Future Cities and Environment
ISSN 2363-9075 Volume 11 (2025)

https://doi.org/10.70917/fce-2025-005
Article

Model Prediction of Land Surface Temperature from Satellite Data with Machine Learning

N. Chanpichaigosol [1], C. Chaichana [2], C. Y. Chu [3], M. Srivanit [4], D. Rinchumphu [5,*], T. Y. Chou [6] and M. L. Yeh [6]

[1] Candidate Department of Mechanical Engineering, Chiang Mai University, Thailand
[2] Department of Mechanical Engineering, Chiang Mai University, Thailand
[3] Green Energy and Biotechnology Industry Development Research Center, Feng Chia University, Taiwan
[4] Faculty of Architecture and Planning, Thammasat University, Thailand
[5] Department of Civil Engineering, Chiang Mai University, Thailand
[6] GIS Research Center, Feng Chia University, Taiwan
* Corresponding author**:** Damronksak Rinchumphu Tel +66-959959519; damrongsak.r@cmu.ac.th

Abstract: BACKGROUND AND OBJECTIVES: Urbanization leads to increased building construction, forming Urban Heat Island (UHI). UHI causes heat accumulation in urban areas, making it difficult to ventilate the area. This study aims to analyze Land Surface Temperatures (LST) using Remote Sensing (RS) data to predict UHI development in urban localities in Chiang Mai, Thailand. The study aims to compare the performance of different machine learning (ML) algorithms in predicting LST and assess their potential for future use in mitigating UHI consequences in urban areas. **METHODS:** RS data from Landsat 8 and Sentinel-2 satellites were used to analyze LST from 2016 to 2022. Five different ML algorithms were employed in this study: Random Forest (RF), AdaBoost Regressor (ABR), Artificial Neural Network (ANN), Linear Regression (LR), and Gradient Boosting (GB). The performance of these algorithms was evaluated using statistical variables. **FINDINGS:** The study found that the RF model had the highest precision in predicting LST, with the lowest Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) values among the models. However, all models had relatively low R-squared (R^2)values, indicating room for developing the accuracy of the predictions. **CONCLUSION:** The study demonstrates the feasibility of using RS data and ML algorithms to predict LST and comprehend UHI development in urban localities. The study also emphasizes the importance of using ML techniques to address UHI consequences in urban areas and can apply these data to urban planning to promote sustainable urban development. Further investigation is necessary to improve the accuracy of the models and determine effective strategies to mitigate UHI effects in urban areas.

Keywords: Land Surface Temperature; Machine Learning; Remote Sensing; Urban Heat Island; Prediction; Climate Action

1. Introduction

The growth of urban regions has resulted in multiple difficulties, one of which is the UHI phenomenon. UHI pertains to the rise in temperature in urban areas compared to the adjacent rural áreas (Almeida et al., 2021), brought about by heat accumulation on building surfaces due to inadequate ventilation and lack of green spaces (Zhang et al., 2022). The UHI phenomenon has direct and indirect consequences on the urban populace, including amplified energy consumption for cooling (Fan et al., 2022), reduced comfort levels outside buildings, and inferior air quality. LST has been recognized as a vital functional variable in evaluating the incidence of UHI in urban areas (Avdan and Jovanovska, 2016; Tomlinson et al., 2011). LST parameters can be procured from diverse sources, such as ground sensors

and satellite images furnished with thermal sensors like MODIS and Landsat 8 (Tomlinson et al., 2011). Nevertheless, these satellites possess relatively low resolution, and thus, the employment of satellites without thermal sensors like Sentinel-2, with more excellent image resolution, has become increasingly favored. This article presents a pattern for projecting LST via data procured from the RS process with the RF algorithm, a type of ML. The study region handpicked for this exploration is Chiang Mai, Thailand, where the UHI phenomenon is widespread due to the city's rapid expansión (Sangnum et al., 2014). The research aims to forecast the future of LST in the study area and provide insights into the UHI phenomenon's impact on urban areas. The results of this article can facilitate urban planning, weather forecasting, and formulating strategies for mitigating the UHI effects on urban areas. The utilization of RS data and ML algorithms in projecting LST can assist in comprehending the UHI phenomenon better and developing practical solutions to alleviate its effects.

The study area is the city of Chiang Mai, Thailand (Figure 1). Chiang Mai is the largest city in northern Thailand, located at 18.788 degrees North and longitude 98.986 degrees East, with a total area of approximately 431.46 km^2. The site has a tropical savanna climate (Srivanit and Iamtrakul, 2019), and research has shown that Chiang Mai is experiencing a UHI effect. From 2000 to 2006, the average temperature in the city increased from 20.52 ± 1.05 °C to 28.08 ± 1.50 °C (Srivanit and Hokao, 2012). In 2014, another analysis of the UHI effect in Chiang Mai found that the UHI value in the city can reach up to 4.35 °C during the hot season (Srivanit and Auttarat, 2016).

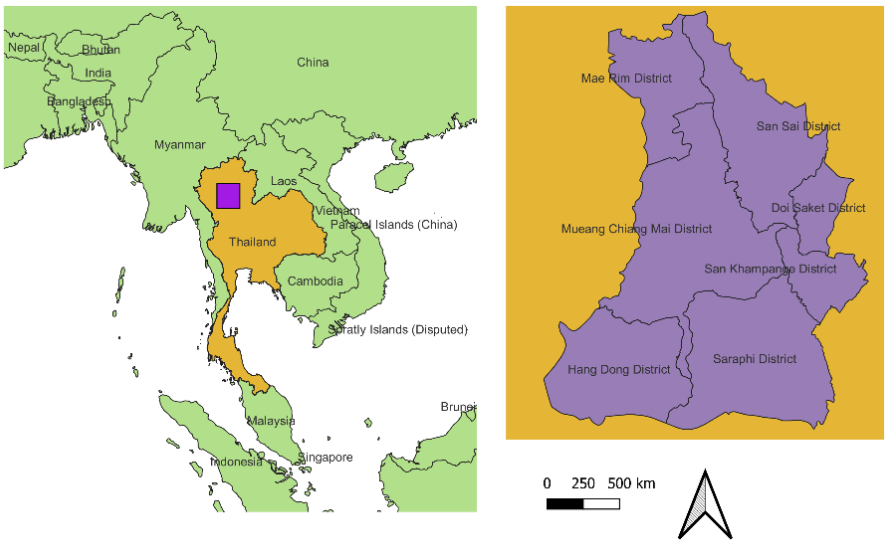

Figure 1. Geographical location of research area Chiang Mai, Thailand (by author).

2. Literature Review

2.1. Urban Heat Island

It is a phenomenon that urban areas have more heat than outer metropolitan areas, which is a problem that has been spread in many places over the past several decades. This phenomenon from the study is divided into 2 periods: 1. Daytime and 2. Night time (Zhou et al., 2011) As shown in Figure 2, finding an area in the city where UHI occurs will inevitably result in higher heat and, consequently, higher building energy consumption. According to the studies of heat islands in Thailand, a study in Bangkok province revealed that UHI directly impacts energy consumption in mining areas (Arifwidodo and Chandrasiri, 2015). On the other hand, in a cold country, the UHI effect can save up to 4.5 kWh/m^2 in heating costs (Boudali Errebai et al., 2022). In terms of the size of the urban area, it affects the UHI by increasing the size of the city, resulting in a higher UHI by the log-linear equation (Imhoff et al., 2010).

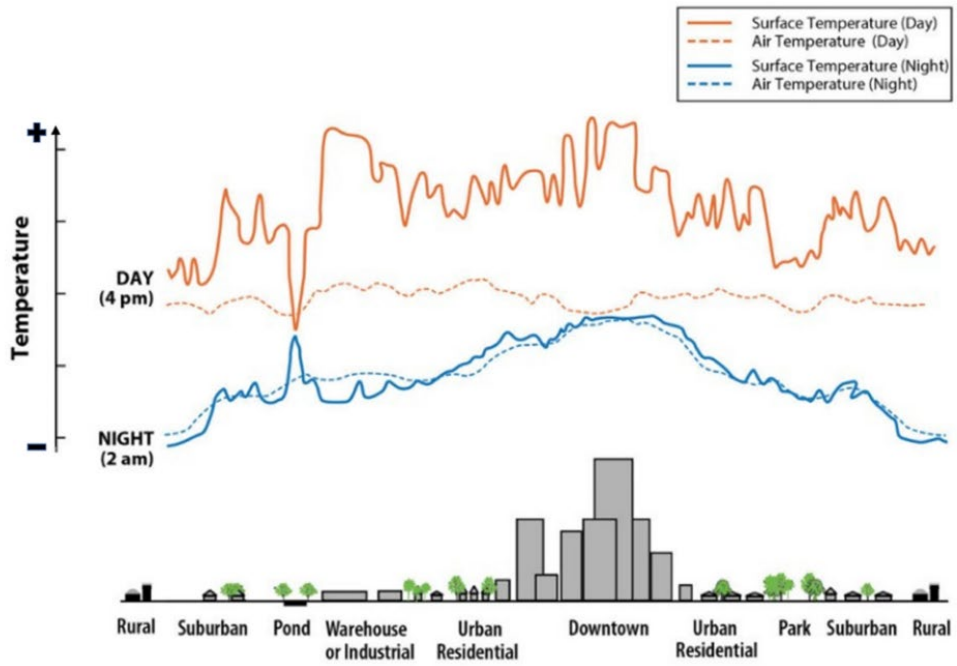

Figure 2. shows the temperature difference within the city and outside the town divided into periods. day and night (Almeida et al., 2021).

Moreover, other UHI impacts exist in China, where many different climate zones (Geng et al., 2023; Stewart and Oke, 2012). Regarding the phenomenon of UHI shift in Thailand, a study has been conducted on the timing of the seasons that affect the UHI phenomenon. It was found that UHI is the lowest in the Rainy Season and returns to the highest In the dry season (Jongtanom et al., 2011). To know the main factors that affect UHI, there is research that will help find ways to reduce the impact of UHI, increasing green areas or designing a city plan to be more suitable for the site In the case of increasing the green space in Chiang Mai, Thailand, it was found that the green area of 60.15 rai can distribute cool air within a radius of 75–100 m. at the same time, the size of 12.42 rai can distribute cool air within a radius 25–75 m (Nantarat et al., 2021).

2.2. Land Surface Temperature

The research findings indicate that Land Use Land Cover (LULC) impacts LST. (Zhou et al., 2011). Various tools are used to study LST, including utilizing data from the Meteorological Department, collecting ground-based data, and using RS with thermal sensors. The data format is typically divided into two main groups: daytime and nighttime. The choice of use depends on the suitability of the study area, the desired level of detail of the data, and access to historical data. One easily accessible source of historical data is RS from satellites. One method for calculating LST using imagery data from Lansat8 is shown in Table 1.

Table 1. shows the calculation process of LST from Landsat 8 satellite data.

Detail	**Equation**	**No. of Equation**
Convert digital number, DN, to radiance value	$\mathrm{TOA} = \mathrm{M}_L\, Q_{Cal} + \mathrm{AL}$	(1)
Convert radiance value to temperature value	$\mathrm{BT} = \frac{\mathrm{K2}}{\ln\left(\left(\frac{\mathrm{K1}}{\mathrm{TOA}}\right) + 1\right)}$	(2)
Convert Kelvin unit to Celsius unit.	$\mathrm{T}_c = \mathrm{BT} - 273$	(3)
Calculate NDVI	$\mathrm{NDVI} = \frac{(\mathrm{NIR} - \mathrm{RED})}{(\mathrm{NIR} + \mathrm{RED})}$	(4)

Calculate the proportion of vegetation (PV)	$PV = \sqrt{\frac{(NDVI - NDVImin)}{(NDVImax - NDVImin)}}$	(5)
Calculate the Emission	$\varepsilon = 0.004 \times P_v + 0.986$	(6)
Calculate LST	$LST = \frac{BT}{(1 + (0.00115 \times \frac{BT}{1.4388}) \times \ln(\varepsilon))}$	(7)

Adapt from (Jiang and Tian, 2010; Peng et al., 2018; Sherafati et al., 2013)

Which

ε	Emissivity
AL	Band-specific additive rescaling factor from the metadata, the band (10) = 0.1
BT	Surface temperature in kelvin
K1	774.89 (Band-specific thermal conversion constant from the metadata)
K2	132.08 (Band-specific thermal conversion constant from the metadata)
M_L	Band-specific multiplicative rescaling factor from the metadata, the band (10) = 3.3420E-04
P_v	The proportion of vegetation
Q_{Cal}	Quantized and calibrated standard product pixel value (DN)
T_C	Surface temperature in celcius
TOA	Top of Atmosphere radiance in (Watts/square meter*srad*um))

2.3. The Variables Affecting LST

This search method of relevant research papers shown in Figure 3 Used the keywords "Land surface temperature" and "Prediction" in the Scopus database, and 73 accessible articles were found. Then, they screened the groups not interested in LST prediction, such as water evaporation, downscale aerial photographs, soil moisture, crop predictions, and insects. They resulted in 34 articles that mentioned predictions. LST, from screening variables shared more than 5 times, can be summarized in Table 2.

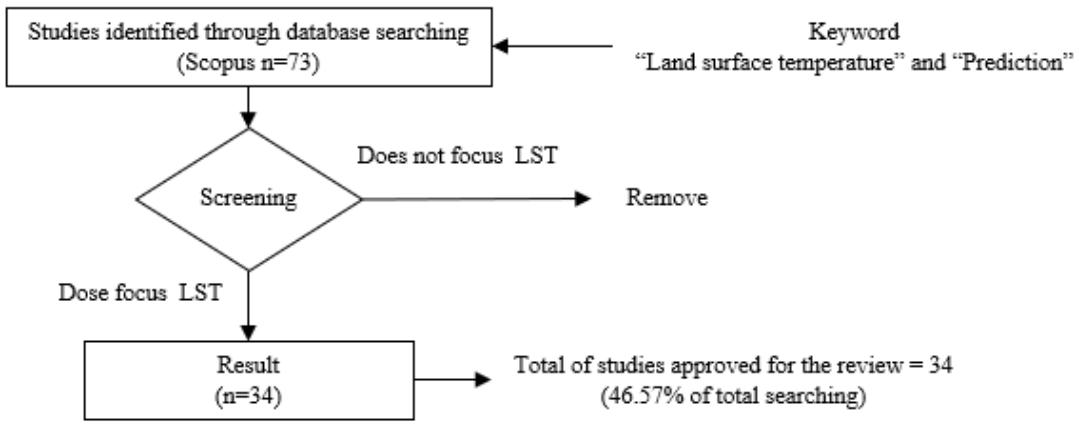

Figure 3. shows the process of identifying variables that affect LST (by author).

Table 2. Variable from review journal.

Variable from Satellite	Description	NO. of Publications That Mentioned The Variable
NDWI	Normalized Difference Water	5
EVI	Enhanced Vegetation Index	5
Lat	Latitude	6

Long	Longitude	6
Elevation	Elevation	6
LULC	Land Use Land Cover	10
NDBI	Normalized Difference Built-up Index	12
NDVI	Normalized Difference Vegetation	19
LST	Land surface Temperature	34

Latitude, Longitude

Latitude is a coordinate that indicates a position on the Earth's surface, with values ranging from 0-90 degrees for North or south. On the other hand, longitude is a coordinate used to indicate a position on the Earth's surface in the west or east direction, with values ranging from 0–180 degrees. In the example of the study area, the center of Chiang Mai province is located at the latitude and longitude coordinates of 18.796143 and 98.979263, respectively. The direct benefit of using latitude and longitude is accurately determining the sample point's position (Bartesaghi-Koc et al., 2022; Kafy et al., 2022; Mohammad and Goswami, 2022).

2.3.1. Normalized Difference Built-Up Index (NDBI)

The NDBI is a variable that can be calculated in various ways from RS studies and indicates the map of built-up areas in urban environments. The calculation of NDBI varies for each satellite, as shown in Equation 8.

$$NDBI = \frac{(SWIR1 - NIR)}{(SWIR1 + NIR)} \quad (8)$$

NDBI is widely used as a variable for urban planning and LULC mapping. Its data range is from −1 to 1 and can be divided into two groups: built-up and non-built-up. Regenerate response (AlDousari et al., 2022). Non-built-up refers to areas that are not developed or constructed, such as natural areas, farmland, or vacant land. Studies have used NDBI as a variable to assess LST, which reveals that a high *NDBI* variable impacts increasing LST (Gao et al., 2022; Kafy et al., 2021; Mushore et al., 2017).

2.3.2. Normalized Difference Water (NDWI)

It is a variable that describes the presence of water bodies from RS, which can be calculated using equation 9. It is widely used in LST prediction, as an increase in the water surface area has a cooling effect, resulting in a decrease in LST. The data range is between −1 and 1, divided into two groups: water and non-water features in satellite or aerial imagery(Gao et al., 2022; Mushore et al., 2022; Mushore et al., 2017).

$$NDWI = \frac{(NIR - SWIR1)}{(NIR + SWIR1)} \quad (9)$$

2.3.3. Normalized Difference Vegetation (NDVI)

It is an essential variable to consider when reviewing all 34 articles. The *NDVI* variable was used in 19 (55.88%) of the articles, and it can be calculated using equation 4, which is one of the variables used to calculate LST directly. The *NDVI* value ranges from −1 to 1, where higher values indicate greater vegetation density. It is commonly used to monitor land cover, agriculture, and natural changes (AlDousari et al., 2022; Karimi et al., 2021).

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \quad (10)$$

2.3.4. Enhanced Vegetation Index (EVI)

The EVI variable is a variable that is computed from MODIS satellite imagery, which is calculated from the red, blue, and near-infrared frequency ranges. This variable is commonly used with MODIS satellite imagery (Mushore et al., 2022). Studies have compared the accuracy of predicting natural vegetation coverage using NDVI and EVI in the MODIS satellite imagery frequency range. These studies have shown that using NDVI for prediction yields greater accuracy than EVI (Alademomi et al., 2020;

Li et al., 2010).Translation: The Standardized Precipitation Index (SPI) was also tested, and the results of both variables differed by only 5% (Farrokhzadeh et al., 2018). As a result, in this research, NDVI was chosen as a representative variable for model development.

2.3.5. Land Use Land Cover

LULC, The study has divided the samples into four groups: 1. Buildings (residential, commercial, and industrial); 2. Water surfaces (rivers, lakes, and others); 3. Green land (trees, parks, grasslands, and other vegetation); and 4. Agricultural areas (arable land, vacant land, sand, and bare lands) (Gao et al., 2022).

The Digital Elevation Model (DEM) digitally portrays a land surface's topography or elevation. It is a 3D model that showcases the height of distinct points on the Earth's surface concerning sea level. DEM information is generally obtained from satellite or airborne images and is frequently utilized in various applications such as land analysis, hydrological simulation, and ecological surveillance.

2.4. Satellite Imagery

The satellite image represents the surface or different parts of the Earth. The imagery is captured using cameras mounted on satellites, each with additional imaging capabilities, as shown in Figure 4. These images can be grouped into two main categories: publicly available and commercial, which can be clearly distinguished based on image resolution and repeat time. The usage of these images can vary depending on the user, such as environmental monitoring (Skelhorn et al., 2016), weather forecasting (Nayak and Ghosh, 2013; Skelhorn et al., 2016), urban planning (Imhoff et al., 2010) agriculture (Allen et al., 2011; Nguyen et al., 2020) and even in disaster management systems. The satellite images are collected in various wavelengths ranging from visible light to microwave.

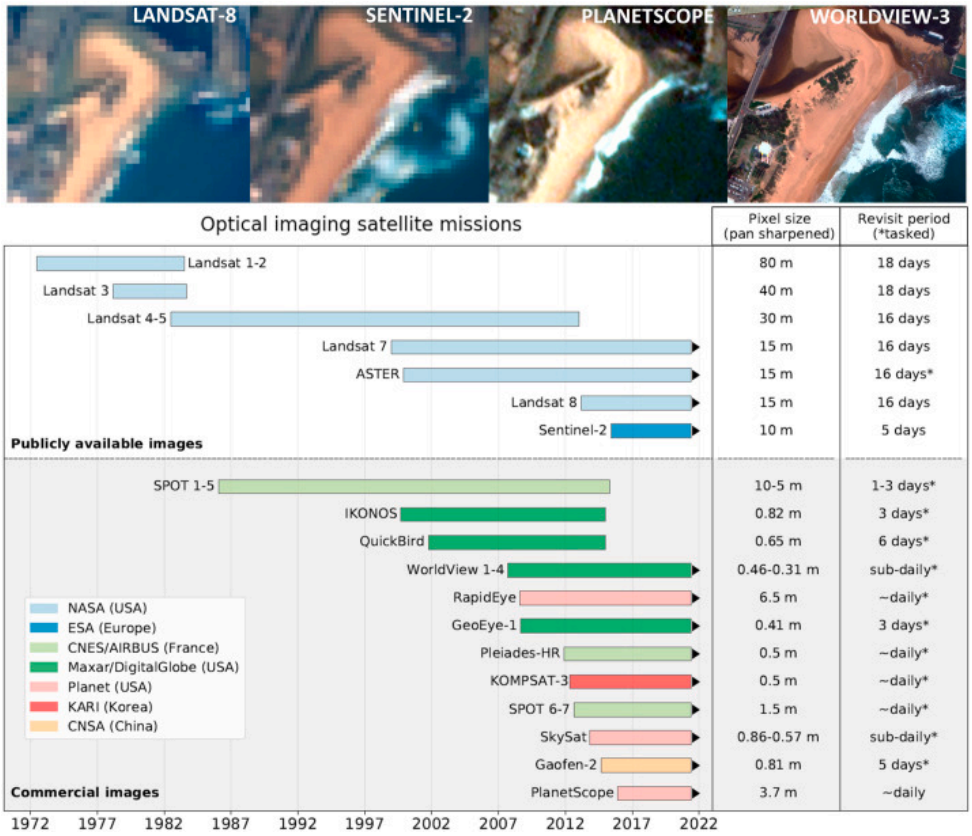

Figure 4. Timeline of public and commercial satellites (Turner et al., 2021).

Table 3. Popular satellites are used in part of UHI.

Thermal Sensor	Satellite	Orbital Frequency	Data Available
ASTER	Terra	2 days	1999
MODIS	Terra	2 days	1999
MODIS	Aqua	2 days	2002
TIRS	Landsat 8	16 days	2013
TIRS 2	Landsat 9	16 days	2021
-	Sentinel-2	5 days	2015

(Adapted from (Shih et al., 2019; Xu et al., 2017; Zhou et al., 2019))

Table 3 shows the differences between each sensor used for measurement and their ability to access the satellite's historical data. Each satellite is often used for further data analysis. For example, Sensor MODIS, which has sample cases in China, analyzes UHI with 253 Chinese cities and 5 climate zones (Geng et al., 2023). Moreover, images taken from Sensor MODIS are also used for urban planning and improving the settlement's environment (Xu et al., 2021). The base of the analysis is quite broad due to the detail of the MODIS images, which are at 1 km. The next group is the Landsat, starting from Landsat 4 to the current Landsat 9, which has continuously improved its sensors. Currently, two sensors, TIRS and TIRS2, are widely used and obtained from Landsat 8 and 9. Landsat 8 is widely used due to its data from 2013, which is ten years old and suitable for predicting various data. Landsat 8 because it is an RS satellite that provides high-resolution images of the Earth's surface. The images captured by Landsat 8 have a resolution of 30 meters for the visible, near-infrared, short-wave infrared, and thermal infrared bands, making it useful for a wide range, including LULC mapping (Jamali, 2019; Jiang and Tian, 2010; Yu et al., 2019), Agricultural monitoring (Dhillon et al., 2022), Natural Resource Management (Singh et al., 2021), Climate change monitoring (Hashim et al., 2019) and UHI (Almeida et al., 2021). Landsat eight images are also often used with other data sources, such as reference data, to support various scientific research and analytical purposes. The satellite's coverage and long-term imagery archive make it a valuable resource for researchers in multiple fields. The limitations of Landsat 9 are that it is a newly launched satellite, meaning the amount of data or the number of satellite images for the data still needs to be increased, resulting in a limited ability to predict. Some case studies in LST analysis in China have used data from 2010 to 2019. Using a more extended range of data in data analysis is necessary (Almeida et al., 2021; Xu et al., 2021).

2.4.1. Landsat 8 Satellite

Landsat Satellite: The Landsat program is a series of Earth observation satellites operated by the United States Geological Survey (USGS) that began in 1972. Satellites carry instruments that capture images of the Earth's surface in various wavelengths, which are used for multiple applications such as LULC mapping and natural disaster response. One of the most important contributions of the Landsat program is the creation of a long-term record of the Earth's surface changes. The Landsat data archive is available to the public free of charge and is widely used by researchers, policymakers, and others worldwide. The newest Landsat satellite, Landsat 9, was launched in September 2021 and continues to provide essential information for managing the Earth's resources.

2.4.2. Sentinel-2 Satellite

Sentinel-2 is a satellite that will record the sun's reflection data from the Earth's surface to examine land and sea surfaces. The satellite will capture satellite images every 5–7 days, with the highest resolution being 10 m. and the lowest solution being 60 m. The data can be accessed through Freely available satellite products, specifically from the Sentinel-2 satellite, which has been open since 2015. However, this satellite does not have a thermal sensor, but it can still be used to calculate the NDVI, which is the main factor used in calculating LST. In this research, the Sentinel-2 satellite, which has the highest image resolution, was chosen in conjunction with the Landsat 8 satellite, which has a thermal sensor installed. Data from both satellites were prepared for further ML processing.

2.4.3. The Program Used to Manage Satellite Images

GIS Application: Most of the geographic information in the field of geography is in map format. The GIS program collects and analyzes data to present it in a more easily understandable format (Siddiqui et

al., 2012)—for example, the programs Esri's ArcGIS and open-source QGIS. ArcGIS is a GIS program used for mapping and working in the raphe, developed by the company "Environmental Systems Research." (Khan and Mohiuddin, 2018). On the other hand, the program QGIS has the full name Quantum Geographic Information System and was developed in Germany. This program was designed in an open-source format, allowing users to use, improve or develop plugins. It is difficult to say which software is better, but in terms of cost, it is advisable to use the QGIS program (Khan & Aaqib, 2017). The literature review found that GIS is used for various types of analysis, including mapping and analyzing temperature using aerial photographs. It is also used to identify the urban area pattern for evaluating the impact of UHI and to study variables that affect the occurrence of heat islands (Jiang & Tian, 2010; Xu et al., 2021). The European example demonstrates that GIS divides cities into three groups (Kasanko et al., 2006).UHI is predicted by referencing the city's expansion using aerial photographs (Mitraka et al., 2015). In addition to using GIS, ENVI-met software is also applied to assist in analyzing changes in outdoor comfort. It helps to visualize the airflow direction and temperature changes in urban areas (Wang et al., 2016).

2.5. Prediction Method

The predictive model is divided into two major groups: LR and non-LR. In the process of LR, there are limitations in terms of variable correlations. In the process of RS, non-LR is commonly used, which provides greater accuracy. An example of a non-LR model is the RF (Liu et al., 2022), ANN (Mohammad et al., 2022), GBR (Zhang et al., 2022), and ABR. In the example mentioned, RF is widely used due to its ability to handle complex variables and provide greater accuracy. It has been accepted that RF can accurately predict NDVI (Ma et al., 2021). In Brazil, RF has been used to predict soybean yield (Schwalbert et al., 2020). RF algorithm is an ensemble learning technique that builds multiple decision trees and aggregates their outputs to generate the final prediction. The algorithm randomly selects subsets of features and data points during tree-building to create diverse trees. The results of each tree are then combined to make the final prediction. This algorithm is renowned for its exceptional accuracy and ability to resist overfitting (Breiman, 2001; Cutler et al., 2007). ABR is a boosting algorithm that combines multiple weak learners to generate the final prediction. The algorithm trains multiple models iteratively by adjusting the weights of the training examples based on the previous model's error. This algorithm typically uses decision trees with shallow depths as weak learners. The ultimate forecast is the combined total of the at-risk students' results. This algorithm is widely appreciated for its ability to handle noisy data and deliver high accuracy (Chen and Guestrin, 2016). ANN is a class of algorithms that takes inspiration from the structure and function of the human brain. ANNs use interconnected nodes (neurons) to model complex relationships between input and output variables. During the training process, the algorithm adjusts the weights of the connections between the neurons to minimize the error between the predicted and actual values. ANNs are known for their ability to model non-linear relationships and handle large datasets (Goodfellow et al., 2016). LR is a straightforward algorithm that models the linear association between input and output variables. It assumes a linear relationship between the variables and attempts to find the best-fit line that minimizes the error between the predicted and actual values. The algorithm uses a closed-form solution to calculate the coefficients of the bar that best fits the data. This algorithm is famous for its simplicity and interpretability (Montgomery et al., 2021). GBR is another boosting algorithm that combines multiple weak learners to generate the final prediction. It is similar to ABR, but it uses a different strategy to adjust the weights of the training examples. Instead of modifying the importance of the criteria, it tries to fit the residuals of the previous model with a new model. The final prediction is the weighted sum of the weak learners' outputs. This algorithm is famous for its high accuracy and ability to handle complex relationships between variables (Friedman, 2001). The evaluation and assessment of ML models use four statistical variables: MAE, MSE (Mean Squared Error), RMSE, and R^2. The MAE denotes the mean absolute deviation between the anticipated and actual values. The model's performance is enhanced with a lesser MAE. The MSE indicates the mean squared deviation between the predicted and actual values. The model's performance is enhanced with a lesser MSE. The RMSE is the square root of the MSE and is also employed to assess the model's performance. The model's performance is enhanced with a lesser RMSE. R^2 calculates the percentage of variation in the dependent variable LST, which is clarified by the independent variables. R^2 ranges from 0 to 1; higher values denote better model performance (Magistrali et al., 2021).

This study aimed to analyze LST using RS data to predict LST in urban areas of Chiang Mai Province, Thailand. Its goal was to learn different machines for predicting LST in urban areas from 2016 to 2022.

3. Materials and Methods

3.1. Preparation of the Image

From a literature review, the research methodology can be summarized as shown in Figure 5. The Images obtained from satellites must undergo preprocessing, which involves image correction to reduce data distortion, noise, and geometric distortion that may occur during image capture. The differences in preprocessing methods can be explained in Figure 6. Each satellite has its preprocessing plans. Landsat 8 Level 2 satellite images preprocessed from Earth Explorer will be used in this research. On the other hand, Sentinel-2 satellite images are divided into two groups: the first group is the data from 2016, which has yet to be preprocessed and, therefore, needs to undergo preprocessing by the researchers. The other group is the data from 2017–2021, which can be selected to use preprocessed data.

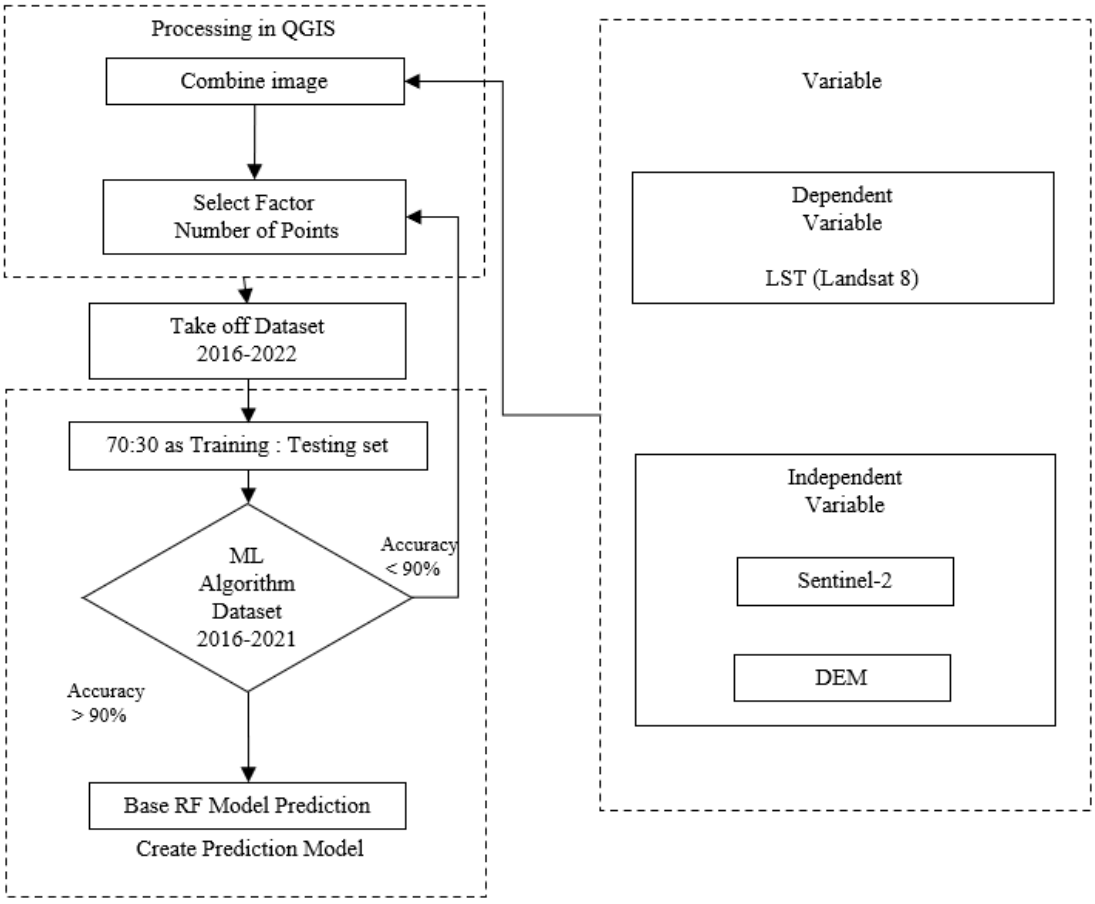

Figure 5. Flowchart of research methodology (by author).

In the preprocessing of Sentinel-2 satellite imagery, a tool called Sen2Core is used, and the Python Scripts are accessed through the Command Prompt. The process involves accessing the Sen2Core folder and initiating the process by specifying the location of the Sentinel-2 satellite imagery files. Once the process is complete, the preprocessed satellite imagery data can be used for further research. Landsat 8, In RS satellite images, is a crucial data source. The Earth Explorer website is used to obtain these images, which involves a few steps. First, the study area is selected, followed by the date and time range for the study. In addition, the cloud cover is limited to no more than 5% to ensure the lowest possible percentage of clouds. Next, the data sets are selected, with the research utilizing Landsat 8 Collection 2 Level-2 satellite imagery. Finally, Band10 data is chosen from the available bands for use in the study. Sentinel-2: For the satellite imagery of Sentinel-2, data will be obtained from the Copernicus Open Access Hub, which provides open access to aerial imagery from the Sentinel-2 satellite. The process for obtaining aerial imagery involves selecting the study area, choosing a date and time range, selecting the Sentinel-2 satellite, and selecting images as close as possible to the images obtained from Landsat 8 to match the days.

3.2. Processing Image in QGIS

Once the satellite images have been acquired, the next step is to process them using QGIS, a powerful open-source Geographic Information System (GIS) software that allows users to manipulate and analyze spatial data. This study will employ QGIS to organize layer data. The initial step in the processing of satellite images involves the organization of the layer data. This procedure entails overlaying the satellite images onto preexisting map layers to create a composite image with all the necessary information. The layer data may include land cover, topography, and other environmental variables that can affect LST. Through layer overlap, researchers can better understand the spatial correlation between LST and various other environmental factors. Generate sampling coordinates for CSV data. After organizing the layer data, the next step is to extract sampling points from the composite image to obtain a CSV data file. Sampling entails the process of choosing particular positions on the image and documenting the corresponding LST value. The sample points are then exported to a CSV file, which can be used to train the ML models. Before training the machine learning models, it is necessary to preprocess the CSV file. This procedure eradicates any missing data, outliers, or errors that may have an adverse effect on the accuracy of the model. Data preprocessing involves the process of scaling and normalizing the data to ensure that all variables are standardized and have the same scale. Following the completion of data preprocessing, the subsequent task involves selecting the most relevant features to train the machine learning models. This procedure entails identifying the variables that substantially impact LST and eliminating any superfluous or inconsequential characteristics.

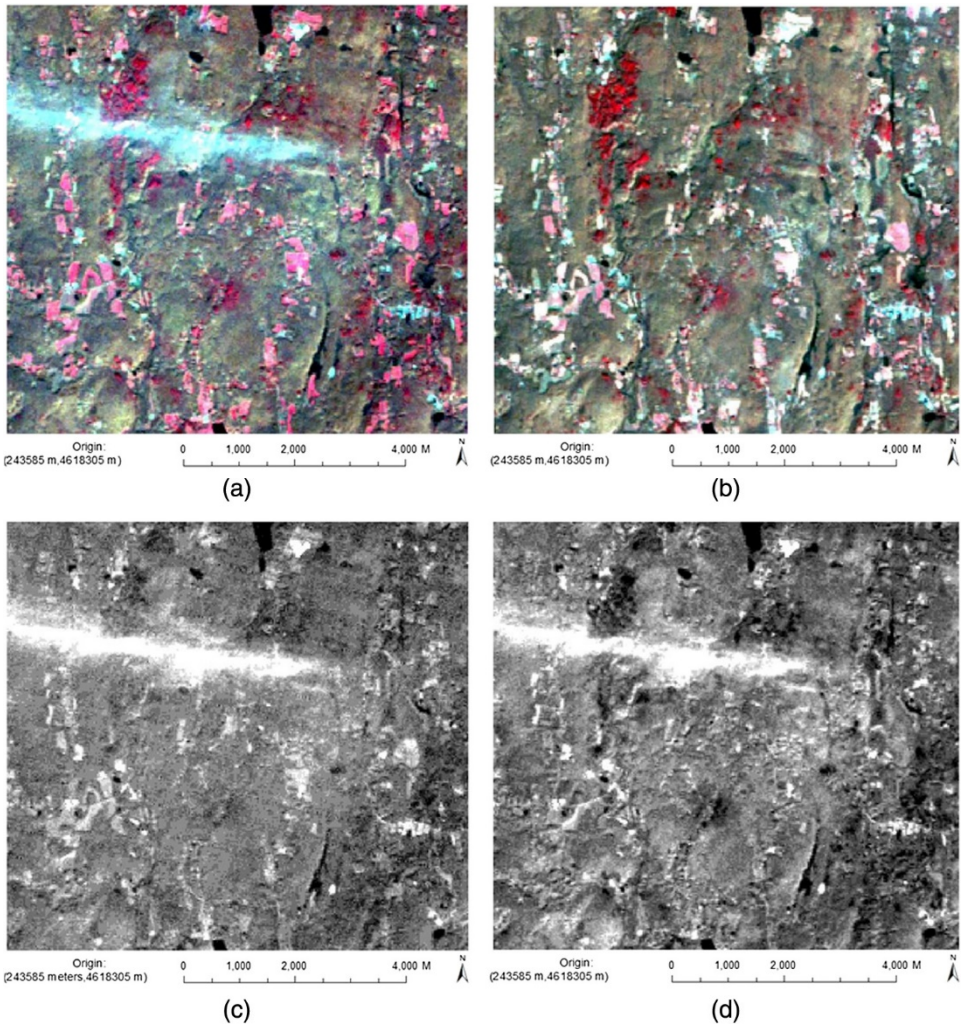

Figure 6. compares the differences between preprocessing files (a) and (c), which have not undergone the process, and (b) and (d), which have undergone the procedure. (Zhang et al., 2014).

3.3. Create Model Prediction ML

ML is utilized to construct predictive models. It entails the creation of code utilizing preexisting algorithms. The programming language chosen for this research will be Python. Utilizing the Spider (Anaconda) platform for software development. The data is imported from QGIS as a .CSV file, with the variables categorized into two groups: the independent variable X and the dependent variable Y. The data is subsequently partitioned into training and testing datasets in a 70:30 ratio to facilitate the modeling process. Ultimately, the data is transmitted to machine learning algorithms. The literature review indicates that the RF, ABR, ANN, GBR, and LR algorithms are the most commonly employed in this research. The code will be implemented using the Spyder software bundled with Anaconda. The code will be implemented in Python, with the approximation parameter set to 100 and the random state parameter set to 42. These values are commonly employed for prediction purposes.

4. Results and Discussion

4.1. Collect the Image

The selection of aerial images is based on the minimum percentage of cloud cover, which falls from 0 to 5%. The research initially chose images from Landsat 8 and subsequently opted for images from Sentinel-2 because Landsat has a longer repeat cycle. The time period for 8 is 17 days, whereas Sentinel-2 has a greater frequency. It enables a broader range of images to be chosen. Table 4 presents exemplars of the datasets gathered for the variables employed in this study. The comprehensive overview can be briefly presented in Table 5, which displays the date disparities. This research period has the most imminent date. However, there were instances when images could not be matched due to excessive cloud cover. Typically, the duration will range from 1 to 3 days, but certain images like Landsat 8 on 09/03/2016 and Sentinel-2 on 26/03/2016 will be discernible images from distinct periods. Many of These issues arose due to an excessive amount of cloud cover.

Table 4. Data sources and descriptions.

Types	**Description**	**Data sources**
RS data	Landsat 8 OLI/TIRS (L2 Cloud cover <5%)	Earthexplorer.usgs.gov
	Sentinel-2 S2A, S2B (The date is close to Landsat 8)	sci-hub.Copernicus. eu
	Digital Elevation Model Chiang Mai Area (2015)	GISTDA
LCLU	Land Cover Land Use Chain Mai area	Digitization in QGIS

Table 5. The Summary of satellite imagery.

Landsat 8 Images		**Sentinel-2**	
Image date	**Scene Identifier (Band 10)**	**Image date**	**Scene Identifier (MA, MB, NA, NB)**
09/03/2016	LC08_L2SP_131047_20160309	26/03/2016	S2A_MSIL1C_N0201_R104_T47Q
12/05/2016	LC08_L2SP_131047_20160309	25/04/2016	S2A_MSIL1C_N0201_R104_T47Q
12/05/2016	LC08_L2SP_131047_20160309	05/05/2016	S2A_MSIL1C_N0202_R104_T47Q
08/02/2017	LC08_L2SP_131047_20170208	09/02/2017	S2A_MSIL1C_N0204_R104_T47Q
12/03/2017	LC08_L2SP_131047_20170312	11/03/2017	S2A_MSIL1C_N0204_R104_T47Q
25/12/2017	LC08_L2SP_131047_20171225	21/12/2017	S2B_MSIL1C_N0206_R104_T47Q
26/01/2018	LC08_L2SP_131047_20180211	01/25/2018	S2A_MSIL1C_N0206_R104_T47Q
11/02/2018	LC08_L2SP_131047_20180211	14/02/2018	S2A_MSIL1C_N0206_R104_T47Q
15/03/2018	LC08_L2SP_131047_20180315	06/03/2018	S2A_MSIL1C_N0206_R104_T47Q
13/01/2019	LC08_L2SP_131047_20190113	20/01/2019	S2A_MSIL2A_N0211_R104_T47Q
29/01/2019	LC08_L2SP_131047_20190129	25/01/2019	S2B_MSIL2A_N0211_R104_T47Q
14/02/2019	LC08_L2SP_131047_20190214	04/02/2019	S2B_MSIL2A_N0211_R104_T47Q
16/01/2020	LC08_L2SP_131047_20200116	15/01/2020	S2A_MSIL2A_N0213_R104_T47Q

01/02/2020	LC08_L2SP_131047_20200201	04/02/2020	S2A_MSIL2A_N0214_R104_T47Q
17/02/2020	LC08_L2SP_131047_20200217	14/02/2020	S2A_MSIL2A_N0214_R104_T47Q
02/01/2021	LC08_L2SP_131047_20210102	04/01/2021	S2B_MSIL2A_N0214_R104_T47Q
03/02/2021	LC08_L2SP_131047_20210203	03/02/2021	S2B_MSIL2A_N0214_R104_T47Q
07/03/2021	LC08_L2SP_131047_20210307	05/03/2021	S2B_MSIL2A_N0214_R104_T47Q
05/01/2022	LC08_L2SP_131047_20220105	09/01/2022	S2B_MSIL2A_N0301_R104_T47Q
14/02/2022	LC09_L2SP_131047_20220214	13/02/2022	S2A_MSIL2A_N0400_R104_T47Q
26/03/2022	LC08_L2SP_131047_20220326	05/03/2022	S2A_MSIL2A_N0400_R104_T47Q

4.2. Image Processing with QGIS

In this process, the images were filtered and calculated using QGIS software. The image was divided into two groups: Landsat 8 and Sentinel-2. For LST images, they were calculated based on Table 1, as shown in Figure 7. On the other hand, the images were merged before calculating various variables such as NDVI, NDBI, NDWI, LULC, EVI, and elevation from GISTDA, as shown in Figure 8.

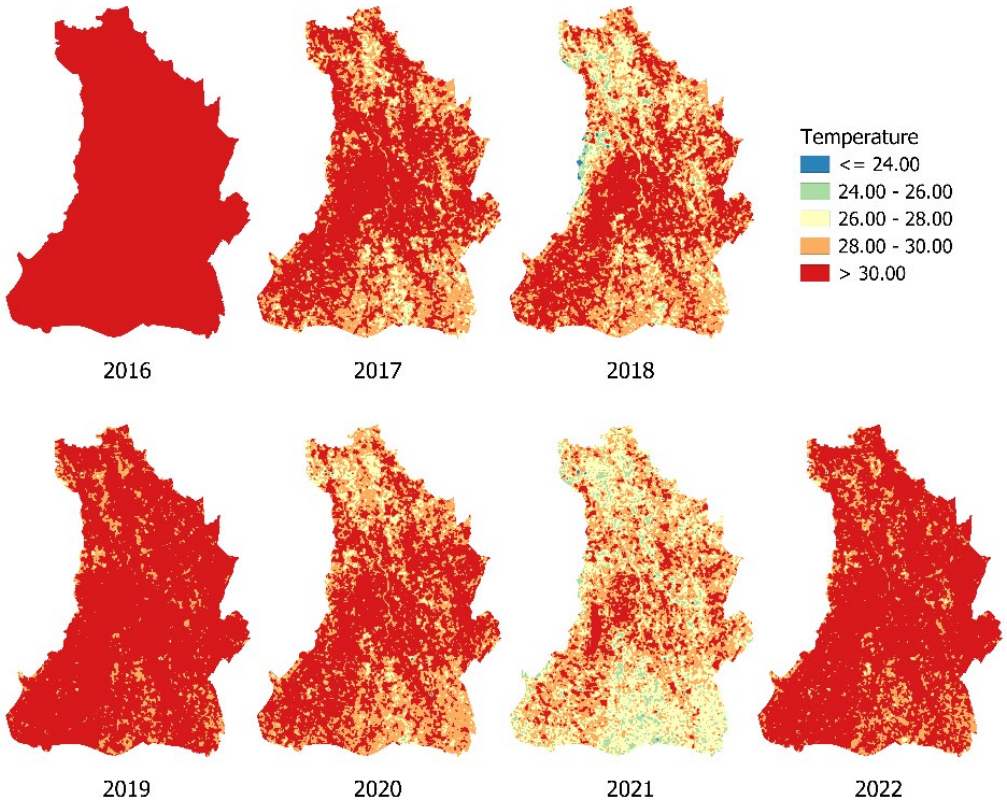

Figure 7. shows an example of the LST result obtained from the calculation process (by author).

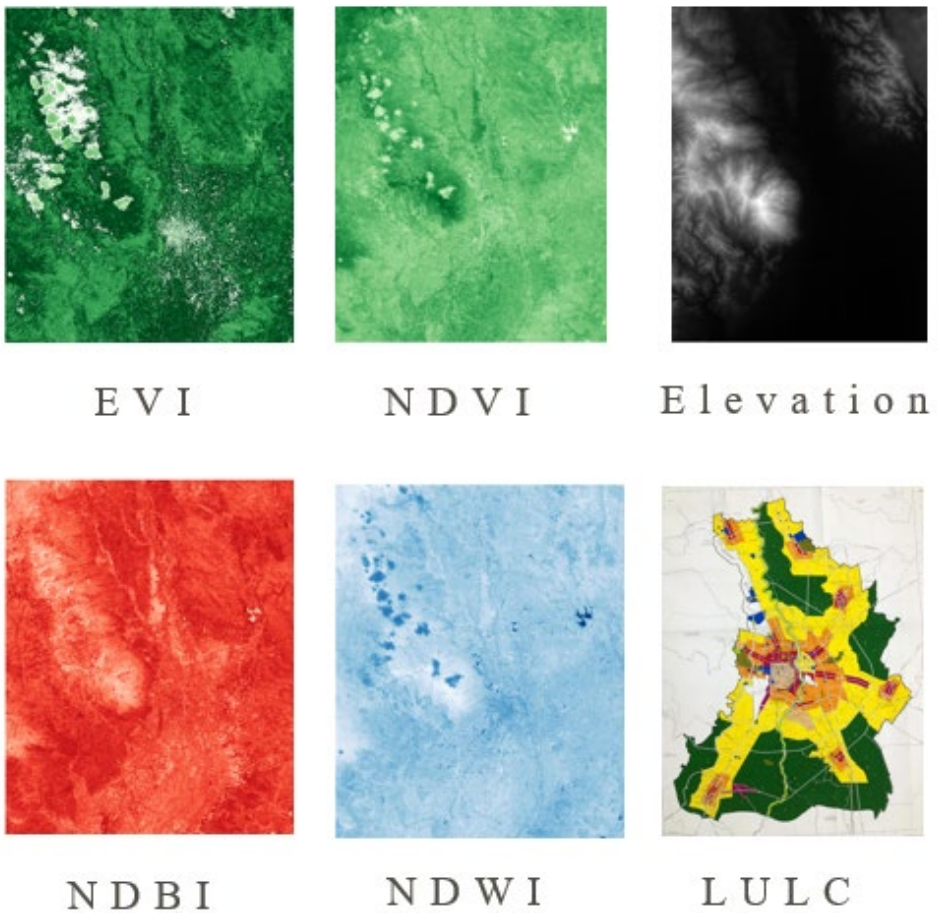

Figure 8. shows an example of the results obtained from processing Sentinel-2 (by author).

4.3. Mapping LULC

Table 6 displays the results of the LULC classification for the years 2016 to 2022. The classification was performed using training data from 2016. The "Area" column displays the cumulative research site area for each year. The "%" column displays the ratio of the total area occupied by each LULC class. The study sites in 2016 encompassed a combined area of 437.4977 square kilometers. The research area is primarily characterized by vegetation covering 59% of the land. The sparse vegetation may be attributed to the arid conditions during the region's dry season. Leading to a reduction in plant growth. The second most prevalent type of land in the research area is built-on land, which accounts for 25% of the total area. The remaining region predominantly comprises soil. Accounting for 15 percent, while water sources contribute to a minor fraction. Subsequent years witnessed variations in LULC patterns. The vegetation coverage experienced a marginal decline from 59% in 2016 to 56% in 2021. The percentage rises to 58% in 2022. The level of urbanized land remains relatively stable. They constitute approximately 8-9% of research sites every year. The prevalence of buildings and soil in 2016 can be attributed to the arid climate conditions in the area. Consequently, there is a reduction in the rate of vegetable proliferation, as evidenced by satellite imagery data. The water coverage in the research area remains consistently at around 5% each year. The primary LULC transformation is characterized by a notable rise in soil occupation, escalating from 15% in 2016 to 31% in 2021, followed by a subsequent decline to 29% in 2022. The high soil coverage can result from a combination of factors, including rice cultivation and the density of vegetation coverage.

In summary, these findings provide a valuable and concise representation of the LULC patterns in the research area throughout the specified time period. By utilizing the training dataset from 2016, it is possible to create maps for subsequent years. These maps can then be employed in various applications, including monitoring and documenting changes in land use patterns over time. Evaluating the effects of land use policies and interventions, as well as predicting future land conditions. changes.

Table 6. The percentage of the area from the LULC mapping process.

Year	2016	2017	2018	2019	2020	2021	2022
Area %	100%	100%	100%	100%	100%	100%	100%
Built-up	17%	8%	8%	8%	8%	9%	9%
Water	1%	5%	5%	6%	5%	5%	5%
Vegetation	30%	58%	60%	62%	58%	56%	58%
Soil	54%	29%	27%	24%	29%	31%	29%

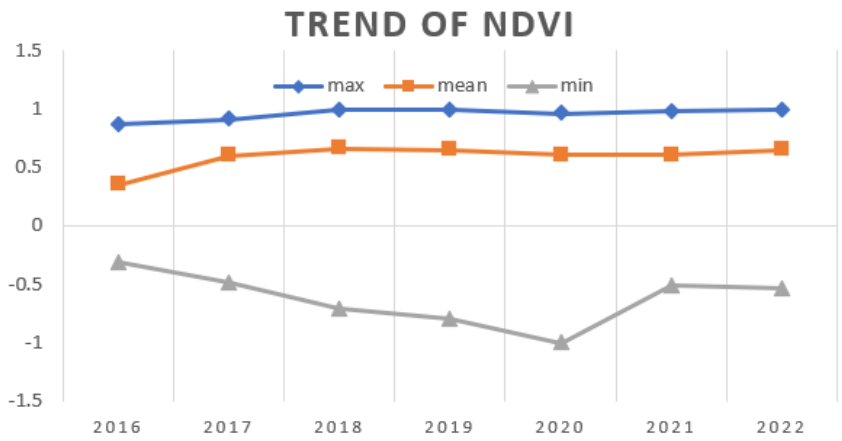

Figure 9. shows the trend of NDVI.

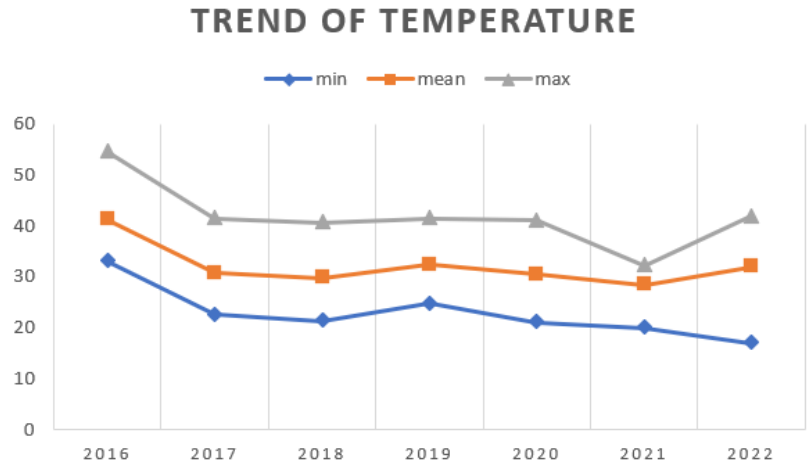

Figure 10. shows the trend Temperature.

Figure 9 shows the trend of NDVI and temperature. The NDVI part describes the vegetation cover and density. The range of NDVI data is from −1 to 1. Figure 10 explains the variability of vegetation density from 2016 to 2022. The vegetation density decreased from 2016 to 2020 and increased again in 2021 and 2022. On the other hand, the temperature or LST obtained from QGIS shows that 2016 had the highest LST, and there was a decreasing trend of LST in the following years.

4.4. The Process Creates the Dataset

A total of 50,000 points are used per day, meaning that there are a total of 1,050,000 datasets that will be used to create a forecasting model. The results from removing the data found that there needs to be more accurate data sets. For example, data from Landsat 8 shows that some LST variables have a value of 0. This point means that the point where the test was performed was in a cloud, and therefore, the LST

could not be measured there. As a result, LST variables need to be filtered with Code Python to eliminate ranges correctly. On the other hand, Sentinel-2 variables appear to have a range of data that exceeds normal. The variable EVI will have a data range of −1 to 1, but for data in this set, there are some points where this variable is more significant than 300, so this data set must be omitted.

4.5. The ML Results

In this section, the study conducted experiments with four ML algorithms: RF, ANN, ABR, GBR, and LR. The first part discusses the correlation coefficients between variables in Table 6.

Table 7. Factor Correlation.

	NDVI	**NDBI**	**LULC**	**DEM**	**EVI**	**NDWI**	**X**	**Y**
NDVI	1	−0.6627	0.2246	−0.0945	0.0802	−0.9479	0.0135	−0.1083
NDBI	−0.6627	1	0.0926	0.1360	−0.0583	0.5588	−0.0660	0.0446
LULC	0.2246	0.0926	1	0.0411	0.0118	−0.2999	−0.0316	−0.0277
DEM	−0.0945	0.1360	0.0411	1	−0.0100	0.0718	−0.5610	0.4601
EVI	0.0802	−0.0583	0.0118	−0.0100	1	−0.0698	0.0079	−0.0126
NDWI	−0.9479	0.5588	−0.2999	0.0711	−0.0698	1	−0.0046	0.1115
X	0.0135	−0.0660	−0.0316	−0.5610	0.0079	−0.0046	1	0.0716
Y	−0.1083	0.0446	−0.0277	0.4601	−0.0126	0.1115	0.0716	1

Table 7 shows the variables with strong correlations. NDVI and NDWI have a correlation coefficient of −0.9479, meaning that the green spaces will decrease if there is an increase in water areas. NDVI and NDBI have a correlation coefficient of −0.6627, indicating that increasing green places decreases building areas. The other variables are classified as having weak correlations. Table 8 shows the measurement of four variables: MAE, MSE, and RMSE. And R^2. The table displays the assessment criteria for five distinct ML models using various input features to anticipate the LST. The evaluation criteria encompass MAE, MSE, RMSE, and R^2. The RF model exhibits the most exceptional performance in predicting LST, with the lowest MAE of 3.63 and RMSE of 5.12. The R^2 0.41 indicates that the RF model can account for 41% of the variability in the LST data.

On the other hand, the ABR model has the poorest performance in predicting LST, with the highest MAE of 5.42, MSE of 42.4, and RMSE of 6.51. The R^2 0.05 indicates that the ABR model can only explain 5% of the variability in the LST data. The ANN and LR models demonstrate similar performance, with an MAE of 4.63 and 4.6, respectively, and an R^2 value of 0.15. The GBR model performs better than the ANN and LR models, with an MAE of 4.12 and an R^2 value of 0.3, but still not as good as the RF model.

Table 8. The Result of Machine Learning.

Model	**MAE**	**MSE**	**RMSE**	**R-squared**
RF	3.63	26.16	5.12	0.41
ABR	5.42	42.4	6.51	0.05
ANN	4.63	38.11	6.17	0.15
LR	4.6	37.88	6.15	0.15
GBR	4.12	31.17	5.58	0.3

5. Conclusion

Urbanization is a worldwide phenomenon swiftly transforming the physical, social, and economic well-being of cities across the globe. An important consequence of urbanization on the environment is the formation of UHI, caused by the reduction of green space and the construction and construction of infrastructure. This results in the buildup of heat in urban regions. This event may entail significant repercussions. The inclusion of increased energy consumption leads to elevated levels of air pollution, resulting in a decline in residents' overall quality of life.

LST is a significant environmental concern that can provide valuable insights into UHI energy usage and global warming. The RS satellite data are renowned in studies of LST due to their extensive coverage, superior spatial resolution, and ability to track changes over long periods of time. Nevertheless,

examining and manipulating substantial quantities of satellite data can be time-consuming and demanding. As a result, machine learning algorithms have emerged as a useful tool for analyzing and making predictions about LST based on available data. A satellite is an artificial object placed in orbit around a celestial body, such as the Earth, for various purposes, including communication, navigation, and scientific research.

This study's objective was to predict LST in urban areas in Chiang Mai, Thailand, using ML models with Sentinel-2 and Landsat 8 data from 2016 to 2022. The data was processed in QGIS, and five diverse models, including RF, ABR, ANN, LR, and GB, were tested.

This research showed that the RF model had the highest accuracy in predicting LST, with the lowest MAE and RMSE values among the models. However, all models had relatively low R^2 values, indicating that there is still room for improvement in the accuracy of the predictions. The findings suggest that further research is essential to refine and improve the accuracy of the ML models for LST prediction. The study's results revealed the capability of using ML to predict LST in urban areas and revealed that RF has the highest capability among the tested models.

Another benefit of investigating the UHI phenomenon is that it can identify areas with high concentrations of UHI. This dataset can be used by urban planners in Chiang Mai Province to determine policies to reduce heat or increase green space in certain areas. This predictive model can help urban planners design and plan better proactive measures.

Despite the limitations of the current investigation, it provides valuable insights into the potential of ML in forecasting LST. It highlights the need for continued research to enhance the precision of these models. The application of ML algorithms to evaluate and anticipate LST from satellite data can significantly contribute to understanding UHI, energy consumption, and climate change. With the rapid pace of urbanization worldwide, this technology is poised to become an increasingly crucial tool in environmental science.

In conclusion, this study demonstrates the viability of utilizing RS data and ML algorithms to predict LST and comprehend UHI development in urban areas. It also underscores the significance of employing machine learning techniques to address the consequences of UHI in urban regions. Further examination is warranted to refine the models' accuracy and determine effective measures to mitigate UHI's effects in urban areas, including developing a UHI Classification to help identify areas more specifically. These findings can assist policymakers and urban planners in Chiang Mai and other cities worldwide in devising effective policies and strategies for reducing the adverse impacts of UHI and promoting sustainable urban development.

Author Contributions

N. Chanpichaigosol conducted the research, designed the methodology, and wrote the original manuscript. C. Chaichana and C. Y. Chu provided technical support and contributed to the data analysis. M. Srivanit, D. Rinchumphu, T. Y. Chou, and M. L. Yeh served as consultants, offering guidance and advice on various aspects of the research project. All authors reviewed and approved the final version of the manuscript.

Acknowledgment

The authors would like to thank the Ministry of Education in Taiwan Experience Education Programs (TEEP) and the Institute of Green Products, Feng Chia University, for their financial support. This research project was supported by Chiang Mai University.

References

Alademomi, A.S., Okolie, C.J., Daramola, O.E., Agboola, R.O. and Salami, T.J. 2020. Assessing the Relationship of LST, NDVI and EVI with Land Cover Changes in the Lagos Lagoon Environment. *Quaestiones Geographicae, 39*(3): 87-109. DOI: https://doi.org/10.2478/quageo-2020-0025

AlDousari, A.E., Kafy, A.A., Saha, M., Fattah, M.A., Almulhim, A.I., Faisal, A.A., Al Rakib, A., Jahir, D.M., A., Rahaman, Z.A., Bakshi, A., Shahrier, M. and Rahman, M.M. 2022.Modelling the impacts of land use/land cover changing pattern on uban thermal characteristics in Kuwait. *Sustainable Cities and Society, 86*: 104107. DOI: https://doi.org/10.1016/j.scs.2022.104107

Allen, R., Irmak, A., Trezza, R., Hendrickx, J.M., Bastiaanssen, W. and Kjaersgaard, J. 2011. Satellite-based ET estimation in agriculture using SEBAL and METRIC. *Hydrological Processes, 25*(26): 4011-4027. DOI: https://doi.org/10.1002/hyp.8408

Almeida, C., Rd, Teodoro, AC and Gonçalves, A. 2021. Study of the Urban Heat Island (UHI) Using Remote Sensing Data/Techniques: A Systematic Review. *Environments, 8*(10): 105. DOI: https://doi.org/10.3390/environments8100105

Arifwidodo, S. and Chandrasiri, O. 2015. Urban heat island and household energy consumption in Bangkok, Thailand. *Energy Procedia, 79*: 189-194. DOI: https://doi.org/10.1016/j.egypro.2015.11.461

Avdan, U. and Jovanovska, G. 2016. Algorithm for automated mapping of land surface temperature using LANDSAT 8 satellite data. *Journal of sensors, 2016*. 1-8. DOI: https://doi.org/10.1155/2016/1480307

Bartesaghi-Koc, C., Osmond, P and Peters, A. 2022. Innovative use of spatial regression models to predict the effects of green infrastructure on land surface temperatures. *Energy and Buildings,* 254. 111564. DOI: https://doi.org/10.1016/j.enbuild.2021.111564

Boudali E.F., Strebel, D., Carmeliet, J. and Derome, D. 2022. Impact of urban heat island on cooling energy demand for residential building in Montreal using meteorological simulations and weather station observations. *Energy and Buildings,* 273. 112410. DOI: https://doi.org/https://doi.org/10.1016/j.enbuild.2022.112410

Breiman, L. 2001. Random Forests. *Machine Learning, 45*(1): 5-32. DOI: https://doi.org/10.1023/A:1010933404324

Chen, T. and Guestrin, C. 2016. XGBoost: A Scalable Tree Boosting System Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA. https://doi.org/10.1145/2939672.2939785

Cutler, D.R., Edwards Jr, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J. and Lawler, J.J. 2007. Random forests for classification in ecology. *Ecology, 88*(11): 2783-2792. DOI: https://doi.org/10.1890/07-0539.1

Dhillon, M.S., Dahms, T., Kübert-Flock, C., Steffan-Dewenter, I., Zhang, J. and Ullmann, T. 2022. Spatiotemporal fusion modelling using STARFM: Examples of Landsat 8 and Sentinel-2 NDVI in Bavaria. *Remote Sensing, 14*(3): 677. DOI: https://doi.org/10.3390/rs14030677

Fan, Y., Wang, Z., Li, Y., Wang, K., Sun, Z., and Ge, J. 2022. Urban heat island reduces annual building energy consumption and temperature related mortality in severe cold region of China. *Urban Climate, 45*: 101262. DOI: https://doi.org/10.1016/j.uclim.2022.101262

Farrokhzadeh, B., Mansouri, S., and Sepehri, A. 2018. Determining the correlation between NDVI and EVI vegetation indices and SPI drought index (Case Study: Golestan rangelands). *Journal of Agricultural Meteorology, 5*(2): 56-65. DOI: https://doi.org/10.22125/agmj.2018.59724

Friedman, J. H. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 29(5): 1189-1232. DOI: https://doi.org/10.1214/aos/1013203451

Gao, J., Gong, J., Yang, J., Li, J., and Li, S. 2022. Measuring Spatial Connectivity between patches of the heat source and sink (SCSS): A new index to quantify the heterogeneity impacts of landscape patterns on land surface temperature. *Landscape and Urban Planning, 217*: 104260. DOI: https://doi.org/10.1016/j.landurbplan.2021.104260

Geng, X., Zhang, D., Li, C., Yuan, Y., Yu, Z., and Wang, X. 2023. Impacts of climatic zones on urban heat island: Spatiotemporal variations, trends, and drivers in China from 2001–2020. *Sustainable Cities and Society, 89*: 104303. DOI: https://doi.org/10.1016/j.scs.2022.104303

Goodfellow, I., Bengio, Y. and Courville, A. 2016. *Deep learning*. MIT press.

Hashim, B.M., Sultan, M.A., Attyia, M.N., Al Maliki, A.A. and Al-Ansari, N. 2019. Change detection and impact of climate changes to Iraqi southern marshes using Landsat 2 Mss, Landsat 8 Oli and sentinel 2 Msi data and Gis applications. *Applied Sciences,* 9(10): 2016. DOI: https://doi.org/10.3390/app9102016

Imhoff, M.L., Zhang, P., Wolfe, R.E and Bounoua, L. 2010. Remote sensing of the urban heat island effect across biomes in the continental USA. *Remote sensing of environment,* 114(3): 504-513. DOI: https://doi.org/10.1016/j.rse.2009.10.008

Jamali, A. 2019. Evaluation and comparison of eight machine learning models in land use/land cover mapping using Landsat 8 OLI: a case study of the northern region of Iran. *SN Applied Sciences,* 1(11): 1-11. DOI: https://doi.org/10.1007/s42452-019-1527-8

Jiang, J. and Tian, G. 2010. Analysis of the impact of land use/land cover change on land surface temperature with remote sensing. *Procedia environmental sciences,* 2: 571-575. DOI: https://doi.org/10.1016/j.proenv.2010.10.062

Jongtanom, Y., Kositanont, C. and Baulert, S. 2011. Temporal variations of urban heat island intensity in three major cities, Thailand. Modern Applied Science, 5(5): 105. DOI: https://doi.org/10.5539/mas.v5n5p105

Kafy, A.A., Abdullah A.l.F., Rahman, M.S., Islam, M., Al Rakib, A., Islam, M.A., Khan, M.H.H., Sikdar, M.S., Sarker, M.H.S., Mawa, J. and Sattar, G.S. 2021. Prediction of seasonal urban thermal field variance index using machine learning algorithms in Cumilla, Bangladesh. *Sustainable Cities and Society,* 64: 102542. DOI: https://doi.org/10.1016/j.scs.2020.102542

Kafy, A.A., Saha, M., Faisal, A.A., Rahaman, Z.A., Rahman, M.T. Liu, D., Fattah, M.A., Al Rakib, A., AlDousari, A.E., Rahaman, S.N., Hasan, M.Z., and Ahasan, M.A.K. 2022. Predicting the impacts of land use/land cover changes on seasonal urban thermal characteristics using machine learning algorithms. *Building and Environment,* 217: 109066. DOI: https://doi.org/10.1016/j.buildenv.2022.109066

Karimi, N., Ng, K.T.W., and Richter, A. 2021. Prediction of fugitive landfill gas hotspots using a random forest algorithm and Sentinel-2 data. *Sustainable Cities and Society,* 73: 103097. DOI: https://doi.org/10.1016/j.scs.2021.103097

Kasanko, M., Barredo, J.I., Lavalle, C., McCormick, N., Demicheli, L., Sagris, V. and Brezger, A. 2006. Are European cities becoming dispersed?: A comparative analysis of 15 European urban areas. *Landscape and Urban Planning,* 77(1), 111-130. DOI: https://doi.org/10.1016/j.landurbplan.2005.02.003

Khan, S. and Aaqib, S.M. 2017. Empirical evaluation of ArcGIS with contemporary open source solutions-A study. *International Journal of Advance Research in Science and Engineering,* 6(1): 724-736. https://www.ijarse.com/images/fullpdf/1511159922_330_IJARSE.pdf

Khan, S. and Mohiuddin, K. 2018. Evaluating the parameters of ArcGIS and QGIS for GIS Applications. *International Journal of Advance Research in Science and Engineering* , 7(3): 582-594. https://www.ijarse.com/images/fullpdf/1562935336_RIMT346ijarse.pdf

Li, Z., Li, X., Wei, D., Xu, X., & Wang, H. 2010. An assessment of correlation on MODIS-NDVI and EVI with natural vegetation coverage in Northern Hebei Province, China. *Energy Policy - ENERG POLICY,* 2: 964-969. DOI: https://doi.org/10.1016/j.proenv.2010.10.108

Liu, M., Zhu, Y., Yang, H., Pu, R., Qiu, C., Zhao, F., Han, S., Xu, W., Meng, Y., Long, H. and Yang, G. 2022. Prediction of apple first flowering date using daily land surface temperature spatio-temporal reconstruction and machine learning. *Computers and Electronics in Agriculture,* 202: 107366. DOI: https://doi.org/10.1016/j.compag.2022.107366

Ma, Y., Zuo, L., Gao, J., Liu, Q. and Liu, L.. 2021. Comparing Four Types Methods for Karst NDVI Prediction Based on Machine Learning. *Atmosphere,* 12(10): 1341. DOI: https://doi.org/10.3390/atmos12101341

Magistrali, I.C., Delgado, R.C., dos Santos, G.L., Pereira, M.G., de Oliveira, E.C., de O. Neves, L., de Souza, L.P., Teodoro, P.E., and Silva Junior, C.A. 2021. Performance of CCCma and GFDL climate models using remote sensing and surface data for the state of Rio de Janeiro-Brazil. *Remote Sensing Applications: Society and Environment,* 21: 100446. DOI: https://doi.org/10.1016/j.rsase.2020.100446

Mitraka, Z., Chrysoulakis, N., Doxani, G., Del Frate, F. and Berger, M. 2015. Urban surface temperature time series estimation at the local scale by spatial-spectral unmixing of satellite observations. *Remote Sensing,* 7(4), 4139-4156. DOI: https://doi.org/10.3390/rs70404139

Mohammad, P. and Goswami, A. 2022. Predicting the impacts of urban development on seasonal urban thermal environment in Guwahati city, northeast India. *Building and Environment*, 226: 109724. DOI: https://doi.org/10.1016/j.buildenv.2022.109724

Mohammad, P., Goswami, A., Chauhan, S. and Nayak, S. 2022. Machine learning algorithm based prediction of land use land cover and land surface temperature changes to characterize the surface urban heat island phenomena over Ahmedabad city, India. *Urban Climate,* 42: 101116. DOI: https://doi.org/10.1016/j.uclim.2022.101116

Montgomery, D.C., Peck, E. A. and Vining, G.G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.

Mushore, T.D., Mutanga, O. and Odindi, J. 2022. Estimating urban LST using multiple remotely sensed spectral indices and elevation retrievals. *Sustainable Cities and Society*, 78: 103623. DOI: https://doi.org/10.1016/j.scs.2021.103623

Mushore, T.D., Odindi, J., Dube, T. and Mutanga, O. 2017. Prediction of future urban surface temperatures using medium resolution satellite data in Harare metropolitan city, Zimbabwe. *Building and Environment,* 122: 397-410. DOI: https://doi.org/10.1016/j.buildenv.2017.06.033

Nantarat, P., Charoentrakulpeeti, W. and Wattanapinyo, A. 2021. The Efficiency of Urban Heat Island Mitigation by Cooling Effects from Greenspace in Chiang Mai Municipality. *Journal of Architectural/Planning Research and Studies (JARS),* 18(1): 131-152. DOI: https://doi.org/10.56261/jars.v18i1.241605

Nayak, M.A. and Ghosh, S. 2013. Prediction of extreme rainfall event using weather pattern recognition and support vector machine classifier. *Theoretical and applied climatology,* 114(3): 583-603. DOI: https://doi.org/10.1007/s00704-013-0867-3

Nguyen, T.T., Hoang, T.D., Pham, M.T., Vu, T.T., Nguyen, T.H., Huynh, Q.T. and Jo, J. 2020. Monitoring agriculture areas with satellite images and deep learning. *Applied Soft Computing,* 95: 106565. DOI: https://doi.org/10.1016/j.asoc.2020.106565

Peng, J., Jia, J., Liu, Y, Li, H. and Wu, J. 2018. Seasonal contrast of the dominant factors for spatial distribution of land surface temperature in urban areas. *Remote sensing of environment*, 215: 255-267. DOI: https://doi.org/10.1016/j.rse.2018.06.010

Sangnum, P., Kammuang-Lue, N., Sakulchangsatjatai, P. and Terdtoon, P., (2014). Urban Heat Island Intensity in Chiang Mai City Using Mobile Surveying Approach. *Advanced Materials Research,* 931-932" 605-613. DOI: https://doi.org/10.4028/www.scientific.net/AMR.931-932.605

Schwalbert, R.A., Amado, T., Corassa, G., Pott, LP, Prasad, PVV and Ciampitti, IA. 2020. Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. *Agricultural and Forest Meteorology,* 284: 107886. DOI: https://doi.org/10.1016/j.agrformet.2019.107886

Sherafati, S., Saradjian, M.R. and Niazmardi, S. 2013. Urban Heat Island Growth Modeling Using Artificial Neural Networks and Support Vector Regression: A case study of Tehran, Iran. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XL-1/W3: 399-403. DOI: https://doi.org/10.5194/isprsarchives-XL-1-W3-399-2013

Shih, H.c, Stow, D.A. and Tsai, Y.H. 2019. Guidance on and comparison of machine learning classifiers for Landsat-based land cover and land use mapping. *International Journal of Remote Sensing,* 40(4): 1248-1274. DOI: https://doi.org/10.1080/01431161.2018.1524179

Siddiqui, S.T., Alam, M.S. and Bokhari, M.U. 2012. Software Tools Required to Develop GIS Applications: An Overview. *Proceedings of the IEEE*, 51-56. DOI: https://doi.org/10.1109/ACCT.2012.106

Singh, S., Sood, V., Taloor, A.K., Prashar, S and Kaur, R. 2021. Qualitative and quantitative analysis of topographically derived CVA algorithms using MODIS and Landsat-8 data over Western Himalayas, India. *Quaternary International,* 575: 85-95. DOI: https://doi.org/10.1016/j.quaint.2020.04.048

Skelhorn, C.P., Levermore, G. and Lindley, S.J. 2016 Impacts on cooling energy consumption due to the UHI and vegetation changes in Manchester, UK. *Energy and Buildings,* 122: 150-159. DOI: https://doi.org/10.1016/j.enbuild.2016.01.035

Srivanit, M. and Auttarat, S. 2016. Classifying Spatial Homogeneity of Thermally-Stabilized Surface to Define Differentiation Local Atmospheric Zones in Chiang Mai City. *Journal of Architectural/Planning Research and Studies (JARS),* 13: 1-22. DOI: https://doi.org/10.56261/jars.v13i1.71583

Srivanit, M. and Hokao, K. 2012. Effects of urban development and spatial characteristics on urban thermal environment in Chiang Mai metropolitan, Thailand. *Lowland Technology International,* 14(2), 9-22. https://cot.unhas.ac.id/journals/index.php/ialt_lti/article/view/436/329

Srivanit, M. and Iamtrakul, P. 2019. Spatial patterns of greenspace cool islands and their relationship to cooling effectiveness in the tropical city of Chiang Mai, Thailand. *Environmental monitoring and assessment,* 191(9): 580. DOI: https://doi.org/10.1007/s10661-019-7749-9

Stewart, I.D. and Oke, T.R.. 2012. Local climate zones for urban temperature studies. *Bulletin of the American Meteorological Society,* 93(12), 1879-1900. DOI: https://doi.org/10.1175/BAMS-D-11-00019.1

Tomlinson, C.J., Chapman, L., Thornes, J.E. and Baker, C. 2011. Remote sensing land surface temperature for meteorology and climatology: a review. *Meteorological Applications,* 18(3), 296-306. DOI: https://doi.org/10.1002/met.287

Turner, I.L., Harley, M.D., Almar, R. and Bergsma, E.W.J. 2021. Satellite optical imagery in Coastal Engineering. *Coastal Engineering,* 167: 103919. DOI: https://doi.org/10.1016/j.coastaleng.2021.103919

Wang, Y., Berardi, U. and Akbari, H. 2016. Comparing the effects of urban heat island mitigation strategies for Toronto, Canada. *Energy and Buildings,* 114: 2-19. DOI: https://doi.org/10.1016/j.enbuild.2015.06.046

Xu, N., Deng, F., Liu, B., Li, C, Fu, H., Yang, H. and Zhang, J., 2021. Changes in the urban surface thermal environment of a Chinese coastal city revealed by downscaling MODIS LST with random forest algorithm. *Journal of Meteorological Research,* 35(5): 759-774. DOI: https://doi.org/10.1007/s13351-021-0023-4

Xu, Y., Ren, C., Cai, M., Edward, N.Y.Y. and Wu, T., 2017. Classification of Local Climate Zones Using ASTER and Landsat Data for High-Density Cities. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 10(7): 3397-3405. DOI: https://doi.org/10.1109/JSTARS.2017.2683484

Yu, Z., Di, L., Yang, R., Tang, J., Lin, L., Zhang, C., Rahman, M.S., Zhao, H., Gaigalas, J. and Yu, E.G. 2019 . Selection of landsat 8 OLI band combinations for land use and land cover classification. *2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics),* DOI: https://doi.org/10.1109/Agro-Geoinformatics.2019.8820595

Zhang, D., Zhang, C., Li, W., Cromley, R., Hanink, D., Civco, D. and Travis, D. 2014. Restoration of the missing pixel information caused by contrails in multispectral remotely sensed imagery. *Journal of Applied Remote Sensing,* 8: 083698. DOI: https://doi.org/10.1117/1.JRS.8.083698

Zhang, M., Zhang, F., Chen, D., Tan, M.L. and Chan, N.W. 2022. Urban local surface temperature prediction using the urban gray-green space landscape and vegetation indices. *Building and Environment,* 226: 109723. DOI: https://doi.org/10.1016/j.buildenv.2022.109723

Zhou, D., Xiao, J., Bonafoni, S., Berger, C., Deilami, K., Zhou, Y., Frolking, S., Yao, R., Qiao, Z and Sobrino, J.A. 2019. Satellite Remote Sensing of Surface Urban Heat Islands: Progress, Challenges, and Perspectives. *Remote Sensing,* 11(1): 48. DOI: https://doi.org/10.3390/rs11010048

hou, J., Chen, Y., Wang, J. and Zhan, W. 2010. Maximum nighttime urban heat island (UHI) intensity simulation by integrating remotely sensed data and meteorological observations. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 4(1), 138-146. DOI: https://doi.org/10.1109/JSTARS.2010.2070871

Zhou, W., Huang, G. and Cadenasso, M.L. 2011. Does spatial configuration matter? Understanding the effects of land cover pattern on land surface temperature in urban landscapes. *Landscape and Urban Planning, 102*(1), 54-63. DOI: https://doi.org/10.1016/j.landurbplan.2011.03.009