10 (2) 2024 175-189 ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
Register: *Jurnal Ilmiah Teknologi Sistem Informasi*

Contents lists available at www.journal.unipdu.ac.id

Register

Journal Page is available to www.journal.unipdu.ac.id/index.php/register

Research article

Improving Urban Heat Island Predictions Using Support Vector Regression and Multi-Sensor Remote Sensing: A Case Study in Malang

Yunifa Miftachul Arif [a, b,*], *Salma Ainur Rohma* [c], *Hani Nurhayati* [d], *Tarranita Kusumadewi* [e], *Fresy Nugroho* [f, g], *Ahmad Fahmi Karami* [h]

[a, c, d, f, h] *Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Jl. Gajayana No.50 Lowokwaru, Kota Malang, 65144 Indonesia,*

[b] *Department of Electrical Engineering, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Jl. Gajayana No.50 Lowokwaru, Kota Malang, 65144 Indonesia,*

[e] *Department of Architecthure Engineering, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Jl. Gajayana No.50 Lowokwaru, Kota Malang, 65144 Indonesia,*

[g] *Department of Mechanical Engineering, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Jl. Gajayana No.50 Lowokwaru, Kota Malang, 65144 Indonesia,*

email: [a, b,*] yunif4@ti.uin-malang.ac.id, [c] salma.ainurrohma@gmail.com, [d] hani@ti.uin-malang.ac.id, [e] tarra_nita@arch.uin-malang.ac.id, [f] fresy@ti.uin-malang.ac.id, [g] hani@ti.uin-malang.ac.id, [h] afkarami@uin-malang.ac.id

** Correspondence*

ARTICLE INFO

Article history:
Received October 1st, 2024
Revised October 7th, 2024
Accepted November 15th, 2024
Available online December 31st, 2024

Keywords:
Urban Heat Island
Land Surface Temperature
Prediction
Machine Learning

Please cite this article in IEEE style as:
Y. M. Arif, S. A. Rohma, H. Nurhayati, T. Kusumadewi, F. Nugroho, and A. F. Karami, "Improving Urban Heat Island Predictions Using Support Vector Regression and Multi-Sensor Remote Sensing: A Case Study in Malang," *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 10, no. 2, pp. 175-190, 2024.

ABSTRACT

The Urban Heat Island (UHI) phenomenon is characterized by higher temperatures in urban areas compared to surrounding rural areas. This condition poses various environmental risks and adversely impacts public health, particularly in Malang, Indonesia. This study aims to predict land surface temperature (LST) in Malang to better understand and mitigate the effects of UHI's. Support Vector Regression (SVR) is employed using remote sensing data from Landsat-8, Sentinel-2, and SRTM. Normalized Difference Vegetation Index (NDVI), Normalized Difference Built-up Index (NDBI), Normalized Difference Water Index (NDWI), Normalized Difference Moisture Index (NDMI), elevation, and LST are calculated and normalized to ensure accurate data representation. Model testing results indicate that the Radial Basis Function (RBF) kernel performs best with hyperparameter settings of C = 10, Epsilon = 0.1, and gamma = 1. This model achieves an R^2 of 0.887, an MSE of 1.625, and a MAPE of 2.71%. These findings confirm that SVR with an appropriately tuned RBF kernel can improve prediction accuracy. Consequently, the study provides a robust foundation for developing more effective predictive models to address UHI management in urban areas.

1. Introduction

The Urban Heat Island (UHI) phenomenon refers to the increase in temperature in urban areas, making them significantly hotter than surrounding rural areas [1]. This creates an index of discomfort characterized by rising global temperatures, climate change, and health risks for city dwellers [2]. UHI has become a global concern due to its extensive environmental impacts. A major factor contributing to UHI is the transformation of natural surfaces into impermeable ones, resulting in substantial changes to land use and land cover in both urban and rural areas [3]. In Indonesia, the UHI effect is especially pronounced in densely populated cities. The transition from vegetated areas to built environments has led to significant temperature differences, negatively impacting the local climate and human health.

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
10 (2) 2024 175-189

Jakarta, the nation's capital, exhibits some of the most severe UHI intensities in the country. Research highlights a direct correlation between urban density and temperature rise [4]. Rapid urbanization and infrastructure development, compounded by insufficient green spaces, have significantly increased temperatures [5]. Malang City, although previously known for its cool climate, has experienced significant UHI impacts in recent years. In the past three years, infrastructure development and urban land expansion in Malang have increased rapidly, causing drastic changes in land use [6]. The construction of high-rise buildings, shopping centers, and residential areas has reduced green spaces, which previously served as natural heat absorbers. The loss of these green areas and their conversion into impermeable surfaces exacerbate the problem, as the remaining vegetation can no longer sufficiently offset the warming effect. Figure 1 illustrates temperature data over the past three years, showing a significant rise in various urban areas of Malang.

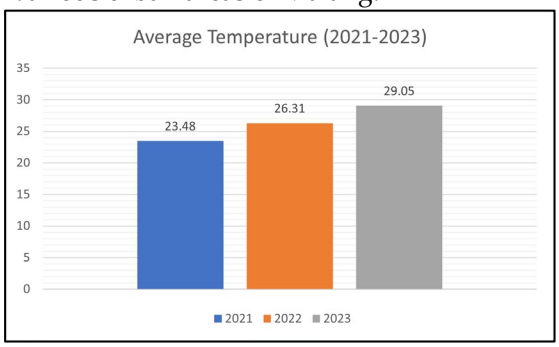


Fig. 1. Average Temperature (2021-2023)

Malang is renowned as the 'City of Education' because it serves as one of the primary educational hubs in East Java Province. Through the Tri Bina Citta slogan, education is one of the city's development plans. Along with the development of educational activity centres, of course, the flow of urbanization in Malang City also increases every year; the flow of urbanization caused by the arrival of tens of thousands of students to this city causes significant temperature changes. According to the Central Bureau of Statistics (BPS), in 2021, the population growth rate of Malang City reached 0.69% per year with a population density of 9,582 people/km^2. In the last 5 years (2015-2020), Malang City has been identified as experiencing an increase in UHI intensity with a rise in land surface temperature that reached 1.7^{0}C. Research conducted by Kusumadewi et al. reveals that this temperature increase has significantly impacted thermal comfort, causing heightened heat stress and a decline in overall livability, especially in densely populated urban areas. It is important to note that these findings are based on scientific studies and not the result of our own research [7].

Various studies have significantly contributed to understanding and predicting the Urban Heat Island (UHI) phenomenon, as summarized in Table 1. One study by Xie and Sun developed a statistical model to predict UHI using air temperature and urban surface temperature [8]. Their approach achieved satisfactory results, particularly in areas with low levels of urbanization, with a simple linear regression model and an iterative procedure minimizing the Mean Absolute Gross Error (MAGE) to 1.18 °C while explaining 69% of the variation. Another study by Furuya et al. revealed the Subsurface Urban Heat Island (SUHI) phenomenon in Milan, Northern Italy [9], , where increased groundwater temperatures in the city center showed a SUHI intensity of 3 °C. This study observed a strong correlation between building and infrastructure density and annual groundwater temperature, with vertical heat flow to the aquifer closely tied to groundwater depth and surface structure density. Groundwater warming trends ranged from +0.1 to +0.4 °C/year, generating 25 MJ/m^2 of thermal energy per year in the SUHI area. In another study, Kumari et al. explored the impact of UHI on electricity consumption in Delhi, India [10]. Analysis using satellite ground surface temperature (LST) data showed a strong relationship between UHI intensity (UHII) per district and electricity consumption. The results showed an increase in electricity consumption of 2600 GWh (11.4%) per year in eight districts of Delhi due to UHI. Similarly, Fitriya et al. used satellite data to analyze UHI and biophysical characteristics, showing higher temperatures in the city core and surrounding areas, particularly in the southern region, while the western region exhibited relatively lower temperatures compared to the east. Their findings highlighted the importance of road-based heat intensity mitigation in urban planning to reduce UHI's negative impacts [11].

 http://doi.org/10.26594/register.v10i2.5022

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
10 (2) 2024 175-189

The utilization of Machine Learning in predicting Urban Heat Island (UHI) effects has proven to be both effective and accurate. One study employed the Cellular Automata (CA) method to analyze UHI growth in Tehran based on urban parameters, including proximity to the nearest road, digital elevation model, slope, and aspect ratio. The study compared two methods, Support Vector Regression (SVR) and Artificial Neural Network (ANN), both of which successfully predicted land surface temperature (LST) with remarkable accuracy. Another study by Addas investigated the conversion of agricultural land to residential areas in Jeddah, Saudi Arabia, and its impact on UHI [12]. Machine learning algorithms, such as bagging and random subspace, were used to map the UHI from 2000 to 2021. The study found that developed urban areas increased UHI, with more than 80% of Jeddah experiencing very high UHI conditions.

Furthermore, research by Varentsov et al. employed machine learning models to analyze the urban-rural temperature difference (ΔT) in the Moscow megacity [13]. Various ML models effectively captured the variation of ΔT from 2001 to 2021, with the best performance achieved using boosting and support vector models. However, the study confirmed that the megacity's growth significantly affects the T trend in the longer term. Similarly, recent research by Mathew et al. focused on developing a predictive model for land surface temperature (LST) around Chandigarh, India, a region undergoing rapid urbanization [14]. The study utilized the Support Vector Regression (SVR) method, incorporating environmental factors such as the Enhanced Vegetation Index (EVI) and road density (RD). This research underscores the comparative effectiveness of SVR and Artificial Neural Network (ANN) models in enhancing LST prediction accuracy.

Previous research has highlighted the significant role of surface temperature in understanding UHI, particularly in defining the surface heat island. Environmental and socio-economic variables have been utilized to explain the surface urban heat island (SUHI) phenomenon and land surface temperature (LST). In one study, LST was derived from 15 Landsat 8 images from 2019 to 2021[15]. Socio-economic data were obtained from official population censuses, while environmental data were obtained from Sentinel-2 and Planet imagery. Using these variables, six algorithms were evaluated for their effectiveness in predicting environmental surface heat gain (ESG). Among them, the DecisionTree algorithm demonstrated superior performance, achieving an R-value of 0.96, an MAE of 1.49°C, and an RMSE of 1.88°C, followed closely by the RDF algorithm. Including data from all seasons and socio-economic variables significantly improved the results. The main contribution of this study is to assess whether these algorithms can improve the characterization process of the SUHI by analyzing the influence of the variables studied. The literature underscores the importance of advanced technologies and integrated data in understanding and managing the UHI phenomenon. Continued adoption of innovative techniques and more precise data can significantly enhance the effectiveness of urban strategies to mitigate UHI effects.

Table 1. Various studies on predictions of the Urban Heat Island (UHI) phenomena

Reference	Year	Object	Region	Method	Dataset
[5]	2020	Statistical Modeling of Urban Heat Island Intensity	Warsaw	Statistic	MODIS
[10]	2021	Impact of urban heat island	Delhi	Statistic	MODIS
[12]	2022	Impact of Urban Heat Island	Jeddah	LR	Landsat-8
[13]	2023	Machine Learning for Simulation of Urban Heat Island	Moscow	CBR SVR	ERA5
[16]	2019	Prediction of Land Surface Temperature for Urban Heat Island	Chandigarh	SVR	Landsat-8
[17]	2023	Machine learning approach for mapping surface urban heat island	Brazil	LR DT SVR RF	Landsat-8 Sentinel-2 ISA Survey
[18]	2021	Characterization of the subsurface urban heat island	Milan	Statistic	MODIS
[19]	2020	Urban heat island intensity and its mitigation strategies	Nagpur	Statistic	Landsat-8

Significant progress has been made in Urban Heat Island (UHI) prediction, mainly through machine learning techniques such as Support Vector Regression (SVR). SVR has gained widespread adoption due to its capability to deliver accurate UHI predictions based on satellite remote sensing data. However, despite these advances, several research gaps remain unresolved, limiting SVR's potential accuracy and scalability in large-scale UHI prediction. While SVR has been successfully implemented, most studies focus on basic model configurations without extensive parameter tuning. Systematic

 http://doi.org/10.26594/register.v10i2.5022

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
10 (2) 2024 175-189

research is therefore required to optimize SVR parameters, such as cost, epsilon, gamma, and degree, across different kernels (linear, RBF, and polynomial) to enhance model performance. This optimization, supported by evaluations using key performance metrics such as the coefficient of determination (R^2), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE), can provide valuable insights into improving prediction accuracy and computational efficiency. In addition to parameter tuning, there is a need for the comprehensive integration of multi-satellite data to enhance spatial and temporal resolution in UHI predictions. Although Landsat 8 and Sentinel-2 are commonly used in UHI studies, integrating additional datasets like Shuttle Radar Topography Mission (SRTM) data has not been widely explored. Combining these datasets could significantly improve spatial resolution and temporal coverage, yielding more accurate UHI predictions and offering deeper insights into urban heat patterns.

Furthermore, addressing the computational challenges associated with large-scale SVR models, particularly when handling extensive datasets, is essential for practical applications. Developing techniques to enhance the scalability and efficiency of SVR models is critical for advancing this field. Overcoming these limitations will not only improve the accuracy of UHI predictions but also provide a more reliable foundation for urban planning and heat mitigation strategies.

This paper aims to address these research gaps by presenting the following key contributions: (a) **SVR Parameter Optimization**: A systematic study is proposed to optimize key SVR parameters such as cost, epsilon, gamma, and degree across various kernel functions (linear, RBF, and polynomial) to improve model accuracy. This optimization is supported by R^2, MSE, and MAPE evaluations to ensure improved prediction performance. (b) **Evaluation of Different Kernel Functions**: A detailed analysis of the impact of various SVR kernel functions (linear, RBF, and polynomial) on prediction accuracy and computational efficiency will be conducted, providing valuable insights into the most effective kernels for UHI prediction. (c) **Multisatellite Data Integration**: The integration of multisatellite data, including Landsat 8, Sentinel-2, and SRTM, is proposed to improve the spatial and temporal resolution of UHI predictions. This approach aims to offer a more comprehensive understanding of urban heat patterns. (d) **Combination of Urban Density Index (UDI) and Elevation**: The inclusion of the Urban Density Index (UDI) and elevation data as additional predictors in UHI models is proposed, providing a more detailed representation of urban structure and topography to further enhance temperature prediction accuracy.

2. Materials and Methods

This research followed a systematic and coherent series of steps, facilitating the identification of subsequent actions needed to achieve the desired outcomes. Each stage was carefully executed to ensure the research's success and minimize potential errors. With this structured approach, the research contributes significantly to understanding and addressing the Urban Heat Island (UHI) phenomenon.

Figure 2 illustrates the research steps, from data collection to evaluation and analysis, aimed at producing accurate predictions. Stage one involved data collection using Google Earth Engine, with satellite data sources such as Landsat 8 OLI/TIRS C2 L1, Sentinel 2 Level 2A, and NASA SRTM Digital Elevation 30m [20]. The collected data was then processed through stage two (pre-processing stage), which included cloud masking and feature calculation. The environmental variables calculated include Normalized Difference Vegetation Index (NDVI), Normalized Difference Built-up Index (NDBI), Normalized Difference Water Index (NDWI), Normalized Difference Moisture Index (NDMI), Urban Development Index (UDI), elevation, and LST. This step aims to clean and prepare the data for further analysis. In stage three, machine learning techniques were applied, including Z-Score normalization, data sharing, and SVR for analysis. Z-Score normalization and data sharing help ensure that the model is unbiased and generalizable. The use of SVR facilitates practical analysis and accurate predictions. Finally, stage four involves evaluating the results using R^2, Mean Absolute Percentage Error (MAPE), and Mean Squared Error (MSE) metrics, with visualizations for further predictive analysis. Each stage is designed to ensure data integrity and the accuracy of analysis results, ensuring that the research produces valid and reliable information.

http://doi.org/10.26594/register.v10i2.5022

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
10 (2) 2024 175-189

Fig. 2. Research Workflow Diagram

2.1. Data Collection and Organization

Based on the research framework, the data for this study is sourced from satellite remote sensing datasets available through the Google Earth Engine platform. The satellite images utilized include data from Landsat, SRTM, and Sentinel. Information on the specific bands to be utilized are provided in Table 2.

Landsat-8, Sentinel-2, and SRTM are essential for monitoring environmental dynamics and analyzing UHI effects. Landsat-8, a collaborative project between NASA and USGS, is critical for surface temperature analysis using Band 10, which is highly sensitive to thermal emissions [15] [16]. Pixel values are calibrated by multiplying by 0.00341802, adding 149 to convert to Kelvin, and then subtracting 273.15 to obtain the temperature in Celsius [17] [18]. Sentinel-2, launched by the European Space Agency (ESA) under the Copernicus program, provides detailed multispectral data across 13 bands, which are essential for NDVI and UHI analysis applications. Pixel values are normalized to a range of 0-1 using a scale factor of 0.0001, offering a spatial resolution of up to 10 meters [4] [22]. SRTM, conducted by NASA and NGA, provides high-resolution digital elevation models using radar interferometry technology, which is helpful for hydrological modelling and environmental planning.

Table 2. Detailed Calibration Information for Landsat-8 OLI/TIRS, Sentinel-2 Level-2A, and NASA SRTM

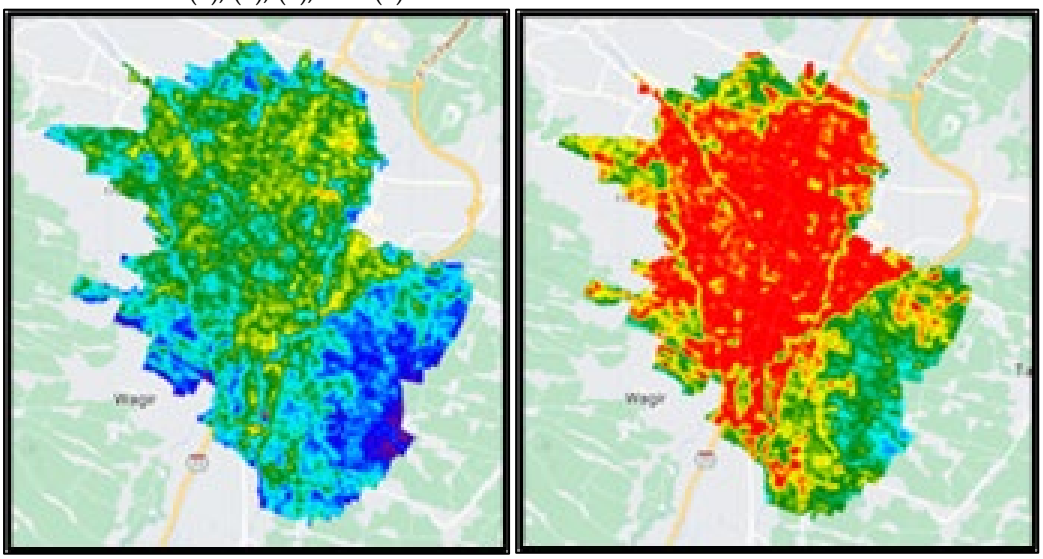
No	Image	Band	Description	Scale	Offset
1.	Landsat-8 OLI/TIRS C2 L1	ST B10	LST	0.00341802	149
2.	Sentinel-2 Level-2A	B3	Green	0.0001	-
		B4	Red	0.0001	-
		B8	NIR	0.0001	-
		B11	SWIR 1	0.0001	-
		B12	SWIR 2	0.0001	-
3.	NASA SRTM Digital Elevation 30m	elevation	-	-	-

Before calculation, the data must first undergo processing, beginning with a cloud masking process to produce cleaner imagery by removing pixels affected by clouds or cloud shadows that could interfere with the results. This step is critical because clouds and their shadows can significantly distort the data, leading to inaccuracies in the analysis. Eliminating these disturbances greatly enhances the quality of the imagery, resulting in more reliable data for modeling and subsequent analysis. The cloud

 http://doi.org/10.26594/register.v10i2.5022

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
10 (2) 2024 175-189

masking process involves identifying and eliminating cloud and shadow-covered pixels using an algorithm to detect these features based on spectral characteristics. Figure 3(a) shows the image before cloud masking, while Figure 3(b) shows the image after cloud masking, demonstrating significant improvements in clarity and accuracy. This improvement facilitates a more precise analysis of the UHI effect. After the cloud masking process, the next step is calculating the features to use. These calculations involve applying simple mathematical formulas to derive the necessary features, including NDVI, NDBI, NDWI, NDMI, elevation, UDI, and LST. The NDVI, NDBI, NDWI, and NDMI features need to be calculated based on (1), (2), (3), and (4) found in Table 3.



(a) (b)

Fig. 3. (a) Before Cloud Masking; (b) After Cloud Masking

Table 3. Remote Sensing Variables: Description and Formulas for Calculations

No	Feature	Description	Calculation	
1.	NDVI	An index used in remote sensing to measure the greenness or presence of vegetation in an area. High values indicate lush vegetation.	$\frac{(NIR - Red)}{(NIR + Red)}$	(1)
2.	NDBI	An index used in remote sensing to measure built-up areas or urbanized land use.	$\frac{(SWIR2 - NIR)}{(SWIR2 + NIR)}$	(2)
3.	NDWI	An index used in remote sensing to measure the water content or moisture in the soil surface.	$\frac{(Green - NIR)}{(Green + NIR)}$	(3)
4.	NDMI	An index is used in remote sensing to measure an area's moisture level or hydration state.	$\frac{(NIR - SWIR1)}{(NIR + SWIR1)}$	(4)
5.	UDI	An index used in remote sensing to measure the degree of urbanization.	-	
6.	Elevate	Earth surface elevation data, measured by the SRTM, provide information on the height of an area above sea level.	-	
7.	LST	The surface temperature value of an area was recorded using ST B10 of the Landsat satellite.	-	

The features obtained in the previous steps are then normalized to ensure they can be effectively analyzed. Normalization is performed to standardize data values and bring them into the same range. The Z-Score normalization method is used, as defined in formula (5), where x is the normalization value, μ is the mean of all data, and σ is the standard deviation of the data used. This calculation aims to produce data with uniform properties for further analysis, ensuring that no variable dominates another solely due to differences in scale. Normalization is particularly critical in predictive models and data clustering, where consistency in variable scales can significantly impact model performance. Additionally, normalization is beneficial when the data has long tails or a highly skewed distribution, as it mitigates the effect of outliers by expressing distances in terms of standard deviation units.

$$Normalization = \frac{(x - \mu)}{\sigma} \quad (5)$$

 http://doi.org/10.26594/register.v10i2.5022

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
10 (2) 2024 175-189

2.2. SVR Method

While SVM is typically used for classification problems, SVR is employed to predict continuous outcomes. This method is particularly effective for handling high-dimensional data and complex non-linear relationships.

In this study, six input variables ($\boldsymbol{x_a}, \boldsymbol{x_b}, \boldsymbol{x_c}, \boldsymbol{x_d}, \boldsymbol{x_e}, \boldsymbol{x_f}$) and one output variable (Y) are used, as in equation (6). The dataset DF incorporates input variables where where $\boldsymbol{x_a}$ represents NDVI variables, $\boldsymbol{x_b}$ represents NDBI, $\boldsymbol{x_c}$ represents NDWI, $\boldsymbol{x_d}$ represents NDMI, $\boldsymbol{x_e}$ represents UDI, dan $\boldsymbol{x_f}$ represents elevation. Meanwhile, Y is the output variable, namely LST. The DF will be implemented using the SVR method by applying a generalized regression (7), where $\boldsymbol{f(y)}$ is the prediction function with input Xr, w vector, and b is the bias. This function will adjust the data points to the primary function. The main function is determined based on equation (7), with the most significant influence being the value of w. The smaller the value of w, the closer the data points will be to the primary function.

$$DF = \{(x_{a1}, x_{b1}, x_{c1}, x_{d1}, x_{e1}, x_{f1})\}) \quad (6)$$

$$f(y) = wX_r + b \quad (7)$$

The constant value *C* aims to determine the control parameter that controls the penalty for model errors. A higher C value imposes a greater penalty for errors, leading to a more precise model with a smaller margin. Besides the Cost value, there are other supporting parameters such as epsilon, degree parameter for polynomial kernel and gamma parameter for RBF kernel. Kernels play a crucial role in SVR by enabling the algorithm to handle non-linear relationships between input and output variables. The kernel allows linear splitting by transforming the input space into a higher dimensional space, equivalent to a non-linear regression function in the real space.

The linear kernel is the most commonly used kernel for linear separation. It has been shown to perform best when a large number of features are present [23] [24]. The polynomial kernel transforms data from a low-dimensional space to a high-dimensional space by applying a polynomial function. This transformation is particularly effective for data that cannot be well-separated in a low-dimensional space but becomes more separable in a high-dimensional space. An additional degree parameter is also applied in the polynomial kernel, which is used as a polynomial degree that will project the data into a dimensional space [24] [25]. The RBF kernel is a kernel that gives weights to data vectors based on how similar they are in higher feature space, and this helps in forming a hyperplane that can separate data well, even if the data cannot be separated linearly in the original space [21][26]. Linear, polynomial and RBF kernels are presented in (8), (9) and (10) respectively [27]. $x_i . x_j$ is the training data to be used, d is the degree value that will be used in the polynomial kernel, and for the RBF kernel, there is σ as a standard deviation that will normalize two data vectors to calculate the similarity or closeness between them in high feature space. The test scenario used in this study is outlined in Table 4.

$$k(x_i, x_i) = x_i . x_j) \quad (8)$$

$$k(x_i, x_j) = (x_i . x_j)^d) \quad (9)$$

$$k(x_i, x_i) = \exp\left(\frac{||X_i - X_j||^2}{2\sigma^2}\right)) \quad (10)$$

Table 4. Testing Scenarios with Kernel Functions and Hyperparameters.

Kernel Function	Hyperplane Parameters	Parameter Value
Kernel Linear	C (*Cost*)	10^{-1}, 1, 10^{1}
	ε (*Epsilon*)	10^{-3},10^{-2}, 10^{-1}
Kernel *Polynomial*	C (*Cost*)	10^{-1}, 1, 10^{1}
	ε (*Epsilon*)	10^{-3},10^{-2}, 10^{-1}
	d (degree)	2, 4, 8
Kernel Radial Basis Function	C (*Cost*)	10^{-1}, 1, 10^{1}
	ε (*Epsilon*)	10^{-3},10^{-2}, 10^{-1}
	γ (*gamma*)	10^{-1}, 1, 10^{1}

 http://doi.org/10.26594/register.v10i2.5022

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
10 (2) 2024 175-189

3. Results and Discussion

In this chapter, we discuss the results of the SVR model testing and analyze the findings. This study evaluated several kernels with various tuning parameters to determine the most accurate prediction model. The model's performance in predicting UHI in Malang City is assessed using R^2, MSE, and MAPE metrics, and the evaluation results will be presented. After testing the model by applying the test scenarios with different kernels and tuning parameters, the results are displayed in Figures 4–6, showing the complete test outcomes and performance evaluation.

Testing with a linear kernel, the performance evaluation of the SVR model based on MAPE, MSE, and R^2 metrics demonstrates consistent results across various hyperparameter configurations. The MAPE values range from 0.747% to 0.751%, with the lowest value achieved at C = 0.1, Epsilon = 0.01, indicating minimal percentage error. Similarly, MSE remains steady at approximately 0.145 for all configurations, reflecting consistent squared error differences between predicted and actual values. The R^2 values are uniformly 0.857, signifying that the model explains the same proportion of variance across all tested configurations. These findings suggest that variations in C and Epsilon within the tested range do not significantly impact model performance.

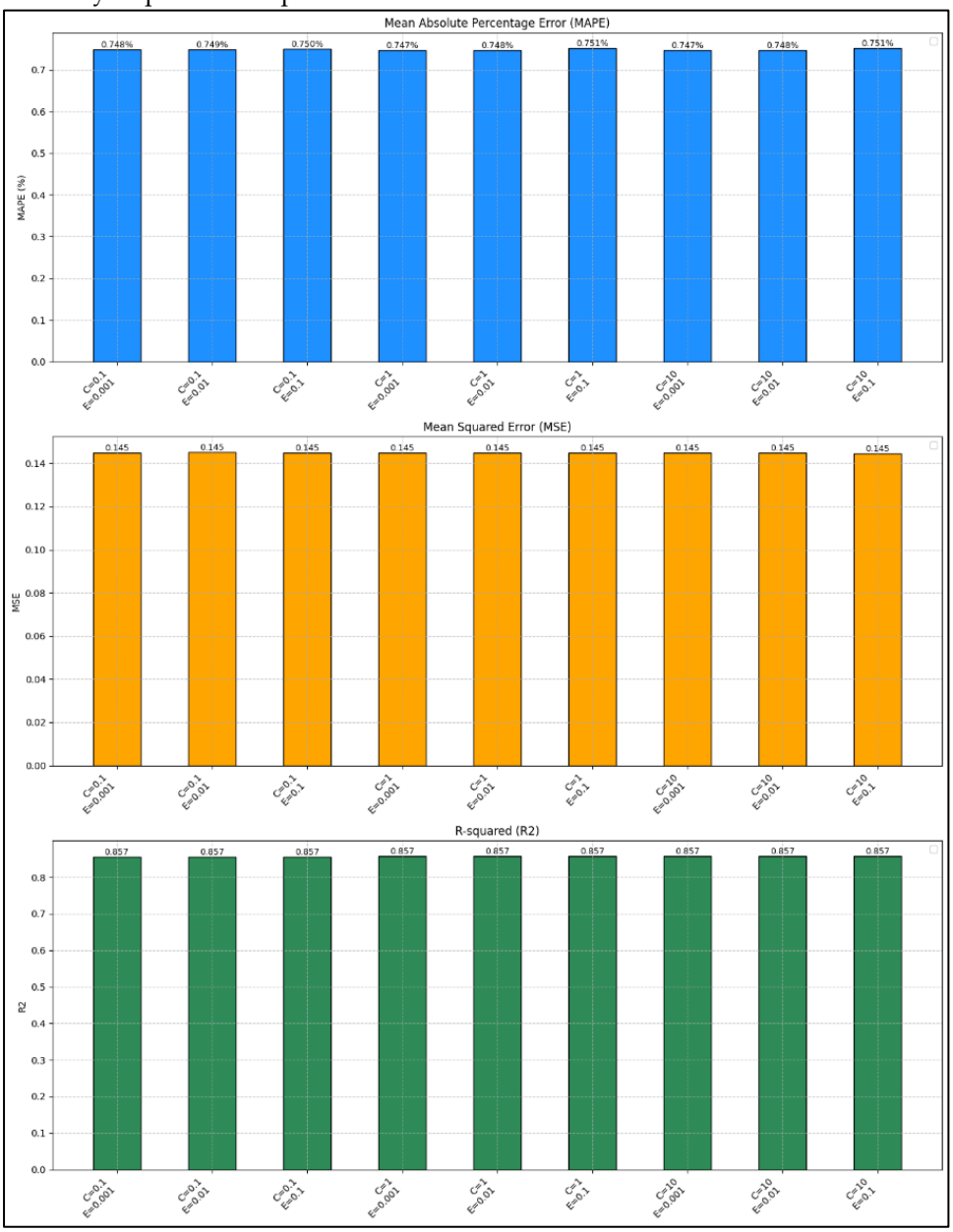


Fig. 4. Performance Evaluation Using R^2, MAPE, and MSE for Linear Kernels

http://doi.org/10.26594/register.v10i2.5022

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
10 (2) 2024 175-189

In contrast, the performance evaluation of the SVR model with the polynomial kernel, based on MAPE, MSE, and R^2 metrics, shows notable variation across different hyperparameter configurations. The MAPE values range from 0.933% to 1.250%, with the lowest value observed at C = 50, Epsilon = 0.1, and degree = 8, reflecting the best percentage error accuracy for this kernel. MSE values vary significantly, from 0.489 to 0.954, with the lowest value achieved under the same configuration that yielded the lowest MAPE, indicating consistent model performance in reducing squared error. The R^2 values range from 0.057 to 0.517, with the highest value of 0.517 corresponding to C = 100, Epsilon = 0.01, and degree = 4, showing that this configuration explains the most significant proportion of variance in the dataset. These results highlight that the polynomial kernel's performance is highly dependent on the choice of degree and regularization parameters.

Fig. 5. Performance Evaluation Using R^2, MAPE, and MSE for Polynomial Kernel

The performance evaluation of the SVR model using the RBF kernel, based on the metrics MAPE, MSE, and R^2, demonstrates consistent and robust results across various hyperparameter settings. The MAPE values range from 0.506% to 0.884%, with the lowest error observed at C = 100 and epsilon = 0.01,

 http://doi.org/10.26594/register.v10i2.5022

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
10 (2) 2024 175-189

indicating the most accurate predictions with minimal percentage error. Similarly, the MSE remains relatively stable, ranging from 0.122 to 0.166, indicating that the squared differences between the predicted and actual values are minor, with the best performance at C = 100 and epsilon = 0.01. The R^2 values range between 0.867 and 0.884, indicating a high level of explained variance, with the highest at C = 100, epsilon = 0.01. These results suggest that the RBF kernel in SVR is highly effective at capturing the complex relationships within the dataset.

Fig. 6. Performance Evaluation Using R^2, MAPE, and MSE for RBF Kernel

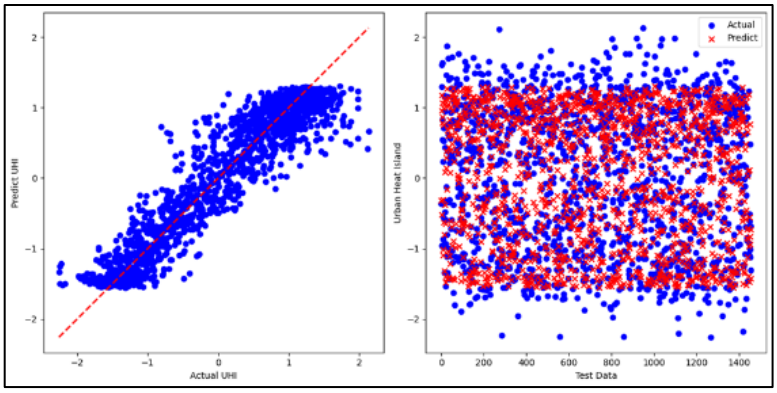
Overall, the evaluation results indicate that the RBF kernel, with a hyperparameter configuration of C = 10, Epsilon = 0.1, and gamma = 1, delivers the best predictive performance, achieving minimal error and strong predictive capability according to the criteria set in this study. The RBF kernel consistently gave the best evaluation metrics (R^2, MSE, and MAPE) results. These overall results confirm that the RBF kernel, optimized with specific hyperparameters, is highly effective in prediction tasks,

 http://doi.org/10.26594/register.v10i2.5022

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
10 (2) 2024 175-189

providing both accuracy and reliability. Coupled with the prediction result graph in Figure 7, it shows that using the best parameters on the RBF kernel produces a good prediction graph. Figure 7(a) compares the actual data and the model prediction results. The blue dots represent the actual data, while the red represents the predicted results. A good match between the two can be observed from the closeness of the blue dots to the red line. Figure 7(b) shows the representation of the predicted data (blue circles) and the test data (red crosses). In the model fit test, the closeness between the blue circles and the red crosses reflects the model's predictive ability to data that has never been seen before. Overall, the RBF kernel model with the specified hyperparameters exhibits an excellent fit to the actual data and demonstrates robust predictive performance.

Based on Table 5, several studies have utilized the SVR method to predict UHI, with variables specifically tailored to the region of interest (ROI). Additionally, using diverse and relevant variables, along with selecting an appropriate model, is crucial for accurate UHI prediction. The choice of variables, including NDVI, NDBI, NDMI, NDWI, UDI, LST, and elevation, provided the model with comprehensive information about environmental conditions. Furthermore, the resulting R^2 value of over 0.85 indicates a strong relationship between the variables in the model. Among the five studies referenced, all employed the SVR method for UHI prediction, which leads to the conclusion that the SVR method is highly suitable for predicting the UHI phenomenon in large urban cities.



(a) (b)

Fig. 7. (a) Scatter Plot of Actual vs. Predicted Data, and (b) Performance Evaluation Based on Comparison of Actual and Predicted Data

Table 5. Results from the Current Study

Reference	Variables	R^2	MSE	MAPE
[16]	RD, elevation, EVI, LST	-	0.394	0.181
[17]	NDVI, NDBI, LST, Socioeconomy	0.96	2.04	1.72
[28]	Air Temperature, Humidity, Wind Speed, and cloud cover	0.8	0.7	-
Kernel RBF (Best Perform)	NDVI, NDBI, NDMI, NDWI, UDI, LST, elevation	0.887	1.625	2.71%

The comparison of studies in Table 5 demonstrates the impact of variable selection and model tuning on UHI prediction using SVR. Zheng's research, which focused on meteorological variables (air temperature, humidity, wind speed, and cloud cover), produced a moderate R^2 of 0.8, indicating a reasonable correlation but lacking a complete representation of UHI dynamics due to the exclusion of land surface characteristics [28]. This led to an MSE of 0.7, highlighting the limitations of relying solely on meteorological data for accurate UHI predictions. In contrast, the research by research Mathew, which incorporated RD, elevation, EVI, and LST, achieved a lower MSE of 0.394. However, the absence of land use indicators like NDBI likely limited the model's ability to fully capture urban heat phenomena, as evidenced by its relatively high MAPE of 0.181 [16]. Furuya's research, which utilized NDVI, NDBI, LST, and socio-economic data, achieved the highest R^2 of 0.96, suggesting that integrating land cover and socio-economic variables provides a more holistic understanding of UHI. However, the

 http://doi.org/10.26594/register.v10i2.5022

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
10 (2) 2024 175-189

study's higher MSE of 2.04 points to potential overfitting, where the model may not generalize well to new data [17].

In comparison, our study, which combined NDVI, NDBI, NDMI, NDWI, UDI, LST, and elevation, produced a robust R^2 of 0.887, an MSE of 1.625, and a MAPE of 2.71%. The integration of diverse variables, particularly the inclusion of UDI and elevation—variables that were rarely used in previous studies—enhanced the model's ability to capture both the surface and structural characteristics of the urban environment. This balanced variable selection allowed for a more comprehensive analysis of UHI, avoiding overfitting while maintaining crucial prediction accuracy. Compared to the other studies, the performance of our model underscores the effectiveness of using a diverse range of environmental and urban form indicators in UHI prediction, reinforcing the suitability of the RBF kernel for analyzing complex urban heat patterns in large cities like Malang.

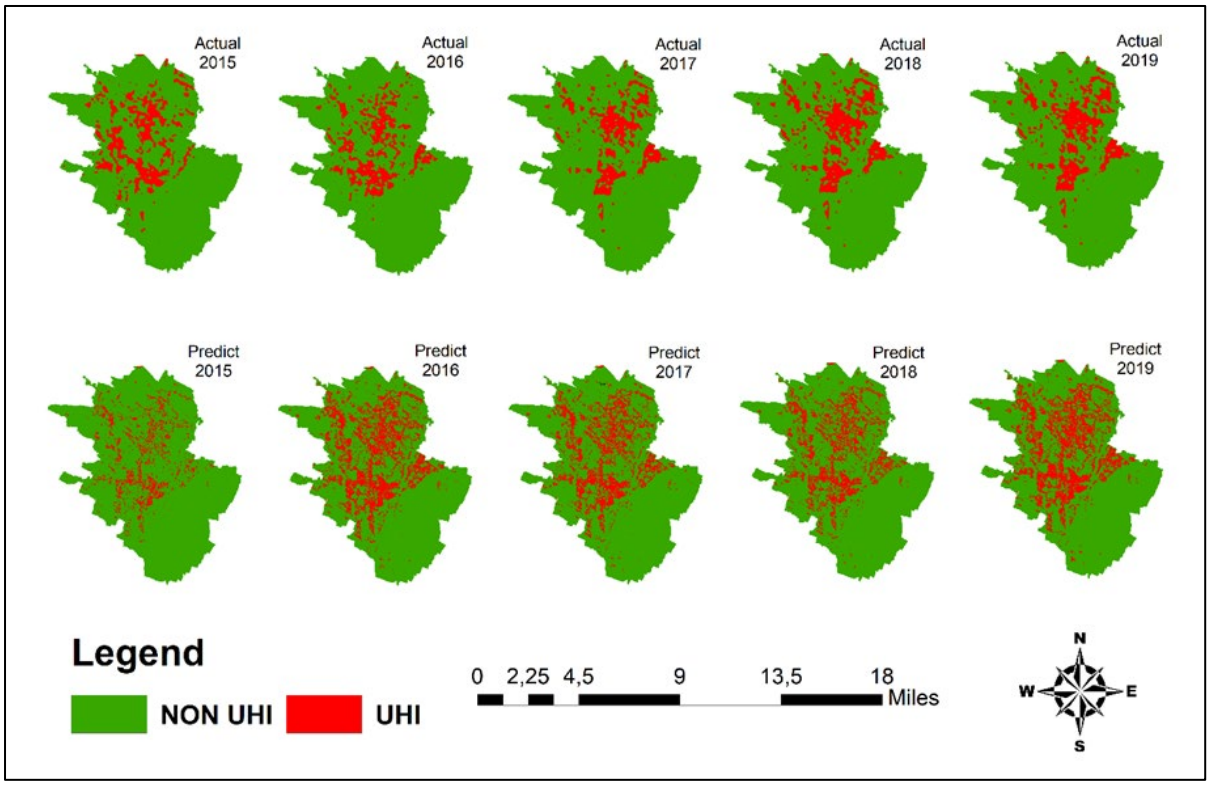


Fig. 8. Analysis of UHI Point Distribution in Malang City: Actual vs Predicted (2015-2019)

4. Conclusion

This research explores the use of SVR to predict LST in Malang as an indicator of the UHI phenomenon. Remote sensing data from Landsat-8, Sentinel-2, and SRTM were utilized, with a cloud masking process applied to enhance image quality. NDVI, NDBI, NDWI, NDMI, elevation, and LST were calculated and normalized. The test results indicate that the RBF kernel with hyperparameters C = 10, Epsilon = 0.1, and gamma = 1 provides the best performance across all evaluation metrics (R^2, MSE, and MAPE), achieving the highest R^2 of 0.887, the lowest MSE of 1.625, and the lowest MAPE of 2.71%. These results align with existing literature, demonstrating the superiority of the RBF kernel in handling complex data.

The minimal prediction error in estimating Urban Heat Island (UHI) effects using SVR is reflected in the low MAPE value of 2.71%. This suggests that the model's predictions closely match actual observations, indicating a reliable estimation of UHI intensity in Malang. However, despite the overall low error rates, certain factors may have contributed to localized discrepancies in prediction accuracy. Urban landscapes are highly heterogeneous, with varying building densities, vegetation cover, and surface materials. These factors can introduce slight prediction challenges, particularly in highly mixed areas where microclimates are present.

Moreover, the reliance on remote sensing data with varying temporal and spatial resolutions may cause minor discrepancies, especially in areas where temperature variations occur rapidly over short distances or time intervals. The RBF kernel's ability to handle nonlinear relationships is pivotal in

http://doi.org/10.26594/register.v10i2.5022

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
10 (2) 2024 175-189

minimizing these errors, making it a suitable choice for UHI prediction. Nonetheless, careful consideration of input data quality and spatial granularity is essential to reduce prediction errors in future UHI models.

Further research prospects include the development of models that integrate multisensor data, such as meteorological and socio-economic data, and the application of advanced machine learning techniques, for example deep learning, to enhance prediction resolution and accuracy. The research can also be extended to cover other regions with different urbanization characteristics, , helping to produce a more generalized model that can be adapted to various urban contexts. In addition, incorporating variables such as land use, buildings, and population mobility patterns would provide greater insight into the factors affecting UHI and enable the development of more effective mitigation strategies. Integrating data from multiple sources and advanced machine learning techniques will open opportunities to create more holistic and accurate UHI prediction models, contributing to improved quality of life in urban environments.

Author Contributions

Y. M. Arif: Conceptualization, formal analysis, funding acquisition, methodology, project administration, writing – original draft, and writing - review and editing. S. A. Rohma: Data curation, formal analysis, software, validation, visualization, and writing – original draft. H. Nurhayati: Data curation, funding acquisition, project administration, validation, and writing – review and editing. T. Kusumadewi: Conceptualization, formal analysis, methodology, resources, supervision, and writing – original draft. F. Nugroho: Formal analysis, investigation, resources, supervision, and writing – review & editing. A. F. Karami: Formal analysis,, methodology, software, validation, and writing – review & editing.

Declaration of Competing Interest

We declare that we have no conflict of interest.

References

[1] A. Al Kafy *et al.*, "The operational role of remote sensing in assessing and predicting land use/land cover and seasonal land surface temperature using machine learning algorithms in Rajshahi, Bangladesh," *Appl. Geomatics*, vol. 13, no. 4, pp. 793–816, 2021, doi: 10.1007/s12518-021-00390-3.

[2] Z. Jandaghian and A. Colombo, "The Role of Water Bodies in Climate Regulation: Insights from Recent Studies on Urban Heat Island Mitigation," *Buildings*, vol. 14, no. 9, 2024, doi: 10.3390/buildings14092945.

[3] M. Unal Cilek and A. Cilek, "Analyses of land surface temperature (LST) variability among local climate zones (LCZs) comparing Landsat-8 and ENVI-met model data," *Sustain. Cities Soc.*, vol. 69, no. 6, p. 102877, 2021, doi: 10.1016/j.scs.2021.102877.

[4] Y. Murayama and R. Wang, "Editorial: Special Issue on Geographical Analysis and Modeling of Urban Heat Island Formation," *Remote Sens.*, vol. 15, no. 18, pp. 1–5, 2023, doi: 10.3390/rs15184474.

[5] L. Gawuc, M. Jefimow, K. Szymankiewicz, M. Kuchcik, A. Sattari, and J. Struzewska, "Statistical Modeling of Urban Heat Island Intensity in Warsaw, Poland Using Simultaneous Air and Surface Temperature Observations," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 2716–2728, 2020, doi: 10.1109/JSTARS.2020.2989071.

[6] M. Dede, C. Asdak, and I. Setiawan, "Spatial dynamics model of land use and land cover changes: A comparison of CA, ANN, and ANN-CA," *Regist. J. Ilm. Teknol. Sist. Inf.*, vol. 8, no. 1, pp. 38–49, 2022, doi: 10.26594/register.v8i1.2339.

[7] T. Kusumadewi *et al.*, "Urban Phytoarchitecture Design Options: Greenspace Orientation and Tree Species Intensification," *J. Adv. Res. Appl. Sci. Eng. Technol.*, vol. 31, no. 1, pp. 183–196, 2023, doi: 10.37934/araset.31.1.183196.

[8] Q. Xie and Q. Sun, "Monitoring thermal environment deterioration and its dynamic response to urban expansion in Wuhan, China," *Urban Clim.*, vol. 39, no. 8, p. 100932, 2021, doi: 10.1016/j.uclim.2021.100932.

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
10 (2) 2024 175-189

[9] Z. A. Rahaman *et al.*, "Assessing the impacts of vegetation cover loss on surface temperature, urban heat island and carbon emission in Penang city, Malaysia," *Build. Environ.*, vol. 222, 2022, doi: 10.1016/j.buildenv.2022.109335.

[10] P. Kumari, V. Garg, R. Kumar, and K. Kumar, "Impact of urban heat island formation on energy consumption in Delhi," *Urban Clim.*, vol. 36, no. 12, p. 100763, 2021, doi: 10.1016/j.uclim.2020.100763.

[11] E. A. Storey, K. R. Lee West, and D. A. Stow, "Utility and optimization of LANDSAT-derived burned area maps for southern California," *Int. J. Remote Sens.*, vol. 42, no. 2, pp. 486–505, 2021, doi: 10.1080/01431161.2020.1809741.

[12] A. Addas, "Machine Learning Techniques to Map the Impact of Urban Heat Island: Investigating the City of Jeddah," *Land*, vol. 12, no. 6, 2023, doi: 10.3390/land12061159.

[13] M. Varentsov, M. Krinitskiy, and V. Stepanenko, "Machine Learning for Simulation of Urban Heat Island Dynamics Based on Large-Scale Meteorological Conditions," *Climate*, vol. 11, no. 10, pp. 1–24, 2023, doi: 10.3390/cli11100200.

[14] A. K. Taloor, Drinder Singh Manhas, and G. Chandra Kothyari, "Retrieval of land surface temperature, normalized difference moisture index, normalized difference water index of the Ravi basin using Landsat data," *Appl. Comput. Geosci.*, vol. 9, no. August 2020, p. 100051, 2021, doi: 10.1016/j.acags.2020.100051.

[15] A. Kristianto, E. Sediyono, and K. D. Hartomo, "Implementation dbscan algorithm to clustering satellite surface temperature data in indonesia," *Regist. J. Ilm. Teknol. Sist. Inf.*, vol. 6, no. 2, pp. 109–118, 2020, doi: 10.26594/register.v6i2.1913.

[16] A. Mathew, S. Sreekumar, S. Khandelwal, and R. Kumar, "Prediction of land surface temperatures for surface urban heat island assessment over Chandigarh city using support vector regression model," *Sol. Energy*, vol. 186, no. 6, pp. 404–415, 2019, doi: 10.1016/j.solener.2019.04.001.

[17] M. T. G. Furuya *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," *Environ. Earth Sci.*, vol. 82, no. 13, pp. 1–14, 2023, doi: 10.1007/s12665-023-11017-8.

[18] A. Previati and G. B. Crosta, "Characterization of the subsurface urban heat island and its sources in the Milan city area, Italy," *Hydrogeol. J.*, vol. 29, no. 7, pp. 2487–2500, 2021, doi: 10.1007/s10040-021-02387-z.

[19] S. Jain, S. Sannigrahi, S. Sen, and S. Bhatt, "Urban heat island intensity and its mitigation strategies in the fast- growing urban area," *J. Urban Manag.*, vol. 9, no. 1, pp. 54–66, 2020, doi: 10.1016/j.jum.2019.09.004.

[20] C. R. de Almeida, A. C. Teodoro, and A. Gonçalves, "Study of the urban heat island (Uhi) using remote sensing data/techniques: A systematic review," *Environments - MDPI*, vol. 8, no. 10. 2021. doi: 10.3390/environments8100105.

[21] K. H. Suradiradja, I. S. Sitanggang, L. Abdullah, and I. Hermadi, "Estimation of biomass of forage sorghum (sorghum bicolor) Cv. Samurai-2 using support vector regression," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 30, no. 3, pp. 1786–1794, 2023, doi: 10.11591/ijeecs.v30.i3.pp1786-1794.

[22] S. Liu and Q. Shi, "Local climate zone mapping as remote sensing scene classification using deep learning: A case study of metropolitan China," *ISPRS J. Photogramm. Remote Sens.*, vol. 164, pp. 229–242, 2020, doi: 10.1016/j.isprsjprs.2020.04.008.

[23] Q. Klopfenstein and S. Vaiter, "Linear support vector regression with linear constraints," *Mach. Learn.*, vol. 110, no. 7, pp. 1939–1974, 2021, doi: 10.1007/s10994-021-06018-2.

[24] D. Parbat and M. Chakraborty, "A python based support vector regression model for prediction of COVID19 cases in India," *Chaos, Solitons and Fractals*, vol. 138, p. 109942, 2020, doi: 10.1016/j.chaos.2020.109942.

[25] I. Izonin, R. Tkachenko, N. Horbal, M. Greguš, V. Verhun, and Y. Tolstyak, "An Approach Toward Numerical Data Augmentation and Regression Modeling Using Polynomial-Kernel-Based SVR," in *Lecture Notes in Networks and Systems*, M. Saraswat, S. Roy, C. Chowdhury, and A. H. Gandomi, Eds., Singapore: Springer Singapore, 2022, pp. 771–781. doi: 10.1007/978-981-

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
10 (2) 2024 175-189

16-5120-5_58.

[26] K. Tscharke, S. Issel, and P. Debus, "Semisupervised Anomaly Detection using Support Vector Regression with Quantum Kernel," *Proc. - 2023 IEEE Int. Conf. Quantum Comput. Eng. QCE 2023*, pp. 611–620, 2023, doi: 10.1109/QCE57702.2023.00075.

[27] A. A. Ajhari, "The Comparison of Sentiment Analysis of Moon Knight Movie Reviews between Multinomial Naive Bayes and Support Vector Machine," *Appl. Inf. Syst. Manag.*, vol. 6, no. 1, pp. 13–20, 2023, doi: 10.15408/aism.v6i1.26045.

[28] Y. Zheng, W. Li, C. Fang, B. Feng, Q. Zhong, and D. Zhang, "Investigating the Impact of Weather Conditions on Urban Heat Island Development in the Subtropical City of Hong Kong," *Atmosphere (Basel).*, vol. 14, no. 2, 2023, doi: 10.3390/atmos14020257.