# Assessment of Urban Heat Island Intensity and Key Drivers in Three Major Indonesian Cities Using Machine Learning

Wulan Salle Karurung [1], Lee Ki Rim [2], Lee Won Hee [3]

[1]Graduate Student, Department of Convergence and Fusion System Engineering, Kyungpook National University, Republic of Korea, wulansalle@knu.ac.kr

[2]Research Visiting Professor, Research Institute of Artificial Intelligent Diagnosis Technology for Multi-Scale Organic and Inorganic Structure, Kyungpook National University, Republic of Korea

[3]Professor, Department of Location-Based Information System, Kyungpook National University, Republic of Korea, wlee33@knu.ac.kr

***Abstract:*** *Urban heat island (UHI) effects in tropical cities are increasingly severe due to urban expansion, reduced green cover, and high population density. Environmental and anthropogenic factors interact to increase the surface temperature, necessitating detailed studies on these factors. This study aimed to quantify and compare the UHI characteristics and driving factors in three metropolitan cities in Indonesia: Jakarta, Surabaya, and Makassar. This study integrated monthly multivariate spatiotemporal data from 2019-2020, covering both natural and anthropogenic factors. The datasets were derived from satellites and included urban heat island (UHI) intensity, land surface temperature (LST), vegetation indices (NDVI), built-up indices (NDBI), PM2.5, $NO_2$, precipitation, wind speed, population density, and land use/land cover data. All datasets were standardized to a spatial resolution of 100 m. Three machine learning models, Random Forest (RF), XGBoost, and LightGBM (LGBM), were compared to evaluate their prediction accuracy. RF outperformed the others, with $R^2$ values of 0.92 for Jakarta and 0.92 for Makassar, whereas LightGBM had a high performance in Surabaya, with an $R^2$ of 0.95. SHAP and partial dependence plots (PDP) were used to interpret the importance of the features and their effects on the interaction. The results showed that Jakarta had the highest UHI intensity, and the top influencing factors were population density and precipitation, similar to Surabaya. In contrast, Makassar was influenced more by CO and precipitation levels. This study demonstrates the importance of integrating multi-feature geospatial data with machine learning to understand the interactions between features and make accurate predictions. This approach can be applied to other study areas, particularly tropical cities. These results can support regional climate resilience actions by identifying vulnerable areas and the key factors for sustainable planning.*

*Keywords: Urban Heat Island (UHI), Machine Learning, Spatiotemporal Analysis, Remote Sensing*

## Introduction

The urban heat island (UHI) phenomenon occurs when an area experiences significantly higher temperatures than its surroundings owing to the replacement of natural surfaces with impermeable urban materials that store heat, and this effect intensifies with increasing urbanization (Marcotullio et al., 2021). Rapid urbanization reduces vegetation cover and increases the inability of surfaces to absorb water, thereby increasing the surface temperature (Ilunga et al., 2024; Jang et al., 2024; Sidiqui et al., 2022). The potential consequences of UHI gradually affect human health and thermal comfort, increase the energy demand for cooling, and worsen air pollution. The concentration of most of the population in urban areas, which accelerates urban development and expansion, could lead to increasingly severe UHI effects in the future.

As a global phenomenon, UHI have significantly different driving factors in different climatic zones. Tropical cities, with their high humidity, consistent solar radiation that exacerbates thermal conditions, and intense and persistent UHI effects, present challenges in analyzing the UHI phenomenon. In tropical cities, consistently high baseline temperatures and rapid urbanization exacerbate the impacts of UHI, impairing human health and degrading water quality. For example, cities such as Jakarta, Surabaya, Makassar, and Bangkok face severe UHI impacts due to urbanization, impermeable surfaces, and a lack of green space (Cetin et al., 2024; Jaelani & Handayani, 2022; Maru et al., 2015; Nandi et al., 2022). These challenges need to be addressed by examining the features that influence UHI dynamics, such as physical, socioeconomic, and air pollution dynamics.

Previous studies have analyzed several key drivers influencing urban heat island dynamics, including increased impervious surfaces, buildings, heat trapping, vegetation, water reduction, and human activities that generate anthropogenic heat (Meng et al., 2025; Vujovic et al., 2021). Climate conditions also influence the UHI index, causing factors such as humidity, wind direction, and precipitation to affect the intensities differently(Al-Obaidi et al., 2021; Park et al., 2024). In addition, a study explored the correlation between $PM_{2.5}$ and temperature in urban areas (Rybarczyk et al., 2025). UHI dynamics can affect air pollution because trapped pollutants can increase their concentration and toxicity. Built environmental structures, such as buildings and roads, can block wind direction and trap pollutants from vehicles, industries, and other sources of pollution. The combination of high temperatures and poor air quality increases the risk of other diseases such as heat-related illnesses, respiratory problems, and cardiovascular diseases. However, research that quantitatively explores and integrates multi-feature effects across multiple tropical cities remains limited.

The integration of remote sensing and geographic information systems (GIS) with advanced data analysis and machine learning techniques has been increasingly used to assess environmental phenomena, particularly in UHI studies (Díaz-Chávez et al., 2024; Hong et al., 2025; Mallick & Alqadhi, 2025). This study analyzed several important factors related to physical and socioeconomic variables and pollutant data to gain a more diverse perspective. Some machine learning methods, such as Random Forest (RF), XGBoost, and LightGBM, perform well in handling complex and multivariate spatiotemporal datasets(Badugu et al., 2024; Liou et al., 2024; Mallick & Alqadhi, 2025). To increase the interpretability of machine learning results, this study used Shapley Additive exPlanations (SHAP) and Partial Dependence Plots (PDP) to get the global importance of each feature contribution, then PDP for visualizing the model's marginal response to features.

This study focuses on Jakarta, Surabaya, and Makassar, three rapidly growing Indonesian cities that face significant challenges related to UHIs. The main objectives were to quantify the UHI intensity using satellite-based land surface temperature (LST) data, evaluate the key factors influencing UHI vulnerability, and compare the UHI dynamics across the three cities. The insights from this study are expected to provide data-driven guidance to urban planners, policymakers, and environmental managers for developing sustainable solutions for UHI mitigation.

## Methodology

### Study Area

This study focuses on three major Indonesian cities, such as Jakarta, Surabaya, and Makassar, marked in Figure 1. These three cities were chosen because of the uniqueness of their urban environments, which are influenced by their geographic locations, economic activities, and demographic pressures. Jakarta is the capital of Indonesia and a prime example of severe UHI impacts in Indonesia, with a high population, building density, lack of green cover, and impermeable land surfaces. Surabaya, the second largest city in Indonesia and the capital of East Java, is famous for its industrial base and has almost the same economic activity as Jakarta. Makassar, a large city outside Java and the capital city of South Sulawesi, has experienced rapid urban expansion and significant land cover changes, including the conversion of natural areas to residential areas, illustrating how UHI risks evolve in a more recent urban context. The characteristics of these three large cities are similar to those of coastal cities, and they offer a

strong framework for investigating the spatial and temporal characteristics of the UHI at various stages of urban development in tropical climates.

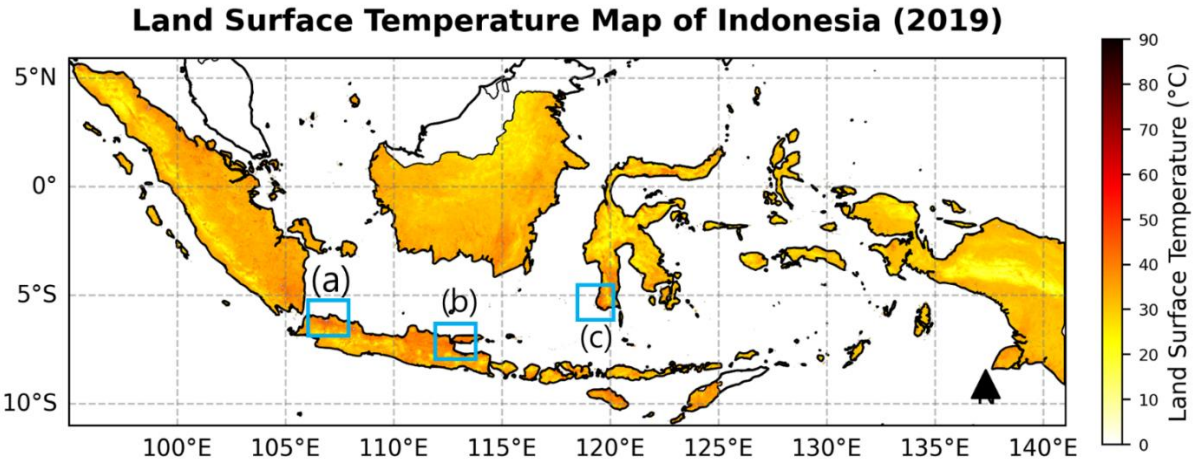


Figure 1: Study areas of (a) Jakarta, (b) Surabaya, and (c) Makassar.

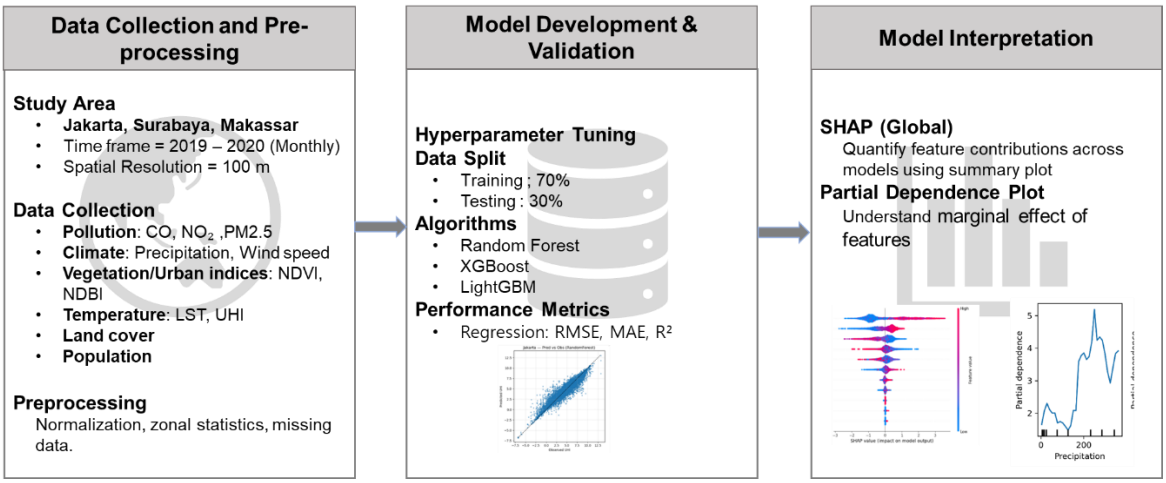


Figure 2: Framework of UHI Data Processing and Modeling

## Data Collection and Preprocessing

The methodological framework is illustrated in Figure 2. In data collections, this study integrated multi-featured data from various types and sources to analyze environmental and urban dynamics from 2019 to 2020. Atmospheric pollution, climate, land use, and socioeconomic data were standardized to a spatial resolution of 100 m and aggregated monthly to obtain both spatial and temporal perspectives. The pollution data used in this study were CO (carbon monoxide) and $NO_2$ (nitrogen dioxide) collected monthly from January 2019 to

December 2020 from the Copernicus Sentinel-5 Precursor (Sentinel-5P) mission data. The Sentinel-5P satellite provides daily global atmospheric coverage, and Level 3 (L3) data products were generated using Google Earth Engine (GEE). The datasets had a resolution of 1113.2 meters. Tropospheric CO concentration was used as a fossil fuel and biomass-burning emission tracer, while vertical column density of $NO_2$ was analyzed as a reactive pollutant that mainly originates from the combustion of fossil fuels. Additionally, $PM_{2.5}$ dataset was obtained from the Global High Air Pollution (GHAP) $PM_{2.5}$ concentration product with a 1 km resolution, which provides temporal daily data on $PM_{2.5}$ concentration ($\mu g/m^3$) (Wei et al., 2023). These pollutant data were selected due to their strong association with urban activities such as traffic emissions and energy consumption, which enhance dynamic information on understanding the urban environment.

Land Use Land Cover (LULC) data collected from the European Space Agency (ESA) LULC product were processed from Sentinel-1 and Sentinel-2 with 10-meter resolution. This study used the 2020 (v100) version that provides the coverage of trees such as tree cover, shrubland, grassland, cropland, built-up, bare/sparse vegetation, snow and ice, permanent water bodies, herbaceous wetland, Mangrove, and moss and lichen. The population data were sourced from the WorldPop Global Project, which estimates the residential population per 100 m Grid Square and is available in the GEE catalog.

Precipitation accumulation (pr) and 10 m wind speed (vs) were obtained from the TerraClimate dataset. This dataset provides high spatial and temporal climate and climatic water balance data for global terrestrial surfaces (Abatzoglou et al., 2018). The monthly data were collected with 4638.3 meters spatial resolution. These meteorological conditions can influence how heat is distributed and mitigated. Wind speed helps to dissipate heat from the city, while precipitation has a cooling effect on surfaces.

The Normalized Difference Built-up Index (NDBI) and Normalized Difference Vegetation Index (NDVI) were generated from Sentinel-2 surface reflectance (Sentinel-2 SR, level-2A) and processed to remove cloud cover of less than 20%. The spectral indices were derived for each monthly composite using the following formula:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR}$$

where near-infrared (NIR), RED, and shortwave infrared (SWIR) correspond to bands 8, 4, and 11, respectively. NDVI is widely used to indicate vegetation greenness and density, whereas NDBI highlights built-up and impervious surfaces.

Monthly LST data were obtained from the MODIS MOD11A2 product with 1 km resolution from January 2019 to December 2020. Daytime LST values were available in Kelvin and then converted to Celsius. The images were then combined to minimize the effects of cloud cover and temporal variability. The UHI intensity was calculated as the difference between the urban LST and the average rural LST for each region (Yang et al., 2024):

$$UHI = LST_{urban} - LST_{rural}$$

Urban and rural areas were distinguished using the ESA WorldCover 2020 dataset, which has a class to separate built-up areas as urban and non-built-up areas as rural. In this process, the LST data were spatially masked and divided based on these predefined urban and rural zones. The masking process helps to ensure each pixel of temperature was assigned consistently to the urban and rural categories, avoiding overlap with the mixed land cover.

After data collection, all datasets were resampled and harmonized to ensure spatial and temporal consistency. Because the datasets originated from various sources with differing spatial and temporal resolutions, preprocessing was required to enable integrated analysis. The meteorological and pollution variables had coarse resolution with ~4.6 km and ~1.1 km, respectively, and were then spatially downscaled to a 100 m grid resolution. In contrast, the higher resolution data from NDVI, NDBI, and LULC were aggregated using mean resampling. Similarly, each data set had various temporal resolutions and was then processed to match the same temporal resolution. Through this integration method, the uncertainty cannot be avoided because of this type of multifeatured assessment. But this type of research allows us to understand more about the interaction between the features.

Table 1: Spatial dataset and data sources.

| Data Type | Variable(s) | Source / Product | Spatial Resolution | Temporal Resolution |
|---|---|---|---|---|
| Pollution Data | CO and $NO_2$ | Sentinel-5P (GEE) | ~1.1 km | Monthly (2019–2020) |
| Air Quality | PM2.5 | GHAP PM2.5 (Wei et al., 2023) | 1 km | Monthly (2019–2020) |
| Climate | Precipitation, Windspeed | TerraClimate (GEE) | ~4.6 km | Monthly (2019–2020) |
| Vegetation / Urban | NDVI, NDBI | Sentinel-2 SR (Level-2A, GEE) | 10 m | Monthly (2019–2020) |
| Urban Heat / Temp | UHI, LST | MODIS MOD11A2 (GEE) | 1 km | Monthly (2019–2020) |
| Land Cover | LULC | ESA WorldCover v100 (2020) | 10 m | Static (2020) |
| Socio-Economic | Population Density | WorldPop Global (GEE) | 100 m | Annual (2020) |

**Machine learning analysis**

An integrated spatiotemporal dataset was prepared for training and testing purposes. The dataset in table format consisted of NDVI, NDBI, population, LULC, pollutant (CO, $NO_2$, $PM_{2.5}$), precipitation, and wind speed as independent variables, and the target variable was UHI intensity. Correlation analysis was conducted before the machine learning analysis to assess the relationship between the UHI and contributing factors using the Pearson correlation matrix.

In this study, three machine learning models were selected because of their robustness and ability to handle complex, nonlinear relationships. The models were Random Forest (RF), which is good for noisy data, and then XGBoost and LightGBM with gradient boosting, with high efficiency in large datasets. The dataset was split into 70% training and 30% testing datasets. Before training, each machine learning model was optimized using hyperparameter tuning to enhance its predictive performance and reduce overfitting. The optimal hyperparameters for the models were as follows: for Random Forest, n_estimators = 500,

min_samples_split = 2, and max_depth = 20; for XGBoost, subsample = 0.6, reg_lambda = 0.5, n_estimators = 300, max_depth = 10, learning_rate = 0.2, and colsample_bytree = 0.8; and for LightGBM, subsample = 0.6, reg_lambda = 0.5, num_leaves = 50, n_estimators = 500, max_depth = 10, learning_rate = 0.2, and colsample_bytree = 1.0.

**Model evaluation**

The regression model performance was evaluated using three error metrics: the coefficient of determination (R2), root mean square error (RMSE), and mean absolute error (MAE). A higher $R^2$ value indicates better estimation performance of the model, whereas lower RMSE and MAE values reflect improved accuracy. The formulas are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

where $\hat{y}_i$ is the predicted value, $y_i$ is the observed value, $\bar{y}$ is the mean of the observed values, and $N$ is the number of observations.

**Model interpretation.**

The results of machine learning models sometimes lack transparency regarding how the model predicts and how each input influences the output. To address this issue, Shapley Additive Explanations (SHAP) were applied to quantify the contribution of each model predictor to the model output. The TreeExplainer was used to compute the SHAP values based on the method used for tree-based models. SHAP values represent the extent to which each feature shifts a prediction from the mean. Summary plots were used to rank the most influential variables, and partial dependence plots (PDPs) were used to explore feature interactions with UHI intensity. PDP can visualize the information on how each feature affects the output model, thus allowing us to interpret the behavior and pattern of the data.

**Results and Discussion**

Figure 3 shows the time-series monthly variations of UHI intensity and other features of environmental, atmospheric, and pollutant factors (LST, NDVI, NDBI, $NO_2$, $PM_{2.5}$, precipitation, wind speed, and CO) in Jakarta, Surabaya, and Makassar from January 2019 to

December 2020. Each line in the graph shows the mean data for each month, with the shaded bands representing the minimum and maximum standard deviations.

Jakarta consistently recorded higher LST and pollutant concentrations ($NO_2$, $PM_{2.5}$, CO) with a UHI range of 3-8°C. Surabaya exhibited moderate levels of UHI in 1-5°C, while Makassar had a lower UHI in 0-6°C and low pollutant concentrations compared to Surabaya and Jakarta. A seasonal cycle was evident, with temperature and pollutant levels peaking during the dry season (April–October) and decreasing with rainfall during the rainy season (October–April). Jakarta showed the highest UHI intensity, with maximum values reaching 8°C in January 2019 and 2020. Surabaya recorded a moderate intensity (2–5°C), whereas Makassar showed the weakest UHI, with an average of 3°C. Across the city, the UHI decreased from July to October 2019 and increased again in early 2020. In July–October 2019 and 2020, precipitation accumulation followed a pattern similar to that of the UHI, whereas wind speed increased. The similar patterns between UHI and precipitation indicate the influence of climatic seasonality, urban characteristics, and human activity.

**Correlation Analysis**

The correlation matrices presented in Figure 4 highlight the associations between UHI intensity and multiple explanatory variables. Each cell shows the correlation value between the crossed variables, ranging from -1 to 1, which provides information on the negative and positive correlations. NDVI showed a negative correlation, indicating that greener areas mitigate UHI effects, whereas the built-up index (NDBI) exhibited a positive correlation, reflecting the support of built-up factors for increasing UHI. However, in Makassar, it showed a negative relationship.

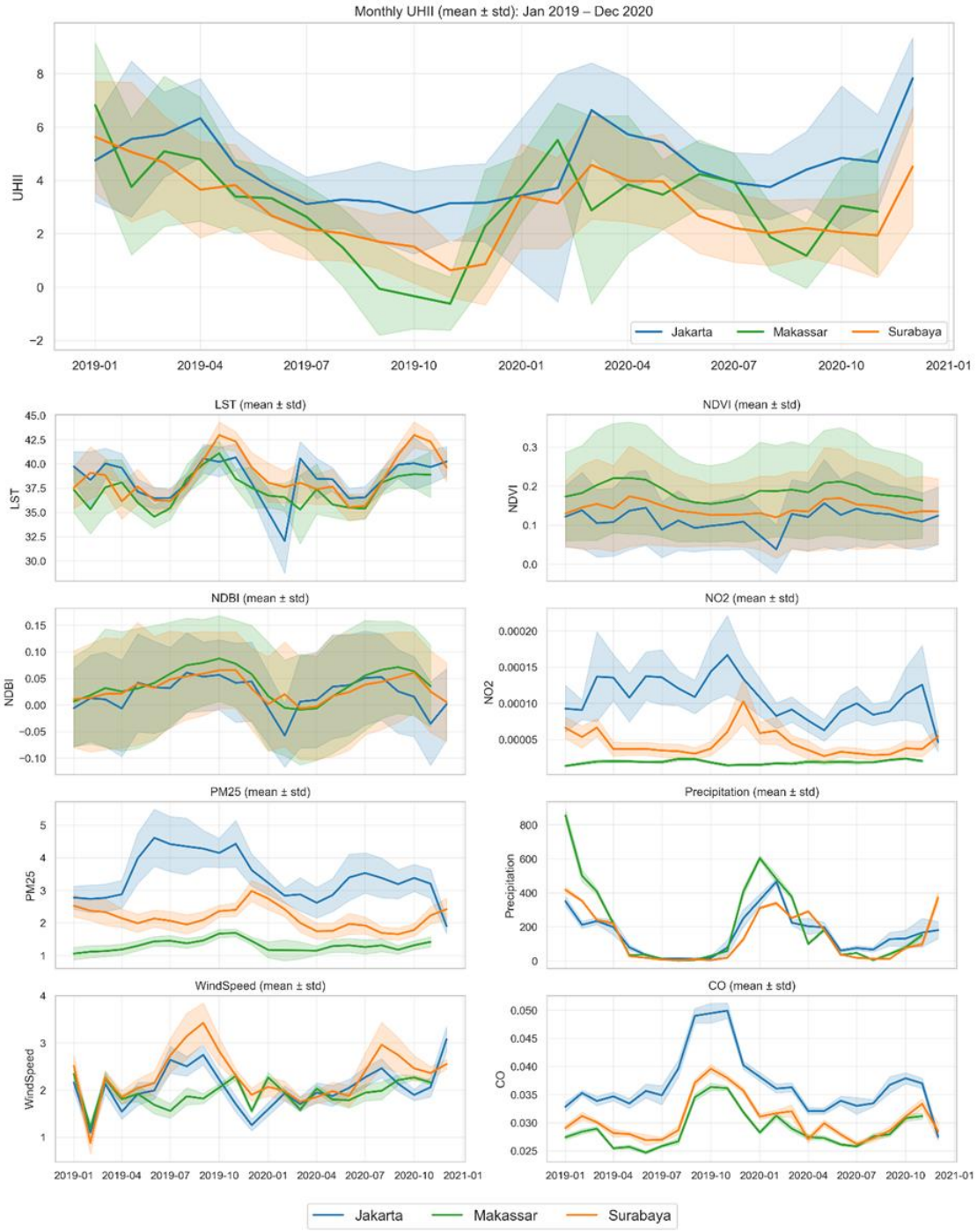


Figure 3: Time series of features.

Air pollutants, particularly PM2.5, $NO_2$, and CO, displayed negative relationship with the UHI and positive correlations in all cities, except for PM2.5 in Surabaya. In terms of climatic features, precipitation was positively correlated with UHI in all cities, whereas wind speed was negatively correlated. In Surabaya, CO was more strongly correlated with the UHI than in Jakarta or Makassar, indicating localized emission and dispersion dynamics. Population density and land-use classes (tree cover, built-up land, and water) also varied in their influence,

with tree cover in all cities being negatively associated with the UHI. In contrast, the built-up fraction and population were positively correlated in Jakarta and Surabaya. In contrast to Makassar, the built-up index and population are negatively correlated, and water coverage is positively correlated. These pattern makes the socio-economic schemes of Jakarta and Surabaya similar, whereas Makassar exhibits an opposite trend.

LST was excluded as an input feature for machine learning training because UHI is directly derived from LST, which would result in high collinearity between the predictor and target variable. Excluding LST avoids redundancy and potential bias and ensures that the contributions of other explanatory variables are assessed independently. However, in this study, LST was used as a comparative feature to distinguish trends, as the monthly patterns of UHI and LST do not always align, as shown in Figure 3.

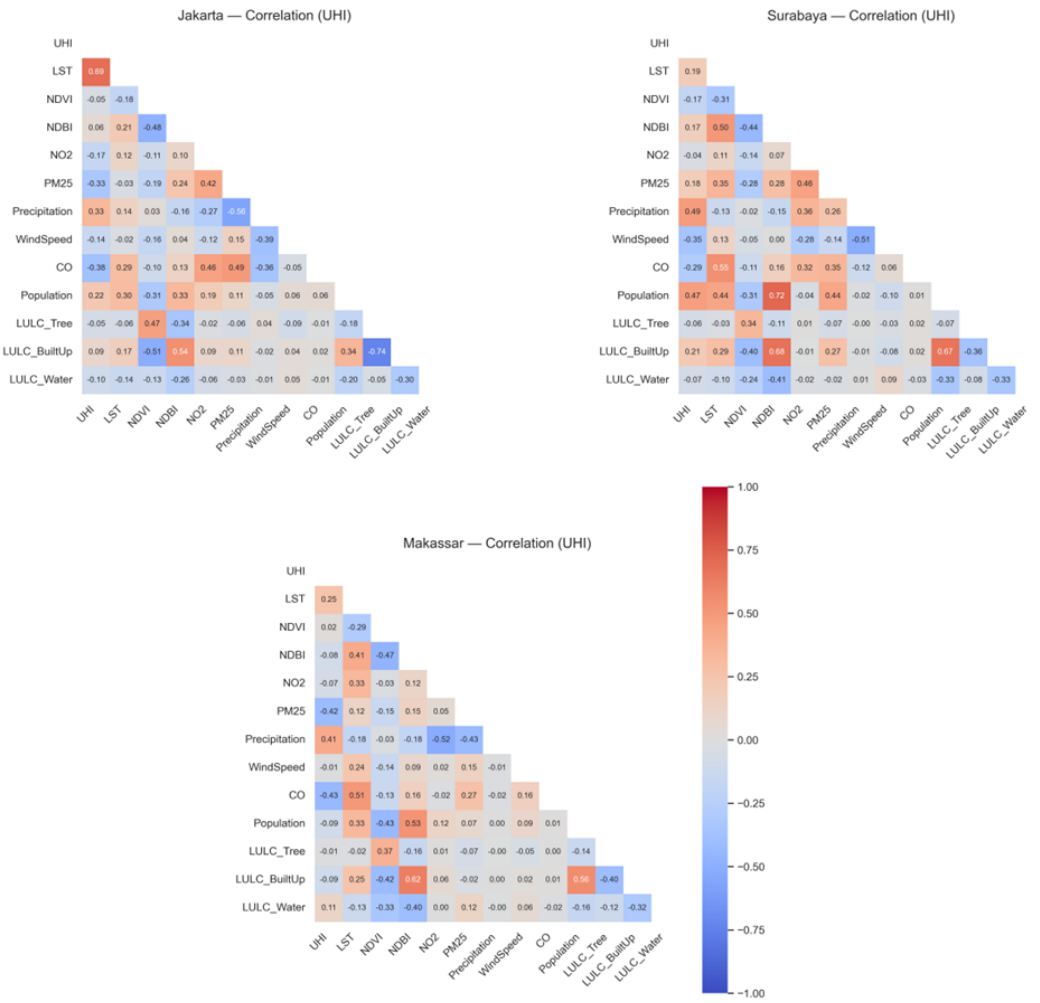


Figure 4: Matrix correlation for all features in the study area.

**Machine Learning Model**

The model performance results are listed in Table 2. In Surabaya, LightGBM slightly outperformed the others, with the highest R2 of 0.951. Makassar and Jakarta had the Random Forest model as the best model, with R2 values of 0.95 and 0.923, respectively. The performance of each city based on the best model is shown in Figure 5.

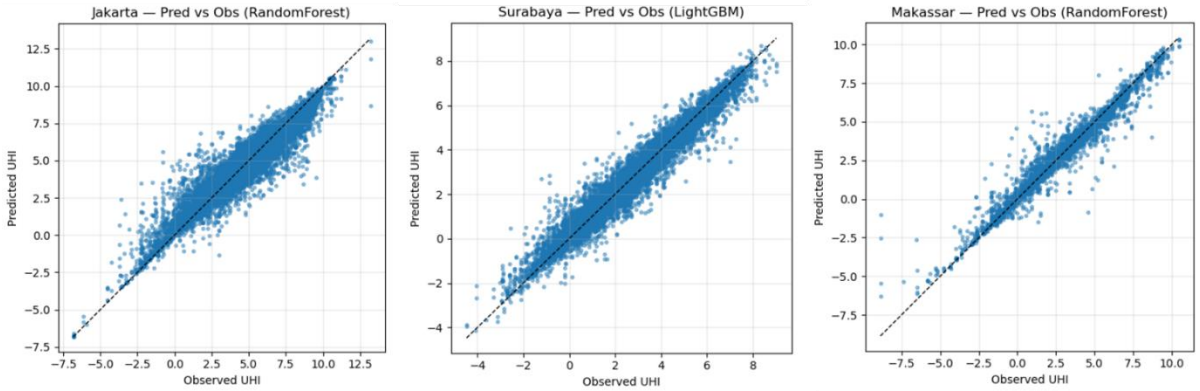


Figure 5: Predicted vs observed scatter plot

Table 2: Model performance metrics

| City | Model | $R^2$ | RMSE | MAE |
|---|---|---|---|---|
| **Surabaya** | XGBoost | 0.948206 | 0.460144 | 0.310600 |
| | LightGBM | 0.951210 | 0.446600 | 0.310289 |
| | Random Forest | 0.946870 | 0.466041 | 0.283379 |
| **Makassar** | XGBoost | 0.946649 | 0.588201 | 0.370512 |
| | LightGBM | 0.950248 | 0.568016 | 0.355977 |
| | Random Forest | 0.953584 | 0.548639 | 0.291595 |
| **Jakarta** | XGBoost | 0.916638 | 0.562636 | 0.381523 |
| | LightGBM | 0.895447 | 0.630104 | 0.446942 |
| | Random Forest | 0.922713 | 0.541748 | 0.338228 |

### SHAP and Feature Importance

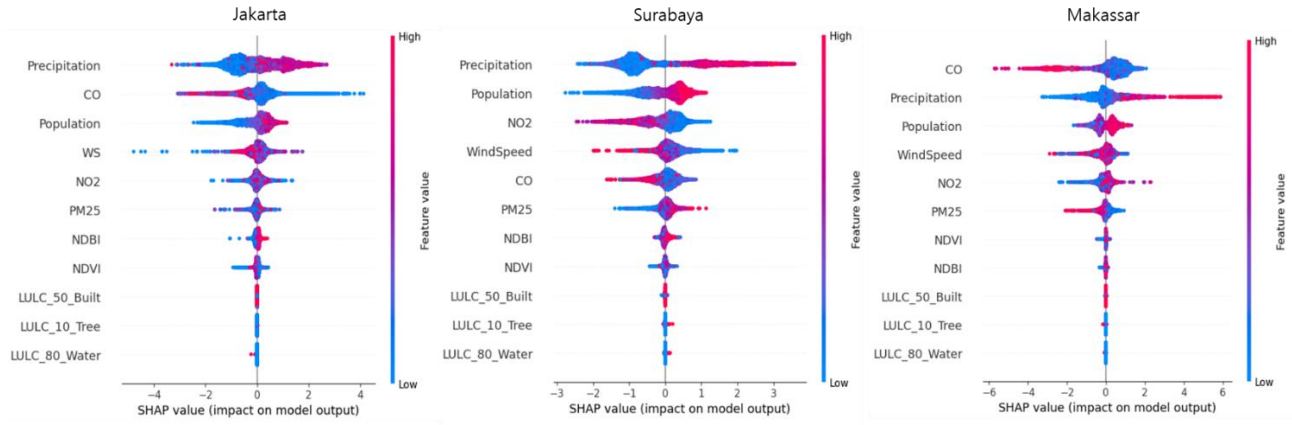


Figure 6: SHAP summary plots for Jakarta, Surabaya, and Makassar.

This study used SHAP to interpret the results of machine learning models to understand the role and influence of each feature on predictions. A summary plot for each city is shown in Figure 6, which displays the global ranking of the feature importance. In Jakarta, precipitation was the most influential feature, with high precipitation associated with a stronger positive contribution to the UHI prediction, followed by CO and population, whereas CO had an inverse contribution. In Surabaya, precipitation and population showed a strong positive effect, whereas $NO_2$ and wind speed had a negative influence. In Makassar, CO was the most influential factor, showing both strong positive and negative effects. Similar to other cities, precipitation and population in Makassar contributed positively to the UHI prediction. Across all cities, the land cover classes, NDBI, and NDVI made relatively minor contributions.

### Partial Dependence Plots (PDP)

The results of the Partial Dependence Plots for the three Indonesian cities are displayed in Figure 7. The top five most influential features, based on the SHAP feature importance results, were selected to examine how these factors affected the model's predicted outcome. All three cities had the same top five features: precipitation, CO, population, wind speed, and $NO_2$ in different orders.

Precipitation in Jakarta and Surabaya showed a fluctuating positive relationship, with the peak of UHI prediction in the range of 150–200 mm of total monthly precipitation. Meanwhile, in Makassar, precipitation has a general positive trend. But if we compare with the LST monthly trend in Figure 3 and the Pearson correlation in Figure 4, in Jakarta, LST and precipitation have both positive relationships, while in Surabaya and Makassar, precipitation has a negative relationship with LST. This difference can be explained by the distinction between LST and

UHI. LST is the absolute thermal condition, while the UHI is the difference between the urban and rural temperatures. The high value of UHI during the high precipitation period can suggest that urban areas cool more slowly than the surrounding rural areas because of the high impervious surface and persistent anthropogenic heat sources in urban areas. These findings are aligned with the previous research about the complex relationship between precipitation, temperature, and UHI (Steensen et al., 2022).

Population density exhibited a generally positive association, with higher values corresponding to higher predicted UHI intensity. Wind speed had a negative trend, with the higher wind speed associated with lower UHI intensity. Although the change was not significant, the UHI partial dependence values were high in all cities. The same case applies to $NO_2$ in Jakarta and Makassar, where the trend line is flat, but the partial dependence value is high. Moreover, the $NO_2$ in Surabaya fluctuated and decreased at high values of $NO_2$. CO had a high prediction of UHI at low levels of CO, but decreased when CO had a high value, and this was similar in Jakarta and Makassar. In Surabaya, the CO concentration was relatively stable, with minor fluctuations.

The results of the UHI intensity comparison across Jakarta, Surabaya, and Makassar revealed important insights into how different aspects of urban form, such as socioeconomic factors, local climate, and physical characteristics, interact to shape urban heat exposure. The results revealed that wind direction and precipitation influenced UHI intensity in all cities, with higher rainfall increasing the temperature gap between urban areas and their surroundings, likely due to the interaction between reduced evapotranspiration and heat retention in developed areas. Among the features analyzed, precipitation and wind speed are beyond human control, but population density and land use patterns amplify their effects. SHAP and PDP analyses confirmed that these factors consistently dominated the UHI predictions across cities, emphasizing their role in urban heat dynamics. These differences in exposure indicate that each city faces similar challenges in reducing green cover areas owing to increasing land demand and population growth. The results showed that rainfall increased the UHI intensity, indicating that the temperature gap between the city and its surroundings became larger.

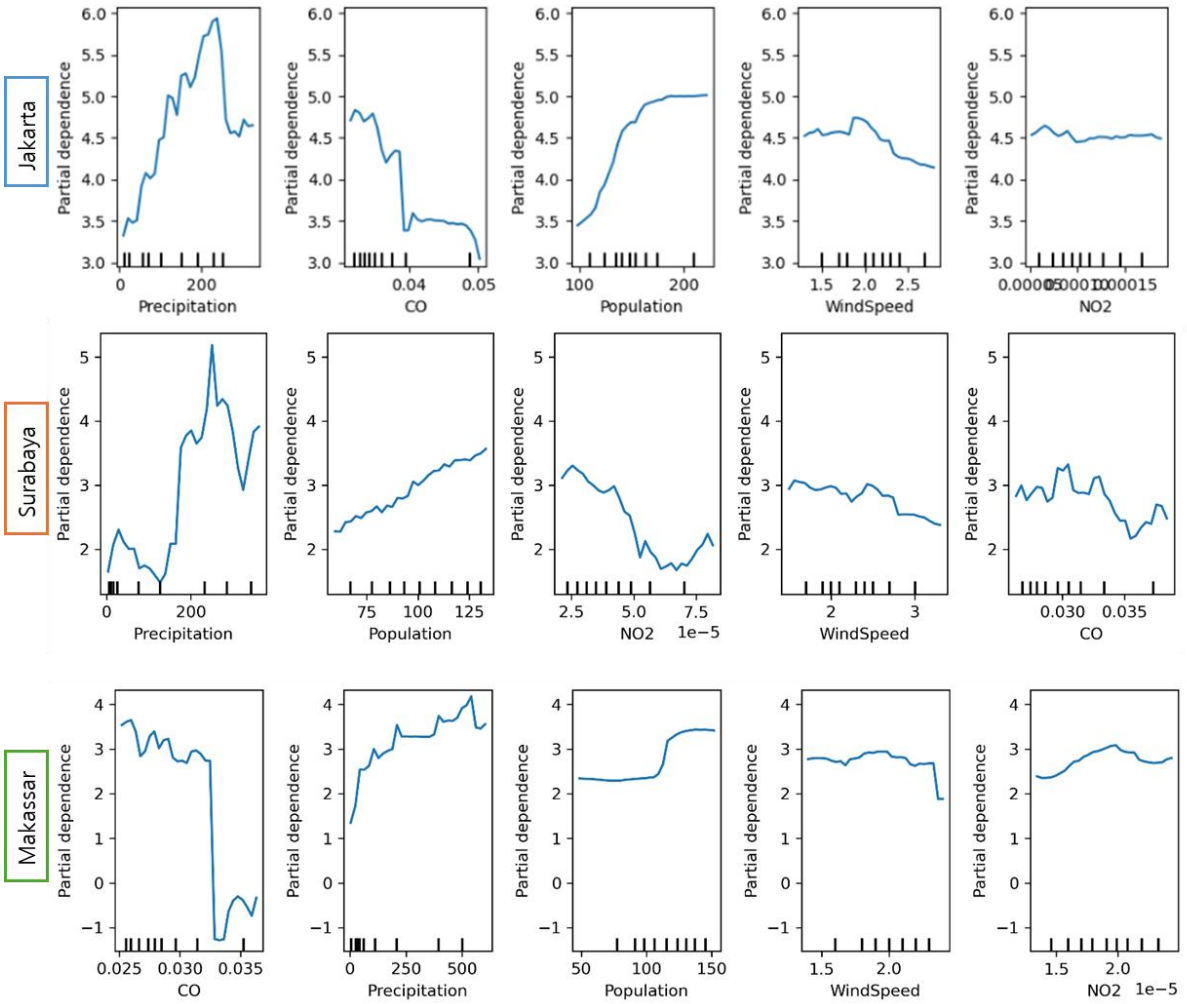


Figure 7: PDP plots of all features.

Given the similar coastal geography and growing economy with business activities in all three cities, it is important to continuously monitor UHI exposure and study the key drivers to identify potential cooling features for sustainable urban living in these cities. The application of remote sensing products, followed by integration with GIS and machine learning, allows researchers and policymakers to monitor and understand the key drivers of UHI and mitigate its impact to support sustainable urban living.

In addition to providing valuable insights, this study has some limitations that should be considered. First, this analysis calculated the UHI variation for monthly data using MODIS temperature maps, which have a low spatial resolution despite their high temporal resolution; also, the downscaling and upscaling processes used to harmonize variables may introduce additional uncertainty. Second, the temporal scope is limited, as only two years of monthly data were analyzed to determine the long-term influence of the UHI in large cities. Third, the machine learning used in this study does not capture temporal dynamics for training because

the training and testing data are the accumulation of all table datasets. For future research, high-resolution UHI data capturing full monthly coverage is needed to enable better visualization and accurate analysis. In addition, advanced analyses that can cover the variation in the temporal and spatial aspects of the analysis should be considered to capture the dynamics of the UHI in urban environments.

**Conclusion and Recommendation**

This study completed an analysis integrating machine learning and XAI approaches to assess the UHI intensity in three Indonesian metropolitan cities: Jakarta, Surabaya, and Makassar. The results show that UHI intensity varies across cities and seasons, with Jakarta exhibiting the strongest UHI, Surabaya exhibiting moderate UHI, and Makassar exhibiting the weakest UHI. The key findings highlight that both climatic and anthropogenic factors drive the UHI dynamics. Jakarta exhibited the most vigorous UHI intensities and had similar patterns in several outputs to Surabaya, particularly in terms of population, precipitation, and wind speed, whereas Makassar presented slightly different outputs. The machine learning models performed well, with Random Forest and LightGBM achieving the best results. The interpretation of the machine learning training results using SHAP and PDP analyses identified precipitation, CO, population, wind speed, and $NO_2$ as the dominant drivers of the UHI among the features used in this study. Overall, the findings of this study can be used to understand the relationships and interactions between climatic and anthropogenic factors that shape UHI dynamics in tropical cities, which can then form the basis for developing predictive modeling to support urban planning, climate adaptation, and evidence-based policymaking in tropical metropolitan areas.

From a policy perspective, these findings can serve as practical guidance for developing local urban heat mitigation strategies. Because the key drivers vary from city to city, this key driver factor assessment allows us to understand the identified primary drivers. For example, increasing green cover and corridor-based development planning can help reduce surface heat accumulation. In this case study, all cities are coastal. Compared to large cities surrounded by land, these three cities integrate emission control with coastal wind management, which can further support urban cooling. Integrating the machine learning and satellite-based monitoring used in this study into local climate adaptation programs and then scaling them up to other urban areas will enable a more data-driven approach that takes a holistic view of global and

local climate influences to identify heat-vulnerable areas, improve early warning systems, and support sustainable urban development in tropical environments.

**Acknowledgements**

This research was supported by the National Research Foundation of Korea (RS-2021-NR060108), the Korea Institute of Energy Technology Evaluation and Planning (KETEP), and the Ministry of Trade, Industry, and Energy (MOTIE) of the Republic of Korea (No. RS-2022-KP002719).

**References**

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015. *Scientific Data*, *5*, 1–12. https://doi.org/10.1038/sdata.2017.191

Al-Obaidi, I., Rayburg, S., Półrolniczak, M., & Neave, M. (2021). Assessing the Impact of Wind Conditions on Urban Heat Islands in Large Australian Cities. *Journal of Ecological Engineering*, *22*(11), 1–15. https://doi.org/10.12911/22998993/142967

Badugu, A., Arunab, K. S., & Mathew, A. (2024). Predicting land surface temperature using data-driven approaches for urban heat island studies: a comparative analysis of correlation with environmental parameters. In *Modeling Earth Systems and Environment* (Vol. 10, Issue 1). Springer International Publishing. https://doi.org/10.1007/s40808-023-01822-2

Cetin, M., Ozenen Kavlak, M., Senyel Kurkcuoglu, M. A., Bilge Ozturk, G., Cabuk, S. N., & Cabuk, A. (2024). Determination of land surface temperature and urban heat island effects with remote sensing capabilities: the case of Kayseri, Türkiye. *Natural Hazards*, *0123456789*. https://doi.org/10.1007/s11069-024-06431-5

Díaz-Chávez, L., Melendez-Surmay, R., & Arregocés, H. A. (2024). Urban heat island intensity in coastal cities of northern Colombia using Landsat data and WRF/UCM model. *Case Studies in Chemical and Environmental Engineering*, *9*(January). https://doi.org/10.1016/j.cscee.2024.100617

Hong, T., Yim, S. H. L., & Heo, Y. (2025). Interpreting complex relationships between urban and meteorological factors and street-level urban heat islands: Application of random forest and SHAP method. *Sustainable Cities and Society*, *126*(April), 106353. https://doi.org/10.1016/j.scs.2025.106353

Ilunga, G., Bechet, J., Linguet, L., Zermani, S., & Mahamat, C. (2024). Spatial and temporal variation of urban heat island in french guiana. *SENSORS*, *24*, 1931.

Jaelani, L. M., & Handayani, C. A. (2022). Spatio-temporal Analysis of Land Surface Temperature Changes in Java Island from Aqua and Terra MODIS Satellite Imageries Using Google Earth Engine. *International Journal of Geoinformatics*, *18*(5), 1–12. https://doi.org/10.52939/ijg.v18i5.2365

Jang, S., Bae, J., & Kim, Y. J. (2024). Street-level urban heat island mitigation: Assessing the cooling effect of green infrastructure using urban IoT sensor big data. *Sustainable Cities and Society*, *100*(October 2023), 105007. https://doi.org/10.1016/j.scs.2023.105007

Liou, Y. A., Tran, D. P., & Nguyen, K. A. (2024). Spatio-temporal patterns and driving forces of surface urban heat island in Taiwan. *Urban Climate*, *53*(December 2023), 101806. https://doi.org/10.1016/j.uclim.2024.101806

Mallick, J., & Alqadhi, S. (2025). Explainable artificial intelligence models for proposing mitigation strategies to combat urbanization impact on land surface temperature dynamics in Saudi Arabia. *Urban Climate*, *59*(August 2024), 102259. https://doi.org/10.1016/j.uclim.2024.102259

Marcotullio, P. J., Keßler, C., Quintero Gonzalez, R., & Schmeltz, M. (2021). Urban Growth and Heat in Tropical Climates. *Frontiers in Ecology and Evolution*, *Volume 9-*. https://www.frontiersin.org/journals/ecology-and-evolution/articles/10.3389/fevo.2021.616626

Maru, R., Baharuddin, I. I., Zhiddiq, S., Arfan, A., Geografi, J., Matematikadan, F., & Alam, I. P. (2015). Trend Analysis of Urban Heat Island Phenomenon in the City of Makassar, South Sulawesi, Indonesia using Landsat. *Asian Journal of Applied Sciences*, *03*(05), 2321–0893. www.ajouronline.com

Meng, Y., Gao, C., Yu, W., Zhao, E., Li, W., Wang, R., Zhao, Y., Zhao, H., & Zeng, J. (2025). The Spatiotemporal Evolution and Driving Factors of Surface Urban Heat Islands: A Comparative Study of Beijing and Dalian (2003–2023). *Remote Sensing*, *17*(10), 1–29. https://doi.org/10.3390/rs17101793

Nandi, N., Indonesia, U. P., Dede, M., Ijost, I., & Indonesia, U. P. (2022). *Indonesian Journal of Science & Technology Urban Heat Island Assessment using Remote Sensing Data in West Java , Indonesia : From Literature Review to Experiments and Analyses*. *April*. https://doi.org/10.17509/ijost.v7i1.44146

Park, K., Baik, J. J., Jin, H. G., & Tabassum, A. (2024). Changes in urban heat island intensity with background temperature and humidity and their associations with near-surface thermodynamic processes. *Urban Climate*, *58*(October), 102191. https://doi.org/10.1016/j.uclim.2024.102191

Rybarczyk, Y., Zalakeviciute, R., Ereminaite, M., & Costa-Stolz, I. (2025). Causal effect of PM2.5 on the urban heat island. *Frontiers in Big Data*, *Volume 8-*. https://www.frontiersin.org/journals/big-data/articles/10.3389/fdata.2025.1546223

Sidiqui, P., Roös, P. B., Herron, M., Jones, D. S., Duncan, E., Jalali, A., Allam, Z., Roberts, B. J., Schmidt, A., Tariq, M. A. U. R., Shah, A. A., Khan, N. A., & Irshad, M. (2022). Urban Heat Island vulnerability mapping using advanced GIS data and tools. *Journal of Earth System Science*, *131*(4), 266. https://doi.org/10.1007/s12040-022-02005-w

Steensen, B. M., Marelle, L., Hodnebrog, & Myhre, G. (2022). Future urban heat island influence on precipitation. *Climate Dynamics*, *58*(11–12), 3393–3403. https://doi.org/10.1007/s00382-021-06105-z

Vujovic, S., Haddad, B., Karaky, H., Sebaibi, N., & Boutouil, M. (2021). Urban Heat Island: Causes, Consequences, and Mitigation Measures with Emphasis on Reflective and Permeable Pavements. *CivilEng*, *2*(2), 459–484. https://doi.org/10.3390/civileng2020026

Wei, J., Li, Z., Lyapustin, A., Wang, J., Dubovik, O., Schwartz, J., Sun, L., Li, C., Liu, S., & Zhu, T. (2023). First close insight into global daily gapless 1 km PM2.5 pollution, variability, and health impact. *Nature Communications*, *14*(1), 8349. https://doi.org/10.1038/s41467-023-43862-3

Yang, Q., Xu, Y., Chakraborty, T. C., Du, M., Hu, T., Zhang, L., Liu, Y., Yao, R., Yang, J., Chen, S., Xiao, C., Liu, R., Zhang, M., & Chen, R. (2024). A global urban heat island intensity dataset: Generation, comparison, and analysis. *Remote Sensing of Environment*, *312*(July). https://doi.org/10.1016/j.rse.2024.114343