

Article

Machine Learning in Urban Tree Canopy Mapping: A Columbia, SC Case Study for Urban Heat Island Analysis

Grayson R. Morgan ^{1,*}, Alexander Fulham ² and T. Grant Farmer ²¹ Department of Geography, Brigham Young University, Provo, UT 84602, USA² Department of Geography, University of South Carolina, Columbia, SC 29208, USA

* Correspondence: grmorgan2013@gmail.com

Abstract: As the world's urban population increases to the predicted 70% of the total population, urban infrastructure and built-up land will continue to grow as well. This growth will continue to have an impact on the urban heat island effect in all of the world's cities. The urban tree canopy has been found to be one of the few factors that can lessen the effects of the urban heat island effect. This study seeks to accomplish two objectives: first, we examine the use of a commonly used machine learning classifier (e.g., Support Vector Machine) for identifying the urban tree canopy using no-cost high resolution NAIP imagery. Second, we seek to use Land Surface Temperature (LST) maps derived from no-cost Landsat thermal imagery to identify correlations between canopy loss and temperature hot spot increases over a 14-year period in Columbia, SC, USA. We found the SVM imagery classifier was highly accurate in classifying both the 2005 imagery (94.3% OA) and the 2019 imagery (94.25% OA) into canopy and other classes. We found the color infrared image available in the 2019 NAIP imagery better for identifying canopy than the true color images available in 2005 (97.8% vs. 90.2%). Visual analysis based on the canopy maps and LST maps showed temperatures rose near areas where tree canopy was lost, and urban development continued. Future studies will seek to improve classification methods by including other classes, other ancillary data sets (e.g., LiDAR), new classification methods (e.g., deep learning), and analytical methods for change detection analysis.

Keywords: urban tree canopy; urban heat island; machine learning; visual analysis

Citation: Morgan, G.R.; Fulham, A.; Farmer, T.G. Machine Learning in Urban Tree Canopy Mapping: A Columbia, SC Case Study for Urban Heat Island Analysis. *Geographies* **2023**, *3*, 359–374. <https://doi.org/10.3390/geographies3020019>

Academic Editor: Pedro Cabral

Received: 14 March 2023

Revised: 24 April 2023

Accepted: 28 April 2023

Published: 16 May 2023



Copyright: © 2023 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/>).

1. Introduction

Rapid urbanization is challenging our ability to create sustainable and safe cities for the 55% of the world's population that live there. It is predicted that by 2050 nearly 70% of the world's population will live in an urban area [1]. Urbanization has a widely recognized impact on the local weather and climate, and one of the most common ways is the Urban Heat Island (UHI) effect. An UHI can be defined as an area where the urban build up creates a microclimate that is warmer than the surrounding rural areas. UHIs are caused by several factors, including the large extent of impervious surface materials covering a large portion of the area in a given city, the layout of urban environments, anthropogenic activities that create heat, and the removal of natural landscapes [2–4]. The UHI consequently contributes to several environmental changes, such as vegetation growth and water and air quality. The increased heat can also severely impact the health of all those who live in the city [5,6]. In response to the historic growth and in hope for a more sustainable future for our cities, the UN established the sustainable development goal 11 in 2015 [7]. This requires a better understanding of how the heat landscape of a city, as well as the mitigating factors such as tree canopy, are changing over time so that we can adjust course and be more protective and sustainable in our approaches to development.

In the many studies that have been conducted thus far to better understand the UHI phenomenon, UHI has been separated into two classes: Atmospheric UHI and Surface UHI. Atmospheric UHI is monitored with in situ sensors that can be expensive to install and use.

Surface UHI can be observed using remote sensing data, as described by [8]. Observing the surface UHI phenomenon using surface temperatures collected by remote sensing instruments (e.g., Landsat, MODIS) allows for spatial and temporal analysis of cities all around the globe from historic and modern imagery collections. Extensive work has been completed to examine the UHI phenomenon across the globe [9–13]. While there are many benefits to examining UHI via remote sensing technologies, such as large areal coverage and low cost, there are also well documented limitations. For example, the longwave emissions recorded by the typical sensor on a satellite or aircraft are only a small fraction of the complex radiation and processes that impact UHI [14,15]. Furthermore, when daytime remote sensing analysis is used rather than nighttime, spatial patterns can be unexpected due to the UHI being attributed to the heat storage in urban materials and impervious surfaces [14,15]. Despite these limitations and others, remote sensing technologies are able to provide a view of UHI that is unique and can be useful [15].

Vegetation has been identified as a mitigating factor of UHI and the heat impacts on humans living in the cities. For example, a tree canopy can affect building energy use [16], reduce daytime urban heat in the summer [17], and urban stormwater management [18,19]. Therefore, a tree canopy is an important attribute of the urban environment that requires regular monitoring. As such, government organizations in cities are often formed to regulate tree removal and planting throughout urban areas. Despite the efforts of these organizations to control canopy gain and loss, city growth has inevitably led to a decrease in a canopy over time for many cities. Obtaining and managing up-to-date tree canopy datasets can be costly and time consuming for entire cities. Many studies where mapping the canopy was the first objective relied on LiDAR data to map the tree canopy, but LiDAR data can be costly and a limited resource [20–22]. High resolution historical aerial or satellite imagery can be used to identify tree canopy changes over time [23–25].

While several studies have investigated the UHI effect in a city or worked to map the urban tree canopy, few studies have brought both elements together to investigate spatial patterns and correlations to show changes over time. Many studies have acknowledged the connection between UHI effect and vegetation cover [26–28]. More recent studies have made further progress, but often using expensive LiDAR data or high-resolution satellite data rather than aerial imagery. Furthermore, [29] found that even just a 10% increase in forest can decrease the UHI effect by 0.83 °C, and [30] examined the role of the canopy structure on both nighttime and daytime UHI effects by using daytime and nighttime MODIS heat map data and a 2-m satellite and LiDAR-derived tree canopy and found that the tree canopy showed a stronger influence in the day. One study also examined change detection of canopy loss and heat change over an 8-year period for a moderately sized city in Massachusetts, but used costly LiDAR and satellite imagery to create the canopy [20]. Many other studies regarding the effect of trees on elements of the UHI effect have been carried out on a very large scale, rather than city wide [17–19]. This study proposes the use of each of the aforementioned remote sensing resources to look at the effect canopy changes have had on the UHI of a moderately sized city, Columbia, SC, USA, over a 14-year period. The objectives of this study were two-fold:

- A. Examine the accuracy of using a machine learning (ML) object-based image classifier for creating a high resolution (1- meter) tree canopy estimation from free NAIP imagery for a moderately sized city.
- B. Identify spatial patterns and correlations between tree canopy change and surface heat change over 14 years in Columbia, SC.

To accomplish these objectives, we propose the use of a support vector machine (SVM) ML classification algorithm to create a tree canopy map for two dates of high-resolution aerial imagery. Furthermore, we will use 30-m spatial resolution Landsat satellite data for obtaining surface temperature information to inform our visual analysis. The results of the study will produce change maps that can be used for visibly identifying areas of change and concern.

2. Materials and Methods

2.1. Study Area

Columbia, South Carolina (SC), USA, is the capital city in the state of South Carolina and home to over 130,000 people [31]. The present study examines the city of Columbia according to its present limits rather than its metropolitan area (Figure 1). The city of Columbia rests at the fall line, or the boundary, of the piedmont region and the Atlantic coastal plane. The city was built along the Congaree river, which forms at the confluence of the Broad and Saluda Rivers. Columbia has a humid subtropical climate with hot and humid summers. Sitting at the center of the state in between the blue ridge mountains and the Atlantic coast, the city encounters some of the state of South Carolina's hottest temperatures.



Figure 1. Location of Columbia, SC in the southeastern United States of America.

2.2. Data

Two sets of open-source remote sensing data were used for each objective, respectively. First, National Agriculture Imagery Program (NAIP) Images were used for classifying the tree canopy, among other land cover classes. This program began in 2002 and is administered by the U.S. Department of Agriculture (USDA) Farm Service Agency. Aerial imagery is collected during growing seasons, always with leaf-on conditions. The digital sensors used for NAIP imagery collection meet rigid calibration specifications [32]. NAIP imagery is generally collected at a 1-m spatial resolution (50–60 cm in some areas) across the conterminous United States. In order to provide a more accurate comparison between the 2005 (1 m) and 2019 (60 cm) data, the 2019 data were resampled to a 1-m spatial resolution using the *Resample* tool in ArcGIS Pro before processing and analysis. NAIP imagery has previously been shown to be an effective remote sensing resource for mapping urban tree canopy [33].

Temperature data for the Columbia, SC area were derived from band 6 of the Landsat 7 Enhanced Thematic Mapper Plus (ETM+) sensor. Landsat 7 was launched in 1999 and notably experienced a failure of the Scan Line Correction (SLC) in 2003 that affects the use of Landsat 7 data. However, tools have been created to correct for the scan lines in ArcGIS software, making the imagery available for other purposes. Thermal imagery data for July of 2005 and 2019, the same years as the NAIP imagery being used for the tree canopy estimation, were chosen to examine spatial heat changes in Columbia. The imagery data were originally collected at a 120-m resolution but were resampled to 30 m

using the *Resample* tool in ArcGIS Pro for comparative analysis. The data collection occurred on 26 July 2005 at 11:49 AM local time with the sun azimuth at 117.34 degrees and sun elevation at 62.64 degrees for the first collection. The second collection occurred on 1 July 2019 at 11:49 AM local time with the sun azimuth at 111.18 degrees and sun elevation at 65.34 degrees. Other information regarding the Landsat and NAIP data is presented in Table 1.

Table 1. Imagery Characteristics.

Imagery Name	Imagery Date	Imagery Type	Imagery Purpose	Imagery Resolution	Image Spectral Characteristics
NAIP 2005	23 June and 23 July 2005	Aerial	Tree Canopy Mapping	1 m	RGB only
NAIP 2019	27 July–30 July 2019	Aerial	Tree Canopy Mapping	60 cm	CIR
LANDSAT 7 2005	26 July 2005	Satellite	Heat Mapping	120 (30) m	10.40–12.50 μm
LANDSAT 7 2019	1 July 2019	Satellite	Heat Mapping	120 (30) m	10.40–12.50 μm

2.3. Methods

2.3.1. Canopy Classification

Canopy classification was completed using National Agriculture Imagery Program (NAIP) images from 2005 and 2019. Each image was first clipped to the extent of the city of Columbia shapefile. Fort Jackson and much of the Congaree, Broad, and Saluda rivers were removed. Figure 2 depicts the general process for obtaining a canopy classification.

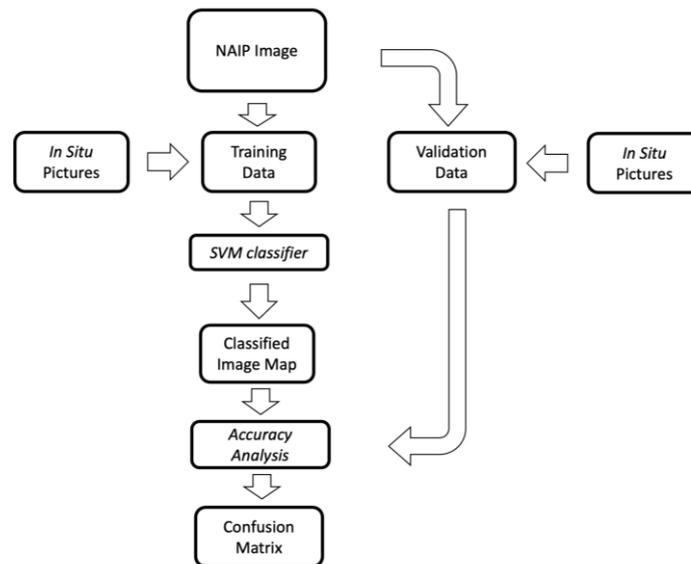


Figure 2. General process for obtaining canopy classification.

Canopy classification was performed using SVM object-based classification in ESRI’s ArcGIS Pro software. A major advantage of object-based classifications is the ability to overcome the “salt and pepper” effect that can occur in pixel-based classifications. This phenomenon refers to rogue single pixels or small pixel clusters that are classified as a particular class, though the pixel is nowhere near other pixels of that class. Support Vector Machine classifiers use a supervised classification method based on statistical learning theory. It was developed in the computer science community in the 1990s and has since been adapted to remote sensing image classification [34], pg. 337. A SVM classifier computes a maximal margin hyperplane where the hyperplane separates the classes (two in the case of Figure 3), and is the furthest away possible from any of the training data. The boundary observations are transformed into a slab and termed ‘support vectors’. SVM classifiers allow nonlinear decision boundaries in more complex datasets than the one shown in Figure 3.

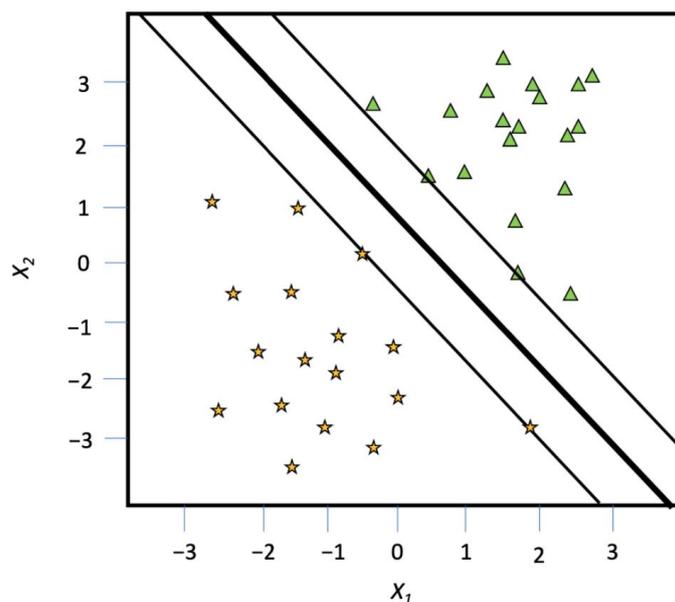


Figure 3. Example of using support vectors to separate two classes.

Training data polygons were digitized throughout each image for the following classes: tree canopy, urban, water, grass/agriculture, shadow, and barren. Each of the polygons were digitized from visual interpretations of the NAIP imagery or Google Earth Imagery. Visits to and pictures from various sites throughout the study area were only used for the 2019 training data, as there have been several changes since 2005. For the 2005 image, 1202 total polygons were digitized for training across all 6 target classes. For the 2019 image, there were 1253 total training polygons for all 6 target classes. A map showing the makeup of the training data can be seen in Figure 4. The goal in digitizing training data was to maintain as accurate of a database as possible. This was partly accomplished by focusing on areas that the authors were most familiar with, and also by spreading the training data as evenly as possible. The authors believe an appropriate balance was struck.

After preparing extensive training data, the image was segmented into objects for classification. In ArcGIS Pro, segmentation parameters include a 1–20 scale for spatial characteristics of the data and a 1–20 scale for spectral characteristics, as well as a minimum segment size. Spectral detail was attributed the highest importance, with spatial detail coming in second. In the range of 1.0 to 20.0, spectral detail was placed at 18, while spatial detail was placed at a 16 in an effort to smooth out the image while also maintaining canopies as best as possible. The minimum segment size was 5 pixels (5×5 m) to accommodate smaller buildings and single trees.

The SVM ML classifier was trained using all of the training polygons, focusing on the following segment or object properties: active chromaticity color, mean digital number, standard deviation, and compactness. Each image used the same characteristics. While only an RGB image was available for 2005, a 2019 color infrared (CIR) image was available and used for classification. With the CIR image, trees and tree canopies were visually more easily distinguishable from other classes. After classification was complete, classes that did not include a tree canopy were combined to become a “no canopy” class, while the canopy class was left alone. This was carried out to provide a focus on the spatial distribution of the tree canopy.

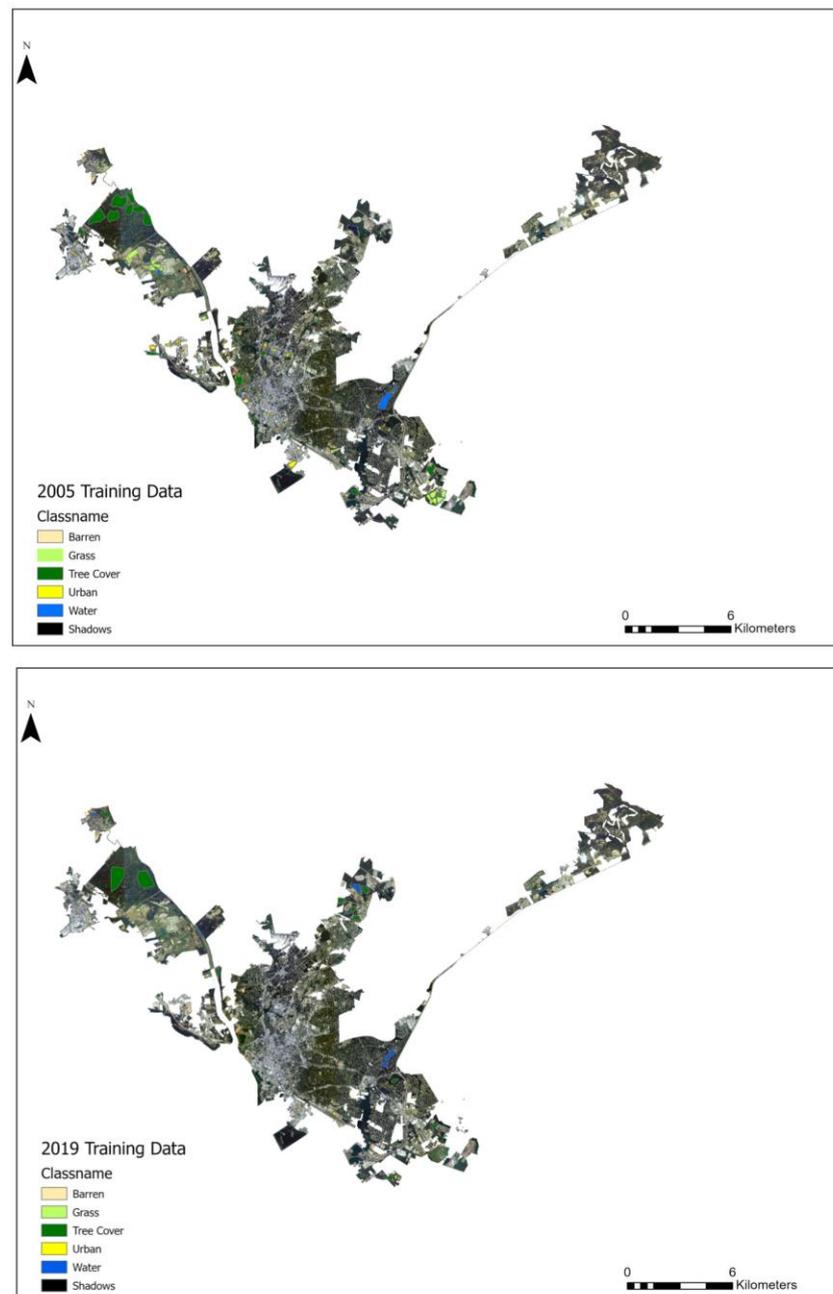


Figure 4. Training polygons used for 2005 and 2019 tree canopy classifications.

2.3.2. Canopy Classification Accuracy Assessment

Accuracy assessment metrics were used to compare the accuracy of the 2005 and 2019 classifications. A confusion matrix was calculated using ArcGIS Pro's *Compute Confusion Matrix* tool. Four main statistics were used to assess accuracy: the producer's accuracy, the user's accuracy, overall accuracy, and kappa. The producer's accuracy is the total number of pixels correctly classified for a class divided by the total number of pixels in that class as known from the ground truthing data. The user's accuracy is the total number of pixels correctly classified into a class, divided by the total number of pixels classified into that class. The overall accuracy (OA) percentage was calculated as follows:

$$OA = \frac{\sum_{i=1}^K x_{ii}}{N} \quad (1)$$

where x_{ii} represents a correctly classified pixel, and N is the total number of pixels being assessed.

Kappa analysis is a multivariate technique for accuracy assessment first published in [35]. A Kappa statistic is similar to overall accuracy, but each considers slightly different elements of data. A kappa estimate (\widehat{K}) was determined as described in [36]:

$$\widehat{K} = \frac{N \sum_{i=1}^k x_{ii} - \sum_{i=1}^k (x_{i+} \times x_{+j})}{N^2 - \sum_{i=1}^k (x_{i+} \times x_{+j})} \quad (2)$$

where N is the total number of samples, k is the number of rows in the confusion matrix, x_{ii} is the number of observations in row i and column i , and x_{i+} and x_{+j} are the marginal totals for row i and column j .

The ground truth data used to validate the classification in the aforementioned statistics came from digitized polygons of visually determined obvious land cover truths. The authors took exceptional care to only use areas that they were confident only contained the class to be validated. The validation was completed with the two classes only- tree cover and the joined class from the several other classes including urban, agriculture/grass, shadows, bare ground, and water. Using the digitized polygons, 600 random validation points were created in the polygons per class. The validation points saved the ground truth value (tree canopy or now), as well as the classified value for comparison and calculation.

2.3.3. Surface Temperature Mapping

The surface temperature maps were generated from Landsat 7 Collection 2 Level 2 satellite data captured on 26 July 2005 and 1 July 2019. As mentioned above, the SLC failed in 2003, creating missing data errors in each recorded image until the mission's conclusion in 2022. In order to use the Landsat 7 datasets for visual analysis, Landsat Toolbox (<https://drive.google.com/file/d/1FKc-G1vMVtWXi66hh15VKoj-zk3UPWgX/view> (accessed on 20 April 2023)) was incorporated into ArcGIS Pro. The tool *Fix Landsat 7 Seamline Errors* was used to interpolate between the known values and the no data pixels. Data were geometrically and radiometrically corrected by the data provider (USGS) before use [37–39]. Land surface temperature data were calculated and derived by the USGS before data download and are described below. A general depiction of the process can be found in Figure 5.

To complete the conversion of the raw values to surface temperature in degrees Celsius, several steps were required. First, the raw values were converted into radiance [37]. This was achieved using the equation:

$$L_{\lambda} = \left(\frac{LMAX_{\lambda} - LMIN_{\lambda}}{QCALMAX - QCALMIN} \right) * (QCAL - QCALMIN) + LMIN_{\lambda} \quad (3)$$

where L_{λ} is the spectral radiance, $QCAL$ is the quantized calibrated pixel value in digital number (DN), $LMAX_{\lambda}$ is the spectral radiance scaled to $QCALMAX$, $LMIN_{\lambda}$ is the spectral radiance scaled to $QCALMIN$, $QCALMIN$ is the minimum quantized calibrated pixel value in DN, and $QCALMAX$ is the maximum quantized calibrated pixel value. These values were obtained from the metadata file provided with the Landsat imagery [37].

To convert the radiance values to brightness temperature values in degrees Kelvin, the radiance values for every pixel were passed through the following equation:

$$T = \frac{K2}{\ln\left(\frac{K1}{L_{\lambda}} + 1\right)} \quad (4)$$

where T is the temperature in degrees Kelvin, $K1$ is calibration constant 1 provided in the metadata, $K2$ is calibration constant 2 also provided in the metadata, and L_{λ} is the spectral radiance in (Watts/(m²*sr*μm)) [37]. From here, further analysis was needed to translate these values into LST. [15,37,40–43] described the use of a single channel

algorithm to convert the brightness temperatures at the sensor to surface temperatures on the earth. Landsat 7 data have only one thermal band, and therefore only a single channel algorithm would suffice. Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Emissivity Dataset (GED) data, ASTER Normalized Difference Vegetation Index (NDVI) data, and atmospheric profiles of geopotential height, specific humidity, and air temperature extracted from reanalysis data were combined in a final calculation to obtain LST values in Kelvin. Finally, the LST was computed using something similar to Equation (5) [15,37,40,41].

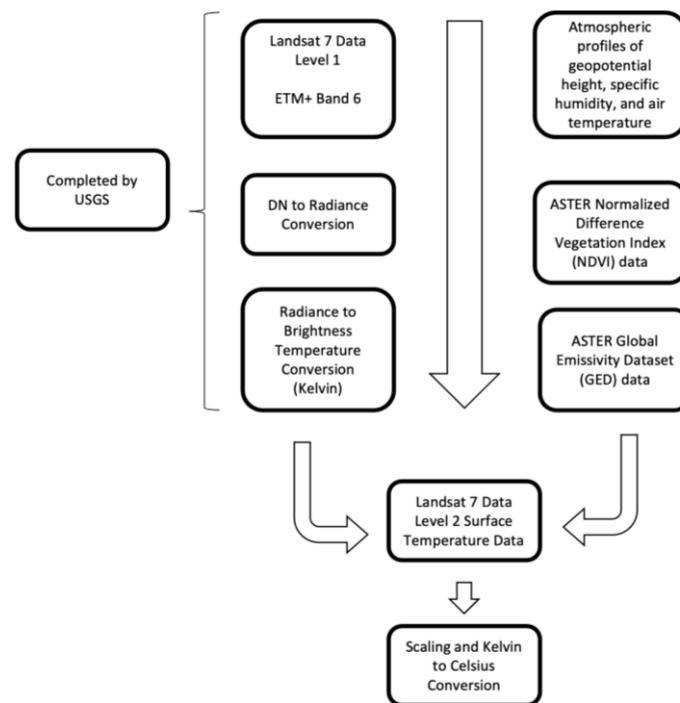


Figure 5. General description of the process for obtaining LST.

$$LST = (T / (1 + (\lambda * T / \rho) * \ln(E))) \quad (5)$$

where LST is the temperature, with correction by emissivity ($^{\circ}\text{K}$); T is the temperature of the brightness at the sensor ($^{\circ}\text{K}$); λ is the wavelength of the emitted radiance; E is the emissivity; ρ is deduced from Equation (6) [15,40–42].

$$\rho = h \frac{c}{\sigma} = (1.438 \times 10^{-2} \text{ mk}) \quad (6)$$

where σ is the Boltzmann constant ($1.38 \times 10^{-23} \text{ J/K}$), h is Planck's constant ($6.626 \times 10^{-34} \text{ Js}$), and c is the speed of light ($2998 \times 10^8 \text{ m/s}$).

When the data were obtained from USGS EarthExplorer (<https://earthexplorer.usgs.gov> (accessed on 20 April 2023)), the steps above were already completed. To scale the values into their original Kelvin values, pixel values were multiplied by 0.00341802 and added to 149 [44]. The new Kelvin values were easily transformed into degrees Celsius by subtracting 273.15.

Quality assessment data were available for the Landsat Level 2 Surface Temperature Data in a separate band where each pixel value describes uncertainty in the temperature values. Pixels closer to clouds and with higher uncertainty had higher overall values in the quality assessment band. The band was examined to identify regions where temperature values were less likely to be correct [38,39].

Each map was visualized using the natural breaks statistical classification for the data. While each data class did not match between the two dates, the natural breaks method was

employed rather than a manual method to preserve the areas of hot and cold across the region. The city and surrounding areas experienced a unique temperature range each day the data were captured; therefore, the authors believe that “hot areas” should be compared in relation to other “hot areas” between the dates.

3. Results

3.1. Tree Canopy Mapping

Tree canopy mapping using the two NAIP images resulted in two land cover maps, one displaying the estimated tree canopy from 2005 and the other estimating the tree canopy from 2019. The results of the classifications can be seen in Figures 6 and 7. Training for each of the images took less than 4 h. After training, classification only required 2.5 h using a Dell Inspiron 5680 6-core intel i7 CPU with 16 gb Ram and Nvidia GTX 1060 3 gb GPU.

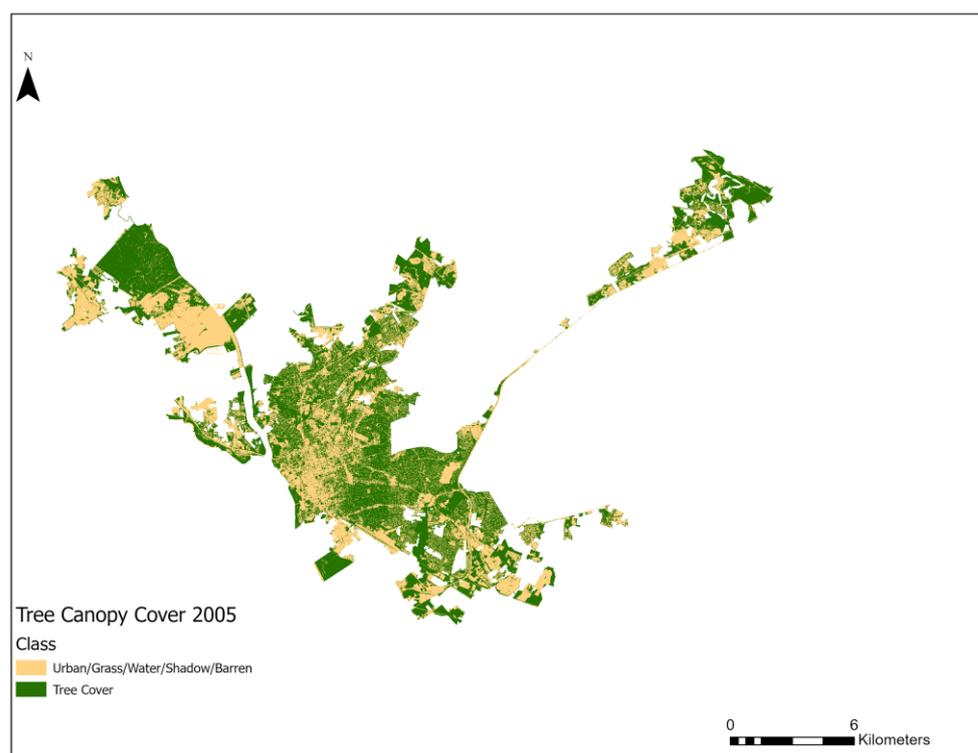


Figure 6. Tree canopy maps derived from the NAIP imagery from 2005.

The accuracy assessment conducted for each classification showed that the canopy classifications performed quite well (Tables 2 and 3). The overall accuracies for both classifications were just over 94% and the kappa statistics were very similar as well. For the 2005 NAIP image classification, the ‘other’ category had a much better users accuracy than producers accuracy, and the tree canopy class had a better producers accuracy than users accuracy. For the 2019 NAIP image classification, we found the exact opposite to be true for each of the ‘other’ class and tree canopy class. Overall accuracy presented in the accuracy assessment suggests confidence in the overall classification.

Visibly, however, there were still some discrepancies in the classification. The canopy class, when it did get confused, was often mistaken for shadows and grass, or vice versa. This was especially apparent in the 2005 image where a near infrared (NIR) band was not included during data capture. This did not seem to be an issue with the 2019 map. The canopy cover map for 2019 showed a significant loss of canopy in Columbia. Some of the changes in canopy could be attributed to development outside of the center of the city, and other tree removal projects that have occurred, including a Bull Street development project in downtown Columbia. Others, including the thinning of canopy throughout the entirety of downtown Columbia, are more difficult to account for. Some of the changes may be

attributed to the classification difficulties discussed above in 2005 and not in 2019. Others are more likely to be trees being cut down in yards as nuisances.

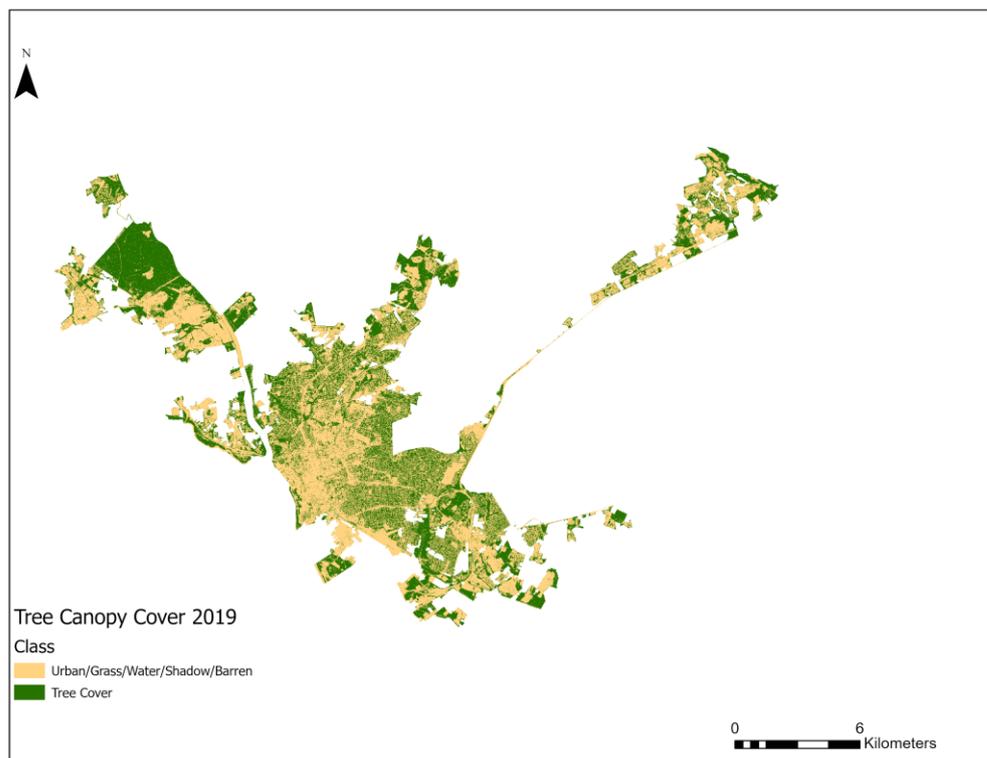


Figure 7. Tree canopy maps derived from the NAIP imagery from 2019.

Table 2. 2005 Classification Confusion Matrix.

	“Other”	Tree Canopy	Total	Users Accuracy	Kappa
“Other”	535	3	538	0.994	-
Tree Canopy	65	597	662	0.902	-
Total	600	600	1200	-	-
Producers Accuracy	0.892	0.995	-	0.943	-
Kappa	-	-	-	-	0.887

Table 3. 2019 Classification Confusion Matrix.

	“Other”	Tree Canopy	Total	Users Accuracy	Kappa
“Other”	588	57	645	0.912	-
Tree Canopy	12	543	555	0.978	-
Total	600	600	1200	-	-
Producers Accuracy	0.98	0.905	-	0.942	-
Kappa	-	-	-	-	0.885

3.2. Surface Temperature Maps

The land surface temperature (LST) maps made from the Landsat 7 tiles are shown in Figure 6. Each of the maps show interesting surface temperature patterns over the 14-year period. Many of the assumed patterns hold true: surface temperature is highest in both maps where concrete and other impervious surfaces are in place throughout the urban environments. Furthermore, known forests within city limits and the tree canopy dispersed

throughout the city show relatively cool temperatures in comparison with the impervious surfaces. As described in Section 2.3.3, the ranges of values between the two map dates are different and the statistical makeup of temperatures in the area are also different. It is important to remember while interpreting the surface temperature maps that while the temperature breakdowns are different, we are able to visibly determine “hot spots” of urban heat from the current color scheme for each date.

Quality assessment maps were created from the Landsat collection quality assessment bands for surface temperature. These maps were overlaid on top of the heat maps to show where the largest areas of uncertainty were by using a highlighter effect (Figure 8). These maps indicate that the region where the Saluda and Broad Rivers join to make the Congaree river provides consistent low-confidence temperature values. Other regions around the city and study area changed in quality from 2005 to 2019. Overall, most of the study area had low quality assessment values that represented trustworthy temperature data.

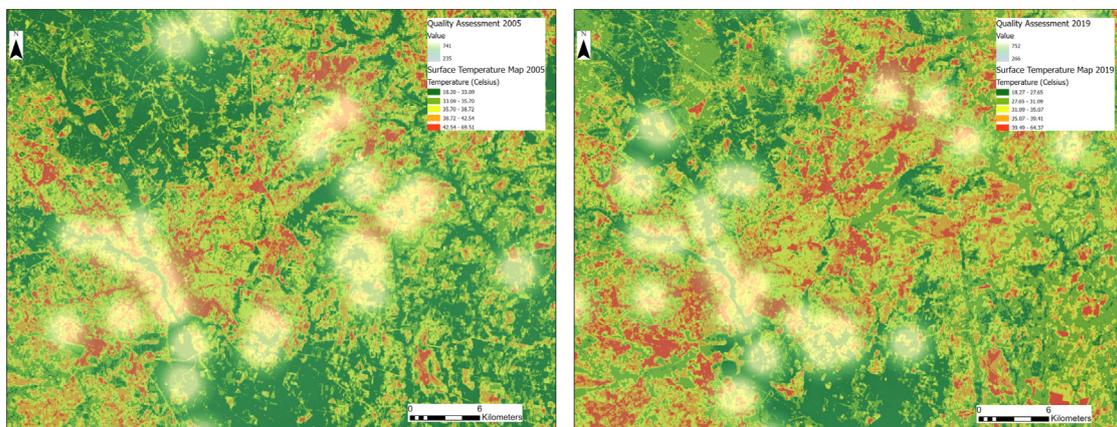


Figure 8. 2005 and 2019 Surface Temperature Quality Assessment maps.

Between the 2005 and 2019 heat maps (Figure 9), it is possible to discern that an apparent cooling of downtown Columbia has occurred (in pink). This apparent cooling has occurred despite the apparent loss of tree canopy. Several areas that showed high surface heat levels in 2005 once again showed at least a portion of that heat again in 2019. Several areas within our study area and just beyond, where other urban growth has occurred, experienced an increase in surface heat levels (in blue circles).

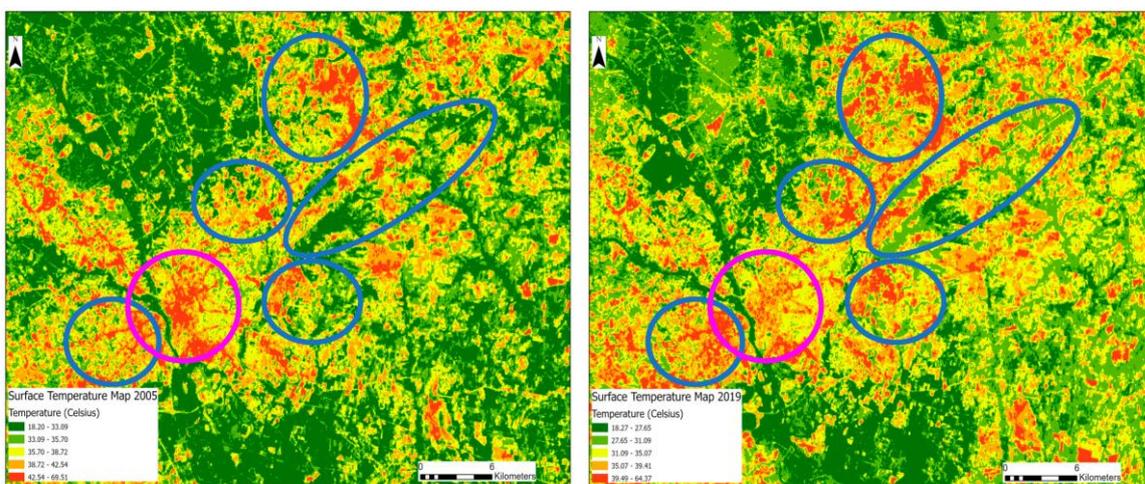


Figure 9. 2005 and 2019 Surface Temperature maps for the Columbia Area.

3.3. Spatial Patterns and Correlations

Because the canopy data were only computed for Columbia city and not the surrounding towns, we can only speak to the Columbia heat patterns in relation to the tree canopy changes. Some larger canopy losses correlated well with tree canopy loss (Figure 10). Two small areas in the Harbison State Forest (indicated in Purple) lost tree canopy and showed an increase of surface heat. Furthermore, an area of northeast Columbia experienced tree canopy loss and a subsequent increase in surface temperature (indicated in black).

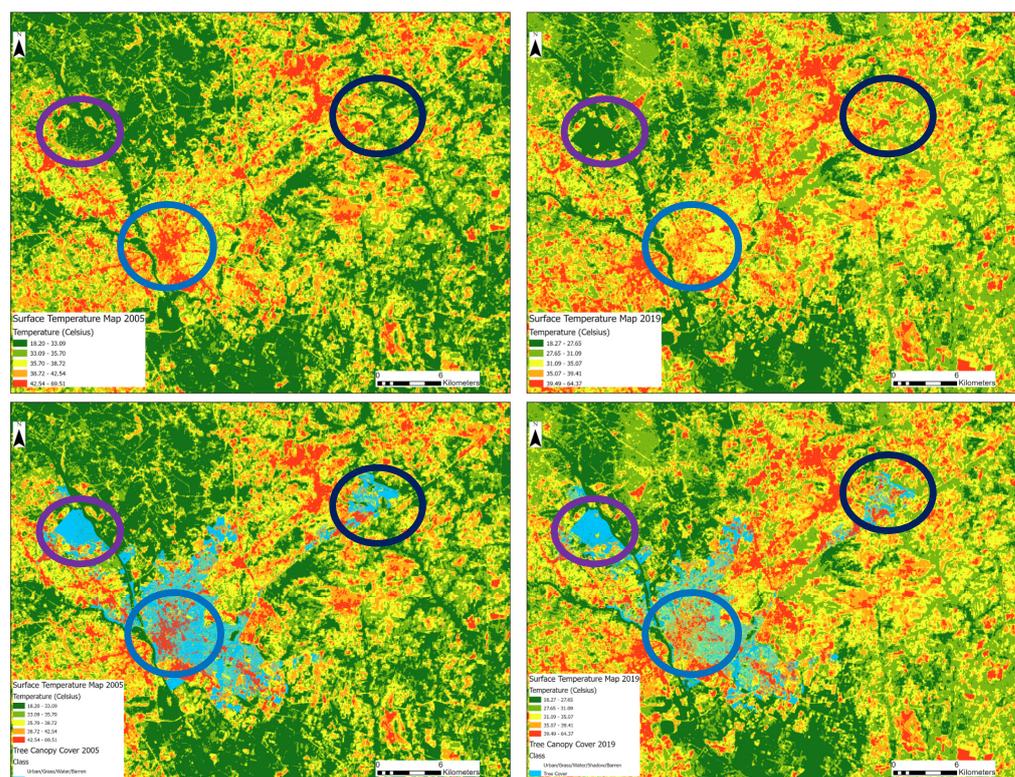


Figure 10. Heat maps for 2005 and 2019 with the tree canopy for Columbia. Circles represent areas of interest: purple: Harbison State Forest; Blue: Downtown Columbia; Black: Loss of canopy with temperature rise.

Negative correlations were also apparent upon visual investigation. For example, the thinning of tree canopy throughout downtown Columbia and many of the suburbs included in the city limits were spatially related to a decrease in surface temperature hot spots. These areas are shown in a dark blue circle.

4. Discussion

The spatial analysis of the LST heat maps and success of the tree canopy mapping followed the successes of other studies. For example, the Harbison State Forest area (Figure 7 circled in purple) is a significant cooling area, similar to the areas [29] discovered that had decreased in overall temperature for that area. Large urban forests are likely to prevent nearby hot spots; indeed, [20] found similar results. A high overall accuracy was achieved in their study using Worldview 2 high resolution satellite imagery to determine urban tree canopy gains and losses, and they used LST maps derived from similar sources to discover hot spots that were highly correlated with new developments and tree canopy loss. Furthermore, [13] found that new built-up land in India was becoming more correlated with urban heat islands, or hot spots, around the city area from 1990 to 2020. Our visual findings linking the tree canopy losses to new hot spots in the Columbia, SC area support the previous findings by [13,20].

The visible LST decrease in the center of Columbia, while the suburbs and surrounding areas became warmer, is an interesting phenomenon. While no definitive conclusion has been drawn, this is possibly due to an increase in vegetation in the downtown areas of Columbia. As regions of the city have undergone growth, the Columbia Tree and Appearance Commission has sought to ensure there is adequate vegetation placed throughout the city [45]. There are several regulations prohibiting the removal of certain trees as well. These policies appear to have impacted the heat island at the center of the city. Furthermore, the warming outside of central Columbia is consistent with the growth the city has encountered. As Columbia grows, developments continue within and outside of the city limits to accommodate the increase in population. While shadows could also provide a less thrilling reason for the decreases in downtown temperature, we argue this is less likely than our other proposition because of the similarities in the built structures in downtown Columbia between 2005 and 2019. There has been some growth to the skyline of downtown Columbia, but the differences are not significant enough to change the amount of ground covered by shadow. Furthermore, the data were collected near peak solar azimuth and elevation to limit the influence of shadows on pixel values.

Each objective of the study encountered challenges while also providing insight into the Columbia, SC, UHI and canopy loss in the area. Canopy mapping required extensive training datasets and significant processing hours, but produced high accuracy tree canopy maps for 2005 and 2019 in Columbia, SC. The accuracy of each classification reached 94%. This is reflective of a high-quality classification and fares well against other urban canopy mapping efforts. For example, for a similar environment in Madison, Wisconsin, ref. [33] achieved 85% accuracy with a SVM classifier. The SVM classifier outperformed other Random Forest (RF) and Boosted Regression Tree (GBM) classifiers. Other studies have found success as well, including ref. [23] achieving 80% OA with Neural Network classification, ref. [46] achieving 89% OA with a Neural Network Classification, and ref. [24] achieving 96% with LiDAR data, ancillary data, and aerial photography including an NIR band.

An urban environment is very complex and introduces several elements for a classifier to overcome, especially when there are only two classes. Visual investigation showed the estimated canopy maps also performed fairly well, though improvements can be made. While the accuracy was remarkably high, it was determined that several of the changes observed in the change detection referred to small canopy changes in residential areas. In these cases, a tree has grown, been cut down, or perhaps there was a slight misalignment to the images and a change was detected, though there was no real change. The 2019 tree canopy classification looks to have fared better than the 2005 tree canopy. The CIR visually and substantially improved the classification. However, the 2005 image unfortunately did not include an NIR band during image capture. In order to overcome these challenges, we believe a few improvements can be made.

Improving the classification for tree canopy can be achieved by including other ancillary data, such as LiDAR data. Furthermore, a different classifier may do a better job at separating the tree canopy class from shadows, grass, and other similar classes. Deep learning remote sensing classifiers are becoming a larger topic of interest and many are performing remarkably well for large scale image analysis [47–50]. Future work should investigate a myriad of classifiers to find one ideal for classifying urban tree canopy. Some of the possible classifiers include the one used in this study, SVM, as well as Random Forest (RF), Deep Lab V3, U-NET, and many others. [51] showed the utility of utilizing several datasets, including LiDAR and vegetation indices, in conjunction with deep learning neural networks to obtain a very high accuracy (96% and 98%) urban canopy map. ArcGIS Pro offers several deep learning and machine learning classifiers using a graphical interface for the use of all GIScientists, regardless of having a technical background or not.

5. Conclusions

This study accomplished two objectives: first, we examined the use of a commonly used machine learning classifier for identifying urban tree canopy using no-cost high resolution NAIP imagery. Second, we used LST maps derived from no-cost Landsat thermal imagery to identify correlations between canopy loss and temperature hot spot increases over a 14-year period in Columbia, SC, USA. The support vector machine imagery classifier was highly accurate in classifying both the 2005 imagery (94.3% OA) and the 2019 imagery (94.25% OA) into canopy and other classes. We found the color infrared image available in the 2019 NAIP imagery better for identifying canopy than the true color images available in 2005 (97.8% vs. 90.2%). Visual analysis based on the canopy maps and LST maps showed temperatures rose near areas where tree canopy was lost and urban development continued. Future studies will seek to improve classification methods by including other classes, other ancillary data sets (e.g., LiDAR), new classification methods (e.g., deep learning), and analytical methods for change detection analysis.

Author Contributions: G.R.M., A.F. and T.G.F.; methodology, G.R.M. and A.F.; software, G.R.M.; validation, G.R.M.; formal analysis, G.R.M.; investigation, G.R.M.; resources, T.G.F.; data curation, G.R.M. and A.F.; writing—original draft preparation, G.R.M.; writing—review and editing, G.R.M.; visualization, G.R.M.; supervision, G.R.M.; project administration, G.R.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NOAA RISA/CAP program, grant number NA16OAR4310163.

Acknowledgments: We would like to acknowledge the guidance and vision provided by Kirstin Dow of the University of South Carolina.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Urban Development. Available online: <https://www.worldbank.org/en/topic/urbandevelopment/overview> (accessed on 28 July 2022).
2. Oke, T.R. The heat island of the urban boundary layer: Characteristics, causes and effects. In *Wind Climate in Cities*; Springer: Dordrecht, The Netherlands, 1995; pp. 81–107. [\[CrossRef\]](#)
3. Nuruzzaman, M. Urban heat island: Causes, effects and mitigation measures—A Review. *Int. J. Environ. Monit. Anal.* **2015**, *3*, 67. [\[CrossRef\]](#)
4. EPA. Heat Island Effect. 2023. Available online: <https://www.epa.gov/heatislands> (accessed on 9 March 2023).
5. Shahmohamadi, P.; Che-Ani, A.I.; Etessam, I.; Maulud, K.N.A.; Tawil, N.M. Healthy environment: The need to mitigate urban heat island effects on human health. *Procedia Eng.* **2011**, *20*, 61–70. [\[CrossRef\]](#)
6. Heaviside, C.; Macintyre, H.; Vardoulakis, S. The urban heat island: Implications for health in a changing environment. *Curr. Environ. Health Rep.* **2017**, *4*, 296–305. [\[CrossRef\]](#)
7. The 17 Goals | Sustainable Development. Available online: <https://sdgs.un.org/goals> (accessed on 28 July 2022).
8. Voogt, J.A.; Oke, T.R. Thermal remote sensing of urban climates. *Remote Sens. Environ.* **2003**, *86*, 370–384. [\[CrossRef\]](#)
9. Kim, H.H. Urban heat island. *Int. J. Remote Sens.* **1992**, *13*, 2319–2336. [\[CrossRef\]](#)
10. Streutker, D.R. A Remote Sensing Study of the urban heat island of Houston, Texas. *Int. J. Remote Sens.* **2002**, *23*, 2595–2608. [\[CrossRef\]](#)
11. Nurwanda, A.; Honjo, T. Analysis of land use change and expansion of surface urban heat island in Bogor City by Remote Sensing. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 165. [\[CrossRef\]](#)
12. Zhang, Y.; Cheng, J. Spatio-temporal analysis of urban heat island using Multisource Remote Sensing Data: A case study in Hangzhou, China. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3317–3326. [\[CrossRef\]](#)
13. Halder, B.; Bandyopadhyay, J.; Banik, P. Monitoring the effect of urban development on urban heat island based on remote sensing and geo-spatial approach in Kolkata and adjacent areas, India. *Sustain. Cities Soc.* **2021**, *74*, 103186. [\[CrossRef\]](#)
14. Parlow, E. Regarding some pitfalls in urban heat island studies using Remote Sensing Technology. *Remote Sens.* **2021**, *13*, 3598. [\[CrossRef\]](#)
15. Almeida, C.R.; Teodoro, A.C.; Gonçalves, A. Study of the urban heat island (UHI) using remote sensing data/techniques: A systematic review. *Environments* **2021**, *8*, 105. [\[CrossRef\]](#)
16. Heisler, G.M. Effects of individual trees on the solar radiation climate of small buildings. *Urban Ecol.* **1986**, *9*, 337–359. [\[CrossRef\]](#)
17. Ziter, C.D.; Pedersen, E.J.; Kucharik, C.J.; Turner, M.G. Scale-dependent interactions between tree canopy cover and impervious surfaces reduce daytime urban heat during summer. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 7575–7580. [\[CrossRef\]](#)

18. Berland, A.; Shiflett, S.A.; Shuster, W.D.; Garmestani, A.S.; Goddard, H.C.; Herrmann, D.L.; Hopton, M.E. The role of trees in urban stormwater management. *Landsc. Urban Plan.* **2017**, *162*, 167–177. [[CrossRef](#)]
19. Kuehler, E.; Hathaway, J.; Tirpak, A. Quantifying the benefits of urban forest systems as a component of the Green Infrastructure Stormwater Treatment Network. *Ecolhydrology* **2017**, *10*, e1813. [[CrossRef](#)]
20. Elmes, A.; Rogan, J.; Williams, C.; Ratick, S.; Nowak, D.; Martin, D. Effects of urban tree canopy loss on land surface temperature magnitude and timing. *ISPRS J. Photogramm. Remote Sens.* **2017**, *128*, 338–353. [[CrossRef](#)]
21. Guo, L.; Chehata, N.; Mallet, C.; Boukir, S. Relevance of airborne lidar and multispectral image data for urban scene classification using random forests. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 56–66. [[CrossRef](#)]
22. Chuang, W.-C.; Boone, C.G.; Locke, D.H.; Grove, J.M.; Whitmer, A.; Buckley, G.; Zhang, S. Tree Canopy Change and neighborhood stability: A comparative analysis of Washington, D.C. and Baltimore, MD. *Urban For. Urban Green.* **2017**, *27*, 363–372. [[CrossRef](#)]
23. Cleve, C.; Kelly, M.; Kearns, F.R.; Moritz, M. Classification of the wildland–urban interface: A comparison of pixel- and object-based classifications using high-resolution aerial photography. *Comput. Environ. Urban Syst.* **2008**, *32*, 317–326. [[CrossRef](#)]
24. MacFaden, S.W.; O’Neil-Dunne, J.P.M.; Royar, A.R.; Lu, J.W.T.; Rundle, A.G. High-resolution tree canopy mapping for New York City using LIDAR and object-based image analysis. *J. Appl. Remote Sens.* **2012**, *6*, 063567. [[CrossRef](#)]
25. Zhou, W.; Troy, A. An object-oriented approach for analysing and characterizing urban landscape at the parcel level. *Int. J. Remote Sens.* **2008**, *29*, 3119–3135. [[CrossRef](#)]
26. Weng, Q.; Lu, D.; Schubring, J. Estimation of land surface temperature–vegetation abundance relationship for urban heat island studies. *Remote Sens. Environ.* **2004**, *89*, 467–483. [[CrossRef](#)]
27. Yuan, F.; Bauer, M.E. Comparison of impervious surface area and normalized difference vegetation index as indicators of surface urban heat island effects in landsat imagery. *Remote Sens. Environ.* **2007**, *106*, 375–386. [[CrossRef](#)]
28. Howe, D.A.; Hathaway, J.M.; Ellis, K.N.; Mason, L.R. Spatial and temporal variability of air temperature across urban neighborhoods with varying amounts of tree canopy. *Urban For. Urban Green.* **2017**, *27*, 109–116. [[CrossRef](#)]
29. Kong, F.; Yin, H.; Wang, C.; Cavan, G.; James, P. A satellite image-based analysis of factors contributing to the green-space cool island intensity on a city scale. *Urban For. Urban Green.* **2014**, *13*, 846–853. [[CrossRef](#)]
30. Chen, J.; Jin, S.; Du, P. Roles of horizontal and vertical tree canopy structure in mitigating daytime and Nighttime Urban Heat Island effects. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *89*, 102060. [[CrossRef](#)]
31. U.S. Census Bureau Quickfacts: Columbia City, South Carolina. Available online: <https://www.census.gov/quickfacts/columbiacitysouthcarolina> (accessed on 29 July 2022).
32. Davis, D. National Agriculture Imagery Program (NAIP) Information Sheet. Available online: https://www.fsa.usda.gov/Assets/USDA-FSA-Public/usdfiles/APFO/support-documents/pdfs/naip_infosheet_2016.pdf (accessed on 9 March 2023).
33. Erker, T.; Wang, L.; Lorentz, L.; Stoltman, A.; Townsend, P.A. A statewide urban tree canopy mapping method. *Remote Sens. Environ.* **2019**, *229*, 148–158. [[CrossRef](#)]
34. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning: With Applications in R*; Springer: New York, NY, USA, 2013.
35. Congalton, R.; Oderwald, R.G.; Mead, R. Assessing Landsat classification accuracy using discrete multivariate statistical techniques. *Photogramm. Eng. Remote Sens.* **1983**, *49*, 1671–1678.
36. Jensen, J.R. *Introductory Digital Image Processing: A Remote Sensing Perspective*, 4th ed.; Pearson Education: Glenview, IL, USA, 2016.
37. Department of the Interior U.S. Geological Survey. Landsat 7 (L7) Data Users Handbook. 2019. Available online: https://d9-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/s3fs-public/atoms/files/LSDS-1927_L7_Data_Users_Handbook-v2.pdf (accessed on 20 April 2023).
38. Department of the Interior U.S. Geological Survey. Landsat Enhanced Thematic Mapper Plus (ETM+) Collection 2 (C2) Level 2 (L2) Data Format Control Book (DFCB). 2020. Available online: https://d9-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/s3fs-public/atoms/files/LSDS-1337_Landsat7ETM-C2-L2-DFCB-v5.pdf (accessed on 20 April 2023).
39. Department of the Interior U.S. Geological Survey. Landsat 4-7 Collection 2 (C2) Level 2 Science Product (L2SP) Guide. 2021. Available online: https://d9-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/s3fs-public/media/files/LSDS-1618_Landsat-4-7_C2-L2-ScienceProductGuide-v4.pdf (accessed on 20 April 2023).
40. Avdan, U.; Jovanovska, G. Algorithm for automated mapping of land surface temperature using Landsat 8 satellite data. *J. Sens.* **2016**, *2016*, 1480307. [[CrossRef](#)]
41. Sobrino, J.A.; Jiménez-Muñoz, J.C.; Paolini, L. Land surface temperature retrieval from Landsat TM 5. *Remote Sens. Environ.* **2004**, *90*, 434–440. [[CrossRef](#)]
42. Sun, Q.; Tan, J.; Xu, Y. An erdas image processing method for retrieving LST and describing urban heat evolution: A case study in the Pearl River Delta region in South China. *Environ. Earth Sci.* **2009**, *59*, 1047–1055. [[CrossRef](#)]
43. Price, J.C. Estimating surface temperatures from satellite thermal infrared data—A simple formulation for the atmospheric effect. *Remote Sens. Environ.* **1983**, *13*, 353–361. [[CrossRef](#)]
44. How Do I Use a Scale Factor with Landsat Level-2 Science Products? Available online: <https://www.usgs.gov/faqs/how-do-i-use-scale-factor-landsat-level-2-science-products> (accessed on 21 April 2023).
45. Columbia Tree and Appearance Commission. 19 September 2022. Available online: <https://boards.columbiasc.gov/columbia-tree-and-appearance-commission/> (accessed on 21 April 2023).

46. Al-Kofahi, S.; Steele, C.; VanLeeuwen, D.; St. Hilaire, R. Mapping land cover in urban residential landscapes using very high spatial resolution aerial photographs. *Urban For. Urban Green.* **2012**, *11*, 291–301. [[CrossRef](#)]
47. Li, H.; Wang, C.; Cui, Y.; Hodgson, M. Mapping salt marsh along coastal South Carolina using U-Net. *ISPRS J. Photogramm. Remote Sens.* **2021**, *179*, 121–132. [[CrossRef](#)]
48. Yuan, Q.; Shen, H.; Li, T.; Li, Z.; Li, S.; Jiang, Y.; Xu, H.; Tan, W.; Yang, Q.; Wang, J.; et al. Deep Learning in Environmental Remote Sensing: Achievements and challenges. *Remote Sens. Environ.* **2020**, *241*, 111716. [[CrossRef](#)]
49. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A Meta-analysis and Review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [[CrossRef](#)]
50. Cheng, G.; Xie, X.; Han, J.; Guo, L.; Xia, G.-S. Remote Sensing Image Scene Classification meets Deep learning: Challenges, methods, benchmarks, and opportunities. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 3735–3756. [[CrossRef](#)]
51. Timilsina, S.; Aryal, J.; Kirkpatrick, J.B. Mapping urban tree cover changes using object-based convolution neural network (OB-CNN). *Remote Sens.* **2020**, *12*, 3017. [[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.