



## Article

# Modeling and Estimating the Land Surface Temperature (LST) Using Remote Sensing and Machine Learning (Case Study: Yazd, Iran)

Mohammad Mansourmoghaddam <sup>1</sup>, Iman Roustae <sup>2,3</sup>, Hamidreza Ghafarian Malamiri <sup>2</sup>,  
Mostafa Sadeghnejad <sup>4</sup>, Jaromir Krzyszcak <sup>5</sup> and Carla Sofia Santos Ferreira <sup>6,7,8,\*</sup>

- <sup>1</sup> Center for Remote Sensing and GIS Research, Shahid Beheshti University, Tehran 1983969411, Iran; m\_mansourmoghaddam@sbu.ac.ir
- <sup>2</sup> Department of Geography, Yazd University, Yazd 8915813135, Iran; irousta@yazd.ac.ir (I.R.); hrghafarian@yazd.ac.ir (H.G.M.)
- <sup>3</sup> Institute for Atmospheric Sciences-Weather and Climate, University of Iceland and Icelandic Meteorological Office (IMO), Bustadavegur 7, IS-108 Reykjavik, Iceland
- <sup>4</sup> Department of Geography and Geospatial Sciences, Kansas State University, 920 N17th Street, Manhattan, KS 66506-2904, USA; sadghnejad@ksu.edu
- <sup>5</sup> Institute of Agrophysics, Polish Academy of Sciences, Doświadczalna 4, 20-290 Lublin, Poland; j.krzyszcak@ipan.lublin.pl
- <sup>6</sup> Department of Physical Geography and Bolin Centre for Climate Research, Stockholm University, SE-10691 Stockholm, Sweden
- <sup>7</sup> Polytechnic Institute of Coimbra, Applied Research Institute, Rua da Misericórdia, Lagar dos Cortiços, S. Martinho do Bispo, 3045-093 Coimbra, Portugal
- <sup>8</sup> Research Centre for Natural Resources, Environment and Society (CERNAS), Polytechnic Institute of Coimbra, Bencanta, 3045-601 Coimbra, Portugal
- \* Correspondence: carla.ferreira@natgeo.su.se



**Citation:** Mansourmoghaddam, M.; Roustae, I.; Ghafarian Malamiri, H.; Sadeghnejad, M.; Krzyszcak, J.; Ferreira, C.S.S. Modeling and Estimating the Land Surface Temperature (LST) Using Remote Sensing and Machine Learning (Case Study: Yazd, Iran). *Remote Sens.* **2024**, *16*, 454. <https://doi.org/10.3390/rs16030454>

Academic Editors: Francesco Marchese, Nicola Genzano and Carolina Filizzola

Received: 27 November 2023

Revised: 18 January 2024

Accepted: 20 January 2024

Published: 24 January 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** The pressing issue of global warming is particularly evident in urban areas, where urban thermal islands amplify the warming effect. Understanding land surface temperature (LST) changes is crucial in mitigating and adapting to the effect of urban heat islands, and ultimately addressing the broader challenge of global warming. This study estimates LST in the city of Yazd, Iran, where field and high-resolution thermal image data are scarce. LST is assessed through surface parameters (indices) available from Landsat-8 satellite images for two contrasting seasons—winter and summer of 2019 and 2020, and then it is estimated for 2021. The LST is modeled using six machine learning algorithms implemented in R software (version 4.0.2). The accuracy of the models is measured using root mean square error (RMSE), mean absolute error (MAE), root mean square logarithmic error (RMSLE), and mean and standard deviation of the different performance indicators. The results show that the gradient boosting model (GBM) machine learning algorithm is the most accurate in estimating LST. The albedo and NDVI are the surface features with the greatest impact on LST for both the summer (with 80.3% and 11.27% of importance) and winter (with 72.74% and 17.21% of importance). The estimated LST for 2021 showed acceptable accuracy for both seasons. The GBM models for each of the seasons are useful for modeling and estimating the LST based on surface parameters using machine learning, and to support decision-making related to spatial variations in urban surface temperatures. The method developed can help to better understand the urban heat island effect and ultimately support mitigation strategies to improve human well-being and enhance resilience to climate change.

**Keywords:** land surface temperature modeling; land surface parameters; machine learning; gradient boosting method

## 1. Introduction

The planet's surface plays a crucial role in earth sciences, as it is influenced by factors such as energy input, surface discharge, humidity, and atmospheric air movements, which collectively determine the net energy. Understanding the surface energy budget and accurately calculating surface temperature is essential for various earth science studies, including those related to urban planning, water resource management, natural disasters, and climate change [1–4].

An increase in urban heat storage capacity creates heat islands, which make the urban areas warmer than the surrounding rural areas [5–7]. This local difference in temperature harms people and the environment as it hinders air quality, and increases energy consumption. It also causes a loss of biological control and affects people's health [8–11]. The temperature in the urban environment is influenced by a wide range of climatic, geographic, and human factors, reflecting local climate characteristics that differ from atmospheric temperature. As a result, recent research has increasingly focused on identifying these temperature changes in urban areas [12,13]. Land surface temperature (LST) is a key determinant of the energy balance in physical processes occurring on the earth's surface [14,15]. The study of LST, along with other factors like evapotranspiration and soil salinity, proves valuable in examining the effects of global warming on food security, energy consumption, carbon sinks [16,17], and finally, human well-being [18,19].

In general, two terms are used to describe urban landscapes. Landscape composition, which refers to the number (or proportion) of land use categories in a defined unit (such as a path, pixel, or region), while landscape arrangement refers to the spatial arrangement of those units [20]. It has already been shown that LST is directly connected to land use/land cover (LU/LC) [21]. It is understood that changes in land use and climate have contributed to the loss of global biodiversity [22]. For instance, Lin et al. [23] investigated the relationship between the morphological characteristics of built-up areas and stressed their influence on urban thermal environments and the intensity of surface urban heat islands (SUHIs). Therefore, to elucidate the effect of LU/LC on LST, it is vital to study the relationship between them, especially for various climatic zones.

The study of the spatial pattern of urban heat islands is critical in discovering the effects of dispersion and changes in the impact of LU/LC on LST. Furthermore, the subsequent LST simulations for the upcoming years, based on LU/LC, play an important role in predicting the effects of heat islands on the future city environment. Such modeling can lead to the adoption of new strategies and policies for controlling the LU/LC changes and support the design of urban areas in such a way that they could reduce the effects of urban heat islands [24,25]. Temperature regulation in cities is one of several ecosystem services provided by green spaces in urban areas. In this regard, two approaches are widely discussed, the first is based on the division of land (lower density and more dispersion), whereas the second assumes saving land (higher density and lesser dispersion) [26,27], but their impact on LST is not fully understood.

Extensive advances in thermal remote sensing, Geographic Information Systems (GIS), and statistical methods have enabled new possibilities for the scientific community for assessing the relationships between landscapes and the urban heat island effects [24]. To optimally use land resources, it is necessary to obtain information about LU/LC changes. As they mainly happen on a large scale, ground truth data collection is very time- and labor-consuming, thus costly, and sometimes even impossible to perform. Multi-spectral satellite imagery and remote sensing techniques can be used as important tools in studying these landscape changes [28], as well as the effects on vital factors, including water sources [29–32].

Recently, the application of remote sensing in urban areas has been mainly focused on aspects related to the LST and spatial patterns of urban heat islands and their relationship with surface parameters, balance, and energy fluxes in the urban surface [33,34], and the relationship between atmospheric temperature and LST [35]. Peng et al. [36] investigated the relationship between LST and topographic elements in Hangzhou, China. The study

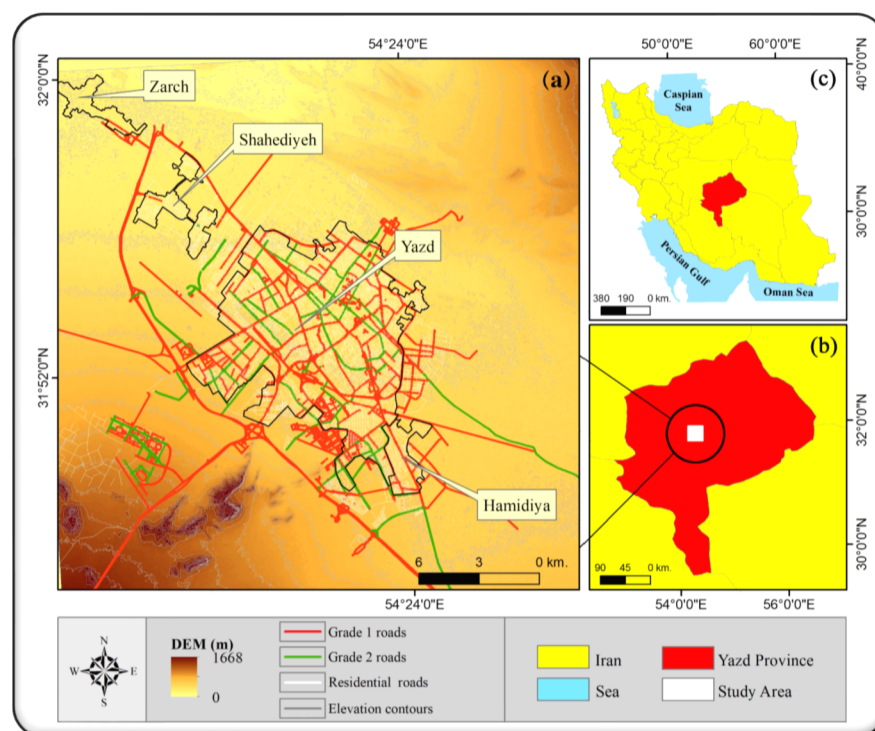
showed a negative correlation between LST, altitude, and slope. Although this study presented a weak relationship between LST and aspect, a positive and strong correlation between the shaded relief map and LST was found. Alavipanah et al. [37] performed a spatio-temporal analysis of the thermal islands in Mashhad, Iran, with a focus on urban development and LU/LC changes based on the data from Landsat-5 (TM), Landsat-7 (ETM+), and Landsat-8 (OLI and TIRS) images. They used the NDVI threshold method, Planck's law, and two separate window algorithms. The results indicated that the land use change from agricultural to urban areas is the main factor leading to an increase in LST and the creation of urban heat islands.

In recent years, machine learning is increasingly being used in remote sensing studies. Machine learning (ML) is a branch of artificial intelligence and computer science that emphasizes the use of data and algorithms to learn and gradually improve the performance and accuracy of results. It is an important part of the growing field of data science [38]. First, cellular automata (CA), Markov chain (MC), multiple-layer perceptron (MLP), and artificial neural networks (ANN) are used to predict LU/LC changes by identifying patterns from historical land use data. Conventional machine learning (ML) methods, on the other hand, frequently use linear regression analysis to predict future LSTs [39,40]. The most popular linear models that, in some cases, can produce respectable results in LST prediction are the multiple linear regression (MLR) and ordinary least squares (OLS) models [41]. The combination of remote sensing and machine learning can lead to the use of more advanced and accurate mathematical algorithms in recognizing and solving environmental problems. Studies such as [20] have investigated the effect of urban landscape composition and configuration on LST using landscape segmentation indicators and machine learning, but the quantification and investigation of the effect of urban parameters on LST and its modeling with a thermal remote sensing-machine learning approach have received less attention. Therefore, this study aims to evaluate machine learning models to estimate the LST using the landscape's physical parameters and remote sensing indicators from Landsat-8 images for contrasting seasons (i.e., summer and winter). This study is developed in Yazd County, Iran, where data availability is scarce, including both field weather data and high-resolution thermal images. Machine learning algorithms were used to generate LST data as an output based on spectral bands and different physical variables by choosing an optimal model, training it carefully, and evaluating the accuracy of its output. Urban decision-makers and researchers can use the results of this research when there is no access to field or high-resolution thermal data or for the purpose of gap-filling thermal data for a better understanding of urban heat island dynamics and the future planning of the city's management and development. The estimated LST can be used for controlling soil moisture, the city's thermal islands, and evapotranspiration rates [42].

## 2. Materials and Methods

### 2.1. Study Area

The study area is Yazd County, located in central Iran, with a total area of 676.5 km<sup>2</sup>. This county includes the cities of Yazd, Hamidia, Shahedieh, and Zarch (Figure 1). The average altitude of the county is 1128.7 m, and the altitude of Yazd's city ranges from 1199 m to 1276 m, increasing from the north and northeast directions to the south and southwest. The average slope of this city is 2.1%, which covers 95.4% of the city in the north-to-south direction [43]. Yazd County has a hot and dry climate, with July, June, and August being the hottest months (average summer temperature ranging from 23 °C to 40 °C), and December, January, and February being the coldest (average winter temperature ranging from 0 °C to 17 °C). With an annual average temperature of 22 °C and an annual average precipitation of 57.8 mm, Yazd is one of the driest cities in Iran [44].



**Figure 1.** The digital elevation model (DEM) and elevation contours of the study area (Yazd County) with cities outlined together with intercity (grade 1) and urban (grade 2) road map (a), the location of Yazd County within the Yazd province (b), and the Yazd province location in Iran (c).

## 2.2. Methodology

### 2.2.1. Input Data: Satellite Images

In this study, 30 m multi-spectral and 100 m thermal bands of Landsat-8 images (including OLI and TIRS sensors) were downloaded from the USGS website of the United States of America ([www.earthexplorer.usgs.gov](http://www.earthexplorer.usgs.gov)) to conduct the research. As this study aims to model the parameters affecting the LST and to estimate it using machine learning, images from the coldest (February) and the hottest (August) months in Yazd regarding 2019 and 2020 were used [44]. The characteristics of the images used in this research are presented in Table 1.

**Table 1.** Information about the remote sensing images used in the research. Note that the overall image includes a larger area than investigated in this study.

| Satellite/Sensor | Spatial Resolution (m) | Date/Hour (GMT)        | Image ID                                 | Cloud Cover (%) |
|------------------|------------------------|------------------------|------------------------------------------|-----------------|
| Landsat-8/OLI    | 30                     | 2019-02-07<br>06:56:42 | LC08_L1TP_162038_20190207_20200829_02_T1 | 31.0            |
|                  |                        | 2019-08-18<br>06:57:05 | LC08_L1TP_162038_20190818_20200827_02_T1 | 0.0             |
|                  |                        | 2020-02-12<br>06:56:59 | LC08_L1TP_162038_20200210_20200823_02_T1 | 36.4            |
|                  |                        | 2020-08-18<br>06:56:59 | LC08_L1TP_162038_20200820_20200905_02_T1 | 27.0            |
|                  |                        | 2021-02-12<br>06:57:01 | LC08_L1TP_162038_20210212_20210302_02_T1 | 19.0            |
|                  |                        | 2021-08-23<br>06:57:07 | LC08_L1TP_162038_20210823_20210831_02_T1 | 22.4            |



The multi-spectral images were prepared for performing subsequent calculations and extracting spectral indices, after geometric and atmospheric corrections. For the atmospheric correction, the fast line-of-sight atmospheric analysis of hypercubes (FLAASH) algorithm [45–48] was used. Also, the atmospheric errors in the thermal bands were reduced using the thermal atmospheric correction algorithm [49,50].

### 2.2.2. Disaggregation of Radiometric Surface Temperature (DisTrad)

The DisTrad technique is used in remote sensing to combine thermal infrared (TIR) images with higher spatial resolution visible and near-infrared (VNIR) images. DisTrad aims to enhance the interpretability and accuracy of TIR imagery by fusing it with VNIR imagery. The original DisTrad approach is based on the correlation between LST and the normalized difference vegetation index (NDVI). Details about the DisTrad model are provided in [51]. In order to match the pixel size of thermal data with multi-spectral data and to increase the accuracy of subsequent calculations, the thermal data were downsampled in our research. The DisTrad method has shown high performance and accuracy in downscaling the thermal bands in several studies [51–55]. In this study, NDVI was combined with the normalized difference built-up index (NDBI) for downscaling the thermal bands, given the improvement in performance accuracy for the study area [51]. To validate the downsampled images, the results of the DisTrad method were resampled for comparison with the original unfused images that have also been used, for example in [56], in other research. The performance of the DisTrad method was assessed with the root mean square error (RMSE).

### 2.2.3. Land Surface Parameters

#### Calculation of Land Surface Temperature (LST)

After applying the necessary corrections and pre-processing (Section 2.2.1), the data were used to calculate the parameters (indices) of the land surface. LST, as one of the most important parameters, was calculated from the brightness temperature using emissivity correction [57–64], following Equation (1):

$$LST = \left[ \frac{\tau}{1 + w \left( \frac{\tau}{p} \right) \ln(\epsilon)} \right] \quad (1)$$

where  $\tau$  is at-sensor brightness temperature;  $w$  is the wavelength of emitted radiance (10.8  $\mu\text{m}$  Landsat-8 TIRS 10th band);  $p = h \times c / s (1.438 \times 10^{-2} \text{ m} \cdot \text{K})$ , with  $h$  being the Planck's constant ( $6.626 \times 10^{-34} \text{ J} \cdot \text{s}$ );  $s$  is the Boltzmann Constant ( $1.38 \times 10^{-23} \text{ J/K}$ );  $c$  is the velocity of light ( $2.988 \times 10^8 \text{ m/s}$ ); and  $\epsilon$  is the land surface emissivity.

The temperature value at the sensor (brightness) was extracted using Equation (2) [57–64]:

$$\tau = \left[ \frac{K_2}{\ln \left( \frac{K_1}{L_\varphi} + 1 \right)} \right] \quad (2)$$

where  $K_1$  and  $K_2$  are the thermal conversion constants taken from Landsat-8 Thermal Infrared Sensor (TIRS) metadata of the 10th band (Table 2). In the above equation,  $L_\varphi$  is spectral radiance values.

**Table 2.** Landsat thermal band conversion constants.

| Sensor | Band | $K_1$ [ $\text{W}/(\text{m}^2 \cdot \text{sr} \cdot \mu\text{m})$ ] | $K_2$ [K] |
|--------|------|---------------------------------------------------------------------|-----------|
| TIRS   | 10   | 774.8                                                               | 1321      |

To calculate at-sensor brightness temperature ( $\tau$ ) from the thermal band, raw data were transformed into spectral radiance values using Equation (3) [65]:

$$L_\varphi = M_L \times Q_{Cal} + A_L \quad (3)$$

where  $L_\varphi$  is the top of atmosphere (TOA) spectral radiance [ $W/(m^2 \cdot sr \cdot \mu m)$ ];  $M_L$  is a multiplicative rescaling factor dependent on the metadata for a particular band;  $Q_{Cal}$  is the quantized and calibrated standard product's pixel values (digital number); and  $A_L$  is the additive rescaling factor dependent on the metadata for a particular band.

The land surface emissivity ( $\epsilon$ ) was calculated using Equation (4) [57–61,63,64]:

$$\epsilon = n P_v + m \quad (4)$$

where  $n = 0.004$  and  $m = 0.986$ ; and  $P_v$  denotes the vegetation proportion, also referred to as fractional vegetation cover. The vegetation proportion ( $P_v$ ) was calculated following Equation (5) [57–64]:

$$P_v = \left[ \frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \right]^2 \quad (5)$$

where  $NDVI_{min}$  and  $NDVI_{max}$  are the minimal and the maximal values of the NDVI (calculated according to Equation (7)).

#### Calculation of Urban Land Surface Features

Before LST modeling, it was necessary to select surface features for machine learning model training that would allow the determination of the optimal model, and check the effect and importance of each of these features on the LST from the sensitivity analysis approach. The following surface geographical features of Yazd city were selected, based on a literature review of the relevant ones, calculated and used in the modeling:

(a) Albedo to provide information about the surface's reflective properties, which can affect the absorption and re-emission of heat. High albedo surfaces reflect more solar radiation and tend to have lower LST values [66]. Albedo was calculated for each pixel of the image.

(b) The NDVI index to assess the cooling effect of vegetation on the surface. Higher NDVI values indicate more vegetation cover, which can lead to lower LST values due to evapotranspiration and shading [67].

(c) The NDBI index to provide information about the spatial distribution of urban areas, which tend to have higher LST values due to the urban heat island effect [68].

(d) The normalized difference bareness index (NDBaI) to provide information about the presence of bare soil or bare land surfaces, which can have higher LST values due to their lower heat capacity compared to vegetated areas.

(e) Elevation, given its effect on the air temperature lapse rate, which can influence LST. Higher elevation areas tend to have lower LST values due to the decrease in atmospheric pressure and temperature with altitude [69]. The elevation of each pixel was taken from DEM.

(f) Distance from water bodies (for each pixel) to account for the influence of water on surface temperature. Areas closer to water bodies tend to have lower LST values due to evaporative cooling [70].

(g) Distance from mountains (for each pixel) to account for the potential shading and cooling effects of nearby mountainous terrain [71].

(h) Distance from grade 1 (highways and intercity roads) and 2 (boulevards and main streets in the city) roads, extracted from OpenStreetMap road network data [72] for each pixel to account for the potential influence of nearby roads on surface temperature.

Each one of the eight surface features was calculated as follows.

### Albedo index

Several methods created by Liang [66] allow for the determination of albedo from various satellite sensors. In this research, the normalized version of Smith's (2010) Landsat method was used to calculate the Landsat shortwave albedo [73] (Equation (6)):

$$Albedo = \left( \frac{0.356b_2 + 0.130b_4 + 0.373b_5 + 0.085b_6 + 0.072 + b_7 - 0.0018}{1.016} \right) \quad (6)$$

where  $b_x$  represents Landsat-8 bands 2 (Blue), 4 (Red), 5 (NIR), 6 (SWIR1), and 7 (SWIR2).

### Normalized Difference Vegetation Index (NDVI)

NDVI is the (normalized) ratio between the Landsat-8 red ( $R$ ) and near-infrared ( $NIR$ ) bands (Equation (7)) [59,74]:

$$NDVI = \left( \frac{NIR - R}{NIR + R} \right) \quad (7)$$

Negative values of NDVI ( $-1$  to  $\sim 0$ ) indicate water, ice, and snow. The values close to zero indicate rock, sand, and soil, whereas small positive values ( $\sim 0$  to  $0.2$ ) indicate vegetation, and large positive values ( $\sim 0.2$  to  $1$ ) indicate dense vegetation [75,76].

### Normalized Difference Built-up Index (NDBI)

The NDBI is a build-up index created to decrease the reflectance of the NIR band in order to identify vegetation and moist surroundings, while at the same time optimizing the reflectance of the SWIR band to detect built-up land [77]. The NDBI is calculated as:

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

where SWIR1 represents short-wavelength infrared band-1 and NIR represents the near-infrared band.

### Normalized Difference Bareness Index (NDBaI)

NDBaI is an index to identify bareness and is calculated following Equation (9). Values  $> -0.150$  indicate bare land [78]:

$$NDBaI = \left( \frac{SWIR1 - TIR1}{SWIR1 + TIR1} \right) \quad (9)$$

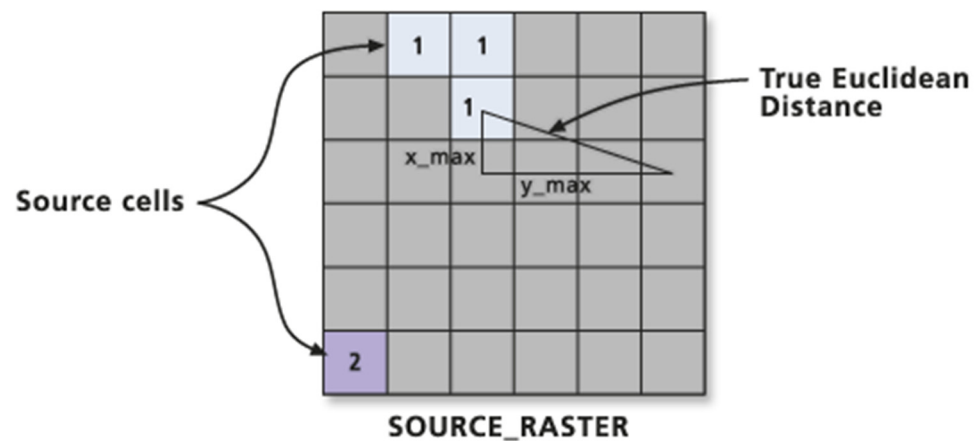
where TIR1 is the thermal infrared (10th) band of Landsat-8.

### Elevation by Digital Elevation Map (DEM)

To obtain the elevation of each pixel as one of the input parameters to the model, a 30 m ASTER GLOBAL DEM image of the study area, derived from the USGS website, was used. The image was first stacked to definitely be resampled to the pixel size of the spectral bands and finally used as one of the input parameters.

### Distances

The distance to (1) heights (mountains), (2) primary and secondary roads, and (3) water bodies was calculated using the Euclidean distance of each pixel from the studied area to these features. To be more precise, the Euclidean distance between each cell's center and the source cell's center was determined based on true Euclidean distance.  $X_{\max}$  and  $Y_{\max}$  serve as the other two sides of the triangle, and the hypotenuse is calculated for each cell to establish its distance to each source cell (Figure 2). This is how the Euclidean algorithm conceptually operates. Instead of the cell distance, the true Euclidean distance is determined via this computation. When determining the shortest distance to a source, if it is less than the maximum distance allowed, the value is set to the cell location on the output image [79].



**Figure 2.** Method used for determination of the true Euclidean distance (adapted from [79]).

To calculate the distance from water bodies, first, the water bodies in the study area were extracted for each image using the modified normalized difference water index (MNDWI). MNDWI uses the SWIR band to replace the green band in the normalized difference water index (NDWI) (Equation (10)). This allows it to restrict both vegetation and impervious surface objects, while, at the same time, revealing subtle water characteristics like suspended silt and pollution problems [80]:

$$MNDWI = \left( \frac{G - SWIR1}{G + SWIR1} \right) \quad (10)$$

where  $G$  is green (third) band of Landsat-8.

#### SHapley Additive exPlanation (SHAP)

The impact of the explanatory variables was assessed using the Shapley value approach [81]. This approach divides the prediction variability among all possible covariates [82]. In this manner, in a model-agnostic way, irrespective of the underlying model, the impact of all explanatory variables on all point predictions can be evaluated [83]. The computational viewpoint of the SHapley Additive exPlanation (SHAP) framework displays the Shapley values, which represent model predictions as linear combinations of binary variables that indicate the presence or absence of each covariate in the model. By performing this, the computational time constraints are evaded that come with kernel-based SHAP estimation [84]. In this study, the contribution of each feature for each instance is displayed on the SHAP summary plot. The sum of the feature contributions and the bias term is equal to the raw prediction of the model, i.e., prediction before applying the inverse link function [85].

#### 2.2.4. Optimal Model Selection

To select the optimal model, the automated machine learning (AutoML) function was used in the R programming environment. AutoML has made it simple for non-experts in the field of machine learning science to be hands-on with machine learning, even though high-performing learning models still require some data science expertise. In particular, deep neural networks are notoriously hard for a non-expert to adjust correctly. AutoML machine learning tool has been built with an intuitive user interface that streamlines the process of training a huge number of candidate models, making it really accessible to non-experts [86]. AutoML includes automatic training and tuning of many models within a user-specified time-limit and selects the appropriate model using cross-validation. This allowed testing of machine learning models to estimate LST. In this study, six machine learning models were selected based on their accuracy in image processing tasks; based on a literature review and limitations of existing algorithms in the H2O model-building phase.

Since the search-stopping criteria were set to be allowed, all appropriate H2O algorithms were allowed to participate in the modeling process:

(1) Random forest (RF) which is commonly used in image processing tasks such as image classification and object detection. RF can handle high-dimensional data and provides accurate results by combining multiple decision trees [87].

(2) Extremely randomized trees (XRT) which is similar to RF but with a higher level of randomness. This is also used in image processing tasks such as image classification and feature selection. XRT can provide robust performance and handle noisy or incomplete image data [88].

(3) Gradient boosting machine (GBM) is widely used in image processing for tasks like image segmentation and object recognition. It can handle large datasets and provide high accuracy [89].

(4) Generalized linear model (GLM) is a flexible statistical model that can handle various types of data distribution. In image processing, GLM is often used for tasks such as image denoising or image reconstruction. GLM can effectively model the relationship between input and output variables in image data [90].

(5) StackedEnsemble that combines multiple base models to improve predictive performance. It is commonly used in image processing for tasks like image recognition and object detection. StackedEnsemble can leverage the strengths of different models and provide enhanced accuracy [91].

(6) Artificial neural networks (ANNs) are deep learning models widely used for tasks like image classification, object detection, and image generation. ANNs consist of multiple layers of interconnected neurons that can learn complex patterns from image data. They can automatically extract features and achieve state-of-the-art performance in various image processing applications [92].

These six machine learning models were also used/tested to estimate LST in different tunings, i.e., the experimental process of finding the optimal values of hyperparameters to maximize model performance. For optimal model selection, aggregated data from two years from each season (in order to increase data variance) were divided into training, testing, and validation datasets. The separation of the data into training and testing sets (ratio 85:15) was performed using the R package Split (version 4.0.2) [93] to obtain a trustworthy result after a few runs [94]. By using a probabilistic splitting approach, this method can perform well on large data as well as unordered [95] and unweighted [96] databases according to the computing of an approximate splitting vector by sampling the aforementioned datasets [95]. Each machine learning model was then applied to the validation dataset, and the most accurate model for predicting the new dataset was then chosen by comparing the model performance. Hyperparameter tweaking is used to determine which train and test datasets are best suited for each sort of machine learning model. The performance of the machine learning models was assessed using statistical metrics, namely root mean absolute error (RMSE), root mean square logarithmic error (RMSLE), and mean absolute error (MAE). Also, in order to evaluate the estimated LST map more accurately, the distance of each pixel from the estimated image to the 45° line was calculated from the actual values. To achieve this objective, the LST values were computed for every pixel row and utilized to ascertain the distance from the 45° polynomial line (DFPL) for each pixel using Equation (11). Using this equation to find the distance between two points, the length of the line segment that connects the two points is measured. The DFPL is commonly indicative of the residual or variance between the observed value and the estimated value derived from the polynomial regression model:

$$d = \sqrt{(x - x_0)^2 + (y - y_0)^2} \quad (11)$$

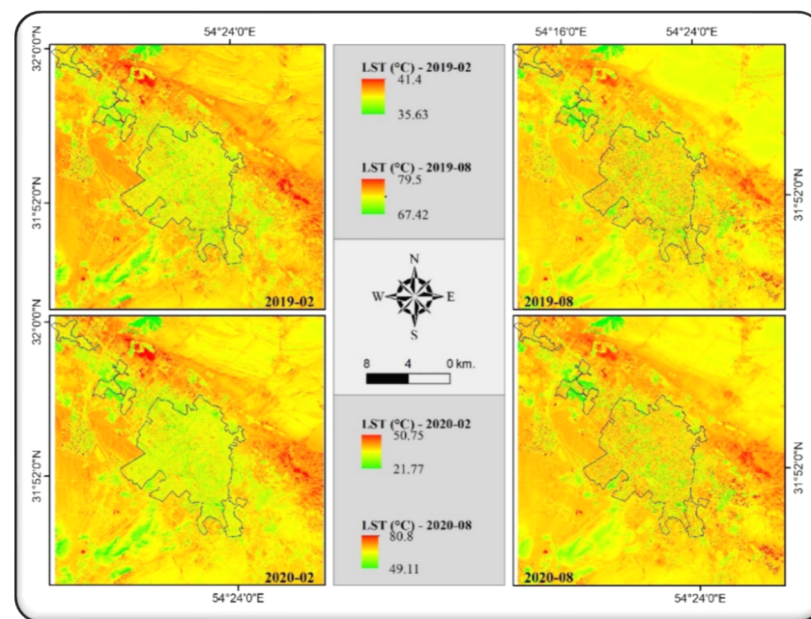
where,  $d$  is the distance of each pixel; and  $x$  and  $y$  are the coordinates of each pixel in the assumption that  $(x_0, y_0)$  and  $(x, y)$  are two pixels in a two-dimensional space.



### 3. Results

#### 3.1. Validation of LST Maps

Before calculating the LST maps, the remote sensing images were downsampled to the pixel size of the Landsat-8 spectral bands (30 m) using the DisTrad method. The LST maps were prepared for the two seasons (winter and summer) for both 2019 and 2020, and used as the basis for models' learning (Figure 3). In general, the LST results in the summer ranged from 79.5 °C to 67.42 °C in 2019, and from 80.8 °C to 49.11 °C in 2020. In winter, the maximum and minimum LST were 41.4 °C and 35.6 °C in 2019, and 50.75 °C and 21.77 °C in 2020, respectively (Figure 3). The low minimum values in 2020 were confirmed by a local unofficial weather station.



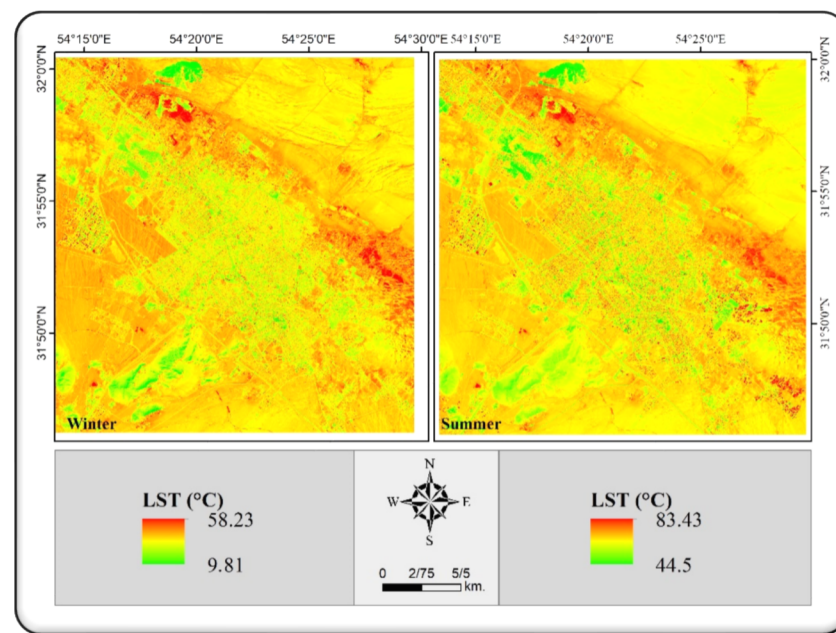
**Figure 3.** The LST maps for the two seasons (winter in the **left** panels and summer in the **right** panels) in 2019 (**top** panels) and 2020 (**bottom** panels).

The results of the performance metrics regarding the resampled downsampled images and the unfused images (see Section 2.2.2) are presented in Table 3. The results of evaluating the accuracy of LST 30 m range between 2.0 °C and 4.0 °C, with average values of 2.7 °C, which indicate that the method used provides acceptable results.

**Table 3.** The mean RMSE results of the comparison between downsampled and actual LST images used in the present study.

| LST Date   | RMSE (°C) |
|------------|-----------|
| 2019-02-07 | 2.08      |
| 2019-08-18 | 4.001     |
| 2020-02-12 | 2.97      |
| 2020-08-18 | 2.01      |

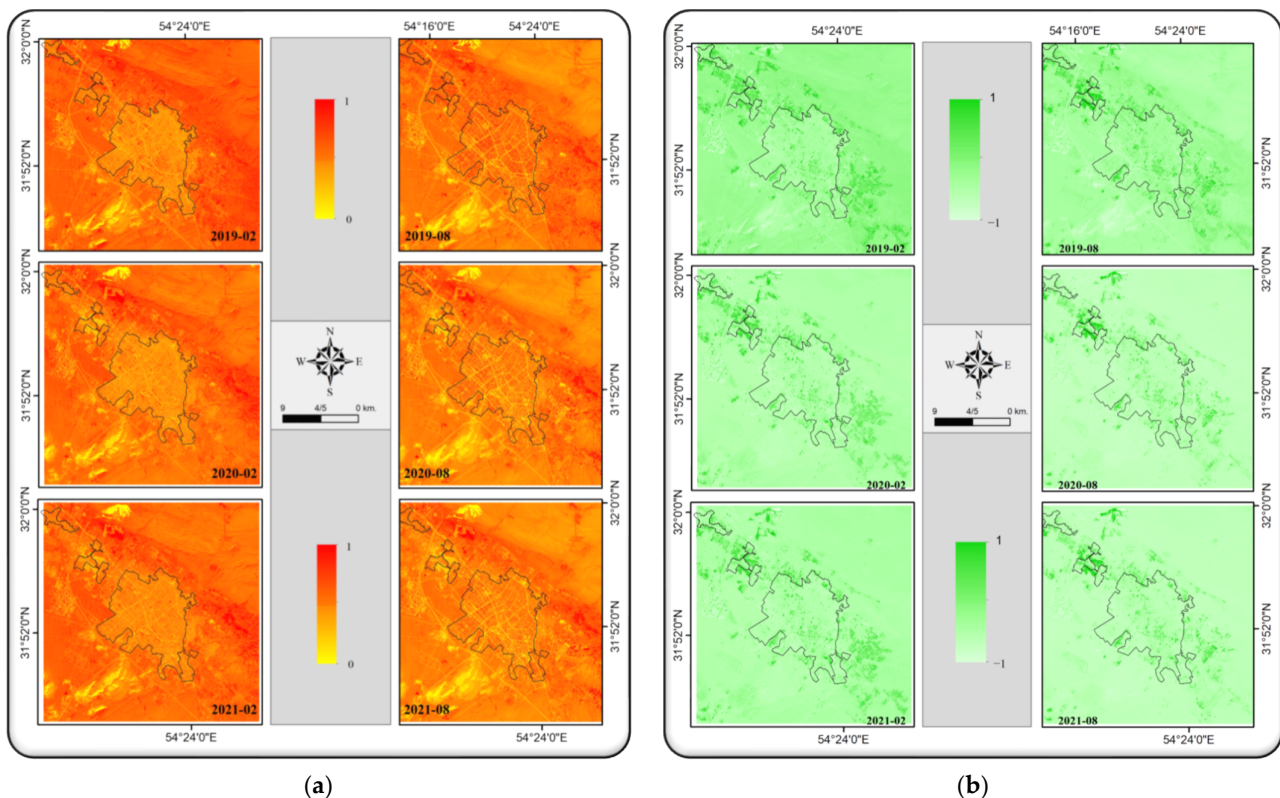
Based on model predictions, the LST images of the winter and summer of 2021 are shown in Figure 4. In 2021, the predicted values for the summer range between 83.4 °C and 44.5 °C, and for the winter between 58.23 °C and 9.81 °C.



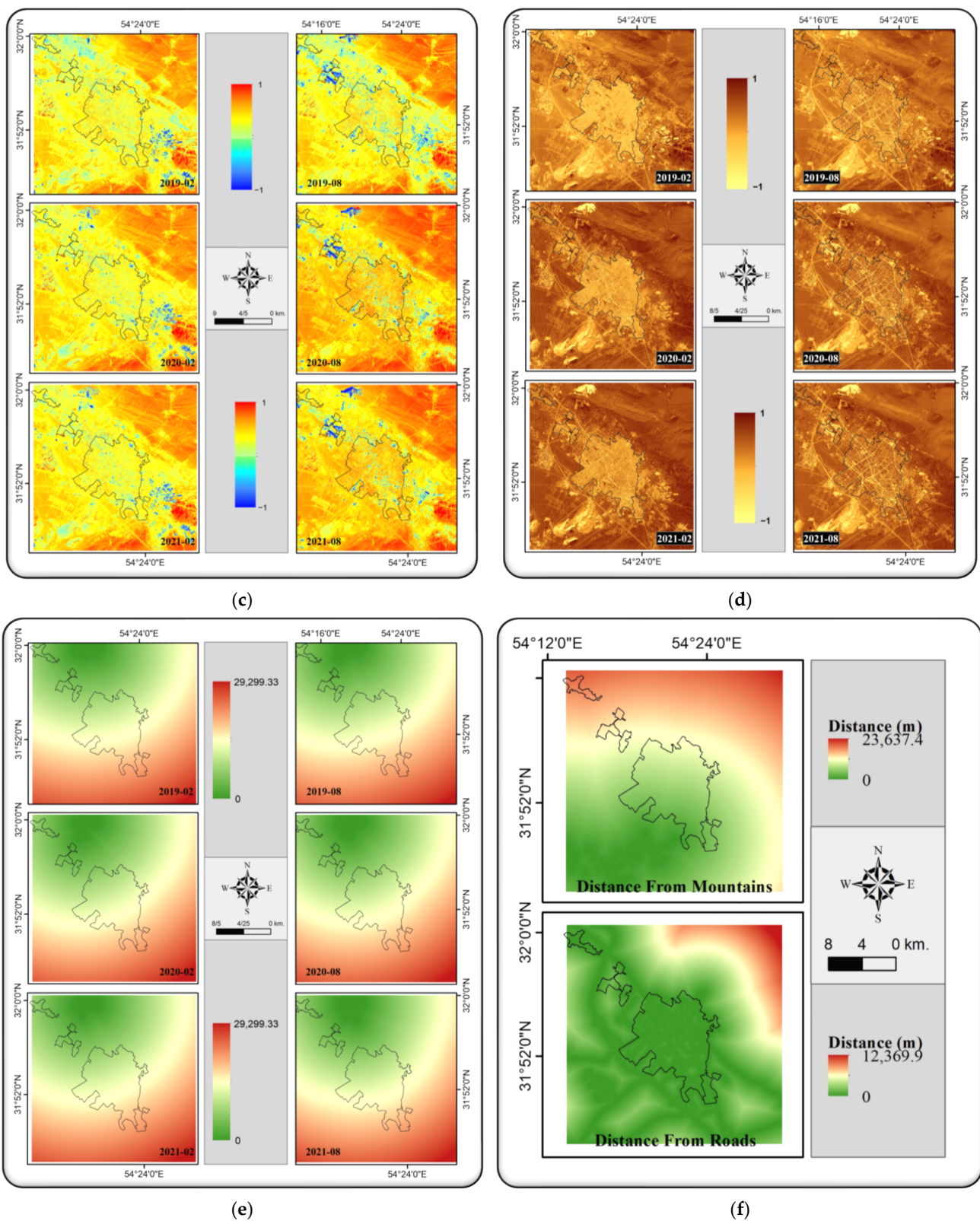
**Figure 4.** The LST images of the 2021 winter (left) and summer (right) based on model predictions.

### 3.2. Model Input Data

The ground surface features of the Yazd County at the time of taking the thermal images (Table 1) were calculated. They include albedo for each pixel (Figure 5a), NDVI (Figure 5b), NDBI (Figure 5c), NDBaI (Figure 5e), distance from water bodies for each pixel (Figure 5f), elevation of each pixel (from DEM) (Figure 1), distance from mountains for each pixel (Figure 5f, top panel), and distance from grade 1 and 2 roads for each pixel (Figure 5f, bottom panel).



**Figure 5.** Cont.



**Figure 5.** Spatial variation in calculated ground surface features: (a) albedo, (b) NDVI, (c) NDBI, (d) NDBaI, (e) distance (m) from water bodies for each pixel, and (f) distance from mountains (top) and grade 1 and 2 roads (bottom) for each pixel map. Each figure includes maps for 2019 (top), 2020 (middle), and 2021 (bottom) for the winter (left panels) and summer (right panels) seasons.



### 3.3. Data Frame Creation, Test and Selection of the Optimum Model

After calculating the ground surface features utilized as the input for the models, the implementation of AutoML facilitated the selection of the most suitable model and the fine-tuning of its parameters. The results of the statistical metrics for the top 17 models investigated are shown in Table 4. Among the tested models, the “GBM\_4\_AutoML\_1\_20220802\_231417” model had the best performance for summer, given the high accuracy of the results (RMSE of 0.32 °C, MAE of 0.17 °C, and RMSLE of 0.00 °C), and the “GBM\_1\_AutoML\_1\_20220803\_104005” was the optimum model for winter, with RMSE of 0.27 °C, MAE of 0.14 °C, and RMSLE of 0.01 °C.

**Table 4.** Statistical measures for model accuracy of the top 17 tested models for summer and winter.

| Rank | Season | Model ID                                                | RMSE (°C) | MAE (°C) | RMSLE (°C) |
|------|--------|---------------------------------------------------------|-----------|----------|------------|
| 1    | Summer | GBM_4_AutoML_1_20220802_231417                          | 0.32      | 0.17     | 0.00       |
|      | Winter | GBM_1_AutoML_1_20220803_104005                          | 0.27      | 0.14     | 0.01       |
| 2    | Summer | GBM_3_AutoML_1_20220802_231417                          | 0.33      | 0.16     | 0.00       |
|      | Winter | DeepLearning_1_AutoML_1_20220803_104005                 | 0.32      | 0.16     | 0.01       |
| 3    | Summer | GBM_5_AutoML_1_20220802_231417                          | 0.34      | 0.17     | 0.00       |
|      | Winter | StackedEnsemble_BestOfFamily_3_AutoML_1_20220803_104005 | 0.34      | 0.22     | 0.01       |
| 4    | Summer | GBM_2_AutoML_1_20220802_231417                          | 0.34      | 0.17     | 0.00       |
|      | Winter | StackedEnsemble_AllModels_2_AutoML_1_20220803_104005    | 0.34      | 0.22     | 0.01       |
| 5    | Summer | GBM_1_AutoML_1_20220802_231417                          | 0.34      | 0.16     | 0.00       |
|      | Winter | StackedEnsemble_AllModels_3_AutoML_1_20220803_104005    | 0.34      | 0.22     | 0.01       |
| 6    | Summer | DRF_1_AutoML_1_20220802_231417                          | 0.34      | 0.16     | 0.00       |
|      | Winter | StackedEnsemble_AllModels_1_AutoML_1_20220803_104005    | 0.35      | 0.22     | 0.01       |
| 7    | Summer | XRT_1_AutoML_1_20220802_231417                          | 0.35      | 0.17     | 0.00       |
|      | Winter | StackedEnsemble_BestOfFamily_1_AutoML_1_20220803_104005 | 0.35      | 0.22     | 0.01       |
| 8    | Summer | DeepLearning_1_AutoML_1_20220802_231417                 | 0.39      | 0.19     | 0.01       |
|      | Winter | StackedEnsemble_BestOfFamily_2_AutoML_1_20220803_104005 | 0.35      | 0.22     | 0.01       |
| 9    | Summer | GBM_grid_1_AutoML_1_20220802_231417_model_1             | 0.43      | 0.26     | 0.01       |
|      | Winter | DRF_1_AutoML_1_20220803_104005                          | 0.39      | 0.22     | 0.01       |
| 10   | Summer | DeepLearning_grid_1_AutoML_1_20220802_231417_model_3    | 0.48      | 0.27     | 0.01       |
|      | Winter | XRT_1_AutoML_1_20220803_104005                          | 0.44      | 0.25     | 0.01       |
| 11   | Summer | DeepLearning_grid_1_AutoML_1_20220802_231417_model_2    | 0.48      | 0.26     | 0.01       |
|      | Winter | GBM_3_AutoML_1_20220803_104005                          | 0.60      | 0.42     | 0.01       |
| 12   | Summer | DeepLearning_grid_1_AutoML_1_20220802_231417_model_1    | 0.51      | 0.29     | 0.01       |
|      | Winter | GBM_2_AutoML_1_20220803_104005                          | 0.60      | 0.42     | 0.01       |
| 13   | Summer | StackedEnsemble_AllModels_4_AutoML_1_20220802_231417    | 0.64      | 0.52     | 0.01       |
|      | Winter | GBM_4_AutoML_1_20220803_104005                          | 0.67      | 0.47     | 0.02       |
| 14   | Summer | StackedEnsemble_AllModels_3_AutoML_1_20220802_231417    | 0.65      | 0.53     | 0.01       |
|      | Winter | GLM_1_AutoML_1_20220803_104005                          | 0.68      | 0.51     | 0.02       |
| 15   | Summer | StackedEnsemble_AllModels_2_AutoML_1_20220802_231417    | 0.65      | 0.53     | 0.01       |
|      | Winter | DeepLearning_grid_1_AutoML_1_20220803_104005_model_1    | 0.82      | 0.63     | 0.02       |
| 16   | Summer | StackedEnsemble_AllModels_1_AutoML_1_20220802_231417    | 0.65      | 0.53     | 0.01       |
|      | Winter | GBM_5_AutoML_1_20220803_104005                          | 1.27      | 0.90     | 0.03       |
| 17   | Summer | StackedEnsemble_BestOfFamily_3_AutoML_1_20220802_231417 | 0.65      | 0.54     | 0.01       |
|      | Winter | GBM_grid_1_AutoML_1_20220803_104005_model_1             | 1.98      | 1.31     | 0.05       |

### 3.4. Modeling of the LST

The extracted optimal model exhibited high accuracy in predicting the LST for the utilized dataset. Table 5 presents the calculated importance and effect of ground surface parameters on LST during the summer and winter seasons. It shows that albedo is the feature with the highest importance for predicting LST, accounting for 80.3% for summer and 72.7% for the winter season. NDVI is the second parameter with the highest importance and influence on LST, accounting for 11.27% for summer and 17.21% for winter seasons. The third most important parameter in the summer season was the NDBAI, with an importance equal to 3.2%, whereas for the winter season, the NDBI was more effective, with an importance equal to 4.45%. Other parameters (fourth to eighth) had an importance (almost) equal to 1%.

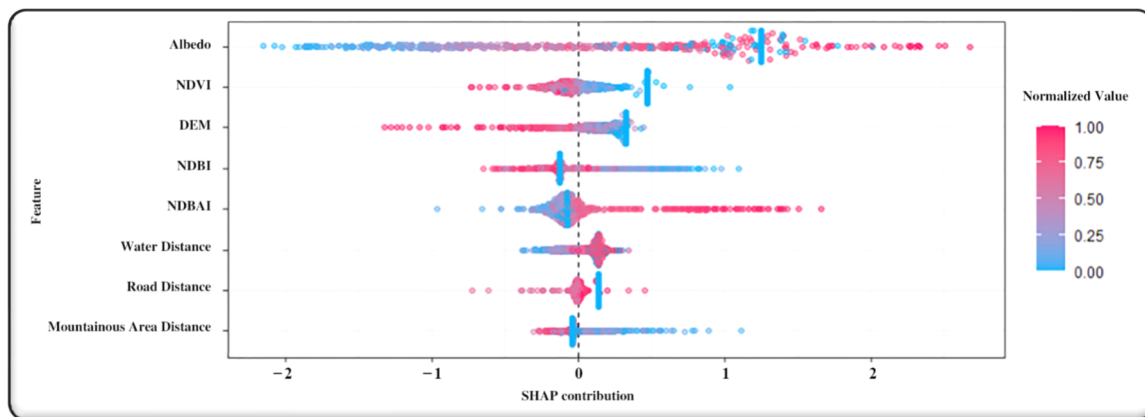
**Table 5.** Importance of the impact of the studied parameters on the LST for the summer and winter seasons.

| Rank | Season | Variable                  | Importance (%) | Scaled Importance |
|------|--------|---------------------------|----------------|-------------------|
| 1    | Summer | Albedo                    | 80.30          | 1.00000           |
|      | Winter | Albedo                    | 72.74          | 1.00000           |
| 2    | Summer | NDVI                      | 11.27          | 0.12769           |
|      | Winter | NDVI                      | 17.21          | 0.21058           |
| 3    | Summer | NDBAI                     | 3.23           | 0.00260           |
|      | Winter | NDBI                      | 4.45           | 0.00548           |
| 4    | Summer | NDBI                      | 1.16           | 0.00183           |
|      | Winter | NDBAI                     | 1.44           | 0.00534           |
| 5    | Summer | DEM                       | 1.02           | 0.00021           |
|      | Winter | Water Distance            | 1.09           | 0.00109           |
| 6    | Summer | Road Distance             | 1.01           | 0.00009           |
|      | Winter | Mountainous Area Distance | 1.04           | 0.00046           |
| 7    | Summer | Water Distance            | 1.01           | 0.00007           |
|      | Winter | Road Distance             | 1.02           | 0.00024           |
| 8    | Summer | Mountainous Area Distance | 1.00           | 0.00003           |
|      | Winter | DEM                       | 1.01           | 0.00018           |

The SHAP summary plot (Figure 6) which shows each variable's contribution, showing its average Shapley value across all inputs, helps to obtain a better view of the importance of each land surface feature and how they affect LST. Albedo has the most expansion in SHAP values and is known as the most important parameter for LST estimation. According to the SHAP values, most of the higher albedo values are positive. Each pixel has a different influence on the model's output, particularly the ones with lower values. On the other hand, lower and higher values related to NDVI, DEM, and NDBI are placed on the positive and negative side of SHAP values, respectively, with the difference that lower values of NDBI are seen on the negative side, and higher values are also seen on the positive side. In the NDVI values, an elongated peak is observed, which is divided into two parts, starting from approximately SHAP = 0, which shows lower values towards SHAP > 0 and higher values towards SHAP < 0. In NDBAI values, a very large peak can be seen, which shows the distribution of its values from less in the negative direction to more in the positive direction of SHAP values. The distance from water has a peak on the positive side of SHAP values, whereas the distance from the road shows two peaks of values, one with higher

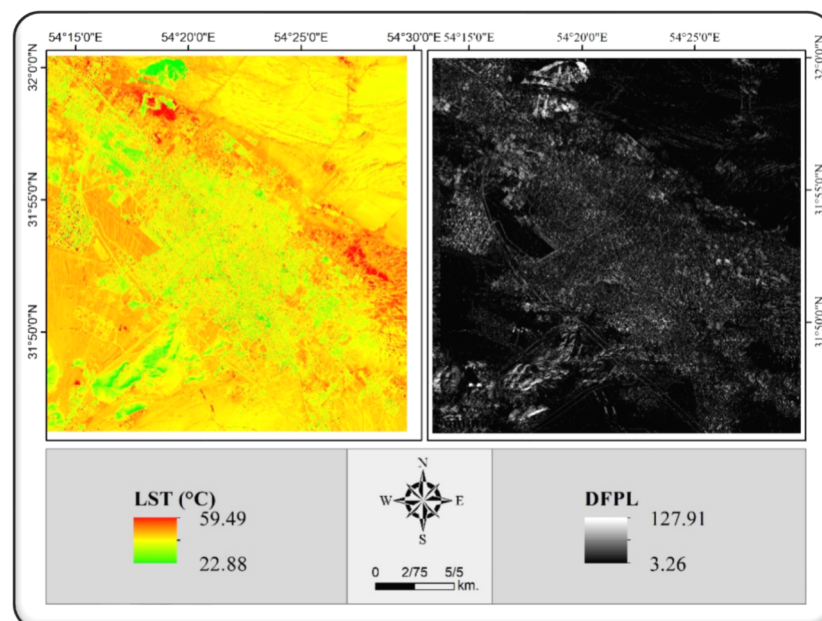


values around zero SHAP (more toward negative SHAP values) and one with lower values, towards positive SHAP values. The distance from mountainous areas has also shown the distribution of its higher values in the negative direction of SHAP values and its lower values starting with a peak in negative SHAP values and towards its positive values.

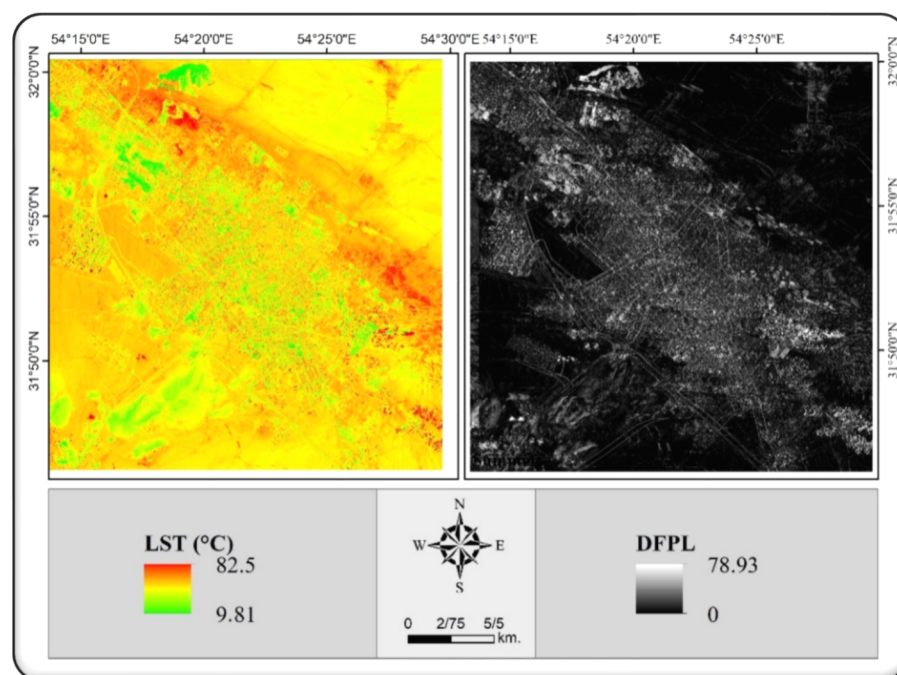


**Figure 6.** Shapley Additive exPlanation (SHAP) summary plot.

After model learning, the maps with estimated LST for the winter and summer of 2021 are presented in Figure 7. The LST values for each row of pixels were calculated and used to determine the distance of each pixel value from the  $45^\circ$  polynomial line (DFPL), which typically represents the residual or the difference between the observed value and the predicted value based on the polynomial regression model. These distance images, showcasing the variation from the actual LST values for 2021, are visually represented on the right side of the top and down panels in Figure 7. Based on DFPL image statistics (Table 6), the LST estimation shows an acceptable difference compared to the actual LST image, with a standard deviation of  $3.3^\circ\text{C}$  and  $3.5^\circ\text{C}$  for winter and summer 2021, respectively. This means that about 76% of pixels (from a total of 3,006,432) are estimated with a difference of  $3^\circ\text{C}$  or less compared to the actual LST, which is an acceptable accuracy [97].



**Figure 7.** Cont.

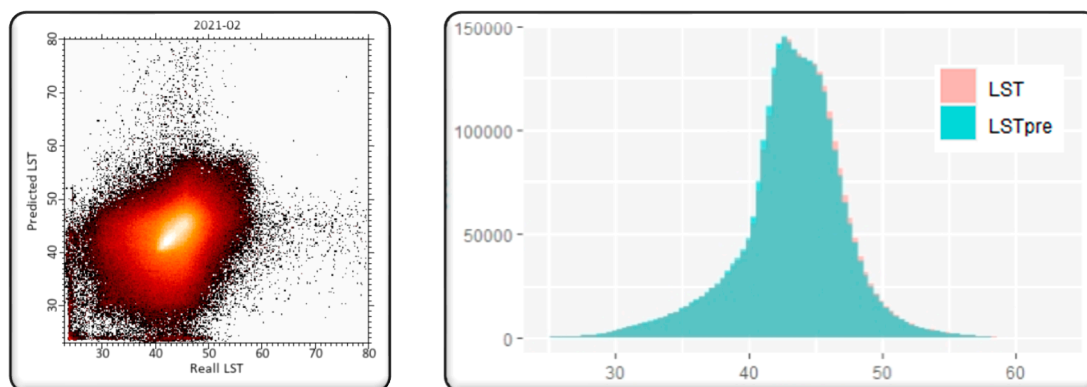


**Figure 7.** The map of the estimated LST for winter 2021 (**top**) and summer (**down**) 2021 for the Yazd cities (**left**) and its distance from the 45° polynomial line (DFPL) image (**right**).

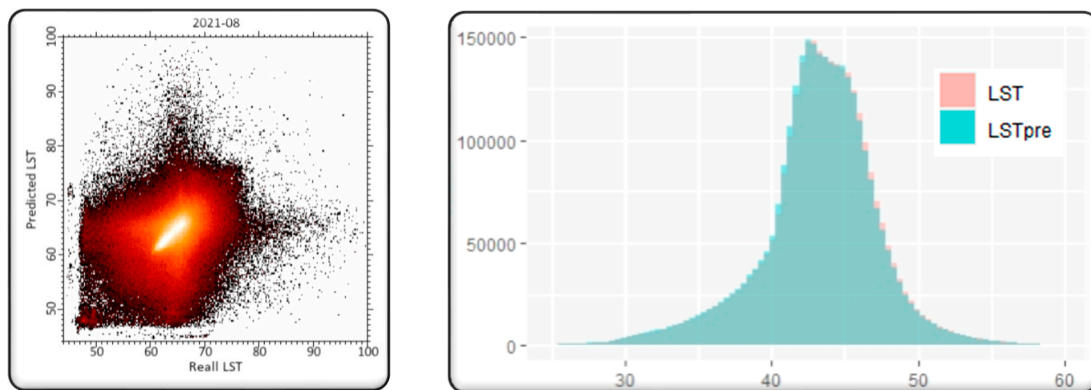
**Table 6.** Mean and standard deviation (STDEV) values of land surface temperature (LST) in the estimated map for winter and summer 2021.

| Index (°C) | Winter | Summer |
|------------|--------|--------|
| Mean       | 3.03   | 2.858  |
| STDEV      | 3.280  | 3.484  |

The scatterplot and histogram of estimated and actual LST images for both winter and summer 2021 (Figure 8) indicate the acceptable accuracy of the model's LST estimation. The histograms (Figure 8, figures on the right) show the differences in the values of the two compared images, which are highlighted by pink and pale blue colors at the edges. Additionally, statistical metrics such as MAE of 3.6 °C for winter and 3.4 °C for summer 2021 also indicate an acceptable accuracy of the models used for LST mapping (Table 7).



**Figure 8.** Cont.



**Figure 8.** Scatterplot for estimated and actual LST values (**left**), and histogram for actual (LST) and estimated (LSTpre) LST values (**right**) for winter (**top**) and summer (**down**) 2021.

**Table 7.** The statistical measures for comparison between the estimated and actual LST maps for the winter and summer of 2021.

| Index | Winter (°C) | Summer (°C) |
|-------|-------------|-------------|
| MAE   | 3.559       | 3.358       |
| RMSE  | 5.247       | 5.296       |
| RMSLE | 0.152       | 0.128       |

#### 4. Discussion

The current research uses machine learning models to estimate LST in Yazd city, which is located in a hot and dry climate zone with limited data availability. In this research, Landsat-8 satellite images were used, given, e.g., their good history in spatial data modeling [98–101]. Based on the obtained results, a fusion of thermal data can be performed with acceptable accuracy using the DisTrad method. The mean RMSE between the fused and unfused thermal data obtained in the study area was 2.7 °C (Table 3), which confirms the proper performance of the DisTrad model in thermal data downscaling. The high accuracy of this method has been confirmed in similar studies elsewhere, especially when compared with other thermal band downscaling algorithms [51–54].

The procedure for the selection of the most suitable machine learning model among the used models indicated that for the input data used, the GBM was the most accurate one, with an RMSE of 0.32 °C and 0.27 °C for the summer and winter seasons, respectively (Table 4). The GBM model, which builds regression trees on all of the features of the dataset in a fully distributed way [102], has been evaluated as an accurate and suitable model in several previous research studies dealing with spatial data [103–106]. In previous studies, other machine learning models, such as generative adversarial networks (GAN) [41], boosted regression tree (BRT), random forest (RF) [107], and support vector machine (SVM) [108], were also studied and evaluated and may be considered in future research. Models such as cellular automata (CA) and Markov models, for example, have been used in a variety of domains such as urban planning, land use forecasting, ecology, and climate change modelling [109–111]. For 2021, the estimation of the LST showed sufficiently high accuracy when compared with the actual LST, with a standard deviation of 3.3 °C and 3.5 °C (Table 6) and MAE equal to 3.6 °C and 3.4 °C (Table 7) in winter and summer, respectively, which is in line with the acceptable error of  $\pm 3$  °C for LST calculation by the algorithms provided [97].

By utilizing spectral indices and machine learning algorithms, the research identified key land surface parameters that significantly influence LST in different seasons. Specifically, albedo and NDVI were found to have a substantial impact on LST (Table 5). Albedo had the highest importance and influenced LST to the highest degree, with 80.3% and

72.7% of importance for summer and winter, respectively. Albedo also showed the largest spread in SHAP values, which can be a double confirmation of the effect and importance of this parameter on LST. The distribution of lower albedo values on both sides (positive and negative) of SHAP values can indicate the greater effect of these values on LST. Probably, it is a confirmation or a response to some city surfaces that have high (concrete) and low (asphalt) albedos [112], and therefore have opposite functions in reflecting or absorbing a higher percentage of solar radiation with different albedo. Those with higher absorption rates absorb more heat from the sun. Consequently, these surfaces tend to have different LST values, probably because they absorb a larger portion of the solar radiation, which is converted into heat energy, so the surface temperature increases. This influence is lower in winter possibly due to the increase in surface moisture which can play a balancing role in addition to the colder air and the effect on LST in the colder season. NDVI was the second most important factor impacting the LST in both summer and winter, having an influence equal to 11.3% and 17.2%, respectively (Table 5). The higher impact of NDVI on LST in winter is possibly because of fewer leaves and less dense vegetation, resulting in lower NDVI values. In winter, an increase in NDVI values can lead to increasing LST [113,114]. A positive correlation between NDVI and LST values in winter has been reported in previous studies [115]. The distribution of higher and lower values of NDVI in SHAP was also almost completely related to the negative and positive values of SHAP, respectively, which support the great effect of NDVI on LST. NDBaI with an influence equal to 3.2% in summer and NDBI with an influence equal to 4.5% in winter also had a considerable impact on LST. This is probably because of the barren lands, which are very hot during the summer in the vicinity of Yazd [116], and also the increase in heat in the residential areas during the winter season due to the use of heating devices. NDBaI values are also relatively widely present in SHAP, with lower values associated with negative SHAPs and higher values associated with positive SHAPs, and this can show the reinforcing role of their presence in increasing LST. Additionally, the reduced vegetation cover may lead to a higher reception of solar energy, and thus increase in LST, in built-up areas in winter than in summer. Other studies have reported a positive relationship between NDBI and LST, especially in the winter season [117]. The NDBI values in SHAP showed the two poles of less in positive and more in negative SHAP in addition to a peak of less in negative SHAP values, which can be due to the role of built-up areas in increasing LST in most cases and vice versa. It also shows its cooling factors (e.g., plants, irrigation, roof gardens) in some areas. Although the values of the distance from water and road showed a small share of influence on LST (~1%), their distribution in SHAP shows their influence in a reverse way. LST increases with the increasing distance to water but decreases with the increasing distance to roads. The two peaks of the distance from the road can be seen from two different perspectives: on the one hand, proximity to roads and the increase in LST due to the increase in density of built-up areas in some areas, and on the other hand, the distance from roads and the decrease in LST due to distance from asphalt and vehicle activity in some other (most) areas.

This study simplifies and develops machine learning models in the field of environmental sciences and thermal remote sensing, establishing a new bridge towards research on their compatibility and improving their accuracy as much as possible in the field of interdisciplinary research in the future. In this connection, data modeling with the approach of prediction and estimation using machine learning and deep learning has been considered in various research studies in the field of spatial sciences, especially agriculture, plants, and other related issues [118,119]. The modeling footprint of spatial thermal data can be used in more studies in the future. However, this study has some limitations. The main limitation was the impossibility of accessing and collecting accurate ground data or high-resolution thermal images. We suggest future research using data collection teams to obtain the field data. Furthermore, additional non-image parameters should be included in LST modeling using machine learning or deep learning models. For example, future research should consider investigating the impact of variables such as soil moisture and soil type, solar radiation, type of air circulation, and cloudiness as considered in a few studies predicting

land surface temperature like [120]. This research also suggests that future research should report the estimated values of LST using machine learning in the peri-urban areas and surrounding villages so that the role of the factors influencing the creation of heat islands in this connection is also considered. This suggestion is based on the scientific background of the average temperature difference between urban and the surrounding rural areas, and its basics have been studied in previous research [121–123].

The findings of this research provide valuable insights for city managers in understanding the factors contributing to urban LST to develop appropriate strategies to mitigate the associated heat stress, and to support effective urban planning. Moreover, the developed models can be applied in the study area for temporal and spatial gap-filling of thermal data. The methodology developed offers a reliable tool for estimating surface temperature and supporting decision-making and planning processes in urban areas with similar climatic characteristics.

## 5. Conclusions

This study estimates land surface temperature (LST) in Yazd City, characterized by a hot and arid climate region, utilizing land surface parameters derived from spectral indices and machine learning algorithms. The analysis employed six machine learning algorithms and incorporated various ground surface features, including albedo, NDVI, NDBI, NDBaI, distance from water bodies, elevation, distance from mountains, and distance from roads, extracted from Landsat-8 imagery during the winter and summer seasons of 2019 and 2020. The best model, selected based on three accuracy metrics (i.e., RMSE, MA, and RMSLE) was implemented and evaluated to estimate LST in 2021. According to the results of this research, the following conclusions can be drawn:

1. Assessing ground surface temperature using machine learning models: the GBM algorithm outperforms the other machine learning models for LST estimation, given its highest accuracy. Due to the high accuracy and the ability to use complex data such as spectral indices, machine learning models can be used to estimate LST in areas where thermal data are limited.
2. Effect of environmental factors in determining Earth's surface temperature: albedo exerts a significant impact on LST, with an importance of 80.30% and 72.74% in summer and winter seasons, respectively. NDVI is the second most important factor in determining LST, accounting for 11.27% and 17.21%, correspondingly. The other six ground surface features investigated have an importance of ~5% in LST for both seasons.
3. Ability to estimate temperature at different spatial scales: spectral indices and machine learning algorithms can be used to estimate LST on a large spatial scale. But for smaller scales, improved and adaptive methods may be needed.
4. The use of spectral indices in time analysis: by considering different spectral indices in different seasons, it is possible to conduct temporal analyses of LST and examine patterns and seasonal changes. These analyses provide a better understanding of the temperature dynamics of the study areas.

The outcomes of this study provide valuable insights into the application of machine learning models for estimating LST using remote sensing spectral indices in areas lacking thermal bands or when finer spatial scales are required. Our findings can be useful in future research and to support decision making regarding practical solutions for urban heat island effect mitigation and environmental management.

**Author Contributions:** Conceptualization, M.M., I.R. and H.G.M.; methodology, M.M.; validation, M.M., H.G.M., I.R., M.S., J.K. and C.S.S.F.; formal analysis, M.M., I.R., M.S. and C.S.S.F.; writing—original draft preparation, M.M.; writing—review and editing, M.M., I.R., H.G.M., J.K. and C.S.S.F.; supervision, I.R. and H.G.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** The work developed by C.F. was funded by the Portuguese Foundation for Science and Technology, through the institutional scientific employment program-contract (CEECINST/00077/2021).



**Data Availability Statement:** Contact [irousta@yazd.ac.ir](mailto:irousta@yazd.ac.ir) or visit <https://mqdm.ir>.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Tong, D.; Chu, J.; Han, Q.; Liu, X. How land finance drives urban expansion under fiscal pressure: Evidence from Chinese cities. *Land* **2022**, *11*, 253. [\[CrossRef\]](#)
2. Liu, X.; Kong, M.; Tong, D.; Zeng, X.; Lai, Y. Property rights and adjustment for sustainable development during post-productivist transitions in China. *Land Use Policy* **2022**, *122*, 106379. [\[CrossRef\]](#)
3. Li, J.; Wang, Z.; Wu, X.; Xu, C.-Y.; Guo, S.; Chen, X. Toward monitoring short-term droughts using a novel daily scale, standardized antecedent precipitation evapotranspiration index. *J. Hydrometeorol.* **2020**, *21*, 891–908. [\[CrossRef\]](#)
4. Wu, X.; Guo, S.; Qian, S.; Wang, Z.; Lai, C.; Li, J.; Liu, P. Long-range precipitation forecast based on multipole and preceding fluctuations of sea surface temperature. *Int. J. Climatol.* **2022**, *42*, 8024–8039. [\[CrossRef\]](#)
5. Oke, T.R. The energetic basis of the urban heat island. *Q. J. R. Meteorol. Soc.* **1982**, *108*, 1–24. [\[CrossRef\]](#)
6. Taha, H. Urban climates and heat islands: Albedo, evapotranspiration, and anthropogenic heat. *Energy Build.* **1997**, *25*, 99–103. [\[CrossRef\]](#)
7. Rizwan, A.M.; Dennis, L.Y.; Chunho, L. A review on the generation, determination and mitigation of Urban Heat Island. *J. Environ. Sci.* **2008**, *20*, 120–128. [\[CrossRef\]](#)
8. Grimmond, S.U. Urbanization and global environmental change: Local effects of urban warming. *Geogr. J.* **2007**, *173*, 83–88. [\[CrossRef\]](#)
9. Kikegawa, Y.; Genchi, Y.; Yoshikado, H.; Kondo, H. Development of a numerical simulation system toward comprehensive assessments of urban warming countermeasures including their impacts upon the urban buildings' energy-demands. *Appl. Energy* **2003**, *76*, 449–466. [\[CrossRef\]](#)
10. Plocoste, T.; Jacoby-Koaly, S.; Molinié, J.; Petit, R. Evidence of the effect of an urban heat island on air quality near a landfill. *Urban Clim.* **2014**, *10*, 745–757. [\[CrossRef\]](#)
11. Meineke, E.K.; Dunn, R.R.; Frank, S.D. Early pest development and loss of biological control are associated with urban warming. *Biol. Lett.* **2014**, *10*, 20140586. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Şimşek, Ç.K.; Ödül, H. A method proposal for monitoring the microclimatic change in an urban area. *Sustain. Cities Soc.* **2019**, *46*, 101407. [\[CrossRef\]](#)
13. Liu, L.; Li, Z.; Fu, X.; Liu, X.; Li, Z.; Zheng, W. Impact of power on uneven development: Evaluating built-up area changes in Chengdu based on NPP-VIIRS images (2015–2019). *Land* **2022**, *11*, 489. [\[CrossRef\]](#)
14. Alavipanah, S.K. *Thermal Remote Sensing and Its Application in the Earth Sciences*; University of Tehran Press: Tehran, Iran, 2006; pp. 903–964.
15. Mansourmoghaddam, M.; Naghipur, N.; Roustia, I.; Alavipanah, S.K.; Olafsson, H.; Ali, A.A. Quantifying the Effects of Green-Town Development on Land Surface Temperatures (LST)(A Case Study at Karizland (Karizboom), Yazd, Iran). *Land* **2023**, *12*, 885. [\[CrossRef\]](#)
16. Jia, B.; Zhou, G. Estimation of global karst carbon sink from 1950s to 2050s using response surface methodology. *Geo-Spat. Inf. Sci.* **2023**, 1–18. [\[CrossRef\]](#)
17. Zhang, S.; Bai, X.; Zhao, C.; Tan, Q.; Luo, G.; Wang, J.; Li, Q.; Wu, L.; Chen, F.; Li, C. Global CO<sub>2</sub> consumption by silicate rock chemical weathering: Its past and future. *Earth's Future* **2021**, *9*, e2020EF001938. [\[CrossRef\]](#)
18. Wang, P.; Yu, P.; Lu, J.; Zhang, Y. The mediation effect of land surface temperature in the relationship between land use-cover change and energy consumption under seasonal variations. *J. Clean. Prod.* **2022**, *340*, 130804. [\[CrossRef\]](#)
19. Kazemi Garajeh, M.; Salmani, B.; Zare Naghadehi, S.; Valipoori Goodarzi, H.; Khasraei, A. An integrated approach of remote sensing and geospatial analysis for modeling and predicting the impacts of climate change on food security. *Sci. Rep.* **2023**, *13*, 1057. [\[CrossRef\]](#)
20. Osborne, P.E.; Alvares-Sanches, T. Quantifying how landscape composition and configuration affect urban land surface temperatures using machine learning and neutral landscapes. *Comput. Environ. Urban Syst.* **2019**, *76*, 80–90. [\[CrossRef\]](#)
21. Quattrochi, D.A.; Luvall, J.C. Thermal infrared remote sensing for analysis of landscape ecological processes: Methods and applications. *Landsc. Ecol.* **1999**, *14*, 577–598. [\[CrossRef\]](#)
22. Li, W.; Wang, W.; Sun, R.; Li, M.; Liu, H.; Shi, Y.; Zhu, D.; Li, J.; Ma, L.; Fu, S. Influence of nitrogen addition on the functional diversity and biomass of fine roots in warm-temperate and subtropical forests. *For. Ecol. Manag.* **2023**, *545*, 121309. [\[CrossRef\]](#)
23. Lin, J.; Wei, K.; Guan, Z. Exploring the connection between morphological characteristic of built-up areas and surface heat islands based on MSPA. *Urban Clim.* **2024**, *53*, 101764. [\[CrossRef\]](#)
24. Tran, D.X.; Pla, F.; Latorre-Carmona, P.; Myint, S.W.; Caetano, M.; Kieu, H.V. Characterizing the relationship between land use land cover change and land surface temperature. *ISPRS J. Photogramm. Remote Sens.* **2017**, *124*, 119–132. [\[CrossRef\]](#)
25. Shang, K.; Xu, L.; Liu, X.; Yin, Z.; Liu, Z.; Li, X.; Yin, L.; Zheng, W. Study of urban heat island effect in Hangzhou metropolitan area based on SW-TES algorithm and image dichotomous model. *SAGE Open* **2023**, *13*, 21582440231208851. [\[CrossRef\]](#)
26. Collas, L.; Green, R.E.; Ross, A.; Wastell, J.H.; Balmford, A. Urban development, land sharing and land sparing: The importance of considering restoration. *J. Appl. Ecol.* **2017**, *54*, 1865–1873. [\[CrossRef\]](#)

27. Stott, I.; Soga, M.; Inger, R.; Gaston, K.J. Land sparing is crucial for urban ecosystem services. *Front. Ecol. Environ.* **2015**, *13*, 387–393. [CrossRef]
28. Amiri, R.; Weng, Q.; Alimohammadi, A.; Alavipanah, S.K. Spatial–temporal dynamics of land surface temperature in relation to fractional vegetation cover and land use/cover in the Tabriz urban area, Iran. *Remote Sens. Environ.* **2009**, *113*, 2606–2617. [CrossRef]
29. Shang, Y.; Song, K.; Lai, F.; Lyu, L.; Liu, G.; Fang, C.; Hou, J.; Qiang, S.; Yu, X.; Wen, Z. Remote sensing of fluorescent humification levels and its potential environmental linkages in lakes across China. *Water Res.* **2023**, *230*, 119540. [CrossRef]
30. Mansourmoghaddam, M.; Ghafarian Malamiri, H.R.; Roustai, I.; Olafsson, H.; Zhang, H. Assessment of Palm Jumeirah Island’s Construction Effects on the Surrounding Water Quality and Surface Temperatures during 2001–2020. *Water* **2022**, *14*, 634. [CrossRef]
31. Wen, Z.; Wang, Q.; Ma, Y.; Jacinthe, P.A.; Liu, G.; Li, S.; Shang, Y.; Tao, H.; Fang, C.; Lyu, L. Remote estimates of suspended particulate matter in global lakes using machine learning models. *Int. Soil Water Conserv. Res.* **2024**, *12*, 200–216. [CrossRef]
32. Li, R.; Zhu, G.; Lu, S.; Sang, L.; Meng, G.; Chen, L.; Jiao, Y.; Wang, Q. Effects of Urbanization on the water cycle in the Shiyang River Basin: Based on stable isotope method. *Hydrol. Earth Syst. Sci. Discuss.* **2023**, *2023*, 4437–4452. [CrossRef]
33. Grimmond, C. Progress in measuring and observing the urban atmosphere. *Theor. Appl. Climatol.* **2006**, *84*, 3–22. [CrossRef]
34. Piringir, M.; Grimmond, C.S.B.; Joffre, S.M.; Mestayer, P.; Middleton, D.; Rotach, M.; Baklanov, A.; De Ridder, K.; Ferreira, J.; Guilloteau, E. Investigating the surface energy balance in urban areas—recent advances and future needs. *Water Air Soil Pollut. Focus* **2002**, *2*, 1–16. [CrossRef]
35. Jiang, J.; Tian, G. Analysis of the impact of land use/land cover change on land surface temperature with remote sensing. *Procedia Environ. Sci.* **2010**, *2*, 571–575. [CrossRef]
36. Peng, X.; Wu, W.; Zheng, Y.; Sun, J.; Hu, T.; Wang, P. Correlation analysis of land surface temperature and topographic elements in Hangzhou, China. *Sci. Rep.* **2020**, *10*, 1–16. [CrossRef] [PubMed]
37. Alavipanah, S.K.; Hashemi Darrehbadami, S.; Kazemzadeh, A. Spatial-Temporal Analysis of Urban Heat-Island of Mashhad City due to Land Use/Cover Change and Expansion. *Geogr. Urban Plan. Res.* **2015**, *3*, 1–17.
38. IBM. What is Machine Learning? Available online: <https://www.ibm.com/cloud/learn/machine-learning> (accessed on 19 January 2024).
39. Ozturk, D. Urban growth simulation of Atakum (Samsun, Turkey) using cellular automata-Markov chain and multi-layer perceptron-Markov chain models. *Remote Sens.* **2015**, *7*, 5918–5950. [CrossRef]
40. Musa, S.I.; Hashim, M.; Reba, M.N.M. A review of geospatial-based urban growth models and modelling initiatives. *Geocarto Int.* **2017**, *32*, 813–833. [CrossRef]
41. Li, Q.; Zheng, H. Prediction of summer daytime land surface temperature in urban environments based on machine learning. *Sustain. Cities Soc.* **2023**, *97*, 104732. [CrossRef]
42. Peng, J.; Ma, J.; Liu, Q.; Liu, Y.; Li, Y.; Yue, Y. Spatial-temporal change of land surface temperature across 285 cities in China: An urban-rural contrast perspective. *Sci. Total Environ.* **2018**, *635*, 487–497. [CrossRef]
43. Mansourmoghaddam, M.; Roustai, I.; Zamani, M.S.; Mokhtari, M.H.; Karimi Firozjaei, M.; Alavipanah, S.K. Investigating And Modeling the Effect of The Composition and Arrangement of The Landscapes of Yazd City on The Land Surface Temperature Using Machine Learning and Landsat-8 and Sentinel-2 Data. *Iran. J. Remote Sens. GIS* **2022**, *15*, 1–26.
44. CustomWeather. Climate & Weather Averages in Yazd, Iran. Available online: <https://www.timeanddate.com/weather/iran/yazd/climate> (accessed on 1 February 2022).
45. Adler-Golden, S.M.; Matthew, M.W.; Bernstein, L.S.; Levine, R.Y.; Berk, A.; Richtsmeier, S.C.; Acharya, P.K.; Anderson, G.P.; Felde, J.W.; Gardner, J. Atmospheric correction for shortwave spectral imagery based on MODTRAN4. In Proceedings of the Imaging Spectrometry V, Denver, CO, USA, 19–21 July 1999; pp. 61–69.
46. Berk, A.; Bernstein, L.; Anderson, G.; Acharya, P.; Robertson, D.; Chetwynd, J.; Adler-Golden, S. MODTRAN cloud and multiple scattering upgrades with application to AVIRIS. *Remote Sens. Environ.* **1998**, *65*, 367–375. [CrossRef]
47. Berk, A.; Bernstein, L.; Robertson, D. MODTRAN: A Moderate Resolution Model for LOWTRAN 7; Final Report; Geophysics Laboratory: Las Vegas, NV, USA, 1989.
48. Matthew, M.W.; Adler-Golden, S.M.; Berk, A.; Richtsmeier, S.C.; Levine, R.Y.; Bernstein, L.S.; Acharya, P.K.; Anderson, G.P.; Felde, G.W.; Hoke, M.L. Status of atmospheric correction using a MODTRAN4-based algorithm. In Proceedings of the Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI, Orlando, FL, USA, 24–28 April 2000; pp. 199–207.
49. Johnson, B.; Young, S.J. *In-Scene Atmospheric Compensation: Application to SEBASS Data Collected at the ARM Site*; Space and Environment Technology Center, the Aerospace Corporation: Segundo, CA, USA, 1998.
50. Hernandez-Baquero, E.D. *Characterization of the Earth’s Surface and Atmosphere from Multispectral and Hyperspectral Thermal Imagery*; Air Force Inst of Tech Wright-Pattersonafb Oh School of Engineering: Dayton, OH, USA, 2000.
51. Essa, W.; Verbeiren, B.; van der Kwast, J.; Van de Voorde, T.; Batelaan, O. Evaluation of the DisTrad thermal sharpening methodology for urban areas. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *19*, 163–172. [CrossRef]
52. Bindhu, V.; Narasimhan, B.; Sudheer, K. Development and verification of a non-linear disaggregation method (NL-DisTrad) to downscale MODIS land surface temperature to the spatial scale of Landsat thermal data to estimate evapotranspiration. *Remote Sens. Environ.* **2013**, *135*, 118–129. [CrossRef]

53. Essa, W.; van der Kwast, J.; Verbeiren, B.; Batelaan, O. Downscaling of thermal images over urban areas using the land surface temperature–impervious percentage relationship. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *23*, 95–108. [\[CrossRef\]](#)
54. Sara, K.; Eswar, R.; Bhattacharya, B.K. The Utility of Simpler Spatial Disaggregation Models for Retrieving Land Surface Temperature at High Spatiotemporal Resolutions. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [\[CrossRef\]](#)
55. Kustas, W.P.; Norman, J.M.; Anderson, M.C.; French, A.N. Estimating subpixel surface temperatures and energy fluxes from the vegetation index–radiometric temperature relationship. *Remote Sens. Environ.* **2003**, *85*, 429–440. [\[CrossRef\]](#)
56. Wu, J.; Xia, L.; Chan, T.O.; Awange, J.; Zhong, B. Downscaling land surface temperature: A framework based on geographically and temporally neural network weighted autoregressive model with spatio-temporal fused scaling factors. *ISPRS J. Photogramm. Remote Sens.* **2022**, *187*, 259–272. [\[CrossRef\]](#)
57. Ranagalage, M.; Estoque, R.C.; Handayani, H.H.; Zhang, X.; Morimoto, T.; Tadono, T.; Murayama, Y. Relation between urban volume and land surface temperature: A comparative study of planned and traditional cities in Japan. *Sustainability* **2018**, *10*, 2366. [\[CrossRef\]](#)
58. Bokaie, M.; Zarkesh, M.K.; Arasteh, P.D.; Hosseini, A. Assessment of urban heat island based on the relationship between land surface temperature and land use/land cover in Tehran. *Sustain. Cities Soc.* **2016**, *23*, 94–104. [\[CrossRef\]](#)
59. Avdan, U.; Jovanovska, G. Algorithm for automated mapping of land surface temperature using LANDSAT 8 satellite data. *J. Sens.* **2016**, *2016*, 1480307. [\[CrossRef\]](#)
60. Dos Santos, A.R.; de Oliveira, F.S.; da Silva, A.G.; Gleriani, J.M.; Gonçalves, W.; Moreira, G.L.; Silva, F.G.; Branco, E.R.F.; Moura, M.M.; da Silva, R.G. Spatial and temporal distribution of urban heat islands. *Sci. Total Environ.* **2017**, *605*, 946–956. [\[CrossRef\]](#) [\[PubMed\]](#)
61. Estoque, R.C.; Pontius Jr, R.G.; Murayama, Y.; Hou, H.; Thapa, R.B.; Lasco, R.D.; Villar, M.A. Simultaneous comparison and assessment of eight remotely sensed maps of Philippine forests. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *67*, 123–134. [\[CrossRef\]](#)
62. Sultana, S.; Satyanarayana, A. Urban heat island intensity during winter over metropolitan cities of India using remote-sensing techniques: Impact of urbanization. *Int. J. Remote Sens.* **2018**, *39*, 6692–6730. [\[CrossRef\]](#)
63. Ziaul, S.; Pal, S. Image based surface temperature extraction and trend detection in an urban area of West Bengal, India. *J. Environ. Geogr.* **2016**, *9*, 13–25. [\[CrossRef\]](#)
64. Rousta, I.; Sarif, M.O.; Gupta, R.D.; Olafsson, H.; Ranagalage, M.; Murayama, Y.; Zhang, H.; Mushore, T.D. Spatiotemporal analysis of land use/land cover and its effects on surface urban heat island using Landsat data: A case study of Metropolitan City Tehran (1988–2018). *Sustainability* **2018**, *10*, 4433. [\[CrossRef\]](#)
65. *LANDSAT 8 Data Users Handbook*; Department of the Interior US Geological Survey: Reston, VA, USA, 2015.
66. Liang, S. Narrowband to broadband conversions of land surface albedo I: Algorithms. *Remote Sens. Environ.* **2001**, *76*, 213–238. [\[CrossRef\]](#)
67. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the Great Plains with ERTS. *NASA Spec. Publ.* **1974**, *351*, 309.
68. Zha, Y.; Gao, J.; Ni, S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int. J. Remote Sens.* **2003**, *24*, 583–594. [\[CrossRef\]](#)
69. Oke, T.R. *Boundary Layer Climates*; Routledge: London, UK, 2002.
70. Wang, W.; Liang, S.; Meyers, T. Validating MODIS land surface temperature products using long-term nighttime ground measurements. *Remote Sens. Environ.* **2008**, *112*, 623–635. [\[CrossRef\]](#)
71. Jimenez-Munoz, J.C.; Cristobal, J.; Sobrino, J.A.; Soria, G.; Ninyerola, M.; Pons, X. Revision of the single-channel algorithm for land surface temperature retrieval from Landsat thermal-infrared data. *IEEE Trans. Geosci. Remote Sens.* **2008**, *47*, 339–349. [\[CrossRef\]](#)
72. Zhao, F.; Wu, H.; Zhu, S.; Zeng, H.; Zhao, Z.; Yang, X.; Zhang, S. Material stock analysis of urban road from nighttime light data based on a bottom-up approach. *Environ. Res.* **2023**, *228*, 115902. [\[CrossRef\]](#) [\[PubMed\]](#)
73. Smith, R. *The Heat Budget of the Earth's Surface Deduced from Space*; Yale Center for Earth Observation: New Haven, CT, USA, 2010.
74. Li, X.; Zhou, Y.; Asrar, G.R.; Imhoff, M.; Li, X. The surface urban heat island response to urban expansion: A panel analysis for the conterminous United States. *Sci. Total Environ.* **2017**, *605*, 426–435. [\[CrossRef\]](#) [\[PubMed\]](#)
75. Weng, Q.; Yang, S. Urban air pollution patterns, land use, and thermal landscape: An examination of the linkage using GIS. *Environ. Monit. Assess.* **2006**, *117*, 463–489. [\[CrossRef\]](#) [\[PubMed\]](#)
76. Ya'acob, N.; Azize, A.B.M.; Mahmon, N.A.; Yusof, A.L.; Azmi, N.F.; Mustafa, N. Temporal forest change detection and forest health assessment using Remote Sensing. *IOP Conf. Ser. Earth Environ. Sci.* **2014**, *19*, 012017.
77. Rasul, A.; Balzter, H.; Ibrahim, G.R.F.; Hameed, H.M.; Wheeler, J.; Adamu, B.; Ibrahim, S.; Najmaddin, P.M. Applying built-up and bare-soil indices from Landsat 8 to cities in dry climates. *Land* **2018**, *7*, 81. [\[CrossRef\]](#)
78. As-Syakur, A.R.; Adnyana, I.W.S.; Arthana, I.W.; Nuarsa, I.W. Enhanced built-up and bareness index (EBBI) for mapping built-up and bare land in an urban area. *Remote Sens.* **2012**, *4*, 2957–2970. [\[CrossRef\]](#)
79. esri. Understanding Euclidean distance analysis. Available online: [https://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/understanding-euclidean-distance-analysis.htm#ESRI\\_SECTION1\\_29048F6D811B40D0A0B7E2BA0F36E92E](https://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/understanding-euclidean-distance-analysis.htm#ESRI_SECTION1_29048F6D811B40D0A0B7E2BA0F36E92E) (accessed on 30 August 2022).
80. Xu, H. A study on information extraction of water body with the modified normalized difference water index (MNDWI). *J. Remote Sens.* **2005**, *9*, 595.



81. Shapley, L.S. A value for n-person games. In *Contributions to the Theory of Games II*; Kuhn, H., Tucker, A., Eds.; Princeton University Press: Princeton, NJ, USA, 1953.
82. Gramegna, A.; Giudici, P. Why to buy insurance? an explainable artificial intelligence approach. *Risks* **2020**, *8*, 137. [\[CrossRef\]](#)
83. Joseph, A. Shapley regressions: A framework for statistical inference on machine learning models. *arXiv* **2019**, arXiv:1903.04209v1. [\[CrossRef\]](#)
84. Lundberg, S.M.; Erion, G.G.; Lee, S.-I. Consistent individualized feature attribution for tree ensembles. *arXiv* **2018**, arXiv:1802.03888.
85. h2o. SHAP Summary Plot. Available online: [https://docs.h2o.ai/h2o/latest-stable/h2o-r/docs/reference/h2o.shap\\_summary\\_plot.html](https://docs.h2o.ai/h2o/latest-stable/h2o-r/docs/reference/h2o.shap_summary_plot.html) (accessed on 7 January 2024).
86. h2o. H2O AutoML: Automatic Machine Learning. Available online: <https://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html> (accessed on 13 January 2024).
87. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
88. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [\[CrossRef\]](#)
89. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [\[CrossRef\]](#)
90. McCullagh, P.; Nelder, J. *Generalized Linear Models*; CRC Press: Boca Raton, FL, USA, 1989.
91. Sill, J.; Takács, G.; Mackey, L.; Lin, D. Feature-weighted linear stacking. *arXiv* **2009**, arXiv:0911.0460.
92. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#)
93. Vakayil, A.; Joseph, R.; Mak, S. SPLIT: Split a dataset for training and testing. *R Package Version* **2021**, *1*.
94. Joseph, V.R. Optimal ratio for data splitting. *Stat. Anal. Data Min. ASA Data Sci. J.* **2022**, *15*, 531–538. [\[CrossRef\]](#)
95. DeWitt, D.J.; Naughton, J.F.; Schneider, D.F. *Parallel Sorting on a Shared-Nothing Architecture Using Probabilistic Splitting*; University of Wisconsin-Madison Department of Computer Sciences: Madison, WI, USA, 1991.
96. Chen, J.; Lu, S. Improved parameterized set splitting algorithms: A probabilistic approach. *Algorithmica* **2009**, *54*, 472–489. [\[CrossRef\]](#)
97. Li, Z.-L.; Tang, B.-H.; Wu, H.; Ren, H.; Yan, G.; Wan, Z.; Trigo, I.F.; Sobrino, J.A. Satellite-derived land surface temperature: Current status and perspectives. *Remote Sens. Environ.* **2013**, *131*, 14–37. [\[CrossRef\]](#)
98. Xu, J.; Zhou, G.; Su, S.; Cao, Q.; Tian, Z. The development of a rigorous model for bathymetric mapping from multispectral satellite-images. *Remote Sens.* **2022**, *14*, 2495. [\[CrossRef\]](#)
99. Wan, R.; Wang, P.; Wang, X.; Yao, X.; Dai, X. Modeling wetland aboveground biomass in the Poyang Lake National Nature Reserve using machine learning algorithms and Landsat-8 imagery. *J. Appl. Remote Sens.* **2018**, *12*, 046029. [\[CrossRef\]](#)
100. Zhou, G.; Su, S.; Xu, J.; Tian, Z.; Cao, Q. Bathymetry Retrieval From Spaceborne Multispectral Subsurface Reflectance. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 2547–2558. [\[CrossRef\]](#)
101. Singh, R.K.; Singh, P.; Drews, M.; Kumar, P.; Singh, H.; Gupta, A.K.; Govil, H.; Kaur, A.; Kumar, M. A machine learning-based classification of LANDSAT images to map land use and land cover of India. *Remote Sens. Appl. Soc. Environ.* **2021**, *24*, 100624. [\[CrossRef\]](#)
102. Elith, J.; Leathwick, J.R.; Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **2008**, *77*, 802–813. [\[CrossRef\]](#)
103. Wu, J.; Zhong, B.; Tian, S.; Yang, A.; Wu, J. Downscaling of urban land surface temperature based on multi-factor geographically weighted regression. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 2897–2911. [\[CrossRef\]](#)
104. Ghosh, A.; Joshi, P. Hyperspectral imagery for disaggregation of land surface temperature with selected regression algorithms over different land use land cover scenes. *ISPRS J. Photogramm. Remote Sens.* **2014**, *96*, 76–93. [\[CrossRef\]](#)
105. Liu, F.; Wang, X.; Sun, F.; Wang, H.; Wu, L.; Zhang, X.; Liu, W.; Che, H. Correction of Overestimation in Observed Land Surface Temperatures Based on Machine Learning Models. *J. Clim.* **2022**, *35*, 5359–5377. [\[CrossRef\]](#)
106. Keramitsoglou, I.; Kiranoudis, C.T.; Weng, Q. Downscaling geostationary land surface temperature imagery for urban analysis. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1253–1257. [\[CrossRef\]](#)
107. Han, D.; An, H.; Cai, H.; Wang, F.; Xu, X.; Qiao, Z.; Jia, K.; Sun, Z.; An, Y. How do 2D/3D urban landscapes impact diurnal land surface temperature: Insights from block scale and machine learning algorithms. *Sustain. Cities Soc.* **2023**, *99*, 104933. [\[CrossRef\]](#)
108. Khan, M.; Qasim, M.; Tahir, A.A.; Farooqi, A. Machine learning-based assessment and simulation of land use modification effects on seasonal and annual land surface temperature variations. *Heliyon* **2023**, *9*, e23043. [\[CrossRef\]](#) [\[PubMed\]](#)
109. Luo, J.; Wang, Y.; Li, G. The innovation effect of administrative hierarchy on intercity connection: The machine learning of twin cities. *J. Innov. Knowl.* **2023**, *8*, 100293. [\[CrossRef\]](#)
110. Ali, S.; Eum, H.-I.; Cho, J.; Dan, L.; Khan, F.; Dairaku, K.; Shrestha, M.L.; Hwang, S.; Nasim, W.; Khan, I.A. Assessment of climate extremes in future projections downscaled by multiple statistical downscaling methods over Pakistan. *Atmos. Res.* **2019**, *222*, 114–133. [\[CrossRef\]](#)
111. Hussain, S.; Mubeen, M.; Nasim, W.; Mumtaz, F.; Abdo, H.G.; Mostafazadeh, R.; Fahad, S. Assessment of future prediction of urban growth and climate change in district Multan, Pakistan using CA-Markov method. *Urban Clim.* **2024**, *53*, 101766. [\[CrossRef\]](#)
112. NASA. Albedo Values. Available online: <https://mydasdata.larc.nasa.gov/basic-page/albedo-values> (accessed on 19 January 2024).
113. Julien, Y.; Sobrino, J.A.; Verhoef, W. Changes in land surface temperatures and NDVI values over Europe between 1982 and 1999. *Remote Sens. Environ.* **2006**, *103*, 43–55. [\[CrossRef\]](#)

114. Kaufmann, R.K.; Zhou, L.; Myneni, R.; Tucker, C.J.; Slayback, D.; Shabanov, N.; Pinzon, J. The effect of vegetation on surface temperature: A statistical analysis of NDVI and climate data. *Geophys. Res. Lett.* **2003**, *30*. [[CrossRef](#)]
115. Sun, D.; Kafatos, M. Note on the NDVI-LST relationship and the use of temperature-related drought indices over North America. *Geophys. Res. Lett.* **2007**, *34*. [[CrossRef](#)]
116. Mansourmoghaddam, M.; Rousta, I.; Zamani, M.; Mokhtari, M.H.; Karimi Firozjaei, M.; Alavipanah, S.K. Study and prediction of land surface temperature changes of Yazd city: Assessing the proximity and changes of land cover. *J. RS GIS Nat. Resour.* **2022**, *12*, 1–5.
117. Guha, S.; Govil, H.; Gill, N.; Dey, A. A long-term seasonal analysis on the relationship between LST and NDBI using Landsat data. *Quat. Int.* **2021**, *575*, 249–258. [[CrossRef](#)]
118. Huang, H.; Huang, J.; Wu, Y.; Zhuo, W.; Song, J.; Li, X.; Li, L.; Su, W.; Ma, H.; Liang, S. The improved winter wheat yield estimation by assimilating GLASS LAI into a crop growth model with the proposed Bayesian posterior-based ensemble Kalman filter. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–18. [[CrossRef](#)]
119. Zheng, H.; Fan, X.; Bo, W.; Yang, X.; Tjahjadi, T.; Jin, S. A multiscale point-supervised network for counting maize tassels in the wild. *Plant Phenomics* **2023**, *5*, 0100. [[CrossRef](#)] [[PubMed](#)]
120. Wang, W.; Samat, A.; Abuduwaili, J.; Ge, Y. Quantifying the influences of land surface parameters on LST variations based on GeoDetector model in Syr Darya Basin, Central Asia. *J. Arid Environ.* **2021**, *186*, 104415. [[CrossRef](#)]
121. Nardino, M.; Laruccia, N. Land Use Changes in a Peri-Urban Area and Consequences on the Urban Heat Island. *Climate* **2019**, *7*, 133. [[CrossRef](#)]
122. Muthiah, K.; Mathivanan, M.; Duraisekaran, E. Dynamics of urban sprawl on the peri-urban landscape and its relationship with urban heat island in Chennai Metropolitan Area, India. *Arab. J. Geosci.* **2022**, *15*, 1694. [[CrossRef](#)]
123. Jato-Espino, D. Spatiotemporal statistical analysis of the Urban Heat Island effect in a Mediterranean region. *Sustain. Cities Soc.* **2019**, *46*, 101427. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.