Harnessing Citizen Science to Enhance Land Surface Temperature Prediction

Kevin I. Diaz, Kandyce Diep, Suhani Dondapati, Maako Fangajei, Anna Felten, Kei Fry, Conor Furey, Aaren George

Mentors: Andrew Clark, Rusty Low, Peder Nelson, Cassie Soeffing,

Peer Mentors: Ollie Snow and Riya Tyagi

STEM Enhancement in Earth Science (SEES) Summer High School Intern Program

United States of America

July 19, 2024

1. Abstract

Rapid urban expansion and the proliferation of heat-retaining surfaces such as asphalt and concrete contribute to elevated temperatures in cities, also characterized as surface urban heat islands (SUHI). To increase the accuracy of surface temperature machine learning models, key for urban planning and disaster management, this study harnessed land cover use data collected by citizen scientists. The data comprised downward photos collected by 2024 NASA SEES Earth System Explorers interns using the GLOBE Observer app at 958 sites across the U.S. After labeling for land cover use via Zooniverse, each site was associated with labeled Sentinel-2 satellite data from Collect Earth Online and a mean land surface temperature (LST) Landsat-8 satellite reading for June 2024. Random Forest and XGBoost models were trained in Python on three distinct datasets – coordinates only, coordinates with GLOBE data, and coordinates with GLOBE and Collect Earth Online data – to develop predictive models for LST. Following Bayesian optimization with 10-fold cross-validation, Random Forest displayed R^2 accuracies of 0.84, 0.78, and 0.79, respectively. XGBoost displayed R^2 accuracies of 0.82, 0.80, and 0.82, respectively. While incorporating land cover data failed to improve predictive accuracy, improving data collection methods and ensuring higher quality data could reveal the true value of citizen-sourced land cover data. This research supports a deeper understanding of the complex relationship between land cover and LST, potentially aiding urban planners in mitigating SUHI effects and fostering community engagement in scientific research.

Keywords: *citizen science, land surface temperature, machine learning, urban heat islands*

1.1 Plain Language Summary

This study looks at how cities are getting hotter than surrounding areas because of more concrete surfaces, like buildings and roads, and fewer types of cooling vegetation, like trees and bushes. To better predict city temperatures around the world for urban planning and disaster management, researchers used information about land cover collected by citizen scientists – volunteer members of the public who contribute to scientific research. In 2024, NASA Earth System Explorers virtual interns took ground photos at 958 locations around the world using the GLOBE app. Citizen scientists helped classify the photos by labeling the different land cover types on the ground, such as grass or concrete. Interns matched the photos with satellite imagery taken in the same location, which citizen scientists labeled for land cover. The interns used this data to test two machine learning models and see if adding land cover information would improve temperature predictions compared to GPS coordinates alone. However, results from the model revealed that adding more land cover information did not improve its predictions, which could be explained by our data collection methods. This research aims to better understand how land cover affects temperatures and how communities can engage in scientific research.

2. Introduction

2.1 Research Question

This study aims to utilize Citizen Science to create a machine-learning model trained to predict land surface temperatures. This research question was addressed using GLOBE Observer down photos in conjunction with Landsat-8 data used to train two machine learning models. The hypothesis is that incorporating vegetation data and Collect Earth Online data into the machine learning models with coordinate data would provide greater accuracy for predicting land surface temperatures.

The citizen science used in the research is GLOBE observer data collected by the Earth System Explorer group of the 2024 STEM Enhancement in Earth Science Summer High School Intern Program. Citizen Science is the act of the public completing design experiments, collecting data, analyzing results, and solving problems. Individuals of any background can contribute and offer an opportunity for the public to become involved in issues in their community (National Park Service, 2024).

This paper utilized GLOBE Observer data to understand land surface temperatures of surface urban heat islands all over the United States and parts of Switzerland and India. Urban heat islands are a global issue that contributes to global warming and the rise of land surface temperatures. This research used Adopt a Pixel 3 km framework to create land cover data for a 9 km^2 plot of land from multiple locations in the GLOBE Observer mobile app (Low et al., 2021). In addition, the coordinates of the land covers used in the data were further analyzed based on vegetation in Collect Earth Online and Zooniverse with the aid of other interns in the program.

2.2 Background Information

Urban heat islands are developed areas indicated by comparatively higher land surface temperatures than surrounding areas. These areas pose a risk to children, seniors, and sick people as heat influxes can lead to health issues, including respiratory problems, heat cramps, and exhaustion. Moreover, heat influxes on urban heat islands will require cities to consume more electricity to cool down systems and infrastructures. This creates an unsustainable cycle that needs to be addressed. This phenomenon can be attributed to the abundance of heat-absorbing impervious surfaces and changes in their radiative and thermal properties (Kasniza Jumari et al.).

This study primarily focuses on surface urban heat islands. They can form at any time throughout the day and night, and they vary in severity due to numerous factors; the spatial distribution of different surfaces greatly impacts the land surface temperatures. For example, impervious surfaces such as asphalt, concrete, and brick absorb heat rather than reflect due to their high thermal storage capacity, resulting in high temperatures. In addition, reduced vegetation in cities also contributes to urban heat islands, as dramatically low evapotranspiration levels cannot counteract high surface temperatures. Anthropogenic heat originating from car exhaust and industrial areas can also contribute to urban heat islands (McCartney, 2023).

With substantial increases in population growth and movement into urban centers, it is becoming more important to understand how the temperature increases caused by SUHIs can be mitigated and how land cover plays a part in reinforcing or reducing such temperature fluctuations. Many major cities are currently experiencing urban heat islands and their negative effects. According to a study by Kasniza Jumari et al. (2023) in Kuala Lumpur, Malaysia, temperatures in select urban and rural areas varied from 10.8 and 25.5 °C in 2013 to a staggering 16.1 and 26.73 °C in 2021. This illustrates the increase in urban heat islands and land surface temperature. If not mitigated, these numbers will continue to climb to reach devastating levels. In the same study, it was determined using remote sensing from the Landsat 8 satellite that using more vegetation and less heat-absorbing materials to design urban centers can mitigate the onset of urban heat islands.

2.3 Literature Review

This literature will explore the methodologies and applications of land surface temperature (LST) prediction, highlighting the integration of citizen science and machine learning (ML) techniques. This review sets the foundation for this study, which aims to enhance LST prediction models by using citizen-sourced land cover data and advanced ML techniques.

2.3.1 Existing Methods of LST Prediction

Various methods have been developed using remote sensing and ML techniques to predict LST. Das and Ghosh (2014) provide Multifractal Detrended Fluctuation Analysis as an effective method for LST prediction, offering nuanced insights into temperature variations. Their results demonstrate the method's potential to find the simple process underneath the larger complex patterns through the theory of fractals, which is crucial for accurate predictions. Bhattacharjee et al. (2020) also introduce spatio-temporal semantic kriging, which they intend to use to address the issues of missing pixels, line drops, and cloud cover. Their results demonstrate that this method, along with incorporating land use/land cover data, produces more accurate LST predictions than other spatio-temporal interpolation methods. This validates the importance of land cover data in predicting LST. However, despite these methods to compensate for missing data, integrating additional data sourced via citizen science could help fill these gaps.

2.3.2 Citizen Science in Environmental Research

Citizen Science (CS) has emerged as a vital component in environmental research, offering non-professional participants to contribute to data collection and scientific advancement (Fraisl et al., 2022). CS applications span various fields, including biodiversity monitoring, land cover assessment, and climate change studies. This approach offers numerous benefits such as cost-effectiveness and increased temporal frequency of data collection (Dickinson et al., 2012; Fritz et al., 2017). However, CS faces challenges related to data quality, participant engagement, and ethical considerations in data sharing (Fraisl et al., 2022). Despite these challenges, CS has demonstrated its potential to provide valuable calibration and validation data for the observation of Earth (Fritz et al., 2017). When coupled with advanced analytical techniques like ML, integrating these rich CS data sets can significantly enhance environmental predictions.

2.3.3 ML in Environmental Prediction

Recent studies have explored machine learning techniques for improving LST prediction. In one such study by Arunab and Mathew (2024), the incorporation of comprehensive spatial information (such as land cover, among many others) into Random Forest and XGBoost models has shown promise in enhancing LST forecasts, with errors within $\pm$2°C. Other studies that focused on specific types of models show the effectiveness of integrating diverse datasets and advanced machine learning techniques. For example, the study by Rengma and Yadav (2023) found that a framework using Random Forest with spectral indices and terrain parameters achieved high accuracy (R^2 = 0.89, RMSE = 0.74°C) in LST prediction. A study by Pande et al.

(2024) found that ensemble models combining XG-Boost, Bagging-XG-Boost, and AdaBoost have demonstrated superior performance in LST forecasting, with Bagging ensemble achieving an R^2 of 0.75. Knowing this, combining the strengths of CS and ML offers a promising approach for more accurate and reliable LST prediction models.

2.4 Community Relevance

Due to the increasing abundance of urban heat islands and their negative effects, it is imperative to understand the scope of the increasing land surface temperature in select areas across the globe. By creating a machine learning model to predict land surface temperatures, researchers, climate activists, and the public can interpret the extent of the negative effects of urban heat islands. By raising concerns backed by accurate and refined predictive machine learning models, mitigation methods can be enacted to lower land surface temperatures down to levels appropriate for the Earth and its inhabitants. Communities that are predicted to experience unusual and significant spikes in land surface temperatures can incorporate restorative measures to reduce those temperatures.

3. Methods and Materials

3.1 Study Site

Directional photos (North, South, East, West) along with zenith (upward) and nadir (downward) were captured using the GLOBE Observer mobile application by the 2024 NASA

SEES Earth System Explorers intern group. These were collected at 37 distinct sites in a 3km x 3km grid for each intern, totaling 1740 photos. As interns were remote, study sites spanned the entire globe, although most were concentrated within the United States. Figure 1 displays the distribution of these sites on an ArcGIS map.

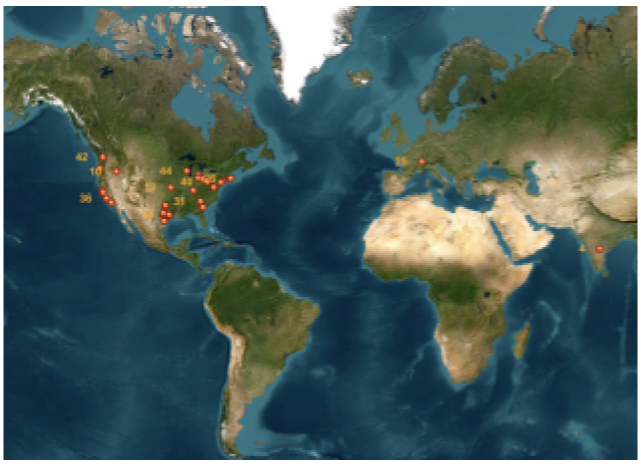

Figure 1. Map illustrating the geographic distribution of GLOBE Observer sample sites utilized by the 2024 SEES Earth System Explorer team. Each red dot indicates a location with 37 distinct points, and adjacent numbers denote the total number of sample units within each region. Credit: Peder Nelson.

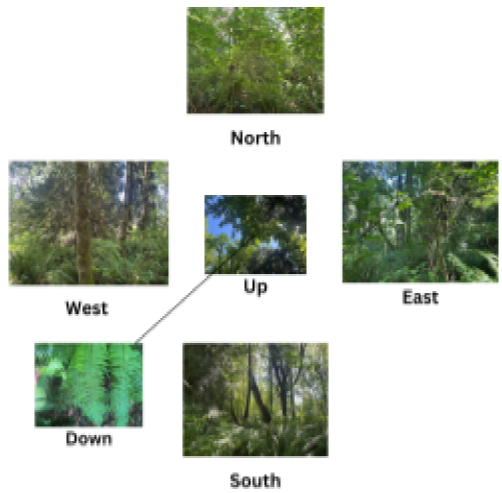

Figure 2. An example of six directional images taken at a GLOBE Observer sample site. Images taken by SEES 2024 interns and sourced from GLOBE database.

3.2. GLOBE data extraction and Zooniverse processing

Downward directional ("down") photos were extracted using the GLOBE API following data collection. After filtering for image quality and availability, 958 sites were deemed fit for analysis. These images were subsequently exported to a Zooniverse project for labeling with the help of citizen scientists; for each downward photo, citizens were given 11 possible choices for land cover as shown in Figure 3, with the option to multi-select. Each image was presented a minimum of three times to citizen scientists before being retired from the dataset to ensure robust data collection. The compiled Zooniverse data was exported to a CSV using a probability distribution for

each land cover option, where each category was represented by the proportion of annotators that selected it. For instance, if 2 out of 3 annotators selected “Bush/Scrub” for a given photo, the input feature for that photo included 0.67. Photos tagged with “Unclear / Other” or “Not a down photo” were manually reviewed by the research team and, if needed, removed from the dataset.

3.3 Land Surface Temperature Data Retrieval

After Zooniverse data exportation, daily mean land surface temperature (LST) measurements in Kelvin were retrieved via the Google Earth Engine methodology developed by Ermida et al. (2020). These measurements were extracted using 30 m multispectral bands from NASA’s Landsat-8 satellite. Temperatures at the coordinates for each image site were found for June 2024, using the first available Landsat-8 temperature reading for a given site after June 1. These temperature readings, which align with the month in which images were captured, also reflect the peak SUHI effects during summer, which are characterized by extended daylight hours and pronounced temperature disparities between urban and rural settings.

3.4 Collect Earth Online Data Collection and Processing

Interns additionally labeled the land cover of Sentinel-2 satellite imagery at each of their 37 sites using the Collect Earth Online (CEO) platform, where a grid of 100 points was generated at each site. Users selected the land cover type for each point from a preselected menu of 13 options: Trees - Canopy Cover, Shadow, Bush/Scrub, Unknown, Grass, Bare Ground, Cultivated Vegetation, Building, Water - lake/pond, Water - river/stream, Water - irrigation ditch, Impervious Surface, or Wetlands. This process quantified land cover distribution; for instance, a site with 86 points labeled as "Bare Ground" was considered to have 86% bare ground coverage. This data was aligned with the closest GLOBE down photo within a 100 m radius.

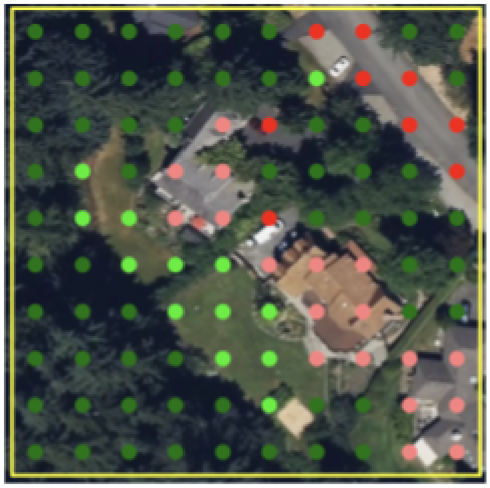

Figure 4. *An example 100 x 100 m grid of 100 labeled points on Collect Earth Online. Different colors represent different types of land cover.*

3.5 Data Preprocessing

Data preprocessing involved removing irrelevant attributes, such as email addresses, which do not provide information about land cover use. Missing values were addressed by imputing the mean for each column using the Python scikit-learn library to ensure the consistency and comprehensive use of available data points. Ultimately, 26 columns remained in the dataset. These columns include Impervious Surface, Grass, Dead Vegetation, Bare Ground, Cultivated Vegetation, Bush/Scrub, Water (river), Water (lake), and various specific land cover elements from Collect Earth Online such as Water (irrigation ditch), Grass, Water (rivers/stream), Impervious Surface (no building), Wetlands, Water (lake/ponded/container), Cultivated Vegetation, Bare Ground, Building, Trees -Canopy Cover, Unknown, Bush/Scrub, Shadow, and Land Surface Temperature (LST). Figure 5 depicts a summary of the processes involved in creating the final dataset.

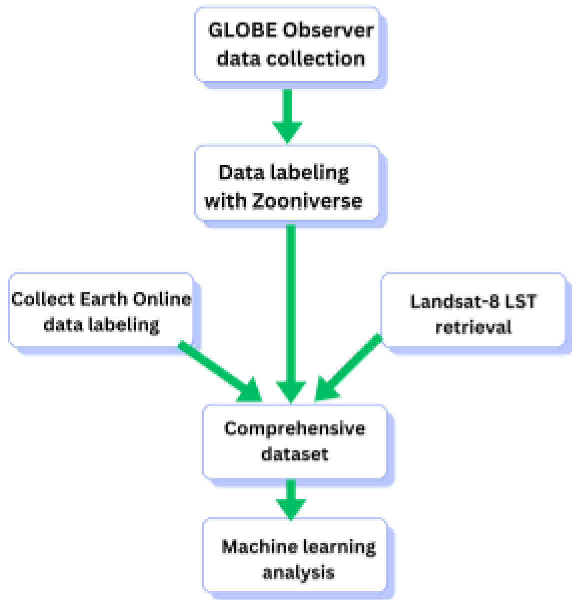

Figure 5. *A graphical overview of the methods used to prepare data for machine learning analysis.*

3.6 Machine Learning Model Training

Following data preprocessing, Random Forest and XGBoost models were trained using three distinct datasets: the first containing only land surface temperature (LST) and site coordinates (latitude and longitude); the second incorporating the aforementioned elements along with Zooniverse-labeled land cover data for GLOBE Observer photos; and a comprehensive dataset that included all previously mentioned elements along with data from Collect Earth Online. The models trained on each dataset are denoted as Model 1, Model 2, and Model 3, respectively. Models were trained using a 90:10 train/test split. Subsequently, K-fold cross-validation was performed with 10 splits to evaluate the performance of each model. Code from the NASA ARSET - Fundamentals of Machine Learning for Earth Science training was used as a reference.

Metrics analyzed included individual feature importances, R-squared values, root mean squared error (RMSE), and mean absolute error (MAE). These metrics are widely used to evaluate machine learning regression models because they provide useful information as to how the model's predictions differ from predicted values (Hodson, 2022). R-squared (R^2), which ranges from 0 to 1, measures the variation in predicted and actual values in a regression model: $R^2 = \frac{1 - Sum\ of\ Squares\ of\ Residuals\ (SSR)}{Total\ Sum\ of\ Squares\ (SST)}$ (Chicco et al., 2021). RMSE is calculated as the average of the squared differences between predicted values and actual values: RMSE = $\sqrt{\frac{1}{n}(\sum_{i=1}^{n}(y_i - \hat{y}_i)^2)}$, where y_i are the actual values and $\hat{y}_i$ are the predicted values (Hodson, 2022). MAE measures the average magnitude of errors in a set of predictions without considering their direction: MAE = $\frac{1}{n}(\sum i = 1^n|y_i - \hat{y}_i|^n$ (Hodson, 2022). After running initial base models, a Bayesian Search was conducted to find the optimal hyperparameters for each model.

4. Results and Data

4.1 Raw GLOBE Observer Data

The 958 "down" images collected using the GLOBE Observer App were generally robust and suitable for analysis, barring occasional quality issues such as blurring and improper angles which were manually screened out through Zooniverse. Many images contained the presence of people or feet, a category that was ignored when training the model. Figure 6 provides examples of photos used in Zooniverse analysis:

Figure 6. *Four examples of "down" images collected by the 2024 Earth System Explorers team, which were subsequently labeled on Zooniverse for their land cover type(s).*

4.2 Zooniverse and Collect Earth Online Data

All 958 images ("subjects") were successfully tagged a minimum of three times by Zooniverse volunteers, for a total of 3,115 classifications. Additionally, 21 out of 46 interns completed Collect Earth Online (CEO) data labeling for each of their 37 sites. Due to incomplete data on the CEO platform and discrepancies in photo-site proximities, only 180 out of 437 rows in the dataset were matched with associated CEO data. Rather than omitting labeled GLOBE photos that lacked Collect Earth Online land cover data, columns were kept blank to maximize dataset utility.

4.3 Processed Data in Random Forest Model

From our Random Forest models, decision trees were developed to visualize our results in a reduced size and showcase example procedures of a prediction based on each evaluated feature that held the highest feature importance. Three visualizations were created to reflect three different cases of training dataset size to compare the model's prediction accuracy. Figure 7 displays Model 1 that incorporates only coordinate data. Figure 8 displays Model 2 that incorporates GLOBE data alongside coordinate data. Figure 9 shows Model 3 that incorporates a comprehensive dataset (coordinates, GLOBE data, and CEO data).

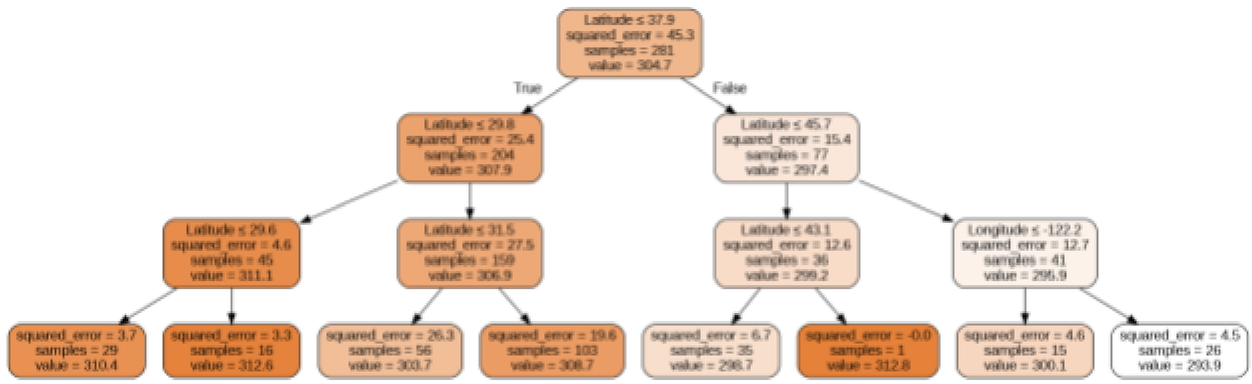

Figure 7. *A single decision tree from the Random Forest model that only incorporates latitude and longitude data at a depth of 3 levels. Darker shades of orange represent higher predicted LST values, while lighter shades of orange represent lower predicted LST values.*

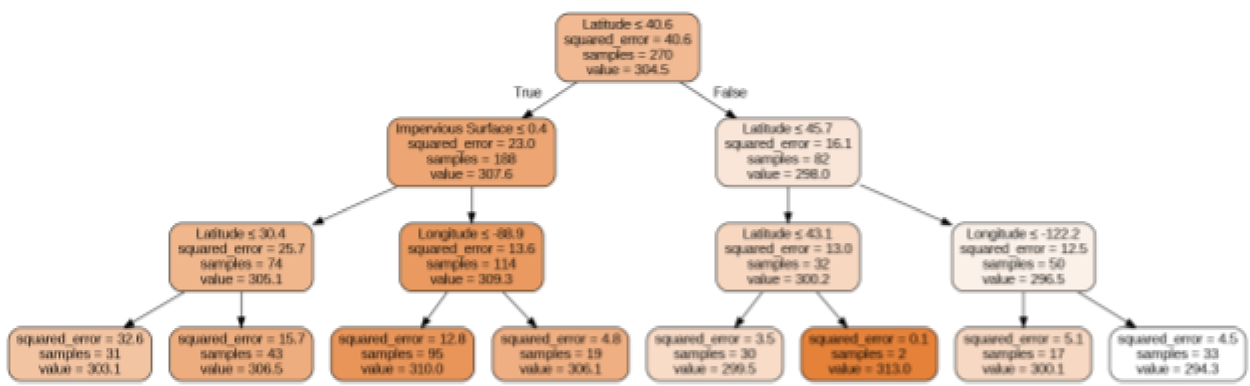

Figure 8. *A single decision tree from the Random Forest model that incorporates vegetation data, as well as latitude and longitude data, at a depth of 3 levels. Darker shades of orange represent higher LST values, while lighter shades of orange represent lower predicted LST values.*

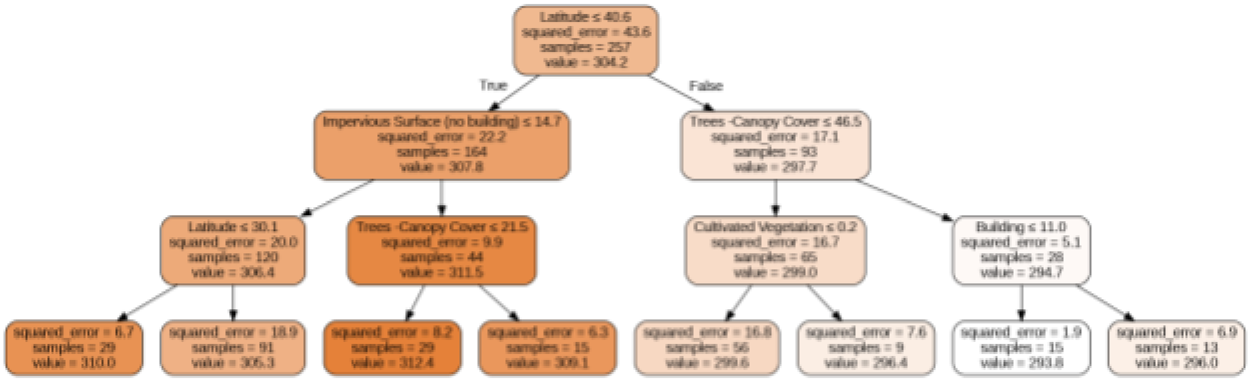

Figure 9. *A single decision tree from the Random Forest model that incorporates the comprehensive dataset with Zooniverse-labeled GLOBE photos, labeled Collect Earth Online data, and site coordinates at a depth of 3 levels. Darker shades of orange represent higher predicted LST values, while lighter shades of orange represent lower predicted LST values.*

As the models begin at the root node, the respective feature is assessed and the tree splits as it moves down. When it reaches the leaf nodes (where it terminates), the squared error is at its minimal threshold. The “squared_error” label represents the mean squared error (MSE) in degrees Kelvin between the actual LST value and predicted LST value. The “samples” label represents how many data items are used within that node. However, it’s important to remember that the number of samples represented in the root node of each single decision tree doesn't necessarily incorporate the overall dataset size as a result of each tree being trained on a bootstrap sample (repeated resampling data with replacement). The “value” label represents the predicted LST.

Using Figure 9 as an example of making a prediction within the decision tree, it is clear how certain features contribute to an increase in urban heat islands and therefore resulting in higher LST. In this case, let’s assign values to each variable for modeling purposes: Latitude = 37, Impervious Surface (no building) = 25.3, and Trees-Canopy Cover = 10. The example scenario begins at the root node and evaluates as “True” as Latitude ≤ 40.6, prompting the model to evaluate further with the Impervious Surface (no building) feature. This evaluates as “False” as our given value of 25.3 does not satisfy Impervious Surface (no building) ≤ 14.7. Thus far, it is apparent that this decision has impacted the predicted LST value. In the context of this feature, when a value is greater than the threshold of 14.7, this indicates that there is a higher presence of impervious surfaces which directly correlates with higher LST values because these areas are known to absorb heat. As the model continues to the next node, it evaluates as "True" as Trees-Canopy Cover ≤ 21.5. This decision leads to the leaf node and the final prediction value to be 312.4 K. Having a low abundance of canopy tree cover results in lower evapotranspiration levels which can't counteract high surface temperatures, increasing the overall predicted LST value. If the Trees-Canopy Cover had a value higher than the threshold of 21.5, the predicted LST value would result in approximately 3 K lower compared to the final prediction given that there is a low density of that specific feature. The same evaluation process can be applied to each model which can determine how each feature affects the predicted LST value based on the surrounding areas.

Shown in Table 1, the root mean squared error (RMSE) in degrees Kelvin, mean absolute error (MAE) in degrees Kelvin, and R-squared values were calculated for each of the three Random Forest models trained on different datasets. Before hyperparameter tuning, preliminary versions of Model 1 (coordinates only), Model 2 (coordinates and GLOBE data), and Model 3 (coordinates, GLOBE data, and CEO data) were trained for baseline performance metrics.

Table 1. *Baseline evaluation metrics for three Random Forest models before tuning.*

Dataset	R^2	RMSE (K)	MAE (K)
Model 1	0.84	2.66	2.02

Model 2	0.80	2.84	2.13
Model 3	0.82	2.66	2.01

Subsequently, a Bayesian Search was performed to obtain optimal hyperparameters for our model using the following parameter distributions:

```
param_space = {
      'n_estimators' : (10, 1000),
      'max_depth: (3, 50),
      'min_samples_split' : (2, 10),
      'min_samples_leaf': (1, 10),
      'max_features': ('sqrt', 'log2')
}
```

Bayesian Search was performed with 10 folds through 100 iterations. Below are the optimal hyperparameters for Model 1:

```
params = {"max_depth": 47, "max_features": "log2",
"min_samples_leaf": 2, "min_samples_split": 2, "n_estimators": 739}
```

The optimal hyperparameters for Model 2:

```
params = {"max_depth": 19, "max_features": "log2", "min_samples_leaf":
1, "min_samples_split": 2, "n_estimators": 1000}
```

The optimal hyperparameters for Model 3:

```
params = {"max_depth": 30, "max_features": "log2", "min_samples_leaf":
1, "min_samples_split": 2, "n_estimators": 328}
```

Bayesian Search was able to slightly improve the evaluation metrics for both Model 1, as listed in Table 2. However, performance metrics decreased for both Model 2 and Model 3, which could potentially result from noise in the dataset or overfitting based on training data.

Table 2. *Evaluation metrics for the Random Forest model for each of the three datasets tested following optimization via Bayesian Search.*

Dataset	R^2	RMSE (K)	MAE (K)
Model 1	0.84	2.49	1.89
Model 2	0.78	3.00	2.30
Model 3	0.79	2.97	2.25

The dataset including only latitude/longitude (Model 1) had the highest performance for our limited sample size, as evidenced by the highest R-squared value and the lowest RMSE and MAE values. The accuracy of Model 2 increased slightly when Collect Earth Online data was incorporated with GLOBE vegetation data (Model 3), suggesting that detailed spatial and vegetation data could provide a more nuanced understanding of LST variations. Interestingly, while the inclusion of CEO data improved model performance compared to GLOBE data alone, it did not surpass the model that relied solely on geographic coordinates.

4.4 Processed Data in XGBoost Model

Three separate XGBoost models were trained using three different datasets, akin to the methodology employed during Random Forest Training. Before performing any hyperparameter tuning, a preliminary version of all three models was trained and evaluated to obtain baseline performance metrics (see Table 3).

Table 3. *Baseline evaluation metrics for three XGBoost models before tuning.*

Dataset	R^2	RMSE	MAE
Model 1	0.75	3.10	2.36
Model 2	0.77	3.34	2.59
Model 3	0.80	2.90	2.13

To obtain optimal hyperparameters for each model, a Bayesian Search was performed using the following parameter distributions:

```
param_space = {
    'learning_rate': (0.01, 0.3, 'uniform'),
    'max_depth': (3, 50),
    'subsample': (0.1, 1.0),
    'min_child_weight': (1, 10),
    'eta': (0, 1),
    'alpha': (0, 10),
    'lambda': (0, 5),
    'gamma': (0, 10)
}
```

The search was performed with 10 folds through 100 iterations. The optimal hyperparameters for Model 1:

```
params = {"objective": "reg:squarederror",
```

```
"learning_rate":0.0990999161353275, "max_depth": 3, "min_child_weight": 10,
"subsample": 1, "lambda": 5, "gamma": 0, "alpha": 5, "tree_method":"exact",
"eta":0}
```

Model 2:

```
params = {"objective": "reg:squarederror", "max_depth": 3, "lambda": 5,
"alpha": 0, "tree_method": "exact", "eta": 1, "gamma": 10, "learning_rate":
0.14107739445372014, "min_child_weight": 10, "subsample":
0.4931393217397809}
```

Model 3:

```
params = {"objective": "reg:squarederror", "max_depth": 3, "lambda":
1, "alpha": 0, "tree_method": "exact", "eta": 0, "gamma": 0,
"learning_rate": 0.09447565846479973, "min_child_weight": 10,
"subsample": 1.0}
```

The Bayesian search was able to moderately improve the output of each model, as shown in Table 4.

Table 4. *Evaluation metrics for all XGBoost models following optimization via Bayesian Search.*

Dataset	R^2	RMSE (K)	MAE (K)
Model 1	0.82	2.86	2.26
Model 2	0.80	3.00	2.36
Model 3	0.82	2.87	2.24

Interestingly, Model 2 showed the greatest increase in performance following optimization with a decrease in RMSE of 0.34 K, a 0.03 increase in R^2, and a 0.23 K MAE decrease. Conversely, Model 3 saw the smallest performance increase with an RMSE decreasing by 0.03 K and R^2 increasing by only 0.02 points. Overall, performance trends here are similar to those from the Random Forest models; Model 2 underperforms compared to Model 1 despite having more features for each sample. Notably, the addition of the CEO data results in a double digit decrease in MAE, along with a 0.13 K decrease in RMSE. This is interesting considering only 41% of samples had complete CEO data. To examine the comparative importances of features from each dataset, we used the F-score metric to create a bar graph of the 17 most important features used in Model 3:

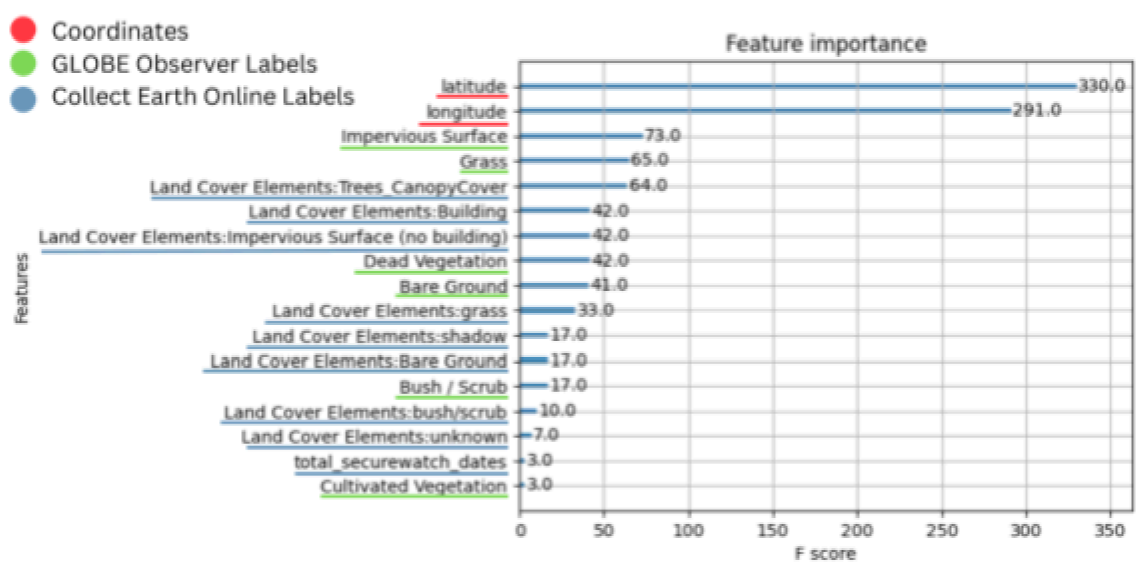

Figure 10. *A bar graph showing the features our models used to predict LST, measured using their F-score. Each feature comes from either coordinate data (red), GLOBE Observer Labels (green), or Collect Earth Online Labels (blue).*

From Figure 10, we can deduce that the most important features (following longitude and latitude) are “Impervious Surface” and “Grass”, from GLOBE Observer and Zooniverse Labeling, followed by “Trees / Canopy Cover” and “Buildings” from Collect Earth Online labels. The least used features (excluding zero-values) are “Cultivated Vegetation”, “Unknown”, and “Bush / Scrub”.

5. Discussion

5.1 Interpretation

Regardless of model, the addition of GLOBE Observer data resulted in a decrease in predictive power in Model 2. The noise in this dataset is most likely due to the quality of the labels assigned to the nadir photos via Zooniverse. Specifically, the labels did not contain quantitative information; participants were simply asked to select all the land cover types that applied to each ground image. It is possible that quantitative labels representing the proportion of each land cover type in each nadir image would not cause the same performance decline that was observed here.

It is interesting that the addition of the CEO data resulted in such a marked improvement in predictive performance, despite only 41% of samples having complete CEO data. This is suggestive that CEO land cover data is a much better predictor of LST than GLOBE Observer nadir image labels; it would be interesting to see how the models perform with even more samples with complete CEO labels.

Additionally, Random Forest models performed slightly better than their XGBoost counterparts across all dataset sizes. This could be due to Random Forest generally performing better on noisy data, leveraging the outputs of multiple decision trees to determine a final prediction. It would be valuable to explore the predictive capabilities of a Random Forest model with an expanded dataset.

Considering the relatively high feature importance of the three highest Collect Earth Online labels, we can assume that greater use of Collect Earth Online labels across all samples would moderately improve accuracy across both models. It would be equally prudent to remove the least important GLOBE Observer features from Zooniverse labeling to allow participants in the survey to more accurately assign quantitative labels to more relevant features.

5.2 Evaluating the Hypothesis

Due to the results of the study, the hypothesis has been rejected for the Random Forest Model because incorporating GLOBE and CEO data to the machine learning models decreased accuracy of the models. However in the XGBoost model, the hypothesis failed to be rejected as the accuracy remained the same in Model 1 and Model 3. In the Random Forest with Bayesian search, the highest R^2 value at 0.84 (84% accuracy) was observed in Model 1, which contained only coordinate data. The XGBoost models with Bayesian search achieved a R^2 value of 0.82 (82% accuracy) in models 1 and 3. Model 1 was solely coordinate data, model 2 had additional GLOBE Observer data, and model 3 had additional Collect Earth Online Data. This suggests that running a model with coordinate and Collect Earth Online Data will garner the highest accuracy.

In addition, the Random Forest model was most accurate overall, achieving an R^2 value of 0.84 for Model 1 in the runs both with and without the Bayesian Search. This indicates that the Random Forest model may be more accurate to predicted land surface temperatures with regards to vegetation in urban heat islands.

5.3 Comparison to Similar Studies

Our research is similar to other studies to utilize the land cover tool in the GLOBE Observer application, machine learning models, and vegetation data. A study by van Jaarsveld et al. (2024) incorporated GLOBE multi directional images into a Random Forest model in order to create reliable information regarding high resolution vegetation drought impacts. This is just one of many scientific studies that use machine learning with citizen science, satellite data, and land cover data.

5.4 Errors

A major source of error in the study's methodology was the focus only on "down" photos. The land cover displayed in one photo of the ground often does not display the full picture of an area's land cover. For instance, urban construction zones can have small patches of grass on the ground, or a pond on the ground may be surrounded by a forest. Incorporating imagery from all six directional photos—north, south, east, west, zenith (upward), and nadir (downward)—into a future model would likely provide a more comprehensive, accurate depiction of land cover variations.

Some sources of error in this study also stem from the application of citizen science into the models. Although citizen science is accessible and can be filtered to only garner qualifying data from larger datasets. There is a certain level of accuracy from the data sources that cannot be confirmed by researchers. Any images collected or Collect Earth Online classifications made are created upon the discretion of the data source and cannot be completely verified by the researchers, contributing to minor faults in the dataset that cannot be detected since there was a large amount of data.

Another error in this study was the lack of adequate data from all the participants of the 2024 SEES Earth System Explorers team. In further detail, it could be said that with more GLOBE Observer down photos and Collect Earth Online classifications, the models could have been trained to a more accurate state. There were only 958 down photos from 1,702 possible locations. The number of viable down photos further decreased due to quality control and other data available to 437 down photos. If all of the down photos were taken and had adequate data, the machine learning models would have been more precise.

6. Conclusion

6.1 Significance and Application

Although the incorporation of land cover data failed to improve predictive accuracy, our team concluded that this was due to a lack of quality data rather than the usefulness of the data itself. Improving data collection methods and ensuring higher quality data could potentially reveal the true value of citizen-sourced land cover data. This study particularly sheds light on the value of Citizen Science. While previous studies have utilized satellite readings to develop predictive models for LST, such as an analysis of the Dhaka metropolitan area using Landsat imagery, none have explored the application of citizen-sourced imagery on a global scale (Faisal et al., 2021).

Through the GLOBE Observer app, anyone can download the GLOBE Observer application to document land cover in their local communities. Both the Zooniverse platform and Collect Earth Online platform also provide free, accessible mechanisms for the public to label citizen and satellite-sourced images. When citizen-sourced data is merged with forms of remote sensing analysis, such as satellite data from Landsat-8 and Collect Earth Online, this enriched data set could significantly enhance model accuracy. Furthermore, involving citizens in collecting and labeling land cover data is a key component of open science that could be used to empower global change (Fraisl et al., 2022). The participatory approach used in this study encourages individuals to engage with issues of climate change and urban sustainability as they contribute to meaningful research.

6.2 Improvements to Study

Some improvements could be made to the study's methodology to improve the output of both models. Firstly, the small sample size (437 samples) contributed to the variance that is present in both Random Forest and XGBoost models. Similarly, it would be useful to include GLOBE Observer images that satisfy the "Snow/Ice" condition and could potentially improve the performance of both models in low-temperature environments. Currently, the models are only able to predict LST in moderate climates during the spring and summer seasons. This is a result of the narrow variety within the data on which the models were trained.

Moreover, there was a significant difference in quality between GLOBE Observer photos. Several samples contained photos which were blurry; excessively magnified; or not true nadir photos. This made it difficult to accurately determine and label the true land cover of each sample. Any incorrect labeling would also contribute to the uncertainty in the predictions provided by the models. Additionally, an inherent challenge with any program relying on the built-in GPS receiver on a user's device is spatial accuracy (Low et al., 2021). GPS sensor readings can be inaccurate, especially with older devices, and some photos taken by SEES interns had a reported accuracy of over 30 m. Future studies would benefit from adopting stricter rules regarding the quality of observations used from the GLOBE Observer database.

Furthermore, many samples lacked their respective CEO land cover labels. This would have significantly increased the number of features available for each sample, potentially improving the predictive power of both models and mitigating bias during training. One suggested improvement would be to take more time for data collection to ensure all samples have their respective CEO labels and GLOBE Observer land covers.

Additionally, this study would benefit from more detailed land cover observations compared to the labeling carried out in the Zooniverse project for this study. A better alternative would be similar to the land cover observations done within the GLOBE Observer app itself, where participants choose which categories of land cover apply to the image and then select the percentages of land cover (out of 100) that apply to a given image. This approach would be advantageous for model training, providing more detailed insight into the composition of each nadir photo.

6.3 Future Work

This study has opened several avenues for future research into the use of Citizen Science to predict LST. Building on these findings, future studies could investigate the use of pre-trained Convolutional Neural Networks to extract relevant features from down photos and train models using the provided feature vectors, as opposed to manual labeling. This would allow for the use of a much larger sample size, as images would be interpreted in code as opposed to manual labeling. This would also allow for the use of North, South, East, and West-facing directional photos, potentially providing more features to enhance the predictive power of a future model. The land cover observations that accompany directional photos in GLOBE Observer could also be extracted and used to train the model, serving as an alternative to manual Zooniverse labeling of ground photos.

Future research could also explore the use of samples from a more diverse array of geographical locations. Samples from colder climates and from a wider range of elevations would enhance the predictive performance of the models on unseen data. Currently, models have not been trained with samples of LSTs below 290 K (17˚C). There is significant possibility for research to expand the capabilities of both models to include sub-290 K samples.

6.4 Mentor Impact

Our research project was heavily influenced by the guidance of our mentors: Andrew Clark, Russanne Low, Peder Nelson, Cassie Soeffing, Ollie Snow, and Riya Tyagi. Rusty provided knowledge about the principles of citizen science, allowing us to cross-validate our data to ensure our findings were reliable and scientifically sound. Peder introduced us to GIS software such as Zooniverse and Collect Earth Online, improving our data labeling and categorization. Cassie advised us on effectively communicating our research goals. Andrew introduced us to ORCiD and open data concepts. Riya and Ollie served as our peer mentors. Riya provided technical advice on using AI models for the project, while Ollie helped with image labeling details on Collect Earth Online. Their combined efforts made this project possible. We extend our thanks to each of them for their contributions.

Datasets and Code

All datasets and code can be found at https://doi.org/10.5281/zenodo.12984362.

Contributions

All contributors created GLOBE Observer and Collect Earth Online data points and edited the paper. A.F. prepared datasets using GLOBE and Collect Earth Online data, trained the Random Forest model, and contributed to the Methods section of the paper. K.D. developed Random Forest tree visualizations, contributed to the Results and Data section, as well as the Abstract section. S.D. contributed to the Introduction, Discussion, Abstract section of the paper, assisted with literature review, and created the References page. M.F. created the Zooniverse labeling project, trained the XGBoost model, and contributed to the Results, Discussion, and Conclusion sections. K.F. created the script for online symposium presentations and helped edit the research paper. K.I.D. read and analyzed previous related research papers and compiled their results in the literature review. A.G. identified variables for consideration, including spatial resolution, seasonal variations, and dataset calibration, while addressing sampling bias.

Acknowledgments

The authors would like to acknowledge the support of the 2024 Earth System Explorers (ESE) Team, NASA Science Mentors, and ESE peer mentors. NASA STEM Enhancement in the Earth Sciences (SEES) Virtual High School Internship program. The NASA Earth Science Education Collaborative leads Earth Explorers through an award to the Institute for Global Environmental Strategies, Arlington, VA (NASA Award NNX6AE28A). The SEES High School Summer Intern Program is led by the Texas Space Grant Consortium at the University of Texas at Austin (NASA Award NNX16AB89A0).

References

ARSET - Fundamentals of Machine Learning for Earth Science | NASA Applied Sciences. (2023, April 20). https://appliedsciences.nasa.gov/get-involved/training/english/arset-fundamentals-machine-learning-earth-science

Arunab, K. S., & Mathew, A. (2024). Exploring spatial machine learning techniques for improving land surface temperature prediction. *Mağallaẗ Al-Kuwayt Li-l-ʿulūm*, *51*(3), 100242. https://doi.org/10.1016/j.kjs.2024.100242

Barry, V. J., Hauswirth, S. M., & Wanders, N. (2024). Machine learning and global vegetation: random forests for downscaling and gap filling. *European Geosciences Union*, *28*(11). https://doi.org/10.5194/hess-28-2357-2024

Bhattacharjee, S., Chen, J., & Ghosh, S. K. (2020). Spatio‐temporal prediction of land surface temperature using semantic kriging. *Transactions in GIS*, *24*(1), 189–212. https://doi.org/10.1111/tgis.12596

Brownlee, J. (2020, November 6). *Scikit-Optimize for Hyperparameter Tuning in Machine Learning*. Machine Learning Mastery. https://machinelearningmastery.com/scikit-optimize-for-hyperparameter-tuning-in-machine-learning/

Brownlee, J. (2021, March 7). *XGBoost for Regression*. Machine Learning Mastery. https://machinelearningmastery.com/xgboost-for-regression/

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ. Computer Science*, *7*, e623. https://doi.org/10.7717/peerj-cs.623

Das, M., & Ghosh, S. K. (2014). Short-term prediction of land surface temperature using multifractal detrended fluctuation analysis. *Institute of Electrical and Electronics Engineers*. https://doi.org/10.1109/indicon.2014.7030684

Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., Phillips, T., & Purcell, K. (2012). The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment*, *10*(6), 291–297. https://doi.org/10.1890/110236

Ermida, S. L., Soares, P., Mantas, V., Göttsche, F. M., & Trigo, I. F. (2020). Google Earth Engine Open-Source Code for Land Surface Temperature Estimation from the Landsat Series. *Remote Sensing*, *12*(9), 1471. https://doi.org/10.3390/rs12091471

Faisal, A. A., Kafy, A. A., Rakib, A. A., Akter, K. S., Jahir, D. M. A., Sikdar, M. S., Ashrafi, T. J., Mallik, S., & Rahman, M. M. (2021). Assessing and predicting land use/land cover, land surface temperature and urban thermal field variance index using Landsat imagery for Dhaka Metropolitan area. *Environmental Challenges*, *4*, 100192. https://doi.org/10.1016/j.envc.2021.100192

Fraisl, D., Hager, G., Bedessem, B., Gold, M., Hsing, P. Y., Danielsen, F., Hitchcock, C. B., Hulbert, J. M., Peira, J., Spiers, H., Thiel, M., & Haklay, M. (2022). Citizen science in environmental and ecological sciences. *Nature Reviews Methods Primers*, *2*(1). https://doi.org/10.1038/s43586-022-00144-4

Fritz, S., Fonte, C., & See, L. (2017). The role of citizen science in earth observation. *Remote Sensing*, *9*(4), 357. https://doi.org/10.3390/rs9040357

Gorodetski, M. (2021, August 28). Hyperparameter tuning methods - grid, random or Bayesian search? | towards data science. *Medium*. https://towardsdatascience.com/bayesian-optimization-for-hyperparameter-tuning-how-and-why-655b0ee0b399

Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development*, *15*(14), 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022

Ismaila, A. B., Muhammed, I., & Adamu, B. (2022). Modelling land surface temperature in urban areas using spatial regression models. *Urban Climate*, *44*, 101213. https://doi.org/10.1016/j.uclim.2022.101213

Kasniza Jumari, N. A. S., Ahmed, A. N., Huang, Y. F., Ng, J. L., Koo, C. H., Chong, K. L., Sherif, M., & Elshafie, A. (2023). Analysis of urban heat islands with landsat satellite images and GIS in Kuala Lumpur Metropolitan City. *Heliyon*, *9*(8), e18424. https://doi.org/10.1016/j.heliyon.2023.e18424

Low, R. D., Nelson, P. V., Soeffing, C., & Clark, A. (2021). Adopt a Pixel 3 km: A Multiscale Data Set Linking Remotely Sensed Land Cover Imagery With Field Based Citizen Science Observation. *Frontiers in Climate*, *3*. https://doi.org/10.3389/fclim.2021.658063

McCartney, S. (2023). Overview and Access of Land Surface Temperature (LST). In *NASA's Applied Remote Sensing Training Program* [Report]. https://appliedsciences.nasa.gov/sites/default/files/2023-05/Day2_P5_Eng.pdf

Pande, C. B., Egbueri, J. C., Costache, R., Sidek, L. M., Wang, Q., Alshehri, F., Din, N. M., Gautam, V. K., & Pal, S. C. (2024). Predictive modeling of land surface temperature (LST) based on Landsat-8 satellite data and machine learning models for sustainable development. *Journal of Cleaner Production*, *444*, 141035. https://doi.org/10.1016/j.jclepro.2024.141035

Rengma, N. S., & Yadav, M. (2023). A GENERIC MACHINE LEARNING-BASED FRAMEWORK FOR PREDICTIVE MODELING OF LAND SURFACE TEMPERATURE. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences/International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLVIII-4/W2-2022*, 95–102. https://doi.org/10.5194/isprs-archives-xlviii-4-w2-2022-95-2023

Tree Methods — xgboost 2.1.0 documentation. (n.d.). https://xgboost.readthedocs.io/en/latest/treemethod.html

Tuychiev, B. (2023, February). *Using XGBoost in Python Tutorial*. Datacamp. https://www.datacamp.com/tutorial/xgboost-in-python

What is Citizen Science? (2024, April 30). U.S. National Park Service. https://www.nps.gov/subjects/citizenscience/citizen-science.htm

Predicting Land Surface Temperature Using Land Cover Data: A Machine Learning Approach with GLOBE Observer and Landsat-8

Students: Kevin Diaz, Kandyce Diep, Suhani Dondapati, Maako Fangajei, Anna Felten, Kei Fry, Conor Furey, Aaren George

Mentors: Andrew Clark, Rusty Low, Peder Nelson, Ollie Snow, Cassie Soeffing, Riya Tyagi

STEM Enhancement in Earth Science (SEES) Summer High School Intern Program

United States of America

July 19, 2024

IVSS Badges

I Am a Problem Solver: This badge applies because our project addresses the problem of UHIs and their increased prevalence due to climate change. Our paper could help city planners and authorities to mitigate the severity of UHIs thanks to a better understanding of the relationship between land cover and LST.

I Am a Data Scientist: Our paper includes student-generated data along with data from Landsat 8. Landsat-8 data was obtained from Google Earth Engine using open-source code developed by Ermida et al. (2020). In combining these data sources, we were able to train a model to make inferences on the land surface temperature at the time that each ground photo was taken. We use the data to develop a better understanding of how land cover can impact LST at a given location. Moreover, in our conclusion, we discuss the limitations of our data.

I Work With a STEM Professional: This badge applies because our project involved collaboration with Peder Nelson, a Senior Instructor and Faculty Research Assistant at Oregon State

University. Peder helped improve the labels and categorization of our Zooniverse data. He suggested that we add some more labeling options for our collaborators that would be more precise, and also assisted us in getting the help of the rest of the Earth System Explorers intern group.

I Am an Engineer: Our report uses a variety of student-generated data to train our ML models. This includes GLOBE Observer photos, student-powered Zooniverse labeling, and CEO labeling for each GLOBE Observer photo. Our paper addresses the real-world problem of UHIs, using student-generated data to engineer a model to provide a better understanding of the relationship between land cover and LST.