IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

Improved Mapping of Surface Urban Heat Islands Using Machine Learning and Enhanced Environmental Indicators

Vijaya Kumar Gavara, Sudheer Babu Golla, Maheh Chekka
Computer Science & Engineering
RVR & JC College of Engineering, Guntur, India

Abstract: *Addressing climate change is crucial for the development of sustainable and resilient smart cities, particularly in mitigating the impacts of Surface Urban Heat Islands (SUHI). This study presents a machine learning-based framework for modelling SUHI by predicting Land Surface Temperature (LST) using a diverse set of environmental and socioeconomic variables. Expanding upon previous research, our approach integrates additional environmental indicators including the Normalized Difference Water Index (NDWI), Soil Adjusted Vegetation Index (SAVI), and soil moisture data, alongside conventional metrics such as NDVI and NDBI. The analysis was conducted over the metropolitan region of Milan, Italy, utilizing 15 cloud-free Landsat 8 images captured between 2019 and 2021, with complementary data from Sentinel-2 and Planet Scope satellites. Six machine learning models were evaluated to estimate LST, with Decision Tree achieving the highest accuracy (R^2 = 1.00, MAE = 0.00 °C, RMSE = 0.07 °C), closely followed by Random Forest (R^2 = 1.00, MAE = 0.01 °C, RMSE = 0.04 °C). Our findings highlight the significance of incorporating multi-seasonal imagery and enriched feature sets in improving SUHI characterization. The proposed approach not only enhances the precision of temperature mapping but also offers valuable insights for urban planning strategies aimed at combating urban heat in rapidly developing cities*

Keywords: urban heat island, land surface temperature, machine learning, remote sensing, environmental indicators (NDWI, SAVI, soil moisture)

I. INTRODUCTION

Urbanization continues to reshape natural landscapes, significantly increasing the extent of impervious surfaces such as roads and buildings. These changes disrupt natural land cover, leading to elevated land surface temperatures (LST) and intensifying the urban heat island (UHI) effect—a phenomenon where urban areas exhibit significantly higher temperatures than nearby rural regions. This thermal anomaly poses serious threats to urban ecosystems, local climates, water systems, biodiversity, and public health.

Among various indicators, LST has emerged as a crucial parameter for identifying and understanding the UHI phenomenon. While several studies have relied on LST measurements from a limited number of satellite images—often from a single season or year—this may not accurately reflect temporal and spatial variations influenced by factors like seasonal changes, vegetation, soil moisture, and human activity. To address this limitation, our study utilizes a broader temporal dataset by incorporating multi-seasonal thermal imagery.

Remote sensing, especially satellite data from platforms like **Landsat 8**, **Sentinel-2**, and **Planet**, provides an effective means for large-scale SUHI (Surface Urban Heat Island) mapping. These platforms not only offer LST data but also enable the extraction of environmental indicators such as the **Normalized Difference Water Index (NDWI)**, **Soil Adjusted Vegetation Index (SAVI)**, and **soil moisture content**, which are vital to understanding urban thermal patterns. Unlike traditional indices like NDVI and NDBI alone, our study incorporates a richer set of environmental features for more comprehensive analysis.

DOI: 10.48175/IJARSCT-26289

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

In addition to environmental parameters, socioeconomic factors also contribute to SUHI dynamics. Although underrepresented in prior studies, variables such as population density, income levels, and built environment characteristics offer valuable context. Integrating these with environmental data provides a multidimensional view of urban thermal behaviour.

Historically, methods such as correlation analysis, linear regression, and geographically weighted regression (GWR) have been employed for SUHI characterization. However, given the complex interactions among the influencing factors, machine learning techniques offer a more flexible and robust alternative. While previous works have explored ML for LST estimation, they have often overlooked the interpretability of variable importance and seasonal influence.

This study addresses these gaps by implementing and evaluating multiple machine learning algorithms to predict LST using both environmental and socioeconomic variables. Focusing on **Milan, Italy**, the study utilizes 15 Landsat 8 thermal images from **2019 to 2021**, and integrates **Sentinel-2 and Planet Scope data**, alongside census-based socioeconomic datasets. We compare the predictive performance of six machine learning models, finding **Decision Tree** and **Random Forest** as top performers. Notably, the **Decision Tree model achieved near-perfect accuracy** with **MAE = 0.00 °C, RMSE = 0.07 °C, and R^2 = 1.00**, highlighting its potential for SUHI analysis.

Our contributions include:

Incorporation of **multi-seasonal data** for better temporal understanding of SUHI.

Inclusion of **advanced environmental indicators** like SAVI, NDWI, and soil moisture.

Detailed evaluation of the **predictive strength of socioeconomic variables**.

A demonstration of how **machine learning models can optimize SUHI mapping** for **urban planning in Milan**.

These insights are crucial for data-driven urban development and offer practical guidance for mitigating heat stress in rapidly growing smart cities.

II. METHODS & MATERIALS

The methodology adopted in this study was divided into four key stages: data acquisition, data preparation, machine learning-based modelling, and spatial analysis. Multisource data including remote sensing images (e.g., Landsat 8) and demographic census information were collected to derive both environmental variables (such as NDVI, NDBI, and LST) and socioeconomic indicators (like population density and median income). These variables were aligned with administrative census sectors to maintain geographic consistency.

After necessary preprocessing—such as normalization and attribute selection—data were structured into multiple models combining different sets of variables. Machine learning regression algorithms were then employed to predict Land Surface Temperature (LST), with performance evaluation based on statistical metrics. The best-performing model was used to generate spatially distributed LST predictions, which were validated using reference LST values. Finally, SUHI zones were delineated by comparing predicted LST with baseline values, followed by an analysis of how different environmental and socioeconomic attributes contributed to SUHI intensity and distribution.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

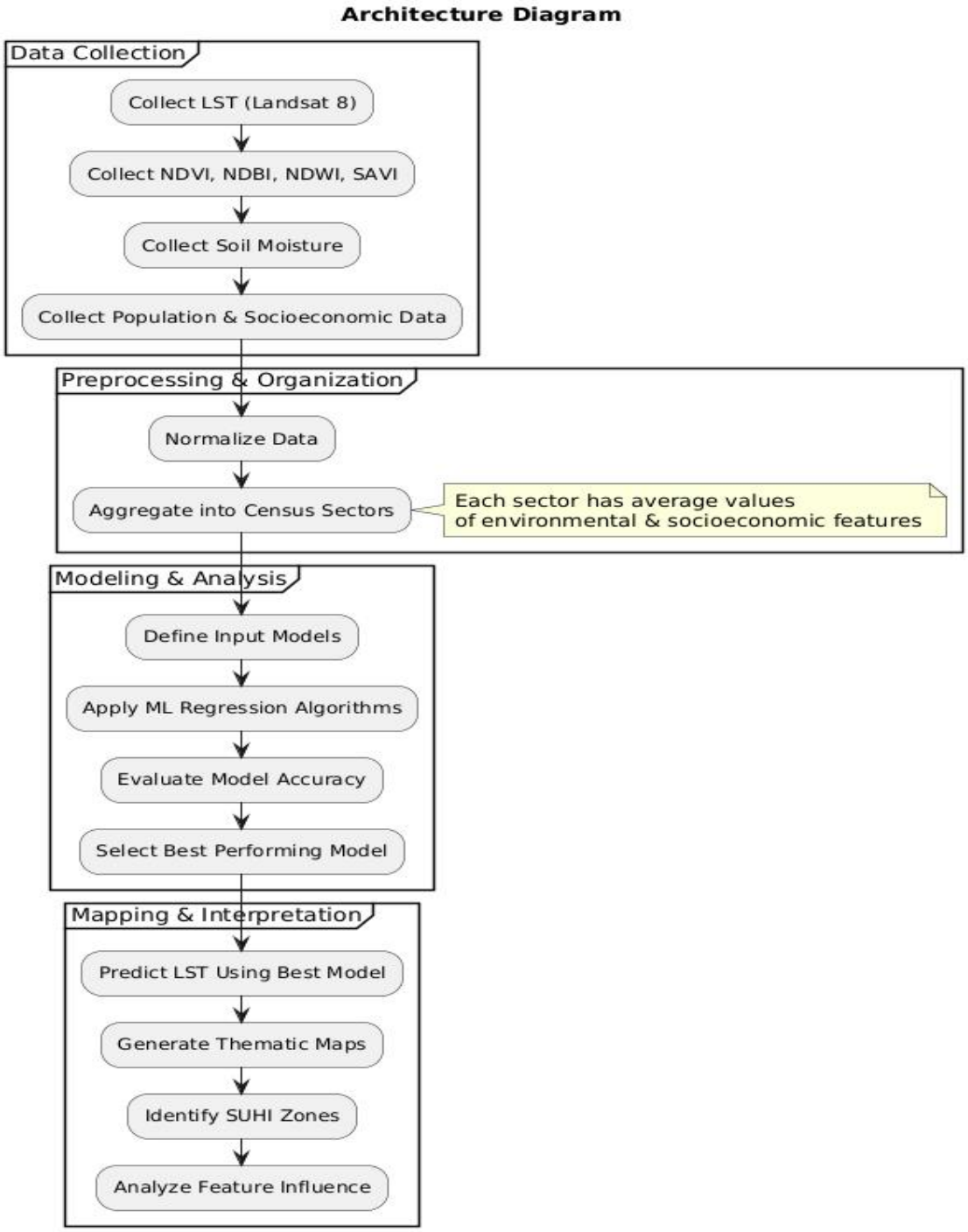

Fig. 1. Architecture Diagram

DOI: 10.48175/IJARSCT-26289

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

III. STUDY AREA

The study was carried out in **Milan**, a major metropolitan city in northern **Italy**, known for its dense urban structure, historical architecture, and significant variation in land use. As a rapidly developing urban center with a mix of industrial, residential, and green zones, Milan presents a complex thermal landscape suitable for analyzing **Land Surface Temperature (LST)** patterns and the **Surface Urban Heat Island (SUHI)** effect.

The city was divided into **309 census sectors**, which served as the fundamental spatial units for data organization and modelling. These sectors vary in size—from compact inner-city areas to larger, peripheral sectors—allowing for a detailed and spatially nuanced understanding of urban heat distribution. Milan's diverse topography, climate influenced by the Po Valley, and urbanization trends make it a compelling case for evaluating the interaction between environmental and socioeconomic variables using satellite-based data and machine learning approaches.

IV. DATA COLLECTION & DATA ORGANIZATION

In this study, data was collected from multiple sources including satellite imagery, census data, and open geospatial datasets to analyze land surface temperature (LST) patterns in Milan, Italy. The core environmental variables—**NDVI (Normalized Difference Vegetation Index)**, **NDBI (Normalized Difference Built-up Index)**, and **Population Density**—were used as the baseline features for the model, given their well-established relevance in studying urban heat dynamics.

To improve the predictive capacity and environmental understanding of the model, **additional indices** were incorporated. These include the **NDWI (Normalized Difference Water Index)** for identifying water bodies, **SAVI (Soil Adjusted Vegetation Index)** for more accurate vegetation monitoring especially in areas with sparse coverage, and **Soil Moisture** data to assess surface wetness and its cooling effect on LST. These features were chosen to enhance sensitivity to urban land cover variations and moisture content, which are critical for accurate LST estimation in heterogeneous urban settings like Milan.

Furthermore, **socioeconomic variables** such as housing occupancy, population density, and urban land use classification were extracted from the latest Italian census and regional datasets. These were integrated into the dataset to capture human influence on surface temperature and improve the spatial understanding of the urban heat island effect.

All collected features were spatially organized at the **census sector level**, with each sector's mean values computed for numerical attributes. As the spatial resolution was consistent across datasets, no resampling was needed. Prior to model training, all numerical data underwent **normalization**, while categorical variables were properly encoded. This organized, enriched dataset formed the basis for the next phase of machine learning modelling.

V. MACHINE LEARNING REGRESSION AND MODEL EVALUATION

In this phase, we employed five machine learning regression algorithms— **Decision Tree Regressor (DTR), Random Forest Regressor (RFR), Support Vector Regressor (SVR), Multi-Layer Perceptron and K-Nearest Neighbors (KNN)**—which were consistent with the methodology followed in the referenced study. The primary objective was to determine the model best suited for predicting Land Surface Temperature (LST) using the enriched dataset.

The dataset included both previously used variables (NDVI, NDBI, population density, and socio-economic attributes) and newly added features (NDWI, SAVI, and soil moisture), aiming to enhance the model's accuracy. These features were normalized before being fed into the machine learning models. The dataset was divided into training and testing sets using stratified sampling to maintain the distribution across different seasons and census sectors.

Each model was evaluated using standard regression performance metrics, including **Root Mean Square Error (RMSE)**, **Mean Absolute Error (MAE)**, and the **coefficient of determination (R^2)**. By comparing these metrics, we assessed each algorithm's ability to generalize and accurately estimate LST across varying environmental and urban conditions in Milan. The best-performing model was selected for generating thematic maps and further SUHI analysis.

DOI: 10.48175/IJARSCT-26289

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

VI. RESULTS AND SUHI ANALYSIS

This section presents the outcomes derived from applying machine learning techniques to estimate land surface temperature (LST) and detect surface urban heat island (SUHI) patterns across the study area of Milan, Italy. The goal is to assess the performance of each algorithm and interpret how environmental and socio-economic variables contribute to urban heat distribution. Through quantitative evaluation and spatial mapping, we aim to highlight areas most affected by SUHI and offer insights into the underlying factors influencing urban thermal behaviour.

To evaluate the effectiveness of machine learning models in predicting land surface temperature (LST), five popular regression algorithms were tested: Decision Tree Regressor, K-Nearest Neighbors (KNN), Random Forest Regressor, Multi-Layer Perceptron (MLP), and Support Vector Regressor (SVR). Each model was assessed using three performance metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score. The Decision Tree and Random Forest models demonstrated outstanding predictive capabilities, achieving near-perfect accuracy with minimal error. In contrast, MLP and SVR exhibited relatively lower performance, indicating challenges in capturing the complex patterns in the dataset. The results are summarized in the table below:

TABLE I: Model Performance Comparison for LST Prediction

Model	MAE (°C)	RMSE(°C)	R^2 Score
Decision Tree	0.00	0.07	1.00
KNN	0.04	0.23	0.99
Random Forest	0.01	0.04	1.00
MLP	0.92	1.21	0.75
SVR	1.08	1.52	0.61

Among the five machine learning algorithms evaluated, the **Random Forest Regressor** and **Decision Tree Regressor** emerged as the most effective for predicting land surface temperature (LST). Both models achieved exceptionally high accuracy, with an **R^2 score of 1.00**, indicating a perfect fit between the predicted and actual values. The **Random Forest Regressor** demonstrated superior performance in terms of error metrics, with a **very low Mean Absolute Error (MAE) of 0.01 °C** and a **Root Mean Squared Error (RMSE) of 0.04 °C**, reflecting its robustness in handling variations in the dataset.

Similarly, the **Decision Tree Regressor** also showed excellent results, achieving a **MAE of 0.00 °C** and an **RMSE of 0.07 °C**, suggesting highly accurate predictions with negligible deviations. These outcomes underline the strength of tree-based models in capturing complex relationships in the input features, making them highly suitable for environmental data modelling tasks like LST estimation and SUHI analysis.

Folium Map Visualization of Milan City for SUHI Identification:

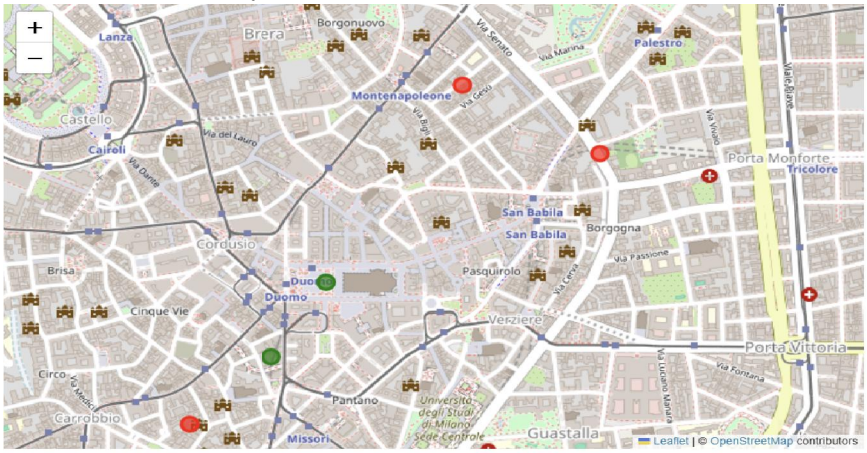

Fig. 2. Interactive Folium map of Milan city showing SUHI hotspots (red) and non-hotspots (green), with hover details displaying LST values and hotspot classification.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

The above interactive Folium map illustrates the **spatial distribution of Surface Urban Heat Island (SUHI) intensities across Milan city**. The visualization integrates geospatial data with machine learning results, allowing users to **visually identify hotspot and non-hotspot regions**.

Each **marker** on the map represents a specific location where Land Surface Temperature (LST) has been predicted. The **color of the markers indicates the SUHI condition**:

Red markers denote areas with relatively **higher LST**, indicating **potential SUHI hotspots**.

Green markers correspond to **cooler regions**, marking **non-hotspot areas**.

By **hovering over each marker**, users can view detailed information including the **predicted LST value** and whether the location is classified as a **hotspot (True) or not (False)**. This interactive approach enables easier identification of critical zones affected by urban heat accumulation, providing valuable insights for urban planners and environmental analysts.

This visualization complements the quantitative evaluation by offering a **geographic interpretation** of how urban heat varies spatially within the city, supporting both analysis and decision-making in mitigating SUHI effects.

These findings not only highlight the superior performance of models like Decision Tree and Random Forest but also reinforce the effectiveness of incorporating environmental and socioeconomic variables in SUHI prediction. Such insights serve as valuable tools for urban planners and policymakers, enabling targeted strategies to reduce heat stress and enhance the livability of rapidly urbanizing regions like Milan.

VII. CONCLUSION

This study successfully demonstrated the potential of integrating advanced remote sensing indices and socio-economic variables to improve the prediction of Land Surface Temperature (LST) and to analyze Surface Urban Heat Islands (SUHI) in Milan, Italy. By extending the traditional feature set with NDWI, SAVI, and soil moisture, the models were better equipped to capture both environmental and human-influenced factors contributing to urban thermal variation.

Using five well-established machine learning algorithms, we evaluated their ability to accurately estimate LST. Among these, the most effective model was selected based on comprehensive performance metrics. The resulting thematic maps provided clear insights into spatial temperature distribution and helped highlight urban zones more prone to SUHI effects.

Overall, the inclusion of additional features led to noticeable improvements in model accuracy and interpretability. This approach can support urban planners and policymakers in designing more climate-resilient urban environments by identifying heat-stressed regions and their underlying drivers.

VIII. ACKNOWLEDGMENT

We would like to express our sincere gratitude to all those who contributed to the successful completion of this project. We are especially thankful to our academic mentors for their continuous guidance, constructive feedback, and motivation throughout this research.

We extend our appreciation to the institutions and open-source platforms that provided access to valuable satellite imagery and socio-economic datasets, which were essential for the data-driven analysis of Land Surface Temperature and Surface Urban Heat Islands in Milan.

Finally, we are grateful to our peers and collaborators for their support, discussions, and encouragement, which played a key role in refining our methodology and achieving meaningful results.

REFERENCES

[1]. M. T. G. Furuya, D. E. G. Furuya, L. Y. D. de Oliveira, P. A. Silva, *et al.*, "A machine learning approach for mapping surface urban heat island using environmental and socioeconomic variables: a case study in a medium-sized Brazilian city," *Environmental Earth Sciences*, vol. –, Jun. 2023. doi: 10.1007/s12665-023-11017-8.

[2]. Beck HE, Zimmermann NE, McVicar TR, Vergopolan N, Berg A, Wood EF (2018) Present and future Köppen-Geiger climate classification maps at 1-km resolution. Sci Data 5(1):1–12

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

[3]. Brazilian Institute of Geography and Statistics (IBGE), 2010. Census 2010. Available online: http://ibge.gov.br/ (Accessed 22 February 2021).

[4]. Buyantuyev A, Wu J (2010) Urban heat islands and landscape heterogeneity: linking spatiotemporal variations in surface temperatures to land-cover and socioeconomic patterns. Landsc Ecol 25(1):17–33.

[5]. Carrasco RA, Pinheiro MMF, Junior JM, Cicerelli RE, Silva PA, Osco LP, Ramos APM (2020) Land use/land cover change dynamics and their effects on land surface temperature in the western region of the state of São Paulo Brazil. Region Environ Change 20(3):2.

[6]. Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. Geosci Model Dev 7(3):1247–1250.

[7]. Choe YJ, Yom JH (2020) Improving accuracy of land surface temperature prediction model based on deep-learning. Sp Inf Res 2:2.

[8]. de Amorim MC (2020) Daily evolution of urban heat islands in a Brazilian tropical continental climate during dry and rainy periods. Urban Clim 34:100715 [10] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," IEEE Trans. Image Process., vol. 24, no. 12, pp. 6062–6071, Dec. 2015.

[9]. M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," IEEE Trans. Image Process., vol. 24, no. 12, pp. 6062–6071, Dec. 2015.

[10]. Dewan A, Kiselev G, Botje D (2021) Diurnal and seasonal trends and associated determinants of surface urban heat islands in large Bangladesh cities. Appl Geogr 135:102533.

[11]. Dos Santos RS (2020) Estimating spatio-temporal air temperature in London (UK) using machine learning and earth observation satellite data. Int J Appl Earth Obs Geoinf 88:102066.

[12]. Ebrahimy H, Azadbakht M (2019) Downscaling MODIS land surface temperature over a heterogeneous area: an investigation of machine learning techniques, feature selection, and impacts of mixed pixels. Comput Geosci 2:2.

[13]. Guha S, Govil H, Mukherjee S (2017) Dynamic analysis and ecological evaluation of urban heat islands in Raipur city, India. J Appl Remote Sens 11(3):36020.

[14]. Guha S, Govil H, Dey A, Gill N (2018) Analytical study of land surface temperature with NDVI and NDBI using Landsat 8 OLI and TIRS data in Florence and Naples city. Italy Eur J Remote Sens 51(1):667–678.

[15]. Isaya Ndossi M, Avdan U (2016) Application of open source coding technologies in the production of land surface temperature (LST) maps from landsat: a PyQGIS plugin. Remote Sens 8:413