Examining the Impacts of Land Use on Air Quality in Chicago: Application of Street View Imagery and Hyperlocal Urban Climate Sensing

Junfeng Jiao, Seung Jun Choi
October, 2024

A publication of the USDOT Tier 1 Center:
Cooperative Mobility for Competitive Megaregions
At The University of Texas at Austin

Technical Report Documentation Page

1. Report No. CM2-77	2. Government Accession No.	3. Recipient's Catalog No. ORCID: 0000-0002-7272-8805
4. Title and Subtitle Examining the Impacts of Land Use on Air Quality in Chicago: Application of Street View Imagery and Hyperlocal Urban Climate Sensing		5. Report Date October, 2024
		6. Performing Organization Code
7. Author(s) Junfeng Jiao, Seung Jun Choi		8. Performing Organization Report No. CM2-77
9. Performing Organization Name and Address The University of Texas at Austin School of Architecture 310 Inner Campus Drive, B7500 Austin, TX 78712		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. USDOT 69A3551747135
12. Sponsoring Agency Name and Address U.S. Department of Transportation Federal Transit Administration Office of the Assistant Secretary for Research and Technology, UTC Program 1200 New Jersey Avenue, SE Washington, DC 20590		13. Type of Report and Period Covered Technical Report conducted September, 2022- August, 2023
		14. Sponsoring Agency Code
15. Supplementary Notes Project performed under a grant from the U.S. Department of Transportation's University Transportation Center's Program.		

16. Abstract

Urban climate sensors are being installed in cities all over the world in order to proactively address the growing issues posed by climate change and to monitor air quality in real time. This study focuses on the City of Chicago and evaluates air quality using Microsoft's recently released Project Eclipse sensors. Using a combination of conventional land use extraction methods and street view data from Google Street View (GSV), land use features close to the sensor locations were recovered. The distinctive qualities of street view photos were examined and spatial data was broken down using principal component analysis (PCA). In order to investigate the variables affecting air quality, the study also used SHapley Additive exPlanations (SHAP) and XGBoost machine learning regression. It is recommended that the built environment and land use in this area be addressed by the city and local authorities in order to mitigate future dangers.

17. Key Words Street view imagery, Air quality, Urban climate sensor, Machine learning		18. Distribution Statement No restrictions.	
19. Security Classif. (of report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of pages 33	22. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Acknowledgements

Table of Contents

Executive Summary

This study investigates the air quality status in Chicago using Microsoft's Project Eclipse urban climate sensors and integrates Google Street View (GSV) imagery with traditional land use data. The research aims to assess the impact of land use on air quality, using machine learning techniques such as XGBoost regression and SHapley Additive exPlanations (SHAP) to analyze the results. Principal component analysis (PCA) was conducted to understand spatial relationships. Key findings revealed that the air quality at most sensor locations aligned with the city's predefined levels, except at the northern intersection of West Belmont Avenue, where significant air quality concerns were detected. The integration of GSV imagery provided a detailed, multi-angle view of urban environments, which enhanced the ability to identify the effects of land use on air quality. The study demonstrated that different land use types have varying impacts on air quality, with transportation and other land use factors playing a crucial role in exacerbating air pollution at the West Belmont Avenue intersection. This research suggests that local authorities should address specific built environment and land use conditions in certain areas, such as West Belmont Avenue, to mitigate potential air quality risks. The findings also highlight the contributions of using street view imagery alongside conventional land use methods to improve air quality assessment and urban planning efforts. Future research may benefit from the use of urban climate sensors and integrating more advanced images.

Chapter 1. Introduction

Rising temperatures, rising sea levels, and an increase in the frequency and intensity of extreme weather events like droughts, flooding, and wildfires are all effects of the gradual increase in the risk of climate change that have an impact on cities around the globe. The health of people is seriously threatened by these risks [1-2]. Apart from these obstacles, air quality is susceptible to the consequences of global warming. Further problems with air quality arise from weaker regional air circulation brought on by higher rates of water evaporation and a declining ozone background [3].

Internet of things (IoT) sensors are being used by cities all over the world to proactively address the climate catastrophe. Usually, these sensors are used to track humidity, temperature, and particle matter in real time. One of the main efforts in smart city design to acquire and use large volumes of data, or big data, is the deployment of IoT networks in urban areas [4-5]. Nevertheless, the data generated by urban sensors is sometimes imprecise, resulting in massive volumes of data that necessitate a specific level of expertise to be exploited efficiently [5–6]. Planners have been confused by the basic nature of processing urban sensor data since the mechanics underlying sensor readings and their implications are not well understood.

Conventionally, studies have attempted to identify the effects of a variety of factors on air quality at a large scale, including weather patterns, air pollutants [7-8], emissions, energy use [9–10], and changes in land use [11–14]. How these factors would affect sensor readings of air quality is unknown, though. Consequently, land use cover maps or satellite image processing have been the preferred methods for extracting land use features via conventional methods [11, 15–17]. The warning is that there is little correlation between the unit of emerging data sources and the unit of analysis of factors derived from traditional means.

In this work, instead of using traditional methods that rely on static two-dimensional data, land use features were extracted using Google Street View (GSV) photos. GSV gathers information in degrees from the perspective of pedestrians, offering a three-dimensional view that provides details on neighborhood environment, urban growth, demographic composition, and landscape assessment [18–25]. Nevertheless, extracting land use information from street

view photography is not frequently used in planning literature because of similar difficulties encountered when employing sensor data.

In the pilot case study, we addressed the current air quality and surrounding land use circumstances in Chicago, United States, by utilizing street view photography and urban climate sensors. The literature now in publication provides extensive documentation of the two sources' relative uses. While street view automobiles have been used to map air pollution in road segments [27], street view photos have been used to measure environmental parameters [18, 26]. Air pollution parameters have been proposed, or measured, using IoT sensors [5, 28, 29]. More research is still needed on the relationship between sensor data and scene elements from street view photos, as well as how to analyze these relationships at the regional level. By comprehending the mechanisms and characteristics of two developing sources, the current study suggests a framework for combining them. Our study's conclusions can steer future planning initiatives.

We hypothesize that, even at the regional level, street view images will differ from traditional land use features, that sensors will offer fresh perspectives on assessing the state of the city's ambient air quality, and that the built environment will influence the operations of the sensors. Our study aims to investigate the following three primary research phases:

- To integrate street view photography with urban climate sensor data from the Microsoft research team in order to solve the current air quality concerns in Chicago.
- To conduct a principal component analysis (PCA) to determine the features of street view imagery and to group the land use variables derived from traditional methods with street view imagery.
- To explore the effects of the determinants on the air quality as measured by sensors, a machine learning regression analysis will be performed, and the SHapley Additive exPlanations (SHAP) value will be computed.

Chapter 2. Literature Review

2.1. Air Pollution and Installation of Sensors in Cities

Significantly harmful health impacts of air pollution include cancer, heart and lung conditions, and even death [30]. As a result of the Clean Air Act of 1990, the EPA established National Ambient Air Quality Standards for six main pollutants that are detrimental to public health: lead, carbon monoxide, nitrogen dioxide, ozone, particle pollution, and sulfur dioxide [31]. Previous research has assessed individual contaminants [17] or modeled spatiotemporal forecasts [15, 32] based on the EPA's criteria.

Sensors offer a novel way to investigate environmental conditions on a hyperlocal scale by providing microscopic real-time climatic data, such as temperature and air pollutants [8]. The micro-to local-geographic scale is referred regarded as "hyperlocal" in smart city literature [33–33]. We use the term hyperlocal to characterize this method since IoT urban climate sensors integrate into cities and measure physical settings like weather conditions on a local scale [33].

2.2. The Determinants of Air Pollution in Cities

A complex interplay between human and environmental factors is the root cause of air pollution in cities [36]. These interactions have different causes and vary in space and time. Most people agree that air pollution and health hazards are mostly caused by traffic emissions [36–37]. Air pollution problems are made worse by vehicle emissions, which scatter and undergo chemical reactions in the atmosphere [37]. Air pollution is caused by factors that worsen traffic emissions, including road proximity [38], road density, associated traffic volumes [37], congestion, and longer commute times [39]. Therefore, air quality is greatly impacted by meteorological and air pollution factors [7-8]. Mitigating air pollution can be aided by reducing greenhouse gas emissions (GHG) [9]. Moreover, there is a correlation between GHG emissions and energy intensity and consumption [10].

Changes in land use might range from minor adjustments to a whole landscape replacement, which can have a substantial effect on air quality [40]. Both locally and globally, these changes

may have an impact on air quality and climate, as well as local meteorological conditions [11]. Although it can be challenging to monitor every change in land use, it is generally acknowledged that varied urban land uses will result in differing air quality in different areas [12–13, 40]. This is due to the fact that modifications in land use have the potential to modify patterns of mobility and modify traffic emissions [41]. While urban green spaces connected to land use can improve air quality in cities, industrial land use, such as trash disposal in landfills, can produce hazardous air pollution [42]. A number of variables, including population density [44], traffic load, build-up density [43], and traffic conditions [45], are related to changes in land use and have an impact on air quality. The path of air pollution in cities is linked, as Mayer [36] noted.

2.3. The Benefits of Using Street View Imagery

Different kinds of information about land usage can be obtained by utilizing GSV and satellite data. Block-level pedestrian perspective is provided by GSV data, and a macroscopic picture is captured by satellite data. When compared to satellite photos, Larkin and Hystad [46] discovered that GSV offered distinctive green space information. But it was discovered that there was little association between GSV and metrics obtained from satellites [46–47]. Although zoning-based land use maps are essential for urban planning research, they might not fully capture the local scale since narrow streets and obscure building blocks are only partially depicted and are instead colored by the various land use categories. However, GSV pictures offer an effective means of gathering complex land use features across large geographic borders and provide a block-level landscape [48–49]. Furthermore, it was discovered that employing land use regression modeling in conjunction with conventional land use datasets enhanced the accuracy of the models for air pollution [50].

Chapter 3. Materials and Methods

3.1. Study Area and Urban Climate Sensor: Project Eclipse

Under the Project Eclipse name, the Microsoft team placed more than a hundred inexpensive air pollution monitors throughout the city of Chicago in 2020. There are 110 sensors installed across the city as of August 2021. At the 21st ACM/IEEE International Conference on Information Processing in Sensor Networks in 2022, Microsoft research teams revealed the idea [51]. In April 2021, the project launched as a pilot program in collaboration with neighborhood organizations. To find more sensor placements, they used community involvement and NYCCAS methodologies [52]. While some devices were installed using a stratified random sampling technique, others were placed by environmental justice organizations. A variety of pollutants, including as CO_2, NO_2, SO_2, O_3, PM1, PM2.5, and PM10, as well as temperature, humidity, and pressure, can be measured by the specially made air quality sensors [53]. They also offer the Air Quality Index (AQI) values.

Higher index readings imply a greater risk of health concern, and the specially built air quality sensor utilized in Project Eclipse complies with EPA air quality guidelines. The six levels of the AQI, which reflect good, moderate, unhealthy for sensitive groups, unhealthy, very unhealthy, and hazardous air quality levels, respectively, are 0 to 50, 51 to 100, 101 to 150, 151 to 200, 201 to 300, and 301 and above [54]. The solar-powered device has a data collection period of between seconds and minutes. Through OpenAPI, users can view the gathered data as an interactive map that is made available to the public.

Variations in air quality are shown in Fig. 1, which also shows the daily average air quality over the study period and the deployment of Project Eclipse sensors. Although the majority of Chicago's air quality remained acceptable or moderate, one sensor, No. 2117, shown in Fig. 1 as red, reported extremely hazardous or dangerous conditions. While Chicago's air quality appears to be acceptable or moderate overall, there have been times when it has reached dangerous or unhealthy levels, requiring emergency health alerts. An illustration of every day AQI that sensors recorded during the investigation is shown in Fig. 2. Fig. 2 presents an overview of all the observation points gathered, as opposed to merely reviewing the air quality at a particular moment in time. The dots in the picture are color-coded: green and yellow denote good to moderate conditions, while orange,

red, and purple signify harmful to hazardous air characteristics. According to the results, Chicago's air quality was primarily good or moderate, as opposed to instances of hazardous, unhealthy, and extremely unhealthy situations, as seen in Fig. 1.
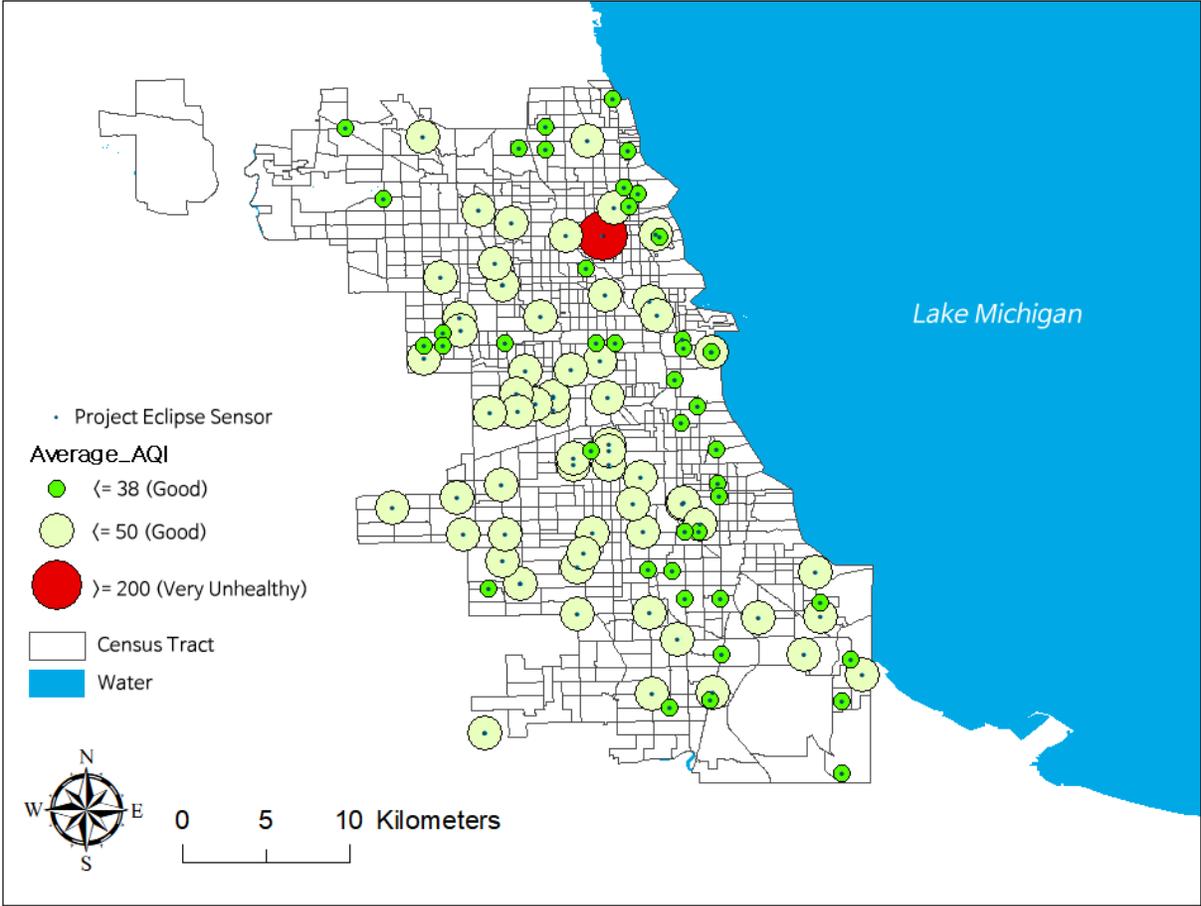

Figure 1. Study Area and Project Eclipse
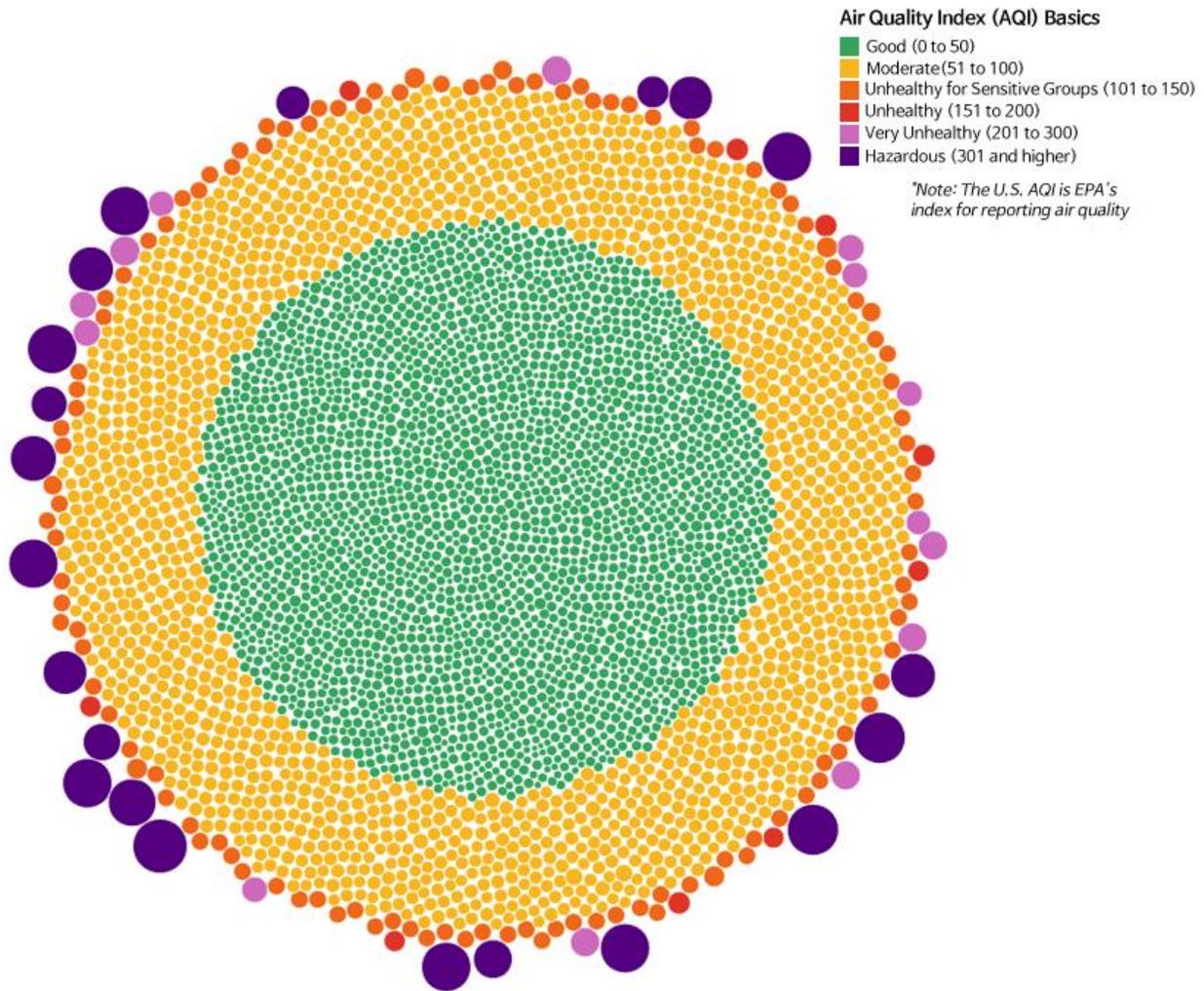
Figure 2. Daily Average Air Quality Index in Chicago

3.2. Research Methods

Fig. 3 shows the process of merging multiple approaches. During the initial stage, a range of factors that could potentially impact Chicago's air quality were gathered and processed. These factors included land use characteristics and data from urban climate sensors that were sourced from Microsoft's Project Eclipse sensors. Landsat8, a map of land use inventory, and GSV were used to extract land use features. The gathered land use features were separated into three groups and PCA was used to break them down in the second step. Biplot was used to locate the variances. In the third stage, we investigated their externalities on air quality using the XGB machine learning

regression model. After integrating the PCA values, the model was optimized and cross-validated. To determine each feature's positive or negative contribution, we computed SHAP values.

An Alienware M17 was used to provide technical help. The device has a 16 GB installed RAM and an Intel Core i7-8750H CPU that runs at 2.20GHz with 2.21 GHz. We processed the model and the street view photos utilizing an Nvidia GeForce RTX 2070 GPU that had 8GB of RAM. Processing street view photos usually takes fifteen minutes or so. We utilized Numpy [55] and Pandas [56], two Python modules, for the preprocessing of the data. For the purpose of data visualization, we used Tableau Desktop, Seaborn [58], and Matplotlib [57].
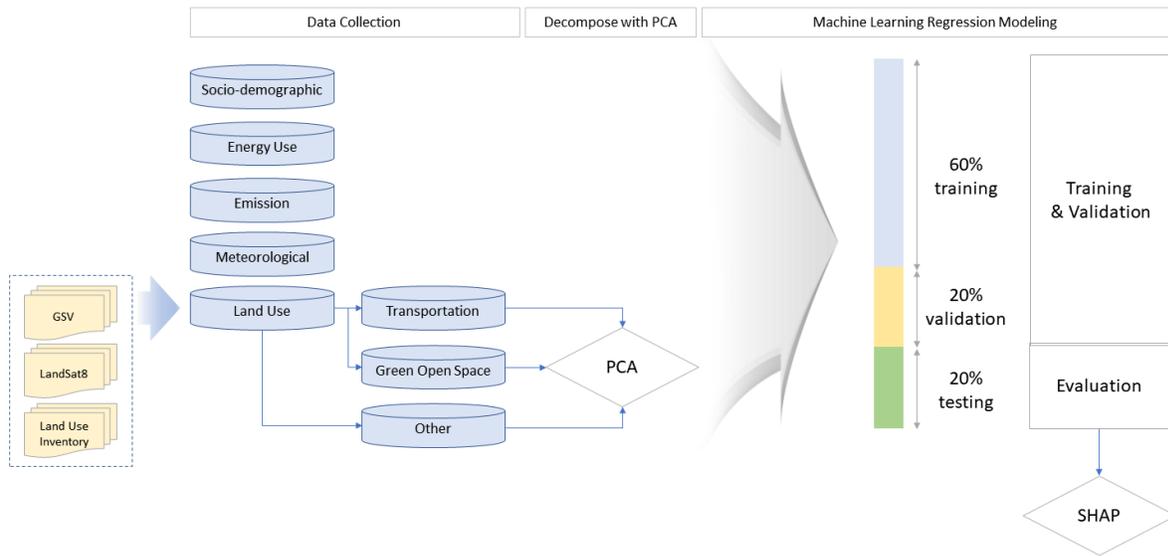
Figure 3. Flow Chart of Combining Multiple Methods

3.3. Land Use Feature Extraction from Street View Imagery

In order to automatically obtain street view photos using address information, the GSV static API was purchased for this investigation. Urban land use data was previously preprocessed using Landsat satellite photos and land use cover maps; however, in this work, we compared the GSV images with these traditional techniques. We captured the surrounding urban surroundings surrounding the Project Eclipse sensors that were mounted using GSV pictures.

To create GSV photos, four pictures at heading angles of 0°, 90°, 180°, and 270° were shot for each street node. Pictures from 2019 that were pertinent to the Chicago Project Eclipse sensor locations were obtained. Fig. 4 displays some of the street view pictures that were downloaded.

Convolutional neural networks (CNNs), like VGG [59], ResNet [60], Yolo [61], and Mask RCNN, have made it possible to automatically extract scene elements from street view photos in an effective manner. For every node in our investigation, we extracted scene features from GSV pictures using ResNet50. Fig. 5 shows our ResNet50's organizational structure.

A CNN with 50 layers is called ResNet50. CNNs are used to process input images, taking into account both the image's width and height. Images with 224 pixels in both height and width are fed into ResNet50. Using taught weights, the 'conv' layers in Figure 5 are in charge of extracting picture characteristics. To improve accuracy, these feature maps are subsequently joined across several layers by training weights. Eventually, the feature map that the convolutional layers produced is classified using a fully connected layer.

Widespread application of neural networks, such as YOLO and ResNet, has been observed in object detection [14, 62–64] and image categorization [65–66]. Convolutional layers use several layers for enhanced accuracy by extracting picture features based on training weights and connecting feature maps. The feature map that was extracted using conventional layers is finally classified by a fully linked layer.

The Places365 dataset, which categorizes photos into 365 categories of scenes like roads, woods, fields, streets, churches, and plazas, is used to pre-train the ResNet50 model. 18 million photos are used to train the model's weights, producing predictions for each image class with a high degree of accuracy. In comparison to VGG networks, ResNet50 requires fewer parameters and requires less computing power during model training since it uses an identity shortcut connection strategy to bypass one or more layers [67]. ResNet beats VGG, GoogleLeNet, and AlexNet, which have top-5 accuracies of 85.01%, 84.01%, and 82.75%, respectively, with an accuracy of 85.07 [68]. Every heading angle generates features:

$$X'_i = \left[X'_{i,0°}, X'_{i,90°}, X'_{i,180°}, X'_{i,270°} \right]$$

where X'_i is the list of features at each of the four heading angles of Node i.

To estimate the scene category for a sensor, we computed the average of the probabilities derived from the four angles, as our model's output provides the likelihood of each scene category. It is hypothesized that the sensor records the scene in four distinct orientations while viewing it in quarter view. Therefore, the following equation was used to calculate the average probability for each scene category:

$$X_{i,j} = \frac{1}{4}\left(X'_{i,0°,j} + X'_{i,90°,j} + X'_{i,180°,j} + X'_{i,270°,j}\right)$$

where $X_{i,j}$ is the feature of Node i at 365 dimensions j, and $i \in N, j \in D$ (categories of scene).

Based on traditional land use characteristics, we categorized the 365 options pertaining to different land use types into more focused groups in this study. Residential, office, industrial, green open space, and transportation are the five primary land use types that we have identified. A potential ratio for each category shows how likely it is that the associated land use type will be present in the scene. In particular, the office land use type includes offices, office buildings, and office cubicles, whereas the residential land use type refers to the potential for housing apartment complexes. The potential for an industrial area is represented by the industrial land use type, whereas the potential for a park, yard, roof garden, botanical garden, and formal garden is combined to form the green open space type. Lastly, the possibility of bridges, bus stops, crosswalks, driveways, parking (including lots and indoor and outdoor garages), train tracks, streets, and highways is taken into account by the transportation land use type. Once more, the 365 dimensions form the basis of the possibility ratio.
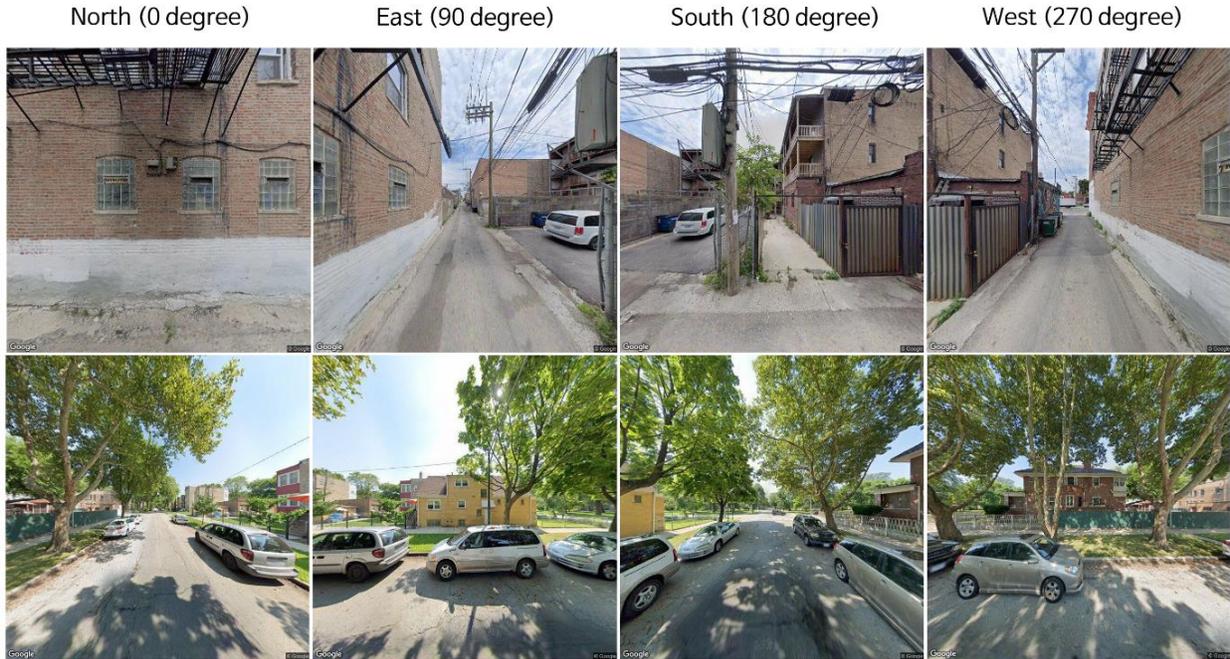
North (0 degree)   East (90 degree)   South (180 degree)   West (270 degree)

Figure 4. Examples of Downloaded Google Street View Images
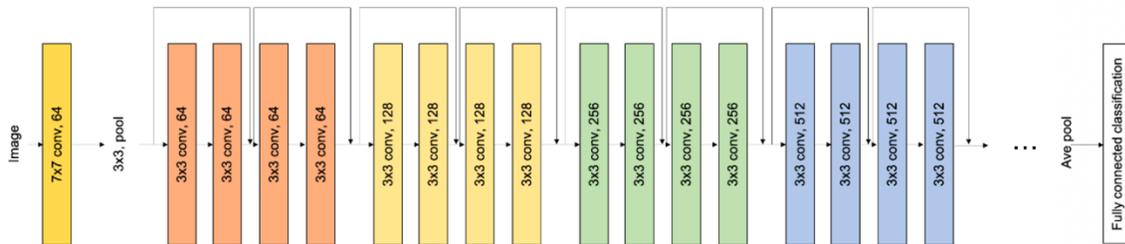
Figure 5. Structure of ResNet50

3.4. Land Use Feature Extraction from Conventional Approaches

Satellite photos and the land use inventory map were processed using GIS for the traditional methods. The ArcMap 10.8 software offered technical assistance. Despite being titled "the land use inventory of 2015," the most recent land use inventory map of Chicago that is currently available was acquired from the Illinois government and published in November 2020. We used a few methods to preprocess the map of land use inventory. First, we established a 400-meter buffer around every Project Eclipse sensor that was installed. Second, we crossed these barriers with the most recent Chicago land use inventory map from 2015. Thirdly, for every Project

Eclipse sensor, we dissolved the intersected layer to determine the ratio of land use types. Fourth, using zoning and building laws as a basis, we combined the computed land use types ratios by principal land use type. Residential, commercial, office, cultural, institutional, industrial, and vacant or under construction sites were among the land use types that were taken into consideration in our study.

We used the September 2021 LandSat8 satellite picture to derive satellite imagery. This picture was selected because it was taken during the day and had very little cloud cover, less than 10%. As seen in Fig. 6, we used ArcMap's image classification toolbar [69] to categorize the different land use types by identifying flora, water features, dense populations, and bare land. Following classification, we extracted the land use inventory and clipped the land use layer to the 400-meter buffers of the deployed Project Eclipse sensors. Ultimately, the three land use features are derived using three distinct sets of data: satellite imagery from LandSat-8, GSV pictures, and land use maps.
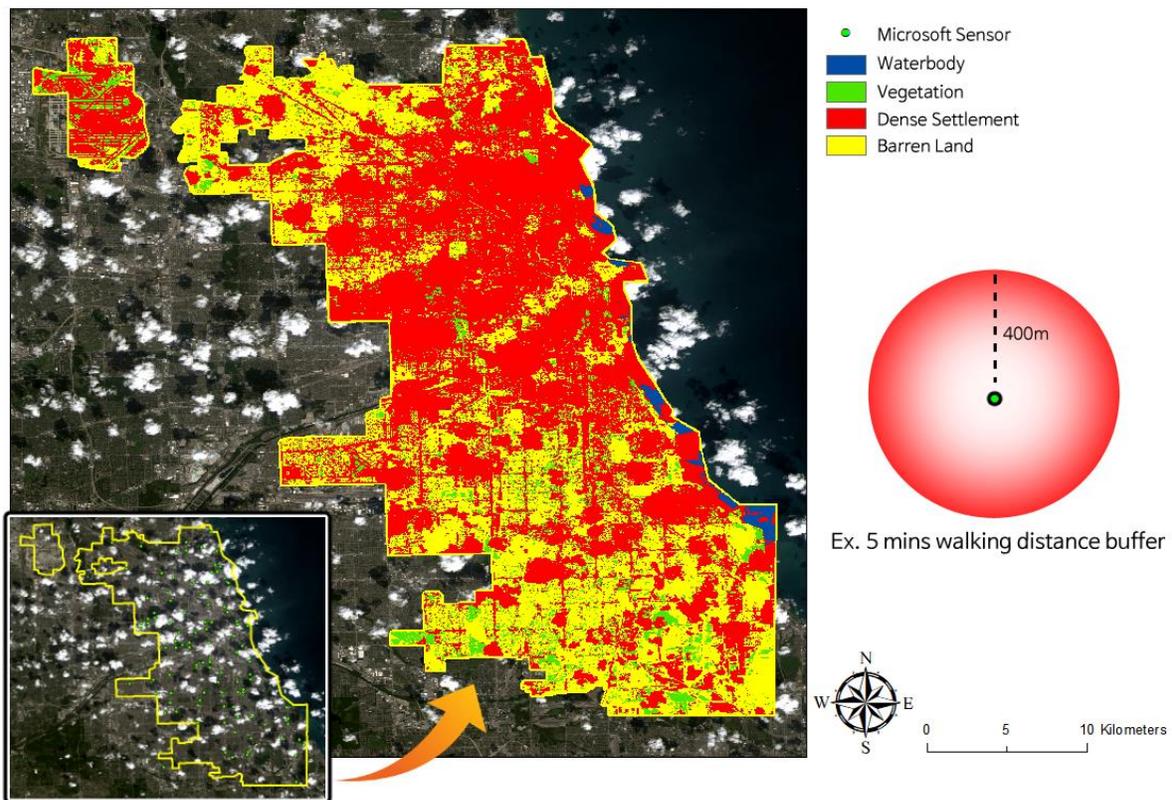

Figure 6. Supervised Land Use Classification Using LandSat8

3.5. Variables

The hourly average AQI recorded by Project Eclipse sensors for the month of August 2021 served as the study's dependent variable. But several sensors were found to be outliers because they either failed to complete the process or kept repeating it, and some were found to be outside of Chicago or to be devoid of street view photos. Consequently, we classified these sensors as anomalies and eliminated them from the examination. Only 96 of the 110 Project Eclipse sensors were thought to be appropriate for the investigation. In addition, we standardized the dependent variable using its natural logarithm. Fig. 7 shows the histogram representing the normalized dependent variable. A histogram is a graphical tool used to visualize the numerical data's frequency distribution. Five criteria were taken into consideration in our study: land use, energy usage, emissions, sociodemographic characteristics, and meteorological conditions. Features derived from Landsat satellite imagery, land use maps, and GSV pictures made up the land use category.

To organize the three distinct sources, we created three land use categories: transportation, open space, and other land uses. The division of traffic and green space was justified by the common belief that they either exacerbate or lessen air pollution. Whereas the land use category for green open space includes LandSat satellite imagery of the waterbody, vegetation, and barren land, the land use category for transportation includes street view imagery of green open space and land uses an inventory of agricultural, open space, and water. The land use inventory of residential, commercial, office, cultural, institutional, industrial, vacant or under construction areas, as well as the dense settlement from LandSat satellite imagery, are all included in the other land use category. Lastly, street view imagery of residential, office, and industrial areas is included. Principal component analysis was used to break down each land use group for two reasons (PCA). First, unlike meteorological variables like temperature, which vary with time, land use is a fixed geographical variable. Due to the underlying nature of geographic data, decomposition is required. PCA is usually chosen, as are Empirical Orthogonal Functions (EOFs) [32]. Prior to performing PCA analysis, the components were normalized using StandardScaler [70]. Regression analysis was conducted using the PCA scores obtained, taking

into account the original data sources. Second, multidimensional variances in two dimensions can be interpreted using the biplot that PCA generates.

Using registered population data from the 2020 decennial census, the socio-demographic component calculates the average population density within 400-meter buffers of the Project Eclipse sensors installed in Chicago. The census population of each Chicago census tract was united with the average, and the appropriate buffer was spatially joined with the average. The average site, source, weather site, and weather Energy Use Intensity (EUI) make up the energy consumption factors, whilst the average greenhouse gas emissions and intensity make up the emission factors. The emission and EUI statistics, which were geocoded in ArcMap above the buffer layer, are provided by the City of Chicago. To determine the mean emission within the relevant buffer, the emission measurements were spatially combined using the 400-meter buffer of the deployed Project Eclipse sensors as the base layer.

As independent variables, the humidity and hourly average temperature recorded by the same Project Eclipse sensor were added. Although the Project Eclipse sensors provide data on pollutants like particulate matter, their direct relationship to the dependent variable, AQI, precluded their inclusion in this analysis as independent variables. Particulate matter as an independent variable would reduce other factors' ability to explain the data. Every deployed Project Eclipse sensor in Chicago serves as the analysis's unit, and it is done hourly. August 2021 is the study period's month.
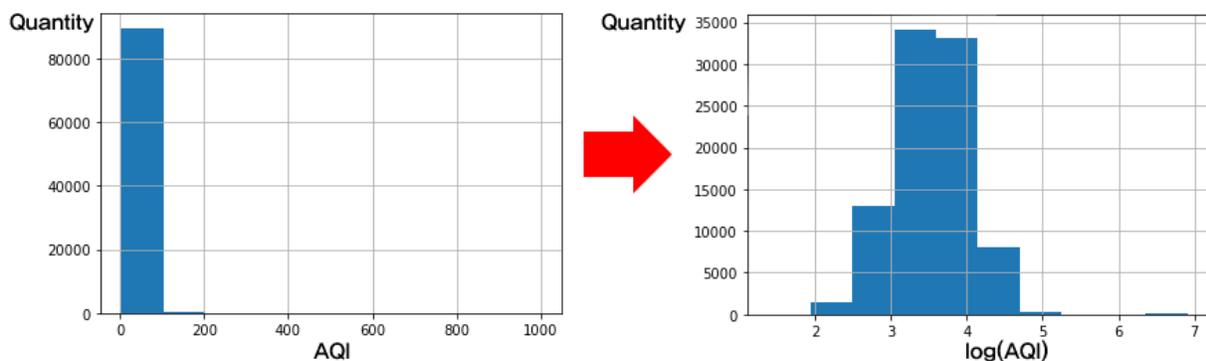

Figure 7. The Normalized dependent Variable of the Number of Observations

3.6. Machine Learning Regression Modeling

This work made use of Chen and Guestrin's [71] XGBoost (XGB), an ensemble learning system that uses boosting techniques. XGB is a scalable paradigm with comparatively little computational overhead [71]. XGB has proven to perform better than neural networks and Random Forest when it comes to land use regression [72]. Multiple learners are used in ensemble learning to tackle complicated issues; voting mixes several methods for its learners, and boosting techniques modifies weights to improve prediction accuracy [73].

The notion of decision trees (DT) serves as the foundation for ensemble machine learning techniques. Consequently, these algorithms' hyperparameters and DTs' hyperparameters are comparable. Usually, these hyperparameters are externally set and constant. It is imperative to use the hyperparameter optimization procedure to stop the model from overfitting.

In order to maximize the hyperparameters in XGB, GridSearchCV was utilized [70]. Four hyperparameters were investigated in our study: the minimum weight, learning rate, maximum tree depth, and number of estimators. Point estimates are referred to as estimators, while decision trees' maximum number of splits is constrained by the max depth hyperparameter. The step size in the computation that minimizes the loss function is controlled by the learning rate hyperparameter. Lastly, throughout the tree-building process, the minimal weight hyperparameter is employed to guarantee that a minimum quantity of samples is needed to divide a node.

In particular, we took into account maximum depths of 6, 8, 10, 12, and 16 in addition to 100, 400, and 800 estimators and 0.05, 0.1, 0.15, and 0.20 learning rates. The minimum weight was also changed in our study, with choices of 1, 10, and 100. Using 5-fold cross-validation, we randomly divided the data into training (60%), cross-validation (20%), and testing (20%) sets. We repeat the sequence with a random state of 1 to guarantee consistency. An observation's sample number is referred to as a random state.

The cross-validation evaluation score that we arrived at was the negative mean absolute error, or NMAE. Higher scores correspond to higher ranks for estimators in GridSearchCV, which ranks estimators according to their error function scores. We set the score to negative in order to be consistent with our objective of minimizing error. We fitted the model 900 times after searching 180 potential hyperparameter settings in 5-folds. With 800 estimators, a maximum depth of 16, a minimum weight of 1, and a learning rate of 0.1 in our optimized hyperparameter setup, we were able to obtain an NMAE of -0.2503.

3.7. Evaluation

In our final analysis, we only used the testing dataset to calculate the mean squared error (MSE), root mean squared logarithmic error (RMSLE), and mean absolute error (MAE). The MSE, RMSLE, and MAE scores that our improved XGB model produced were 0.161, 0.401, and 0.249, respectively.

In order to assess the feature explanations, we lastly computed the SHAP value [74]. To divide the "payout" equitably, the mean marginal contribution of features determined by repeating the computation of feasible coalitions is represented by the SHAP value [75]. The origins of SHAP values can be found in cooperative game theory, which deals with the equitable allocation of costs and profits [76].

We used our trained XGB model to apply to the SHAP module in order to generate SHAP values [69]. However, because it requires computing potential coalitions, determining the SHAP value can be computationally demanding. To calculate the SHAP value, we therefore randomly pulled 500 samples and chose one sensor that reported Chicago's most hazardous air quality state. We used the SHAP decision plot, which shows feature contribution, to illustrate their local observation. The graphic illustrates how feature values lead to the expected value starting at the base value, or output value.

Chapter 4. Results

4.1. Measured Air Quality Status in Chicago

The City of Chicago (COC) determined the Air Quality and Health Index scores for every census block group in 2020 and carried out a community-level analysis of pollution exposure. Greater burdens are indicated by a higher score. According to their findings, the city's west and south bear the brunt of the strain.

In this study, Ordinary Kriging was utilized to extend the point observation of the daily mean AQI acquired by sensors. The generated map was then split into quantile breakpoints that were the same as those utilized in the COC's investigation. A comparison of the two maps of air quality is presented in Fig. 8. According to both assessments, the city's western areas have comparatively low air quality. Nonetheless, compared to the COC's established values, our AQI map indicated that the South regions were comparatively less affected.

This research separated the day's hours into seven-hour segments for the purpose of descriptive spatiotemporal analysis. The hourly mean AQI recorded over these periods by Project Eclipse is displayed in Fig. 9. Chicago's air quality is generally considered to be good. On the other hand, during the late morning, evening, night, and late night hours, the Project Eclipse sensor, which is situated at the intersection of West Belmont Avenue in the northeastern part of Chicago, identified extremely unhealthy or hazardous air quality conditions. These times usually correspond with rush hour, and pollutants produced in the evenings may remain inert until the wee hours of the morning.
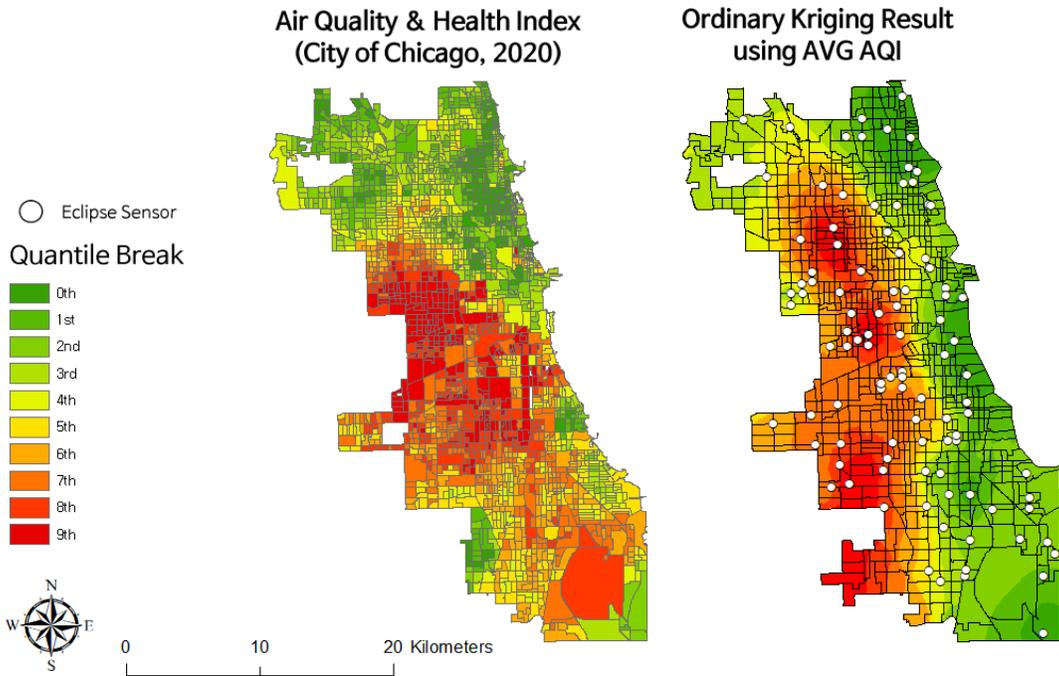
Figure 8. Comparison of Air Quality Analyses

Figure 9. Mean AQI in Hour Periods

4.2. PCA Analysis Result

Fig. 10 shows how the cumulative variance explained grows with the number of principle components. The variation of the data is captured by each main component, with the first one capturing the largest amount. The cumulative variance plot illustrates the proportion of variance in the total that can be attributed to each component as it accumulates. The findings show that with six, five, and seven main components, respectively, transportation, green open space, and other land use categories can explain more than 80% of the overall variance.

Biplots overlaying the two principal components with the highest percentage of variance are displayed in Fig. 11. Every observation in a biplot is shown as a point in the two-dimensional space that the first two principal components define. A line that is most parallel to the X- or Y-axis represents each variable's contribution to the major components. The variation's size is indicated by the length of the red line; the greater the variance, the longer the line. A comparatively higher correlation between the observations is indicated by the proximity of the arrow lines. The street view imaging variables exhibit different characteristics in the biplots of transportation, green open space, and other land use categories. The transportation biplot's street view imagery's arrow line does not align with, nor point in the same direction as, the land use inventory's transportation land use factor (see Fig. 11a.). The street view imagery variables in the green open space biplot are situated farther away from the open space land use factor found in the land use inventory (refer to Fig. 11b.). A waterbody variable from a land satellite image is near the arrow line in the Street View imagery of green open space. The biplot analysis reveals that there are comparatively stronger correlations between the industrial area from street view images and the land use inventory, as well as between industrial and unoccupied or construction land use (see Fig. 11c.). When comparing office and residential land use from street view imagery and dense settlement from satellite pictures, the arrow lines representing office, institutional, and commercial land use from the land use inventory are closer. It should be noted that the residential land use arrow line from the land use inventory points in a different direction.
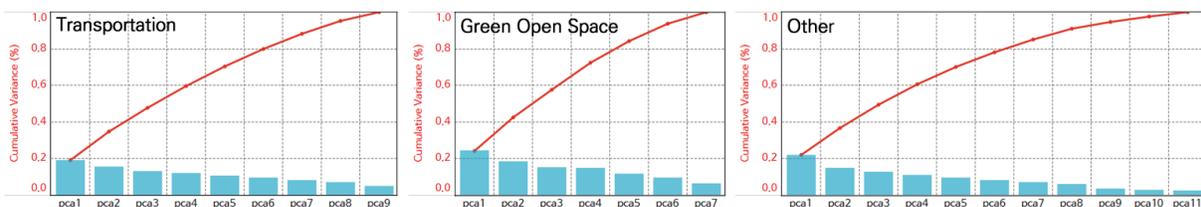

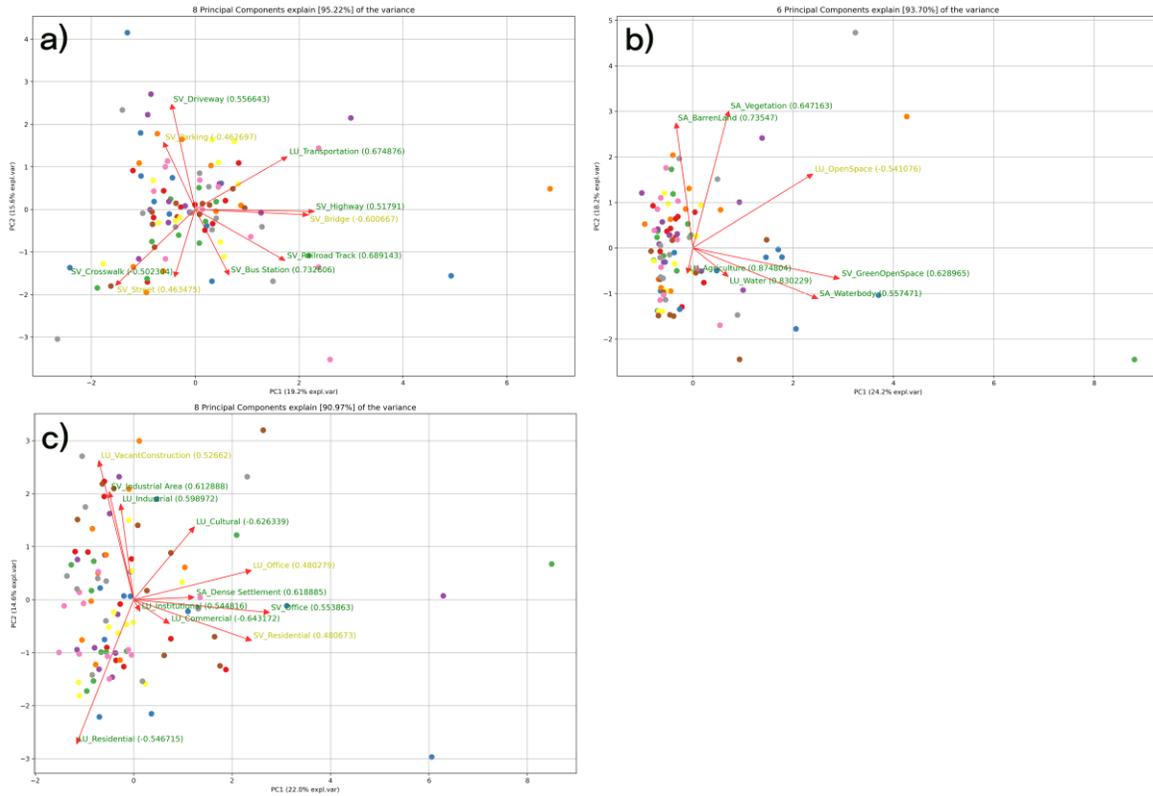*Figure 10. Cumulative Eigenvalue Variances*

Figure 11. Biplot of a) Transportation, b) Green Open Space, c) Other Land Use Category

4.3. SHAP Value Calculation Result

The influence of characteristics on the SHAP values, which quantify each feature's contribution to the final air quality prediction, is depicted in Fig. 12. A higher AQI denotes a more serious issue. The first two observations are from a sensor (No. 2117 at West Belmont Avenue), which reports the most hazardous air quality status in Chicago; the other two are samples taken at random. With the assigned feature values in parenthesis, each graph displays the cumulative effect of features on the final score at the top.

At the West Belmont Avenue intersection, where hazardous air quality was observed, the main elements in the other land use category generally had a greater influence on the air quality than other elements. Population density and the third primary component of the other land use category, in particular, significantly improved the SHAP value. The primary components of the other (second and third) and transportation (fifth) land use categories, as well as meteorological

conditions, had an adverse effect on the SHAP value in the random samples. The dependent variable was marginally impacted by the primary elements of the green open space land use category, emission, and energy factors in both measures.
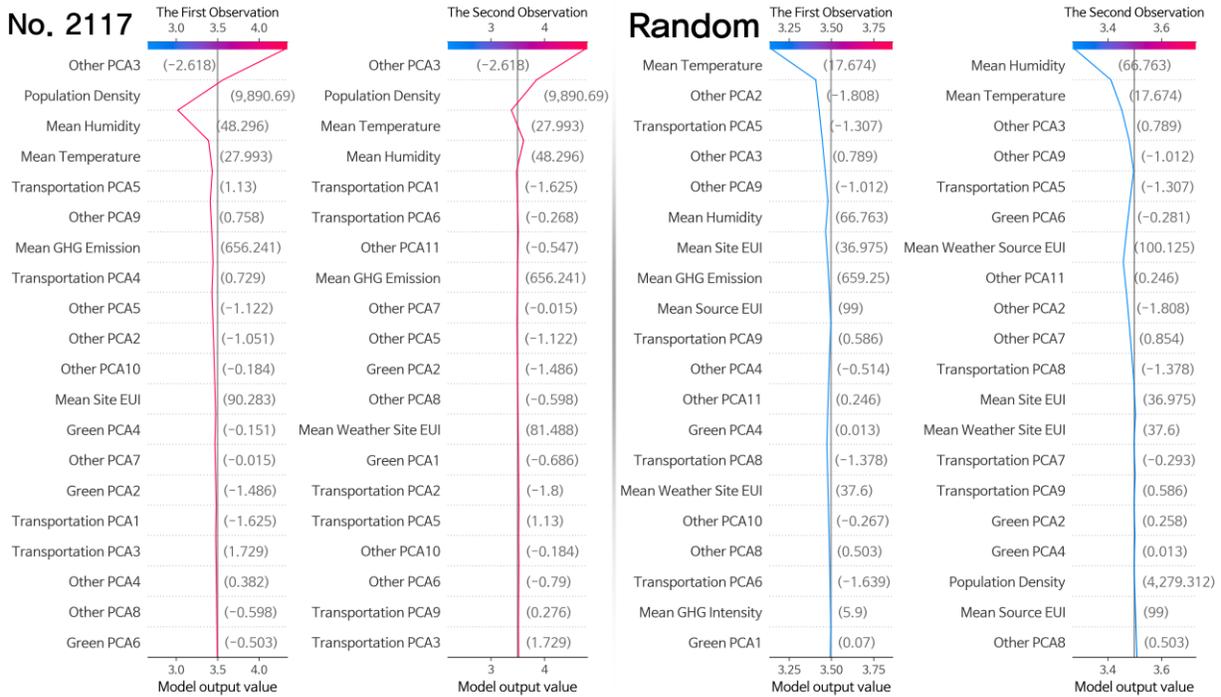
Figure 12. SHAP Decision Plot

Chapter 5. Discussion and Conclusions

In order to improve future urban design, this project introduced street view imagery and urban temperature sensors. In particular, we used GSV pictures in our investigation and contrasted them with land use components that were obtained using conventional land use extraction techniques. The objective was to evaluate the impact of land use on sensor-measured air quality.

The following are the study's main conclusions. First, extracting new land use features can benefit greatly from the usage of street view imagery. The cumulative scenery from the four directions, north, east, west, and south, near the sensor locations is provided by street view photography, and thus may offer more precise urban land use settings. Land use information is divided into more specific categories, such as office buildings, gas stations, campuses, and bus

stations, using land use attributes that are taken from pre-trained datasets, like Places365. Our PCA analysis and biplot demonstrated that, in contrast to traditional land use sources, street view imagery has a unique character. The creation of CNNs like VGG, ResNet, and YOLO has made it possible to swiftly and effectively extract information from street view photos. Rather of arguing for or against a certain typology, we propose to close the current gap in land use feature extraction by combining street view photos with traditional methods. We currently average the scene features from various viewpoints, however adding weights to the scenes or angles would be a good idea if some scene categories or kinds have a bigger influence on the quality of the air.

Second, according to data from urban climate sensors, there are dangerous air quality problems at the intersection of West Belmont Avenue. The EPA has classified these situations as hazardous, meaning that they could pose a health risk to everyone when they get worse in the late morning, evening, and nighttime. More research is necessary because these results align with the COC's previous evaluations of the air quality and health index. In order to detect any discrepancies in the data, it is worthwhile to investigate the possibility of incorporating hyperlocal air quality measures from urban climate sensors into the current measurement techniques used by agencies and the COC. As digitization becomes an unavoidable trend in modern planning practice and IoT sensor installations proliferate throughout cities [77], COC should encourage the involvement of diverse stakeholders and the decentralized placement of sensors. But rather than only promoting private vendor operation, it is crucial to take environmental justice concerns into account and figure out ways to distribute sensors to underprivileged communities while involving community engagement.

Third, our research supports other research [8, 11–14] that found that diverse land uses have variable effects on air quality. It was discovered that the impacts of the land use primary components varied between the samples. The surrounding other land use characteristics considerably worsened the dangerous air quality at the West Belmont Avenue intersection where it was claimed to be. In order to lessen the impact, we advise that the COC handle the built environment and land use features independently.

Using data from recently installed Microsoft urban climate sensors, this study used a novel research methodology that combined traditional methods of quantifying the effects of land use on urban air quality with street view photography. The study adds to the field in a number of ways. First of all, it suggests a novel method for utilizing street view photos to quantify the impact of land use on urban air quality. Second, the study made advantage of Chicago's previously untapped urban climate sensor data. Microsoft intends to increase the number of sensors installed across the country, and this analysis framework offers a starting point for using and validating the data. Thirdly, our research verified that the utilization of street view photos in conjunction with urban land use has diverse effects on air quality at hyperlocal scales. Fourth, the study of Chicago's air quality revealed that there are dangerous air conditions in several areas of the city that call for emergency health alerts. Policymakers and local government organizations can use the findings to better Chicago's air quality conditions by determining how certain factors affect the AQI in the impacted areas.

Our analysis does, however, recognize some restrictions on the application of Project Eclipse sensors. Even with our best efforts to eliminate anomalies, skewed results or problems with the sensors are still possible. It's also unclear how Microsoft calculates the AQI for Project Eclipse. We advise cross-validating the Project Eclipse data using traditional air quality monitoring techniques to guarantee accurate results. Further research could also look at a comparison between Purple Air and the Project Eclipse operation.

It is important to take note of a few other issues. First, Places365 pre-trains the ResNet50 model, which might have trouble adjusting for biases. Secondly, the buffer's absolute threshold distance is not unambiguous. Third, the four perspectives offered by Google Street View photos might not adequately depict the intricacy of the metropolitan setting. Fourth, micro-scale traffic flow and emission parameters that potentially affect air quality were not taken into account in our regression study. These limitations should be addressed in future study.

References

1. Portier, C., Thigpen Tart, K., Carter, S., Dilworth, C., Grambsch, A., Gohlke, J., Hess, J., Howard, S., Luber, G., Lutz, J., Maslak, T., Radtke, M., Rosenthal, J., Rowles, T., Sandifer, P., Scheraga, J., Strickman, D., Trtanj, J. & Whung, P.-Y. (2010). A Human Health Perspective on Climate Change: A Report Outlining Research Needs on the Human Health Effects of Climate Change. Environmental Health Perspectives. https://doi.org/10.1289/ehp.1002272

2. Watts, N., Amann, M., Arnell, N., Ayeb-Karlsson, S., Beagley, J., Belesova, K., Boykoff, M., Byass, P., Cai, W., Campbell-Lendrum, D., Capstick, S., Chambers, J., Coleman, S., Dalin, C., Daly, M., Dasandi, N., Dasgupta, S., Davies, M., Di Napoli, C., … Costello, A. (2021). The 2020 report of The Lancet Countdown on health and climate change: responding to converging crises. The Lancet, 397(10269), 129–170. https://doi.org/10.1016/S0140-6736(20)32290-X

3. Jacob, D. J. & Winner, D. A. (2009). Effect of climate change on air quality. Atmospheric Environment, 43(1), 51–63. https://doi.org/10.1016/j.atmosenv.2008.09.051

4. Plageras, A. P., Psannis, K. E., Stergiou, C., Wang, H. & Gupta, B. B. (2018). Efficient IoT-based sensor BIG Data collection–processing and analysis in smart buildings. Future Generation Computer Systems, 82, 349–357. https://doi.org/10.1016/j.future.2017.09.082

5. Lee, H., Choi, S. & Jiao, J. (2021). Examining the COVID-19 effects on travel behavior using smart IoT sensors: A case study of smart city planning in Gangnam, Seoul. International Journal of Sustainable Building Technology and Urban Development, 12(4), 347–362. https://doi.org/10.22712/SUSB.20210029

6. Lytras, M. D. & Visvizi, A. (2018). Who Uses Smart City Services and What to Make of It: Toward Interdisciplinary Smart Cities Research. Sustainability, 10(6), 1998. https://doi.org/10.3390/su10061998

7. Gao, J. & Zha, Y. (2010). Meteorological Influence on Predicting Air Pollution from MODIS-Derived Aerosol Optical Thickness: A Case Study in Nanjing, China. Remote Sensing, 2(9). https://doi.org/10.3390/rs2092136

8. Iskandaryan, D., Ramos, F. & Trilles, S. (2020). Air Quality Prediction in Smart Cities Using Machine Learning Technologies based on Sensor Data: A Review. Applied Sciences, 10(7), 2401. https://doi.org/10.3390/app10072401

9.   West, J. J., Smith, S. J., Silva, R. A., Naik, V., Zhang, Y., Adelman, Z., Fry, M. M., Anenberg, S., Horowitz, L. W. & Lamarque, J.-F. (2013). Co-benefits of mitigating global greenhouse gas emissions for future air quality and human health. Nature Climate Change, 3(10), 885–889. https://doi.org/10.1038/nclimate2009

10.   Wang, J. C. (2019). Analysis of energy use intensity and greenhouse gas emissions for universities in Taiwan. Journal of Cleaner Production, 241, 118363. https://doi.org/10.1016/j.jclepro.2019.118363

11.   Heald, C. L. & Spracklen, D. V. (2015). Land Use Change Impacts on Air Quality and Climate. Chemical Reviews, 115(10), 4476–4496. https://doi.org/10.1021/cr500446g

12.   McCarty, J. & Kaza, N. (2015). Urban form and air quality in the United States. Landscape and Urban Planning, 139, 168–179. https://doi.org/10.1016/j.landurbplan.2015.03.008

13.   Zou, B., Xu, S., Sternberg, T. & Fang, X. (2016). Effect of Land Use and Cover Change on Air Quality in Urban Sprawl. Sustainability, 8(7). https://doi.org/10.3390/su8070677

14.   Hewitt, C. N., Ashworth, K. & MacKenzie, A. R. (2020). Using green infrastructure to improve urban air quality (GI4AQ). Ambio, 49(1), 62–73. https://doi.org/10.1007/s13280-019-01164-3

15.   Lee, M., Kloog, I., Chudnovsky, A., Lyapustin, A., Wang, Y., Melly, S., Coull, B., Koutrakis, P. & Schwartz, J. (2016). Spatiotemporal prediction of fine particulate matter using high-resolution satellite images in the Southeastern US 2003–2011. Journal of Exposure Science & Environmental Epidemiology, 26(4), 377–384. https://doi.org/10.1038/jes.2015.41

16.   Xu, G., Jiao, L., Zhao, S., Yuan, M., Li, X., Han, Y., Zhang, B. & Dong, T. (2016). Examining the Impacts of Land Use on Air Quality from a Spatio-Temporal Perspective in Wuhan, China. Atmosphere, 7(5), 62. https://doi.org/10.3390/atmos7050062

17.   Henneman, L. R. F., Shen, H., Hogrefe, C., Russell, A. G. & Zigler, C. M. (2021). Four Decades of United States Mobile Source Pollutants: Spatial–Temporal Trends Assessed by Ground-Based Monitors, Air Quality Models, and Satellites. Environmental Science & Technology, 55(2), 882–892. https://doi.org/10.1021/acs.est.0c07128

18.   Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q. & Zhang, W. (2015). Assessing street-level urban greenery using Google Street View and a modified green view index. Urban Forestry & Urban Greening, 14(3), 675–685. https://doi.org/10.1016/j.ufug.2015.06.006

19.　　Yin, L. & Wang, Z. (2016). Measuring visual enclosure for street walkability: Using machine learning algorithms and Google Street View imagery. Applied Geography, 76, 147–153. https://doi.org/10.1016/j.apgeog.2016.09.024

20.　　Richards, D. R. & Edwards, P. J. (2017). Quantifying street tree regulating ecosystem services using Google Street View. Ecological Indicators, 77, 31–40. https://doi.org/10.1016/j.ecolind.2017.01.028

21.　　Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L. & Fei-Fei, L. (2017). Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. Proceedings of the National Academy of Sciences, 114(50), 13108. https://doi.org/10.1073/pnas.1700035114

22.　　Rundle, A. G., Bader, M. D. M., Richards, C. A., Neckerman, K. M. & Teitler, J. O. (2011). Using Google Street View to Audit Neighborhood Environments. American Journal of Preventive Medicine, 40(1), 94–100. https://doi.org/10.1016/j.amepre.2010.09.034

23.　　Hara, K., Le, V. & Froehlich, J. (2013). Combining crowdsourcing and google street view to identify street-level accessibility problems. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 631–640. https://doi.org/10.1145/2470654.2470744

24.　　Isola, P. D., Bogert, J. N., Chapple, K. M., Israr, S., Gillespie, T. L. & Weinberg, J. A. (2019). Google Street View assessment of environmental safety features at the scene of pedestrian automobile injury. The Journal of Trauma and Acute Care Surgery, 87(1), 82–86. https://doi.org/10.1097/TA.0000000000002338

25.　　Ilic, L., Sawada, M. & Zarzelli, A. (2019). Deep mapping gentrification in a large Canadian city using deep learning and Google Street View. PLOS ONE, 14(3), e0212814. https://doi.org/10.1371/journal.pone.0212814

26.　　O'Regan, A. C., Byrne, R., Hellebust, S., & Nyhan, M. M. (2022). Associations between Google Street View-derived urban greenspace metrics and air pollution measured using a distributed sensor network. Sustainable Cities and Society, 87, 104221. https://doi.org/10.1016/j.scs.2022.104221

27.　　Messier, K. P., Chambliss, S. E., Gani, S., Alvarez, R., Brauer, M., Choi, J. J., Hamburg, S. P., Kerckhoffs, J., LaFranchi, B., Lunden, M. M., Marshall, J. D., Portier, C. J., Roy, A., Szpiro, A. A., Vermeulen, R. C. H., & Apte, J. S. (2018). Mapping Air Pollution with Google Street View Cars: Efficient Approaches with Mobile Monitoring and Land Use Regression.

Environmental Science & Technology, 52(21), Article 21.

https://doi.org/10.1021/acs.est.8b03395

28.   Parmar, G., Lakhani, S., & Chattopadhyay, M. K. (2017). An IoT based low cost air pollution monitoring system. 2017 International Conference on Recent Innovations in Signal Processing and Embedded Systems (RISE), 524–528.

https://doi.org/10.1109/RISE.2017.8378212

29.   Zhao, Y.-L., Tang, J., Huang, H.-P., Wang, Z., Chen, T.-L., Chiang, C.-W., & Chiang, P.-C. (2020). Development of IoT Technologies for Air Pollution Prevention and Improvement. Aerosol and Air Quality Research, 20(12), Article 12. https://doi.org/10.4209/aaqr.2020.05.0255

30.   Rovira, J., Domingo, J. L. & Schuhmacher, M. (2020). Air quality, health impacts and burden of disease due to air pollution (PM10, PM2.5, NO2 and O3): Application of AirQ+ model to the Camp de Tarragona County (Catalonia, Spain). Science of The Total Environment, 703, 135538. https://doi.org/10.1016/j.scitotenv.2019.135538

31.   US EPA, O. (2014, 10. April). NAAQS Table [Other Policies and Guidance].

https://www.epa.gov/criteria-air-pollutants/naaqs-table

32.   Amato, F., Guignard, F., Robert, S. & Kanevski, M. (2020). A novel framework for spatio-temporal prediction of environmental data using deep learning. Scientific Reports, 10(1), 22243. https://doi.org/10.1038/s41598-020-79148-7

33.   Marble, S. (2018). Everything That Can Be Measured Will Be Measured. Technology|Architecture + Design, 2(2), 127–129.

https://doi.org/10.1080/24751448.2018.1497355

34.   S. Mora, A. Anjomshoaa, T. Benson, F. Duarte, & C. Ratti. (2019). Towards Large-scale Drive-by Sensing with Multi-purpose City Scanner Nodes. 2019 IEEE 5th World Forum on Internet of Things (WF-IoT), 743–748. https://doi.org/10.1109/WF-IoT.2019.8767186

35.   Venter, Z. S., Brousse, O., Esau, I. & Meier, F. (2020). Hyperlocal mapping of urban air temperature using remote sensing and crowdsourced weather data. Remote Sensing of Environment, 242, 111791. https://doi.org/10.1016/j.rse.2020.111791

36.   Mayer, H. (1999). Air pollution in cities. Atmospheric Environment, 33(24), 4029–4037. https://doi.org/10.1016/S1352-2310(99)00144-2

37.   Holguin, F. (2008). Traffic, Outdoor Air Pollution, and Asthma. Immunology and Allergy Clinics of North America, 28(3), 577–588. https://doi.org/10.1016/j.iac.2008.03.008

38.   McConnell, R., Berhane, K., Yao, L., Jerrett, M., Lurmann, F., Gilliland, F., Künzli, N., Gauderman, J., Avol, E., Thomas, D. & Peters, J. (2006). Traffic, Susceptibility, and Childhood Asthma. Environmental Health Perspectives, 114(5), 766–772. https://doi.org/10.1289/ehp.8594

39.   Lu, J., Li, B., Li, H. & Al-Barakani, A. (2021). Expansion of city scale, traffic modes, traffic congestion, and air pollution. Cities, 108, 102974. https://doi.org/10.1016/j.cities.2020.102974

40.   Mölders, N. (2011). Land-Use and Land-Cover Changes: Impact on Climate and Air Quality. Springer Science & Business Media.

41.   Bandeira, J. M., Coelho, M. C., Sá, M. E., Tavares, R. & Borrego, C. (2011). Impact of land use on urban mobility patterns, emissions and air quality in a Portuguese medium-sized city. Science of The Total Environment, 409(6), 1154–1163. https://doi.org/10.1016/j.scitotenv.2010.12.008

42.   Shen, T. T. (1982). Air quality assessment for land disposal of industrial wastes. Environmental Management, 6(4), 297–305. https://doi.org/10.1007/BF01875061

43.   Weng, Q. & Yang, S. (2006). Urban Air Pollution Patterns, Land Use, and Thermal Landscape: An Examination of the Linkage Using GIS. Environmental Monitoring and Assessment, 117(1–3), 463–489. https://doi.org/10.1007/s10661-006-0888-9

44.   Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P. & Briggs, D. (2008). A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmospheric Environment, 42(33), 7561–7578. https://doi.org/10.1016/j.atmosenv.2008.05.057

45.   Yang, H., Chen, W. & Liang, Z. (2017). Impact of Land Use on PM2.5 Pollution in a Representative City of Middle China. International Journal of Environmental Research and Public Health, 14(5), 462. https://doi.org/10.3390/ijerph14050462

46.   Larkin, A. & Hystad, P. (2019). Evaluating street view exposure measures of visible green space for health research. Journal of Exposure Science & Environmental Epidemiology, 29(4), 447–456. https://doi.org/10.1038/s41370-018-0017-1

47.   Helbich, M., Yao, Y., Liu, Y., Zhang, J., Liu, P. & Wang, R. (2019). Using deep learning to examine street view green and blue spaces and their associations with geriatric depression in Beijing, China. Environment International, 126, 107–117. https://doi.org/10.1016/j.envint.2019.02.013

48. Li, X., Zhang, C. & Li, W. (2017). Building block level urban land-use information retrieval based on Google Street View images. GIScience & Remote Sensing, 54(6), 819–835. https://doi.org/10.1080/15481603.2017.1338389

49. Qi, M., Lu, T. & Hankey, S. (2020). Using Google Street View Imagery in Land Use Regression to Predict Street Level Particulate Air Pollution. ISEE Conference Abstracts. https://doi.org/10.1289/isee.2020.virtual.O-OS-641

50. Lu, T., Marshall, J. D., Zhang, W., Hystad, P., Kim, S.-Y., Bechle, M. J., Demuzere, M. & Hankey, S. (2021). National Empirical Models of Air Pollution Using Microscale Measures of the Urban Environment. Environmental Science & Technology, 55(22), 15519–15530. https://doi.org/10.1021/acs.est.1c04047

51. Daepp, M. I. G., Cabral, A., Ranganathan, V., Iyer, V., Counts, S., Johns, P., Roseway, A., Catlett, C., Jancke, G., Gehring, D., Needham, C., von Veh, C., Tran, T., Story, L., D'Amone, G. & Nguyen, B. H. (2022). Eclipse: An End-to-End Platform for Low-Cost, Hyperlocal Environmental Sensing in Cities. 2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), 28–40. https://doi.org/10.1109/IPSN54338.2022.00010

52. Matte, T. D., Ross, Z., Kheirbek, I., Eisl, H., Johnson, S., Gorczynski, J. E., Kass, D., Markowitz, S., Pezeshki, G. & Clougherty, J. E. (2013). Monitoring intraurban spatial patterns of multiple combustion air pollutants in New York City: design and implementation. Journal of Exposure Science & Environmental Epidemiology, 23(3), 223–231. https://doi.org/10.1038/jes.2012.126

53. Project Eclipse. (n.d.). Project Eclipse. Microsoft Research. Retrieved January 30, 2022, from https://www.microsoft.com/en-us/research/project/project-eclipse/

54. AQI Basics | AirNow.gov. (n.d.). AirNow.gov, U.S. EPA. Retrieved January 30, 2022, from https://www.airnow.gov/aqi/aqi-basics

55. Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. Nature, 585(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

56.   Reback, J., McKinney, W., jbrockmendel, Bossche, J. V. den, Augspurger, T., Cloud, P., gfyoung, Sinhrks, Klein, A., Roeschke, M., Hawkins, S., Tratner, J., She, C., Ayd, W., Petersen, T., Garcia, M., Schendel, J., Hayden, A., MomIsBestFriend, … Mehyar, M. (2020). pandas-dev/pandas: Pandas 1.0.3. Zenodo. https://doi.org/10.5281/zenodo.3715232

57.   Caswell, T. A., Droettboom, M., Lee, A., Andrade, E. S. de, Hoffmann, T., Hunter, J., Klymak, J., Firing, E., Stansby, D., Varoquaux, N., Nielsen, J. H., Root, B., May, R., Elson, P., Seppänen, J. K., Dale, D., Lee, J.-J., McDougall, D., Straw, A., … Ivanov, P. (2021). matplotlib/matplotlib: REL: v3.5.1. Zenodo. https://doi.org/10.5281/zenodo.5773480

58.   Waskom, M. L. (2021). seaborn: statistical data visualization. Journal of Open Source Software, 6(60), 3021. https://doi.org/10.21105/joss.03021

59.   Simonyan, K. & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition (arXiv:1409.1556). arXiv. https://doi.org/10.48550/arXiv.1409.1556

60.   He, K., Zhang, X., Ren, S. & Sun, J. (2015). Deep Residual Learning for Image Recognition (arXiv:1512.03385). arXiv. https://doi.org/10.48550/arXiv.1512.03385

61.   Redmon, J. & Farhadi, A. (2018). YOLOv3: An Incremental Improvement (arXiv:1804.02767). arXiv. https://doi.org/10.48550/arXiv.1804.02767

62.   Hendry, & Chen, R.-C. (2019). Automatic License Plate Recognition via sliding-window darknet-YOLO deep learning. Image and Vision Computing, 87, 47–56. https://doi.org/10.1016/j.imavis.2019.04.007

63.   Huang, R., Pedoeem, J., & Chen, C. (2018). YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers. 2018 IEEE International Conference on Big Data (Big Data), 2503–2510. https://doi.org/10.1109/BigData.2018.8621865

64.   Cui, H., Yuan, G., Liu, N., Xu, M., & Song, H. (2020). Convolutional neural network for recognizing highway traffic congestion. Journal of Intelligent Transportation Systems, 24(3), 279–289. https://doi.org/10.1080/15472450.2020.1742121

65.   Kang, J., Körner, M., Wang, Y., Taubenböck, H., & Zhu, X. X. (2018). Building instance classification using street view images. ISPRS Journal of Photogrammetry and Remote Sensing, 145, 44–59. https://doi.org/10.1016/j.isprsjprs.2018.02.006

66.   Liu, T., Abd-Elrahman, A., Morton, J., & Wilhelm, V. L. (2018). Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small

unmanned aircraft system. GIScience & Remote Sensing, 55(2), 243–264.

https://doi.org/10.1080/15481603.2018.1426091

67.   Bressem, K. K., Adams, L. C., Erxleben, C., Hamm, B., Niehues, S. M. & Vahldiek, J. L. (2020). Comparing different deep learning architectures for classification of chest radiographs. Scientific Reports, 10(1), 13590. https://doi.org/10.1038/s41598-020-70479-z

68.   Zhou, B. (2018). Interpretable representation learning for visual intelligence. 140

69.   Arcmap. (n.d.). Image classification using the ArcGIS Spatial Analyst extension—ArcMap | Documentation. Retrieved January 30, 2022, from https://desktop.arcgis.com/en/arcmap/latest/extensions/spatial-analyst/image-classification/image-classification-using-spatial-analyst.htm

70.   Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12(85), 2825–2830.

71.   Chen, T. & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

72.   Wong, P.-Y., Su, H.-J., Lee, H.-Y., Chen, Y.-C., Hsiao, Y.-P., Huang, J.-W., Teo, T.-A., Wu, C.-D. & Spengler, J. D. (2021). Using land-use machine learning models to estimate daily NO2 concentration variations in Taiwan. Journal of Cleaner Production, 317, 128411. https://doi.org/10.1016/j.jclepro.2021.128411

73.   Freund, Y. (1995). Boosting a Weak Learning Algorithm by Majority. Information and Computation, 121(2), 256–285. https://doi.org/10.1006/inco.1995.1136

74.   Lundberg, S. M. & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, 30. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

75.   Molnar, C. (n.d.). 9.5 Shapley Values | Interpretable Machine Learning. Retrieved January 20, 2022, from https://christophm.github.io/interpretable-ml-book/shapley.html#general-idea

76.   Rozemberczki, B., Watson, L., Bayer, P., Yang, H.-T., Kiss, O., Nilsson, S. & Sarkar, R. (2022). The Shapley Value in Machine Learning. ArXiv:2202.05594 [Cs]. http://arxiv.org/abs/2202.05594

77.   Santarius, T. (2022). Climate justice and digitalization: A plea to consider broader socio-economic implications of digitalization and climate change. GAIA - Ecological Perspectives for Science and Society, 31(3), 146–150. https://doi.org/10.14512/gaia.31.3.4