

Urban Equity Measure of Tree Canopy: A case study of New York City

Yuhao Jia [†], Muyan Duan ^{*,†}

Department of architecture and urban design, University of California, Los Angeles, California.

* Corresponding Author Email: myd2020@g.ucla.edu

[†]These authors contributed equally.

Abstract. The urban tree canopy is widely recognized as an urban resource that contributes to the physical and mental health of residents, yet tree inequality is widespread. To achieve the goal of promoting tree equity and filling the gap in relevant academic research, this study develops a 2-step 9CNN model to obtain highly accurate urban tree canopy (UTC) data from high-resolution satellite images. New York City tree canopy data is obtained by this new method and correlates with New York City socioeconomic data to examine UTC's horizontal and vertical equity. OLS regression correlations of per capita UTC area with the number of elderly, poor, low-educated, and disabled were used to measure vertical equity. To avoid the interference of spatial autocorrelation, GWR models were used to test for spatial differences in vertical equity. The results of the horizontal equity study show that there is far less UTC space per capita in New York's urban centers than in the fringe areas. Vertical equity in UTC is poor across all of vulnerable groups and is particularly pronounced in some regions. These findings can subsequently help to develop policies for UTC equity and environmental justice concerns.

Keywords: Tree Canopy; Vertical Equity; Convolutional Neural Network; Deep Learning; Statistical Analysis.

1. Introduction

1.1. Urban Equity and Tree Canopy

The urban tree canopy (UTC) is widely considered an environmental commodity or amenity that can provide benefits to citizens [1], such as improvement of air quality [2], mitigation of heat island effects [3], and reduction of noise pollution [4]. In this respect, UTC has similar properties to urban public services. It is reasonable that UTC should also have horizontal and vertical equity characters. Examining the spatial equity of UTC has increasingly become an important interdisciplinary focus for ecologists and social scientists [5]. Many studies have focused on the vertical equity of UTC, i.e., by correlating tree canopy cover data with socioeconomic factors to examine the equity of UTC for vulnerable groups [1,6,7]. There has been little research to define the horizontal equity of UTC.

1.2. Tree Canopy Cover Data

Most of the studies have used traditional methods to calculate the canopy cover data, which are based on remote sensing techniques and Satellite images or using ground survey data, most of which can reach over 90% accuracy [8]. However, with the advancement of AI technology, there are possibilities to obtain more accurate canopy cover data through machine learning, image recognition methods. Google Environmental Insights Explorer Studio has used this method to identify and map tree canopy cover in several major U.S. cities [9]. There are no studies using this type of data for UTC equity measurements.

1.3. Tree Canopy Classification Using Satellite Images

In the last decade, many advanced machine learning methods, such as artificial neural networks (ANN), SVM, RF, decision trees and other models, have gained special attention in remote sensing-based applications (e.g., land cover classification) [10,11]. It has been found that ANN, SVM and RF

generally provide better accuracy compared to other traditional classifier techniques [12]. Further, many studies especially focus on convolutional neural networks (CNNs) as the current most advanced technique for satellite classification tasks. Some research analyses have shown that although traditional machine-learning techniques provide good accuracy, CNNs outperform other techniques [12]. Few studies have used machine learning models to identify a single land type.

In this study, a new 2-step method of tree canopy classification based on the CNN model and Random Forest Model is developed to generate the tree canopy data of New York City. Based on the above aspects, this study fills these three gaps: (a) systematically defining the horizontal and vertical equity of UTC, (b) using AI-identified tree canopy cover data to measure the equity, and (c) developing a new method of tree canopy classification by satellite imagery.

2. Data and Methods

CNN models are better at processing images compared to other machine learning models. This is reflected in satellite map recognition where CNN models are good at recognizing the shape of site types. The image data set for training CNN models is often a series of pixel matrices with spectral information, which can be used to identify land types by image shape more accurately, but cannot be used to identify the exact area of a land type. Segmentation models like U-Net [13] are able to do this task but require high computational complexity.

Based on this, this study develops a two-step approach to identify regions of tree canopies using combined CNN and Random Forest model, which has high capacity and requires less computational complexity.

This study uses the high-resolution satellite images of Philadelphia 2018 with the tree canopy distribution data of Philadelphia 2018 as the dataset for training the model. The satellite images are from Google Class XVII, and the tree canopy distribution data are from the Philadelphia Open Data website.

2.1. Step 1: Shape Recognition CNN Model

According to the image resolution and average canopy area size, the satellite image of Philadelphia was divided into 30*30 pixel sliders, and the data of each slider was used as a training sample (Figure 1). The spectral values of each pixel in the slider were transformed into arrays in python. An image was labeled 1 if it contains canopy and labeled 0 if it doesn't. The total number of samples was 12000 (6000 positive and 6000 negative), and 80% of them were selected as training set and 20% as testing set. With AlexNet [14] as CNN architecture, 0.5 as learning rate, "adadelta" as optimizer, after 15 training epochs, the accuracy reached 99% on training set and 98% on testing set.



Figure 1. Step 1 Training & Test Data.

2.2. Step 2: Precise Identification RF Model

Any slider containing the canopy distribution is extracted and used as the dataset for the second part of the training. Each pixel in the slider is treated as a sample in this step with red, green, blue, and alpha 4 spectrums as input data (Figure 2). An unpruned random Forest model was used here.

80% of the pixels were selected as training set and 20% as testing set. Both accuracy (Acc) and intersection over union (IoU) were used to evaluate the final results, which were defined as below:

$$Acc = (True\ Positive + True\ Negative) / (All\ samples) \quad (1)$$

$$IoU = (True\ Positive) / (False\ Negative + False\ Positive + True\ Positive) \quad (2)$$

The Acc of the final result on test set reached 97% on training set and 95% on testing set. The IoU reached 98% on training set and 96% on testing set.

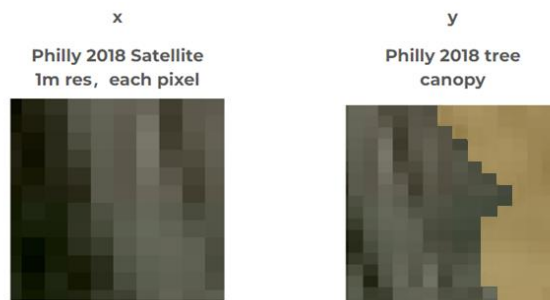


Figure 2. Step 2 Training & Test Data.

2.3. Collection of 2020 NYC Tree Canopy Data

The 2-step model was applied to the 17-level Google satellite images of New York City 2020 to generate the latest tree canopy cover data. The accuracy of the acquired data could not be tested as the NYC open data did not contain the 2020 tree canopy cover data. However, based on the researchers' visual inspection and the accuracy of the model, the data was determined to have a high accuracy.

2.4. Collection of 2020 Census Data

NYC census data was derived from five-year ACS estimates for the year 2020, which includes four areas (Manhattan, Bronx, Staten Island, Brooklyn and Queens). In addition to population, features poverty, age over 65, education less than high school, and disability were also captured to represent socially vulnerable groups.

2.5. Joining and Cleaning the Data

The NYC tree canopy data was joined to the census tracts in the way of intersection. The area of tree canopy in each tract was summed by grouping by census tracts.

The joined dataset had 2327 observations (census tracts) in total. After removing the rows with value 'NA' and with 0 value of population, the final dataset had 2206 observations.

2.6. Vertical and Horizontal Equity Analysis

2.6.1. Horizontal Equity

Based on the overall definition of horizontal equity, this study defines the horizontal equity of tree canopy as an urban facility: The equal distribution of the area of tree canopy cover between distinct individuals and groups that are considered equal in ability or need.

Based on this definition, the area of tree canopy cover per capita was used as an indicator to quantify horizontal equity.

2.6.2. Vertical Equity

Similar to other urban facilities, vertical equity in tree canopy implies that the socially vulnerable groups in the city should enjoy more benefits from the tree canopy.

Since the canopy area per capita and the selected social vulnerability features are not causal, this study used OLS linear regression to test the correlation between canopy area per capita as the dependent variable and each demographic feature as the independent variable.

If the tree canopy area per capita is significantly correlated with a feature in the regression model, it means that the socially vulnerable groups represented by this feature have great vertical equity of tree canopy. If the tree canopy area per capita is significantly uncorrelated or negatively correlated with a feature in the regression model, the socially vulnerable group represented by this feature has less vertical equity in the canopy.

Since four social vulnerability features have data on the number of people, the per capita canopy area of that socially vulnerable group in each census tract is used to characterize spatial differences of equity inside this group. The data analysis and visualization are performed entirely in R Studio. Geographically Weighted Regression was used to explore spatial differences of equity inside a group.

3. Results and Discussion

3.1. Tree Canopy Distribution

Figure 3 shows the NYC canopy distribution data obtained by CNN model. It can be seen that except for a few large parks with large contiguous tree canopies, the tree canopies in urban areas are mostly distributed in a dotted pattern.

3.2. Horizontal Equity

Figure 4 shows the data for the per capita canopy area of each census tract, which also represents the horizontal equity of the tree canopy. The Manhattan Island area has the least amount of canopy area per capita overall, while Staten Island and the entire eastern edge of NYC have a very high amount of canopy area per capita. The horizontal equity of UTC shows polarization in center city and other areas.

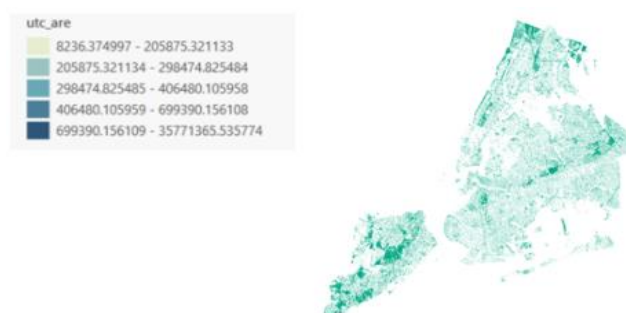


Figure 3. NYC Tree Canopy Distribution.

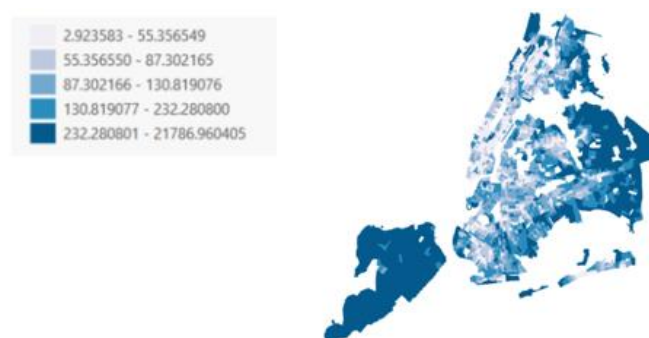


Figure 4. NYC Tree Canopy Horizontal Equity.

3.3. Vertical Equity

3.3.1. OLS Regression

The single variable OLS regression model was used to investigate the correlation between UTC per capita and each feature. The following figure shows the data results of the OLS regression model. According to the result, the corresponding p-value of the poverty and education level is far less than

0.05, which means their correlation coefficients are highly significant with the area of UTC per capita. The corresponding p-value of the elderly and disability is higher than 0.05, which means these two features are not statistically related to the area of UTC per capita (Table 1).

The correlation coefficient of the "Poverty" feature shows that the number of poor people is negatively correlated with the area of UTC per capita. This means that places with more people in poverty have less UTC per capita. This also indicates that the vertical equity of UTC is poor for people in poverty (Figure 5).

The correlation coefficient of the Education level feature shows that the number of people with low education is negatively correlated with the area of UTC per capita. It means that the more people with low education, the less the UTC per capita. This also indicates that the vertical equity of UTC is poor for people with low education (Figure 5).

The elderly and disabled populations are not related to UTC per capita, which means these two groups do not have access to more UTC as urban amenities. That is, the vertical equity of UTC is also poor for both groups (Figure 5).

Table 1. Four Variables OLS Regression Analysis

Independent Variable	Estimate	P value
The Elderly	-0.10557	0.141
The Poverty	-0.14751	1.1e-058***
The Disability	-0.7830	0.0561
Low-education groups	-0.2056	5.52e-08***

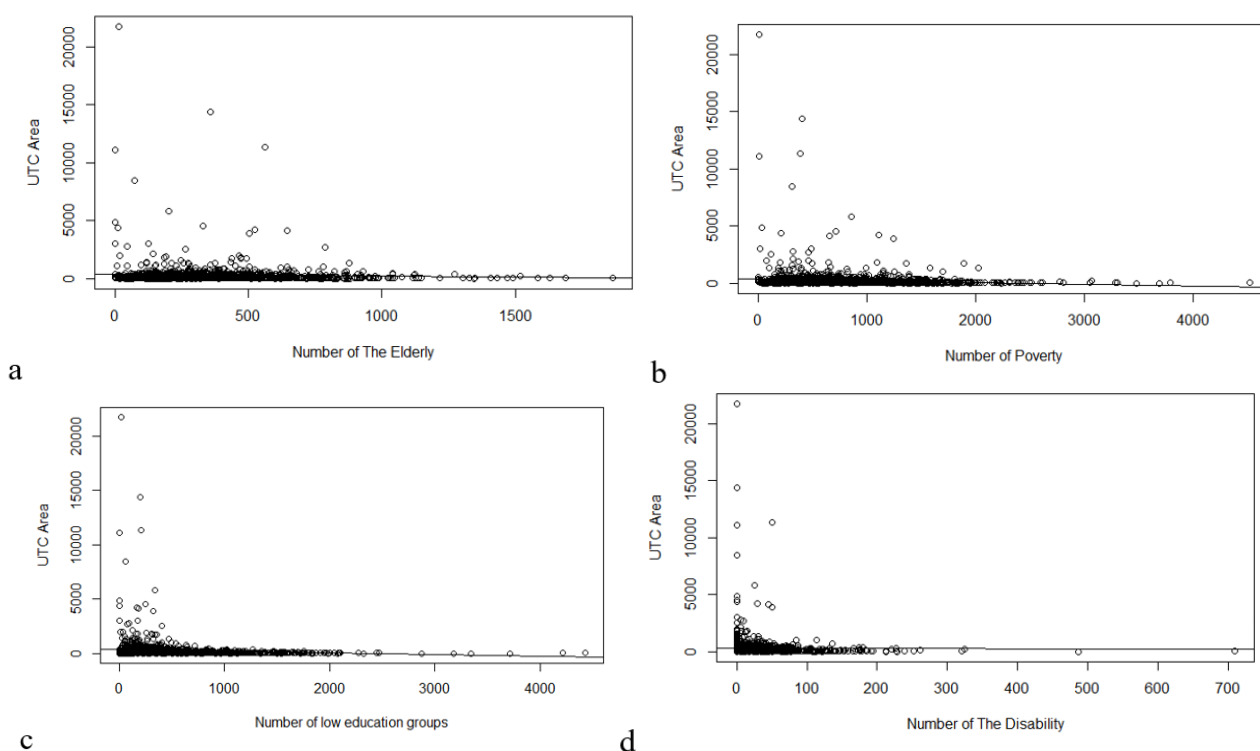


Figure 5. (a) UTC and the Elderly Population. (b) UTC and Population in Poverty. (c) UTC and Population Low Education Level. (d) UTC and Population with Disabilities.

3.3.2. Geographically Weighted Regression (GWR)

Since some vulnerable groups may have spatial aggregation, we test the spatial autocorrelation of the data by Moran's I test.

In the result of Moran's I test, the p-value is less than 0.05 which means there was a significant space autocorrelation in the residual of the previous OLS model. To avoid the interference of autocorrelation, the GWR model was selected for further analysis.

Table 2. Moran's I Test Result

Statistic	Standard Deviate	Moran I statistic	P Value
	2.1358	0.028044	0.01635

According to the GWR results, it can be found that the elderly feature in Staten Island is significantly negatively correlated with the area of UTC per capita. Other places show no significantly correlation with it (Figure 6). In general, the elderly group has poor equity in UTC. Elderly people in Staten Island have poorer vertical equity of UTC than others.

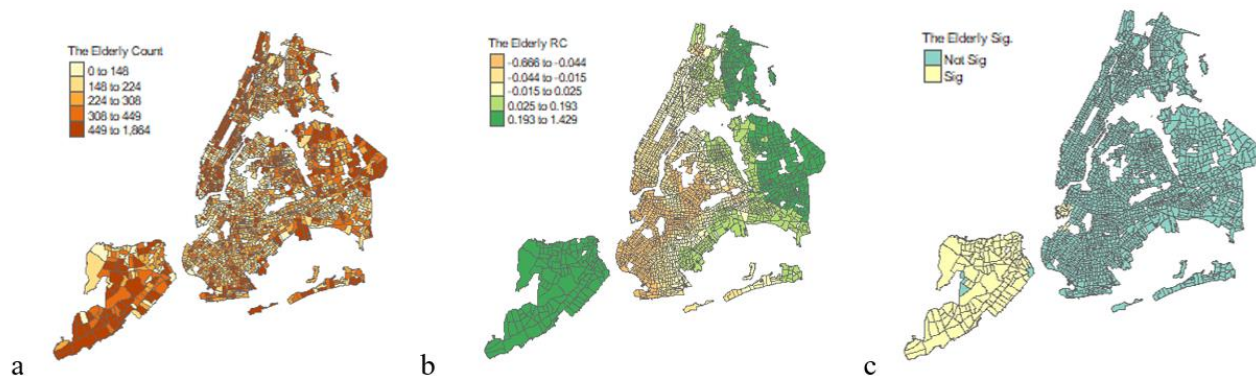


Figure 6. (a) Distribution of the number of elderly. (b) The elderly regression coefficient distribution. (c) The elderly significance test distribution.

For the poverty feature, the GWR result shows that poverty feature in Staten Island and eastern of Bronx are significantly negatively correlated with the area of UTC per capita, other places shows no significantly correlation with it (Figure 6). In general, the poverty group has poor equity of UTC. Poor people in Manhattan, the middle of Brooklyn and the south of Queens have better vertical equity of UTC than others.

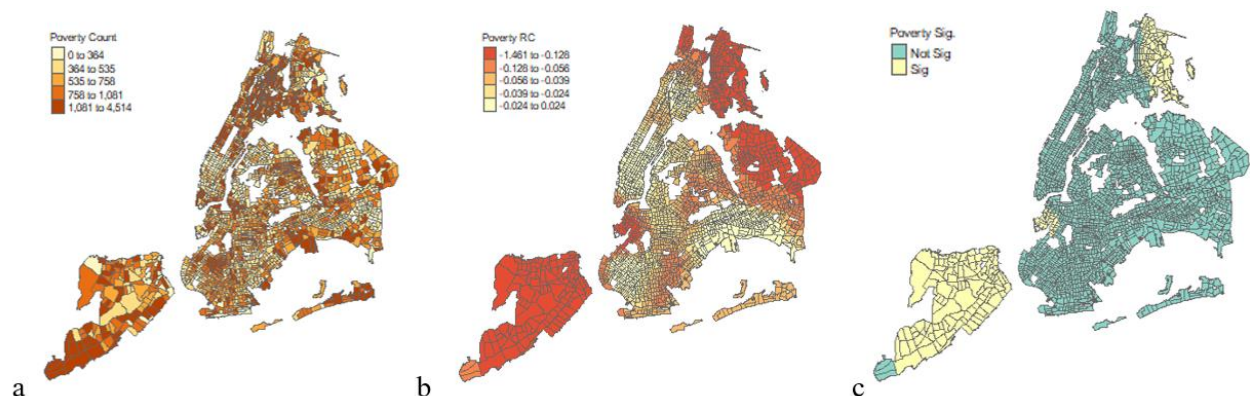


Figure 7. (a) Distribution of the number of poverty. (b) The poverty regression coefficient distribution. (c) The poverty significance test distribution.

For the disability feature, the GWR result shows that disability feature has no significantly correlation with the area of UTC per capita over all of the NYC (Figure 7). In general, the disability group has poor equity of UTC.

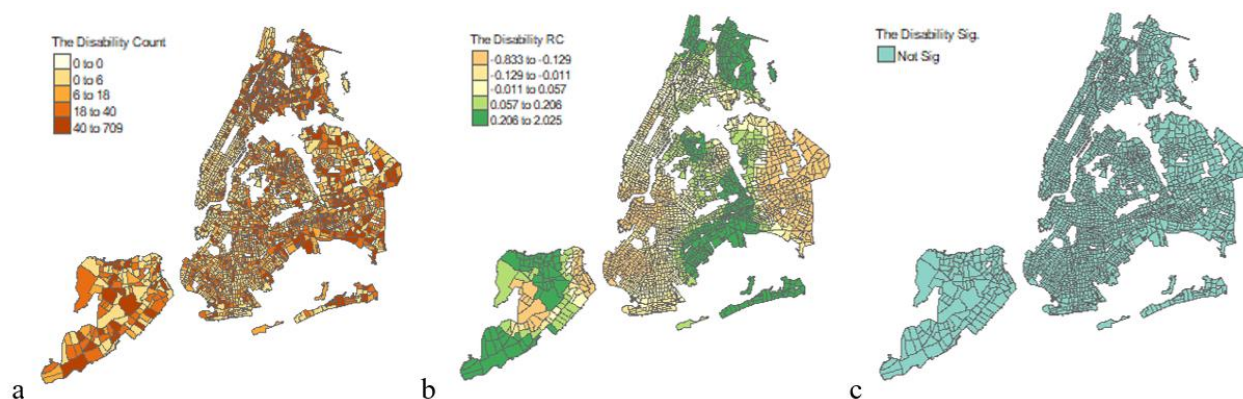


Figure 8. (a) Distribution of the number of disability. (b) The disability regression coefficient distribution. (c) The disability significance test distribution.

For the education feature, the GWR result shows that low education feature in Staten Island and eastern of Bronx are significantly negatively correlated with the area of UTC per capita, other places shows no significantly correlation with it (Figure 8). In general, the low education group has poor equity of UTC. These people in Manhattan, the middle of Brooklyn and the south of Queens have better vertical equity of UTC than others.

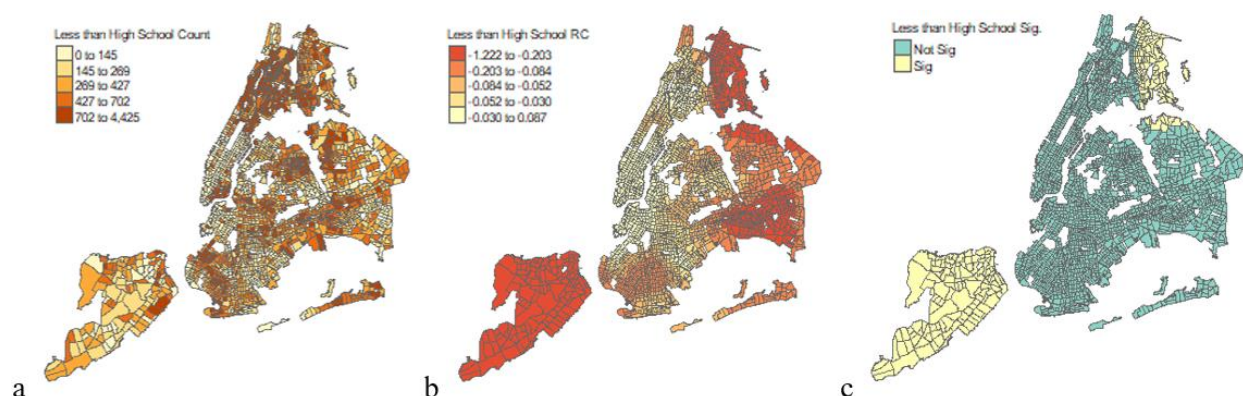


Figure 9. (a) Distribution of the number of low education. (b) The low education regression coefficient distribution. (c) The low education significance test distribution.

In summary, after studying their relationship with UTC area separately, it can be found that the needs of socially vulnerable groups are not taken into account in balancing the canopy area of each region. In other words, the vertical equity of UTC as urban facilities performs poor in all 4 socially vulnerable groups, which are the older people, people in poverty, people with disability and people with low educational level.

3.3.3. Policy Suggestions

In the past, decisions on the spatial layout of urban trees did not pay special attention to socially vulnerable groups with the strongest use needs, and social justice needs to be improved. Establishment of tree canopy system evaluation, clarifying and refining evaluation indicators, and promoting its vertical equity are essential.

For NYC, socially vulnerable groups do not enjoy more UTC as urban amenities, so the distribution of socially vulnerable groups needs to be considered as an important feature in future management planning of the urban canopy to achieve better vertical equity of UTC. According to this study, the planning department can focus on two aspects: the distribution of socially vulnerable groups, and the spatial differences in the enjoyment of UTC within these groups. More UTC can be provided to socially vulnerable groups by increasing the index scheme in some areas with less vertical equity.

4. Conclusion

In New York City, the horizontal equity of UTC exhibits polarization: the urban center has significantly less UTC space per capita than the urban fringe. UTC is more related to its residents' physical and mental health, its vertical equity should be of greater concern. Good vertical equity is reflected in the fact that socially vulnerable groups would enjoy more urban amenities. Vertical equity in UTC is presented poorly in all four categories of socially vulnerable groups explored in this study. Within each category of groups, vertical equity in UTC remains spatially distinct. For urban environmental planning departments, some policies could be developed to improve these issues. Based on the findings of this study, planning departments can focus on some areas with poor vertical equity and provide more UTC for socially vulnerable groups in these areas.

Overall, a new 2-step CNN model was developed in this study to identify satellite images to obtain tree canopy cover data. OLS regression was used to analyze the correlation between tree UTC and individual vulnerable groups, to explore whether vulnerable groups enjoyed more tree canopy area. The analysis showed that UTC, as an urban amenity, does not favor any vulnerable group. This needs to be taken into account in future urban planning of tree canopy cover. Features representing socially vulnerable groups were tested and found to be spatially autocorrelated, and the GWR model was applied to avoid this effect. The results show that there is still a spatial imbalance in UTC vertical equity within a single vulnerable group. This result can help governments to target the development of policies to increase UTC.

References

- [1] Schwarz K, Fragkias M, Boone CG, et al. Trees grow on money: urban tree canopy cover and environmental justice. *PloS one*. 2015, 10(4): e0122051.
- [2] Russo A, Escobedo FJ, Zerbe S. Quantifying the local-scale ecosystem services provided by urban treed streetscapes in Bolzano, Italy. Aims Press; 2016.
- [3] Ziter CD, Pedersen EJ, Kucharik CJ, et al. Scale-dependent interactions between tree canopy cover and impervious surfaces reduce daytime urban heat during summer. *Proceedings of the National Academy of Sciences*. 2019; 116(15): 7575-80.
- [4] De Carvalho RM, Szlafsztein CF. Urban vegetation loss and ecosystem services: The influence on climate regulation and noise and air pollution. *Environmental Pollution*. 2019; 245: 844-52.
- [5] Leets L, Sprenger A, Hartman RO, et al. Promoting tree equity in Washington, DC. *Trees, Forests and People*. 2022; 7: 100209.
- [6] Aguilera SM. Prioritizing Tree Planting in Shade-Deprived Urban Areas as a Response to Climate Change. *Hastings Env't LJ*. 2021; 27: 101.
- [7] Flocks J, Escobedo F, Wade J, Varela S, Wald C. Environmental justice implications of urban tree cover in Miami-Dade County, Florida. *Environmental Justice*. 2011; 4(2): 125-34.
- [8] Riley CB, Gardiner MM. Examining the distributional equity of urban tree canopy cover and ecosystem services across United States cities. *PLoS One*. 2020; 15(2): e0228499.
- [9] Environmental Insights Explorer. Labs: Tree canopy. 2020
- [10] Mountrakis G, Im J, Ogole C. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2011; 66(3): 247-59.
- [11] Deng JS, Wang K, Deng YH, et al. PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data. *International Journal of Remote Sensing*. 2008; 29(16): 4823-38.
- [12] Carranza-García M, García-Gutiérrez J, Riquelme JC. A framework for evaluating land use and land cover classification using convolutional neural networks. *Remote Sensing*. 2019; 11(3): 274.
- [13] Von Eicken T, Basu A, Buch V, Vogels W. U-Net: A user-level network interface for parallel and distributed computing. *ACM SIGOPS Operating Systems Review*. 1995 Dec 3; 29(5): 40-53.
- [14] Krizhevsky A, Sutskever I, Hinton GE. 2012 AlexNet. *Adv. Neural Inf. Process. Syst*. 2012: 1-9.