Energy Proceedings
Vol 44, 2024
ISSN 2004-2965

Understanding the Relationship Between Urban Form Factors and Extreme Heat in Seoul

Danbi Kim [1], Steven Jige Quan [2*]

1 City Energy Lab, Graduate School of Environmental Studies, Seoul National University, Seoul, 08826, South Korea

2 Graduate School of Environmental Studies, Seoul National University, Seoul, 08826, South Korea
(*Corresponding Author: sjquan@snu.ac.kr | stevenchuan6656@gmail.com)

ABSTRACT

Climate change has been a pressing global issue and people are experiencing more frequent and severer extreme weather events. In South Korea, extreme heat has been drawing much attention in recent years due to its significant impact on public health and energy and water consumptions. Extreme heat is particularly exacerbated by the urban heat island (UHI) effect in cities. Many studies have examined the relationship between urban form factors and surface UHI empirically. But few of them have studied how UHI changes in response to an extreme heat event, termed heat resilience in recent studies. Additionally, most of current studies used traditional regression models assuming linear relationships, which may not be the case for UHI effects. To address this gap, this study aims to identify nonlinear relationships between urban form factors and land surface temperature (LST) and heat resilience, using machine learning methods. The study adopted the gradient boosting decision tree (GBDT) models to predict LST and heat resilience and compared the findings with those using spatial regression models. The results suggest that the GBDT model has a higher prediction power than traditional regression, and the GBDT models show that the urban form factors have nonlinear relationships with LST and heat resilience under extreme heat. The findings of the study provide valuable guidance for urban planning practice aimed at prioritizing planning elements in urban form toward heat-resilient cities.

Keywords: extreme heat, urban form, machine learning algorithm, nonlinear relationship

1. INTRODUCTION

Climate change has been a significant global concern since the publication of the first IPCC report in 1990. There has been a growing alarm about the climate crisis as we increasingly encounter situations that directly affect human life, such as extreme weather events, ecosystem collapse, and food crisis. For example, extreme heat in South Korea has become an increasingly important issue in recent years due to the impact of climate change. Summer of 2018 was the hottest on record in Korea, with temperatures consistently soaring above 35℃. In response to the extreme heat in 2018, the Korean government implemented various measures, such as free cooling center, temporary water stations.

However, these phenomena can manifest different impacts depending on the characteristics of the region, especially the physical environment. Urban areas are more vulnerable to extreme heat. Urban areas account for only 2% of the world's land area, yet they contribute to 80% of GDP, 66% of energy consumption, and 75% of greenhouse gas [3, 6]. In urban areas, buildings are densely populated and there is a lack of green spaces compared to the suburbs, a lot of risk occurs during the summer due to extreme heat events, as the urban area absorbs the heat during the day [35]. Therefore, this study aims to understand the impact of the physical urban environment on extreme heat, classifying urban form factors into building density and land cover, and assessing how these variables affect land surface temperature (LST). In addition, the purpose of the study is to answer the research question: how and to what extent do the urban form factors affect LST and heat resilience on extreme heat days?

2. LITERATURE REVIEW

2.1 Urban form factors and urban heat island effect

Urban areas are primarily composed of dense buildings, which typically have lifespan of over 50 years. Thus, it is crucial to analyze the building characteristics in urban areas and discuss how to apply analyzed data in

This is a paper for 15th International Conference on Applied Energy (ICAE2023), Dec. 3-7, 2023, Doha, Qatar.

response to climate change. Previous studies found that urban form factors, including the built environment and land cover, have played a significant role in increasing air temperature in urban areas, well known as Urban heat island (UHI) effect [2, 4, 5, 27].

A closer look at the urban form factors discussed in the previous research, NDVI and NDBI, such as artificial materials and impervious surface are more vital indicators associated with the UHI effect [11]. For example, the road density and the ratio of lack of green space area were highly positively correlated with heat vulnerability and extreme heat risk in the Taipei metropolitan area in Taiwan [2]. According to the studies of Chun, Guldmann, and Guhathakurta, building ground foot area, solar radiation, and sky view factor had more substantial negative effects on the UHI during summer. Also, increased greenery, water bodies, and NDVI reduces temperatures in summer and increases it in winter, the most important mediators of excess heat [3-5]. According to Turner and Galletti, dense urban development does not always indicate a higher UHI effect. They emphasized that well-designed compact development could reduce UHI effect than compared to sprawling development [32].

As for the methodology of analyzing urban form factors and the heat environment, most of the research related to the relationship focused on analyzing the relationship between LST and variables of the built environment [2, 20]. Most of them used methods to identify the relationship through OLS regression analysis, it shows that both green space and river areas contribute to reducing surface temperature. Another study indicates that sprawling cities are more vulnerable to extreme heat than compact cities [29]. These traditional regression models, however, assume that urban form factors are linearly associated with LST, which may not be the case [14].

2.2 Extreme heat event

As mentioned above, the UHI effect exacerbates extreme heat, and the intensity of this effect is influenced by urban form, compared with rural areas [26, 28]. Such an effect should also be paid more attention to under climate change [19]. Therefore, we need to consider the risks to urban residents, such as public health and infrastructure with more frequent and stronger extreme heat events [12, 16]. Extreme heat events can be defined as daily high temperature or average temperature beyond some degree point for at least two or four days [33]. The heat metrics and duration of day vary by country; for example, extreme heat is defined with a daily maximum temperature higher at 33℃ for at least two consecutive days in South Korea. Most studies have been conducted on identifying the spatial distribution of extreme heat, evaluation of damage risk indicators, and heatwave vulnerability. Additionally, from a climate justice perspective, numerous studies have analyzed the spatial relationship between socioeconomic factors such as demographic characteristics (single-person households, low-income households, the elderly population, etc.) [7, 15, 23].

2.3 Urban heat resilience

In recent years, “resilience” has become an essential concept in socio-ecological systems and policy management in urban research [1, 21, 22]. The studies commonly emphasized that resilience is the ability to respond or recover from unpredictable, frequent, and intense extreme weather events. To improve heat resilience in urban areas, urban planners must manage urban form factors including building density and land-use, a crucial component of urban heat planning [21]. For example, urban greening (green area, park, green roofs or wall) also needs to be considered as potential factors that increase “adaptive capacity” by improving resilience in cities [17]. Although the importance of heat resilience has grown, as noted earlier, the measurement of resilience is rare in the literature [35]. For these reasons, a recent study measured heat resilience as the LST difference between normal and extreme heat events in Macau, China. They emphasized that the approach can be useful information to define urban heat resilience in improving the “adaptive capacity” to extreme heat events in urban areas [18, 35].

2.4 Machine learning approach

In recent years, machine learning methods have been widely used in many fields, such as disease diagnosis, stock market prediction, image recognition, and classification. Also, it became one of the compelling methods in territorial science and urban planning field. Gounaridis et al implemented a Random Forest Model to predict future land use/cover under the premise of economic performance scenarios [10]. Additionally, the other authors conducted an analysis predicting urban growth by comparing various machine learning algorithms (LR, RF, ANN, and EGB), and it shows that machine learning algorithms have improved predictive power compared to traditional regression analysis [13, 30]. Also, machine learning algorithms are potent tools for browsing nonlinear relationships and threshold effects [37].

Among machine learning algorithms, the gradient-boosting decision trees (GBDT) model, a tree-based ensemble model, is widely used in transportation. According to previous studies, its prediction is more accurate than the regression model and can handle the multicollinearity issue [8, 9, 36]. More importantly, it better reveals the patterns of nonlinear relationships between variables than traditional linear regression methods [31, 34].

The interpretability analysis of such a machine learning model can further provide a better interpretable result that make easier decision for human [24, 25]. Partial dependence plot (PDP) and feature importance function are often used to visualize and analyze the results of gradient boosting models. PDPs show the dependence between the target function and a set of features of interest, marginalizing the values of all other features (the complement features). The SHAP function was used to measure feature importance, which is based on the Shapley value and derived from coalitional game theory [24].

This study chose the GBDT algorithm implemented as the *GradientBoostingRegressor* class in the scikit-learn library to identify the nonlinear relationship between urban form and UHI, and PDPs and SHAP value-based feature importance for interpretability analyses.

3. METERIAL AND METHODS

Seoul is the capital of South Korea and the hub of politics, economy, society, and culture in the nation. The total area is 605 km², which covers 0.61% of the entire South Korea, consisting of a total of 25 districts and 426 neighborhoods (Fig 1). The scope of this study is based on 3,217 cells in the entire Seoul with 450m x 450m grid. The grid size is decided based on the previous literature and 30m resolution of the LST image. For an accurate result, this study used 2,760 cells, excluding those partial cells on the boundary of Seoul.

The study selected regular heat day and extreme heat day for analysis from a time period between May 1st and September 30th (Summer season in South Korea) over ten years from 2012 to 2023. To extract the LST, Landsat satellite images published on the USGS-Earth Explore website were used. The selected extreme heat day is May 30th, 2020, which had the 30.0℃ air temperature above the 85th percentile of all temperature values during the time period [30, 32]. The regular heat day was selected as September 19th, 2020, with the 26.2℃ air temperature. The cloud covers of both dates are within 20% (regular heat day:16.79% and extreme heat day: 12.81%). ArcGIS pro and Python were used for calculating LST, analyzing urban form factor and implementing machine learning models.

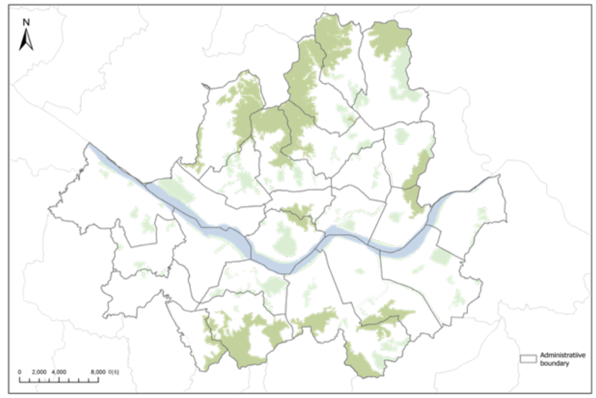

Fig. 1. Study area (Administrative boundary in Seoul)

To measure heat resilience, this study adopted the definition of the LST difference between regular and extreme heat day as a heat resilience temperature from Xi et al.'s study [32]. The LST maps on the regular and extreme heat days and the heat resilience map are shown in Figs 2, 3 and 4. In this study, regular heat day LST (Regular_LST), extreme heat day LST (Extreme_LST), and heat resilience (HR_T) are dependent variables for three models. A total of 8 independent variables are used to measure three types of urban form factors: building density, land cover, and large-scale landscape. They include: building coverage ratio, building height variation, NDVI, waterbody ratio, elevation, distance to park, distance to stream, distance to mountain. Two more variables were added, the latitude (loc.y) and longitude (loc.x) of the center of each grid cell, representing cell locations.

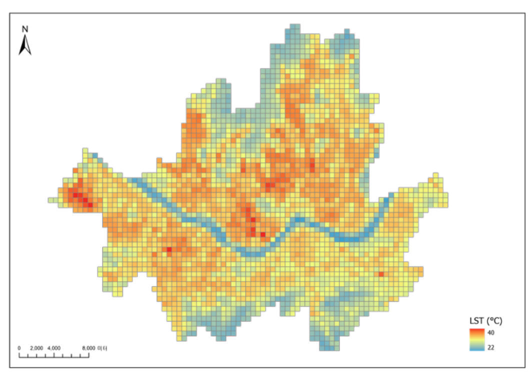

Fig. 2 Distribution of LST on regular heat day

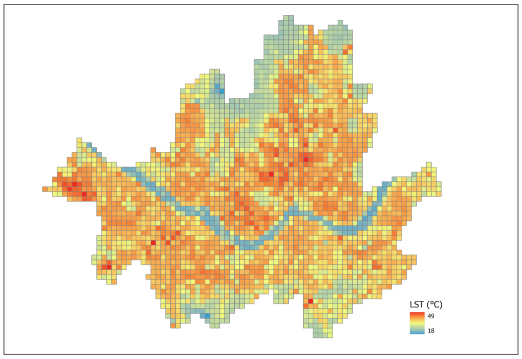

Fig. 3. Distribution of LST on extreme heat day

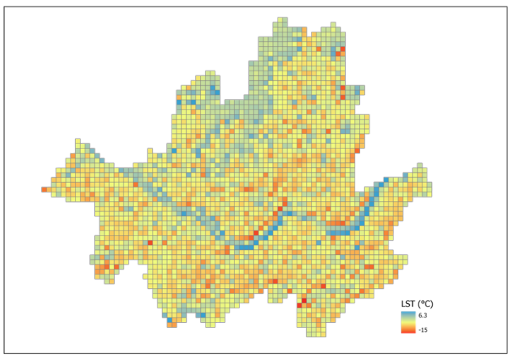

Fig. 4. Distribution of heat resilience temperature

In the training of GBDT models, data are usually split into training and test data sets. However, due to small sample size, the study used all data for cross validation in training. To tune the hyperparameter, the study used grid search to compare performance of hyperparameter combinations in three models: the learning rate in (0.01, 0.1, 0.2, 0.3, 0.4, 0.5), number of trees in (50, 100, 150, 200, 250, 300), and tree depth in (2, 3, 4, 5, 6). The study developed spatial regression models for comparison.

4. RESULTS

The results showed that the most critical variables affecting the LST and heat resilience were the building coverage ratio and NDVI average (Figs. 5-7). In addition, the feature importance plots concluded that the extreme heat LST and heat resilience are relatively more affected by landscape features, such as the distance to mountain, distance to stream, and waterbody ratio, compared with regular heat LST. The results generally support the findings from previous literatures that built-up area and green area are important for UHI. However, distance to park and building height variation are less important variables in all three models. The relationships between dependent and independent variables become better revealed through PDPs, which show the effective range for each variable (Figs. 8-10).

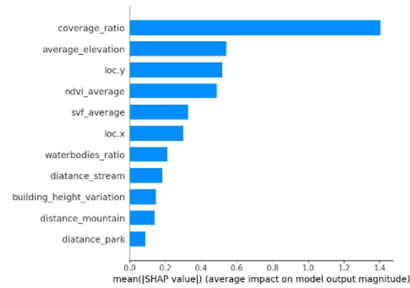

Fig. 5. Feature importance in Regular LST

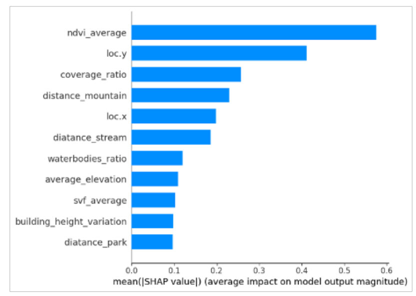

Fig. 6. Feature importance in Extreme LST

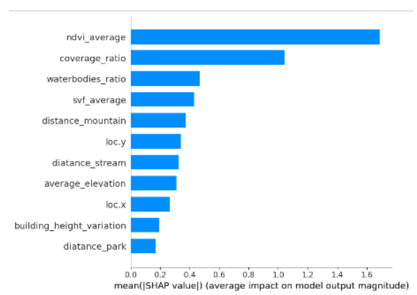

Fig. 7. Feature importance in HR_T

The predicted extreme heat LST and heat resilience show more nonlinear relationship compared with regular LST. For example, in PDPs for regular LST, the predicted LST decreases when the NDVI value increases beyond the threshold 0.1 (Fig 11), and the increased distance to mountain does not show obvious effects until 7000 meters and threshold after that point (Fig 12). But in PDP for extreme heat LST, the increased SVF has greater effects on LST when it is more than 0.7, with a change from around 35.2℃ to reach 38.0℃ (Fig 13). The predicted heat resilience increases with increased SVF until 0.8, after that point the effect becomes negative, showing a complex pattern (Fig 14).

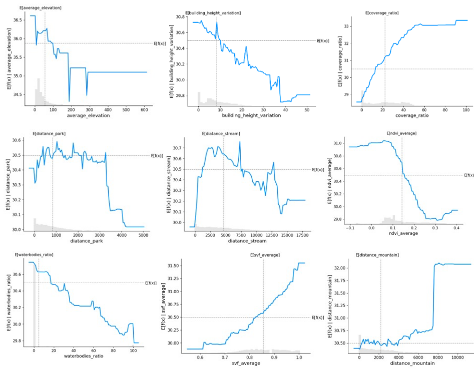

Fig. 8. Partial Dependence Plot for Regular LST

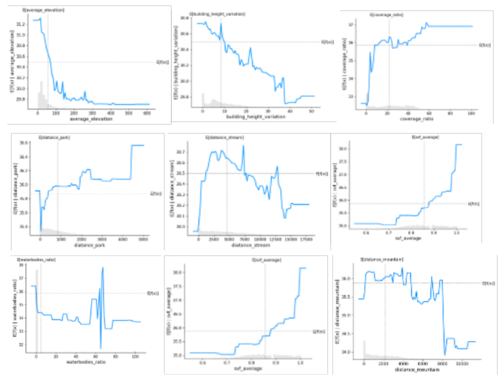

Fig. 9. Partial Dependence Plot for Extreme LST

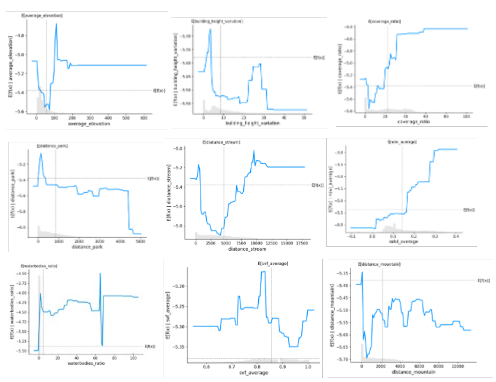

Fig. 10. Partial Dependence Plot for HR_T

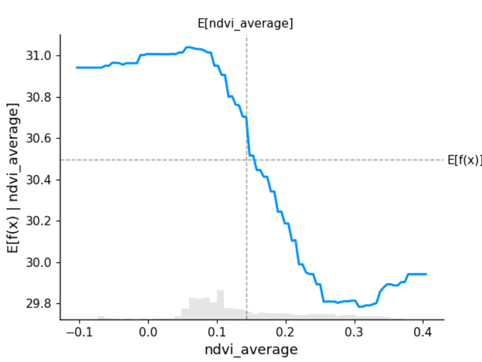

Fig. 11. PDP for NDVI in Regular_LST

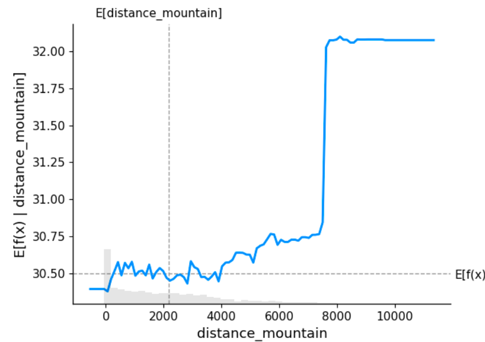

Fig. 12. PDP for distance to mountain in Regular_LST

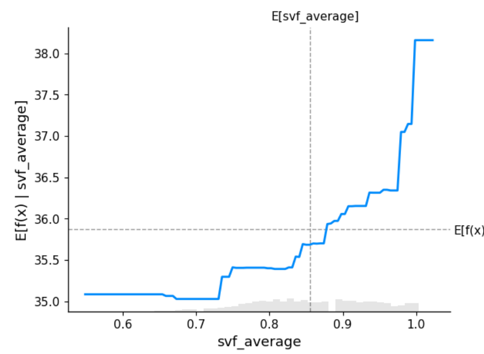

Fig. 13. PDP for sky view factor in Extreme_LST

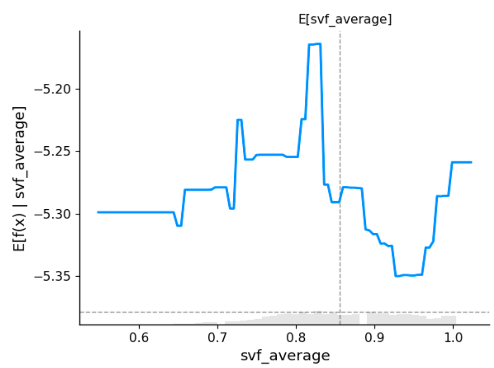

Fig. 14. PDP for sky view factor in HR_T

To understand the model performance in GBDT, this study compared R-squared, mean squared error (MSE), and mean absolute error (MAE) values with spatial lag model (SLM) (Table 1 & Table 2). Based on the results, the GBDT model has higher performance than spatial regression model with higher R-squared, smaller MSE and RMSE.

Table 1. Comparison between SLM and GBDT in LSTs

	Regular_LST			Extreme_LST		
Model	**R^2**	**MSE**	**RMSE**	**R^2**	**MSE**	**RMSE**
SLM	0.894	1.028	1.014	0.718	6.510	2.551

GBDT	0.991	0.080	0.284	0.870	2.990	1.729

Table 2. Comparison between SLM and GBDT in HR_T

	HR_T		
Model	**R^2**	**MSE**	**MAE**
SLM	0.249	6.853	2.617
GBDT	0.553	4.055	2.013

5. CONCLUSIONS

This study examines the relationship between urban form variables and UHI performance in Seoul, including regular LST, extreme heat LST and heat resilience. The overall results of this study indicate that impervious area ratio is the most influential variable for LST and heat resilience. At the same time, building density also affects LST because of the strong correlation between building density and the impervious area ratio. Additionally, building coverage ratio and NDVI are also important factors. The study suggests a nonlinear relationship between urban form factors and LST and heat resilience. Especially PDPs can show the effective range for each variable.

In general, the study contributes to a better understanding of influence of urban form factors on UHI effects in extreme heat events, especially important ones such as building density, land cover, and large-scale landscape. The findings, therefore, inform urban planning and design in adaption to climate change more effectively by prioritizing the factors.

5.1 Policy implication

Urban planners need indices for planning urban spaces that can mitigate or adapt to extreme heat. The results of this study can be a basis for developing a decision-making tool with a better understanding of the importance of variables in affecting heat resilience. The developed models could also be used for evaluating the effect of urban and environmental plans, climate-related projects and environmental policies by urban planners, policy drafters, and various stakeholders.

Urban planners and designers can collaborate to make land-use regulations and streetscape guidelines focusing on heat mitigation strategies by using cooling pavements for pedestrian ways and roads, green walls, and roofs. Firstly, we need to understand the urban form factors in each region. For example, according to the Land Planning Law in Korea, the appropriate area of park and green space for urban development projects should be 6m^2 per resident. However, there is a possibility of a generalization error that does not reflect the characteristics of each region. In that case, this study anticipate that we could predict the appropriate building density suitable based on each religion's characteristics.

5.2 Limitation

The LST can be one indicator defining extreme heat. However, it cannot explain all aspects of extreme heat characterized by many factors such as air temperature, relative humidity, and individual socio-economic situations. Therefore, conducting analysis based on LST is limited in understanding extreme heat. In addition, satellite image has limitations in their quality (season, cloud cover, interval, and time). It is challenging to extract surface temperatures due to a large amount of cloud cover from May to September when extreme heat events are concentrated, because it usually overlaps with the rainy season.

Hence, in future studies, the scope of research can be expanded by conducting a time-series analysis accompanied by collecting temperature and humidity data on an hourly basis based on real-time urban sensor data through machine learning techniques. In addition, this study has a limitation in that it only has only two time points in one year. It is not enough to generalize the whole summer season. If future studies use multi-year data with time-series analysis, it is expected to provide more accurate and meaningful results.

As extreme heat events become increasingly frequent and severe, many cities aim to build urban environments with a strong capability to recover from those events. However, there is still a lack of standards to measure or monitor heat resilience. In future studies, it would be beneficial to have a standard evaluation or monitoring method for comparative analyses across regions. With those improvements, the results of this study would be a better aid for urban planners to identify areas that are particularly vulnerable to climate risks.

ACKNOWLEDGEMENT

This work was supported by the Creative-Pioneering Researchers Program through Seoul National University (SNU), the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT) (No. 2018R1C1B5043758; No. 2022R1C1C1004953), and the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Education) (No. 5120200113713).

REFERENCE

[1] Alberti, Marina, et al. Integrating humans into ecology: opportunities and challenges for studying urban ecosystems. *BioScience*, 2003, 53.12: 1169-1179.

[2] Chen, Tzu-Ling; Lin, Zih-Hong. Impact of land use types on the spatial heterogeneity of extreme heat environments in a metropolitan area. *Sustainable Cities and Society*, 2021, 72: 103005.

[3] Chun, Bumseok; Guhathakurta, Subhrajit. Daytime and nighttime urban heat islands statistical models for Atlanta. *Environment and Planning B: Urban Analytics and City Science*, 2017, 44.2: 308-327.

[4] Chun, Bumseok; Guldmann, J.-M. Spatial statistical analysis and simulation of the urban heat island in high-density central cities. *Landscape and urban planning*, 2014, 125: 76-88.

[5] Chun, Bumseok; Guldmann, Jean-Michel. Impact of greening on the urban heat island: Seasonal variations and mitigation strategies. *Computers, Environment and Urban Systems*, 2018, 71: 165-176.

[6] Churkina, Galina. The role of urbanization in the global carbon cycle. *Frontiers in Ecology and Evolution*, 2016, 3: 144.

[7] Dialesandro, John, et al. Dimensions of thermal inequity: neighborhood social demographics and urban heat in the Southwestern US. *International journal of environmental research and public health*, 2021, 18.3: 941.

[8] Ding, Chuan; CAO, Xinyu Jason; NÆSS, Petter. Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo. *Transportation Research Part A: Policy and Practice*, 2018, 110: 107-117.

[9] Elith, Jane; Leathwick, John R.; Hastie, Trevor. A working guide to boosted regression trees. *Journal of animal ecology*, 2008, 77.4: 802-813.

[10] Gounaridis, Dimitrios; Koukoulas, Sotirios. Urban land cover thematic disaggregation, employing datasets from multiple sources and RandomForests modeling. *International Journal of Applied Earth Observation and Geoinformation*, 2016, 51: 1-10.

[11] Johnson, Daniel P., et al. Developing an applied extreme heat vulnerability index utilizing socioeconomic and environmental data. *Applied Geography*, 2012, 35.1-2: 23-31.

[12] Kalkstein, Laurence S.; Davis, Robert E. Weather and human mortality: an evaluation of demographic and interregional responses in the United States. *Annals of the association of American geographers*, 1989, 79.1: 44-64.

[13] Kim, D., Jang, C. R., Cho, J. Y., Jang, M. Y., & Song, J. Development of heatwave probability prediction model based on recurrent neural network: A case of seoul metropolitan government. *Journal of the Korean Society of Hazard Mitigation*, 2021, 21.5: 49-57.

[14] Kim, Tae-Hoon; Hong, Han-Kuk. A study on apartment price models using regression model and neural network model. *Korea Research Institute for Human Settlements*, 2004, 43: 183-200.

[15] Koman, Patricia D., et al. MI-Environment: Geospatial patterns and inequality of relative heat stress vulnerability in Michigan. *Health & place*, 2019, 60: 102228.

[16] Laaidi, Karine, et al. The impact of heat islands on mortality in Paris during the August 2003 heat wave. *Environmental health perspectives*, 2012, 120.2: 254-259.

[17] Leal Filho, Walter, et al. Coping with the impacts of urban heat islands. A literature based study on understanding urban heat vulnerability and the need for resilience in cities in a global climate change context. *Journal of Cleaner Production*, 2018, 171: 1140-1149.

[18] Li, Chaosu; SONG, Yan. Government response to climate change in China: A study of provincial and municipal plans. *Journal of Environmental Planning and Management*, 2016, 59.9: 1679-1710.

[19] Li, Y., Lee, S., & Han, J. Analysis of the relationship between three-dimensional built environment and urban surface temperature. *Journal of Korea Planning Association*, 2019, 54.2: 93-108.

[20] Luo, Pinyang, et al. How 2D and 3D built environments impact urban surface temperature under extreme heat: A study in Chengdu, China. *Building and Environment*, 2023, 231: 110035.

[21] Merrow, Sara; Keith, Ladd; WAGNER, Tess. Planning for extreme heat: a review. *Journal of Extreme Events*, 2019, 6.03n04: 2050003.

[22] Meerow, Sara; Newell, Joshua P.; Stults, Melissa. Defining urban resilience: A review. *Landscape and urban planning*, 2016, 147: 38-49.

[23] Mitchell, Daniel, et al. Attributing human mortality during extreme heat waves to anthropogenic climate change. *Environmental Research Letters*, 2016, 11.7: 074006.

[24] Molnar, Christoph. *Interpretable machine learning*. Lulu. com, 2020.

[25] Murdoch, W. James, et al. Definitions, methods, and applications in interpretable machine

learning. *Proceedings of the National Academy of Sciences*, 2019, 116.44: 22071-22080.

[26] Oleson, K. W., et al. Avoided climate impacts of urban and rural heat and cold waves over the US using large climate model ensembles for RCP8. 5 and RCP4. 5. *Climatic change*, 2018, 146: 377-392.

[27] Oke, Tim R. City size and the urban heat island. *Atmospheric Environment (1967)*, 1973, 7.8: 769-779.

[28] Sachindra, D. A., et al. Impact of climate change on urban heat island effect and extreme temperatures: a case-study. *Quarterly Journal of the Royal Meteorological Society*, 2016, 142.694: 172-186.

[29] Stone, Brian; HESS, Jeremy J.; FRUMKIN, Howard. Urban form and extreme heat events: are sprawling cities more vulnerable to climate change than compact cities?. *Environmental health perspectives*, 2010, 118.10: 1425-1428.

[30] Shin, Jaewon , Jeong, Sujong & Chang, Dongyeong. Estimation of Forest Carbon Stock in South Korea Using Machine Learning with High-Resolution Remote Sensing Data. *Atmosphere*, 2023, *33*.1, 61-72.

[31] Tao, et al. Exploring the nonlinear relationship between the built environment and active travel in the twin cities. *Journal of Planning Education and Research*, 2023, 43.3: 637-652.

[32] Turner, V. Kelly; Galletti, Christopher S. Do sustainable urban designs generate more ecosystem services? A case study of Civano in Tucson, Arizona. *The Professional Geographer*, 2015, 67.2: 204-217.

[33] Vaidyanathan, Ambarish, et al. A statistical framework to evaluate extreme weather definitions from a health perspective: a demonstration based on extreme heat events. *Bulletin of the American Meteorological Society*, 2016, 97.10: 1817-1830.

[34] Wu, Jingxian, et al. Spatial heterogeneity in the nonlinear impact of built environment on commuting time of active users: A gradient boosting regression tree approach. *Journal of Advanced Transportation*, 2023.

[35] Xi, Zhijie, et al. Built environment influences on urban climate resilience: Evidence from extreme heat events in Macau. *Science of The Total Environment*, 2023, 859: 160270.

[36] Xiao, Yijing, et al. Impacts of working from home during COVID-19 pandemic on physical and mental well-being of office workstation users. *Journal of occupational and environmental medicine*, 2021, 63.3: 181.

[37] Xu, Yiming, et al. Identifying key factors associated with ridesplitting adoption rate and modeling their nonlinear relationships. *Transportation Research Part A: Policy and Practice*, 2021, 144: 170-188.