinel-2, and NASA SRTM (refer to Table 1).

Table 1. Remote sensing datasets

| Dataset | Time period | Bands | Resolution |
|---|---|---|---|
| Landsat 8 Level 2, Collection 2, Tier 1 | 05/01/2024 – 09/30/2024<br>05/01/2025– 08/14/2025 | SR_4, SR_5, and ST_B10 | 30 meters |
| NASA SRTM Digital Elevation 30m | | Elevation | 30 meters |
| Sentinel-2 | 01/01/2024 – 08/14/2025 | B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12 | 10 meters (B2, B3, B4, and B8)<br>20 meters (B5, B6, B7, B8A, B11, and B12) |

### *2.2. Retrieval of urban land surface temperature from Landsat 8 imaginary*

The thermal band of Landsat 8 imaginary was used to obtain the LST data. The data were filtered for the dry seasons of 2024 and 2025 (up to August 14, 2025). The LST map is prepared in QGIS software (available at https://qgis.org/) and is shown in Figure 2. The process of preparing the map involved several steps, which are implemented in GEE's code editor. First, the spectral band values from Landsat 8 were converted into spectral radiance [4]. Next, to obtain the emissivity-corrected LST, land surface emissivity and the Normalized Difference Vegetation Index (NDVI) needed to be calculated [5,6]. The NDVI was calculated using the 4th (red) and 5th (near-infrared) bands of Landsat 8.

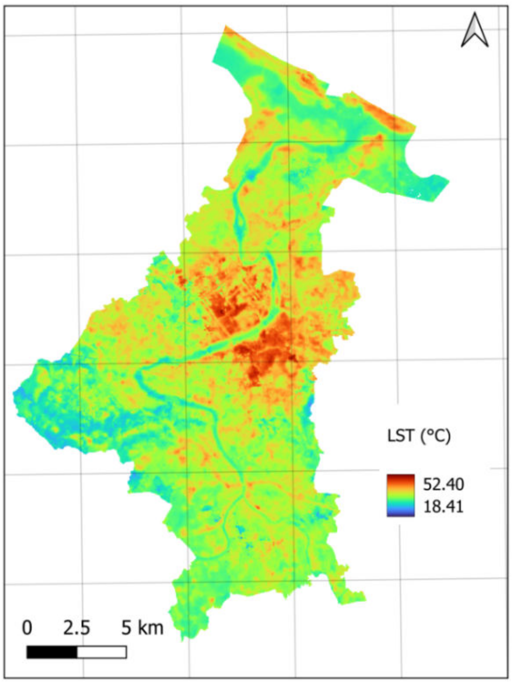


Figure 2. Land surface temperature in the study area

### *2.3. Topographical and urban morphological features*

Topographical features, such as elevation, slope, and aspect, significantly influence the spatial variation of urban LST. Elevation generally exhibits a negative correlation with LST, as higher altitudes tend to associate with cooler temperatures. Slope also affects the LST variation as steeper terrain reduces solar radiation exposure. Aspect influences LST through variations in solar exposure. To effectively model LST, it is required to take into account the LULC in the study area. The LULC map in this study is computed in GEE with the Random Forest classifier and Sentinel-2 data. The areas ($km^2$) and the proportion (%) of each LULC class in the study area are summarized in Table 2. Moreover, urban morphological features, including bare land density, built-up density, green space density, and waterbody density have been demonstrated to be critical LST influencing factors. These density maps are computed via a morphological mean filter with a radius of 3 pixels.

Table 2. Areas and proportions of LULC classes

| LULC Class | Area ($km^2$) | Proportion (%) |
|---|---|---|
| Bare land | 36.97 | 13.87 |
| Built-up | 33.90 | 12.72 |
| Green space | 156.14 | 58.58 |
| Waterbody | 39.53 | 14.83 |

Moreover, for assessing urban sprawl and identifying areas of compactness in urban areas, this study relies on Shannon entropy. This index is calculated from the aforementioned LULC map. Herein, the larger the Shannon entropy, the more dispersed the urban area. Therefore, Shannon entropy can offer more insights into the urban morphological features in Hue City and should be used in the LST model. The reason is that areas with high entropy values (indicating dispersed landscapes) may exhibit different thermal patterns compared to areas with low entropy values (demonstrating urban compactness). The nine explanatory variables are presented in Figure 3. All maps are resampled to the spatial resolution of 30 meters.

### *2.4. Extreme gradient boosting machine (XGBoost) regressor*

The XGBoost [7] is a powerful machine learning model that has shown significant potential in urban LST modeling. This machine learning approach is widely used in geospatial data analysis due to its high accuracy, robustness, and fast computation. XGBoost often excels in handling complex datasets by efficiently learning nonlinear relationships between variables. In the context of LST modeling, XGBoost can be used to generalize the mapping function between LST and its influencing factors. Hence, this method is particularly useful in urban areas with complex topographical and urban morphological features. In this study, XGBoost is implemented with the Python toolbox provided at https://xgboost.readthedocs.io/.

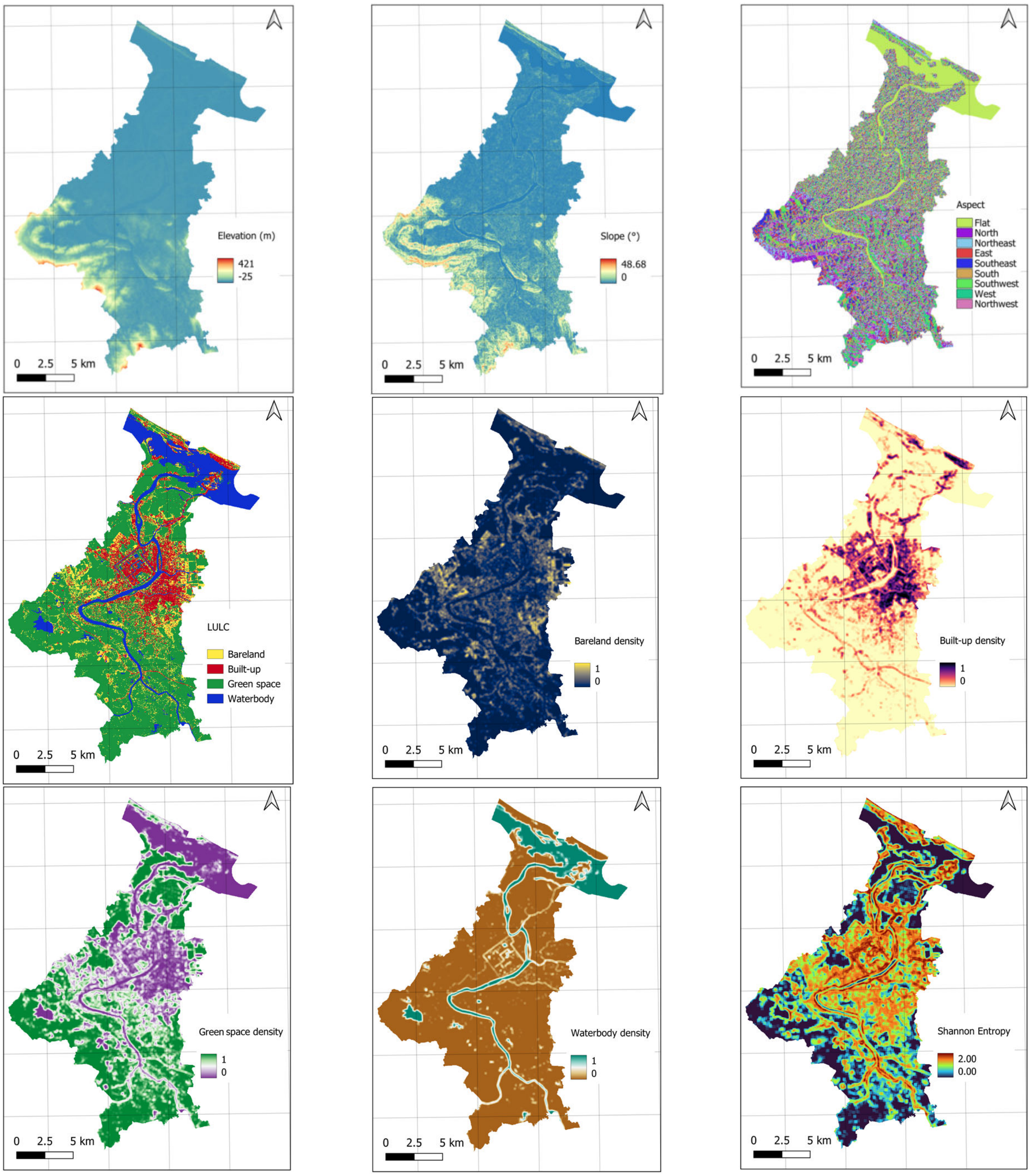


Figure 3. Explanatory variables

## 3. Results and discussion

The XGBoost model is used to construct a functional relationship between LST and the nine explanatory variables. To generate the dataset, 50,000 data points in the study area are randomly sampled. This dataset is then divided into a training (70%) and a testing set (30%). Accordingly, 35,000 samples were used to train the model; 15,000 samples were employed to inspect the generalization properties of the model. In addition, the implementation of the XGBoost regressor requires the setting of the model's hyper-parameters: the number of estimators, the maximum tree depth, the learning rate, the $L_2$-regularization coefficient, and the $L_1$-regularization coefficient. Via several trial-and-error runs, the suitable values for those hyper-parameters are identified and reported in

Table 3. The prediction performance of the machine learning model is reported in Table 4. In the testing phase, XGBoost attains satisfactory outcomes with a RMSE of 1.99, a MAPE of 4.21%, a MAE of 1.50, and a $R^2$ of 0.72. As observed from Table 4, the difference in performance between the training and testing phase is not significant. Therefore, it can be seen that the trained model is not overfitted. The prediction outcomes of XGBoost are further visualized via the scatter plots in Figure 4. As can be seen from the figure, the major proportion of the prediction results is scattered around the line-of-best-fit, demonstrating good predictive performance of the model.

Table 3. Hyper-parameter setting

| Number of estimators | Maximum tree depth | Learning rate | $L_2$-regularization coefficient | $L_1$-regularization coefficient |
|---|---|---|---|---|
| 350 | 6 | 0.15 | 0. 1 | 0.0001 |

Table 4. Prediction performance

| Phase | RMSE | MAPE (%) | MAE | $R^2$ |
|---|---|---|---|---|
| Training | 1.55 | 3.33 | 1.18 | 0.83 |
| Testing | 1.99 | 4.21 | 1.50 | 0.72 |

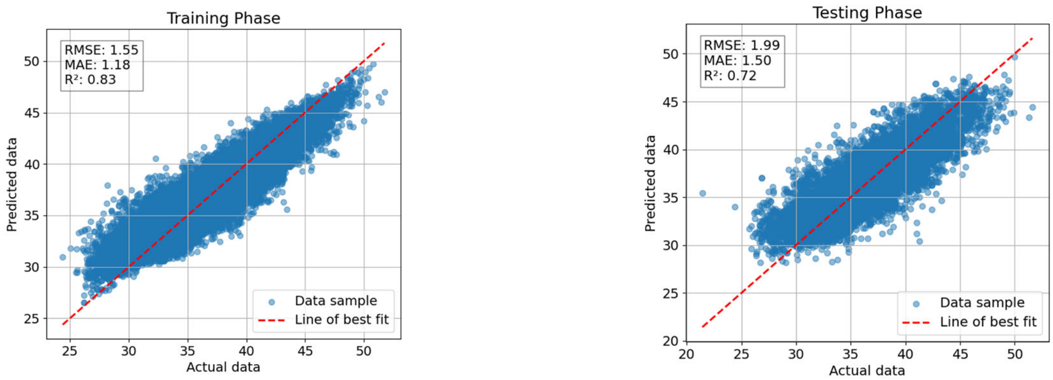


Figure 4. LST prediction results in the training and testing phases

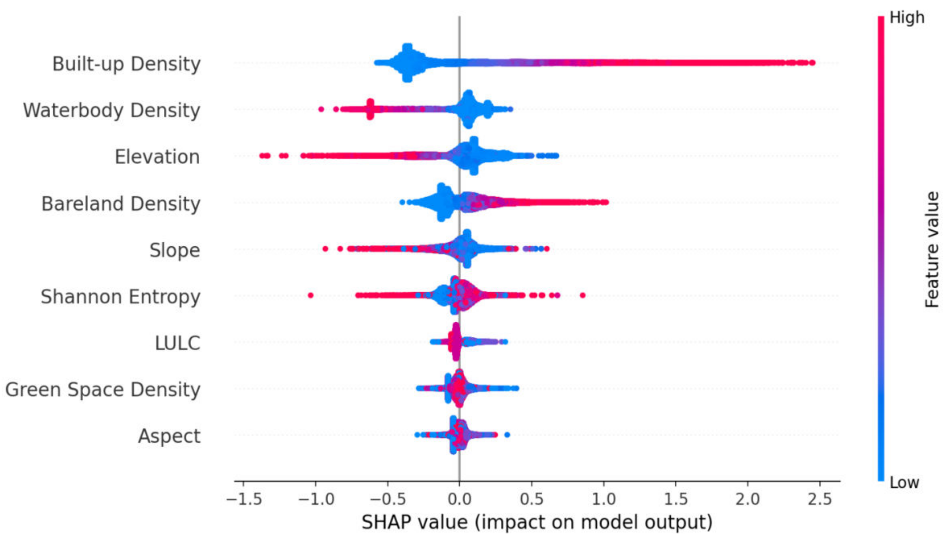


Figure 5. Impact plot obtained from SHAP

To inspect the effect of the input variable on the model's output, this study relies on the Shapley Additive exPlanations (SHAP) coupled with tree-based models [8]. SHAP is a popular method for interpreting the predictions of a machine learning model. It is inspired from the cooperative game theory to explain how each feature contributes to a model's output by comparing predictions with and without that feature. Using SHAP, it is able to obtain detailed, local explanations of individual predictions by quantifying the contribution of each input feature. The impact plot, obtained from SHAP, is presented in Figure 5. Both urban morphology and topography show substantial influences on the variation of LST in the study area with the built-up density obtaining the 1st rank and waterbody density attaining the 2nd rank. The elevation, bare land density, and slope obtain the 3rd, 4th, and 5th ranks, respectively. The factors of Shannon Entropy and LULC are less important than those previously mentioned factors. Meanwhile, the influence of Shannon entropy is not as substantial as that of other variables characterizing the urban morphology. It is also apparent that built-up density and bare land density positively correlate with LST. Meanwhile, water body density and green space density demonstrate negative correlations with the target variable. This fact implies the evidential cooling effect of water bodies and green space patches in the study area.

## 4. Conclusion

This study has presented a machine learning-based approach for spatial modeling of urban LST in Hue City. The nine variables, characterizing the urban morphology and topography, are employed as influencing factors. XGBoost regressor is used to generalize a model capable of predicting the spatial variation of urban LST. Based on XGBoost and SHAP analysis, it can be concluded that the urban morphology influences the urban heat patterns in Hue City with built-up density and waterbody density are the most influential variables. The analysis conducted with SHAP also reveals notable findings regarding the relationship between urban features and LST. The findings in this study can significantly assist urban planners in enhancing sustainable development and mitigating the impacts of urban expansion.

## References

[1] Ullah S, Qiao X, Abbas M. (2024). *Addressing the impact of land use land cover changes on land surface temperature using machine learning algorithms*. Scientific Reports 14 (1):18746. doi:10.1038/s41598-024-68492-7.

[2] Richards DR, Belcher RN. (2020). *Global Changes in Urban Vegetation Cover*. Remote Sensing 12 (1):23.

[3] Duc, H.N., Duc, T.V., Canh, H.T. (2025). *From Data to Insights: Modeling Urban Land Surface Temperature Using Geospatial Analysis and Interpretable Machine Learning*. Sensors 25 (4):1169.

[4] Waleed M, Sajjad M. (2022). *Leveraging cloud-based computing and spatial modeling approaches for land surface temperature disparities in response to land cover change: Evidence from Pakistan*. Remote Sensing Applications: Society and Environment 25:100665. doi:https://doi.org/10.1016/j.rsase.2021.100665.

[5] Hou H, Estoque RC. (2020). *Detecting Cooling Effect of Landscape from Composition and Configuration: An Urban Heat Island Study on Hangzhou*. Urban Forestry & Urban Greening 53:126719. doi:https://doi.org/10.1016/j.ufug.2020.126719.

[6] Sobrino JA, Jiménez-Muñoz JC, Paolini L. (2004). *Land surface temperature retrieval from LANDSAT TM 5*. Remote Sensing of Environment 90 (4):434-440. doi:https://doi.org/10.1016/j.rse.2004.02.003.

[7] Chen T, Guestrin C. (2016). XGBoost: A Scalable Tree Boosting System. Paper presented at the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA.

[8] Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee S-I. (2020). *From local explanations to global understanding with explainable AI for trees*. Nature Machine Intelligence 2 (1):56-67. doi:10.1038/s42256-019-0138-9.