Examining the Interplay Between the Natural and Socio-physical Environment Using Machine Learning Algorithms

Gia Elise Barboza-Salerno, JD PhD[1], Sharefa Duhaney, MPH[2]
[1]College of Social Work, The Ohio State University, [2]College of Public Health, The Ohio State University

Abstract ID 66677

Implications for Policy and Practice

- To effectively prevent child abuse and neglect (CAN), policies must focus on <u>environmental</u> interventions as well as economic, and social interventions
- AI and machine learning algorithms can aid in the identification of risk factors (e.g., air quality monitoring, green space equity, walkability and accessibility, safety) to reduce CAN risk. **<u>Results:</u>**

Summary of Results

Decreases CAN Risk	Increases CAN Risk
Greater Tree Equity	Walkability
Distance from Highway	Area Pollutants
Financial Strength	Poor Mental Health
Access to Green Space	High Temperature

Figure 1. Summary of environmental characteristics contributing CAN risk

Research Question(s)

- What features of the built environment such as proximity, connectivity, and accessibility to green space (e.g., parks, trees), green space equity, vegetation density, proximity to highways, and neighborhood walkability, are associated with CAN risk controlling for social and economic vulnerability?

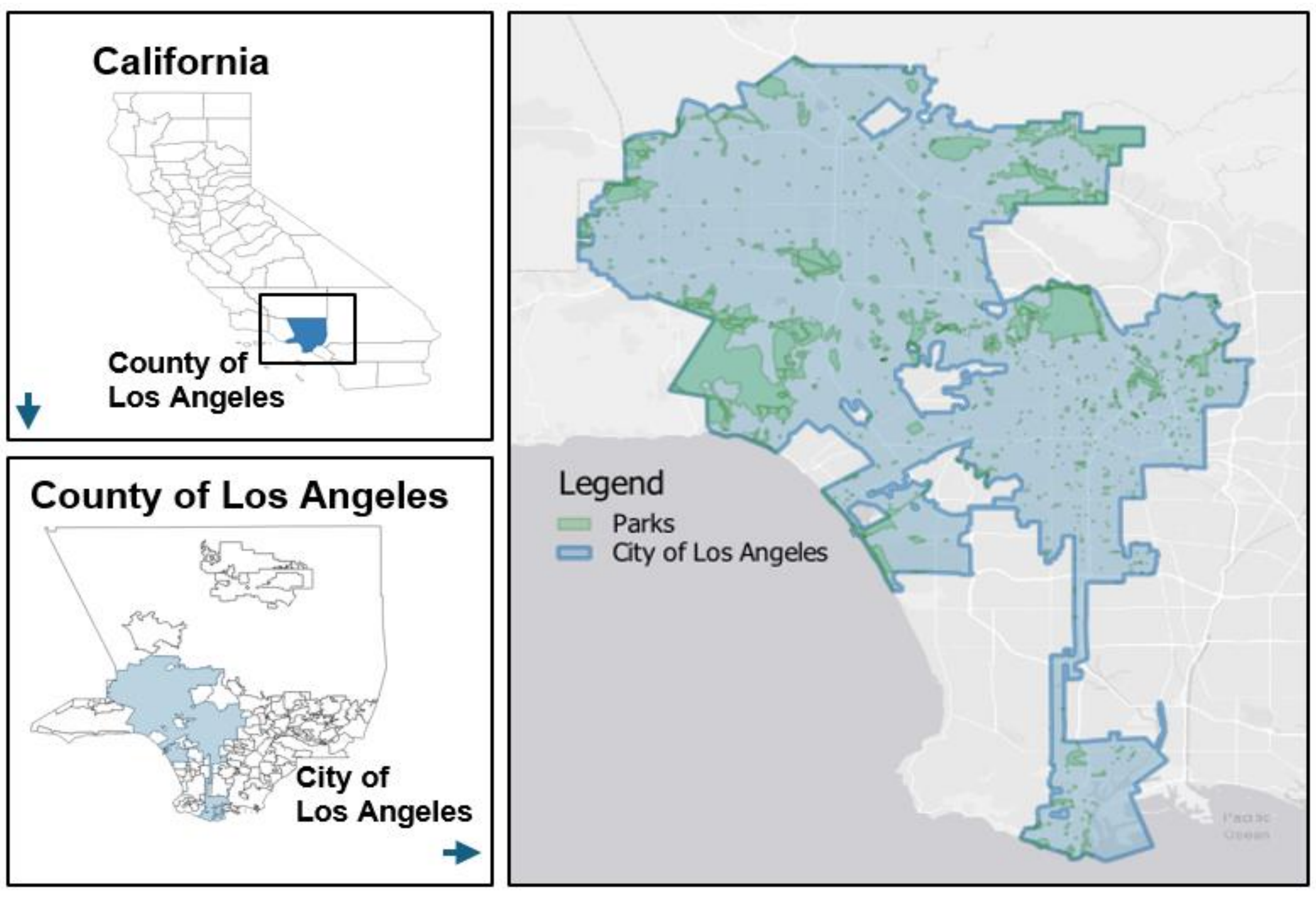

Figure 2. Study area (City of Los Angeles) where children were reported to police for physical abuse or neglect in a public space.

Introduction

- No previous research has comprehensively examined how different aspects of the sociophysical and built environments increase **CAN** risk
- Five possible categories of environmental determinants of child well-being include: (i) urban design, (ii) contaminants, (iii) the parenting environment, (iv) socio-economic conditions, and (v) climate change[1]
- We used machine learning (i.e., Google Street View (GSV), and Google Vision (GV)) and geospatial analyses to extract information from street images in areas that had an incident of CAN

Methodology

1. Downloading, Filtering, Clipping, Cleaning CAN Incident Data and Covariates

2. Analyzing CAN Risk

a. Spatial characteristics and mapping

b. Compare CAN risk in different environmental contexts

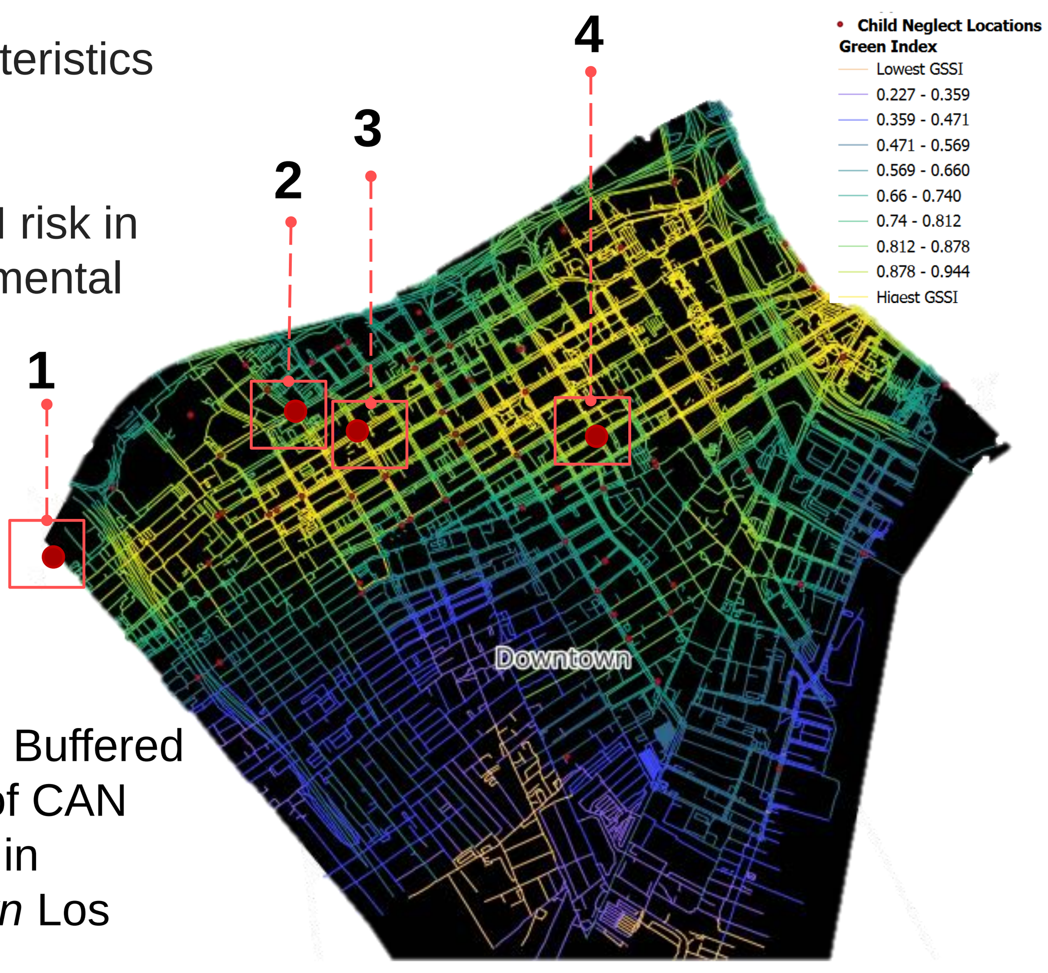

Figure 3. Buffered location of CAN incidents in *Downtown* Los Angeles

3. Feature Extraction from Built Environment

a. Street view Input

b. Image Segmentation

c. Green Pixel Segmentation

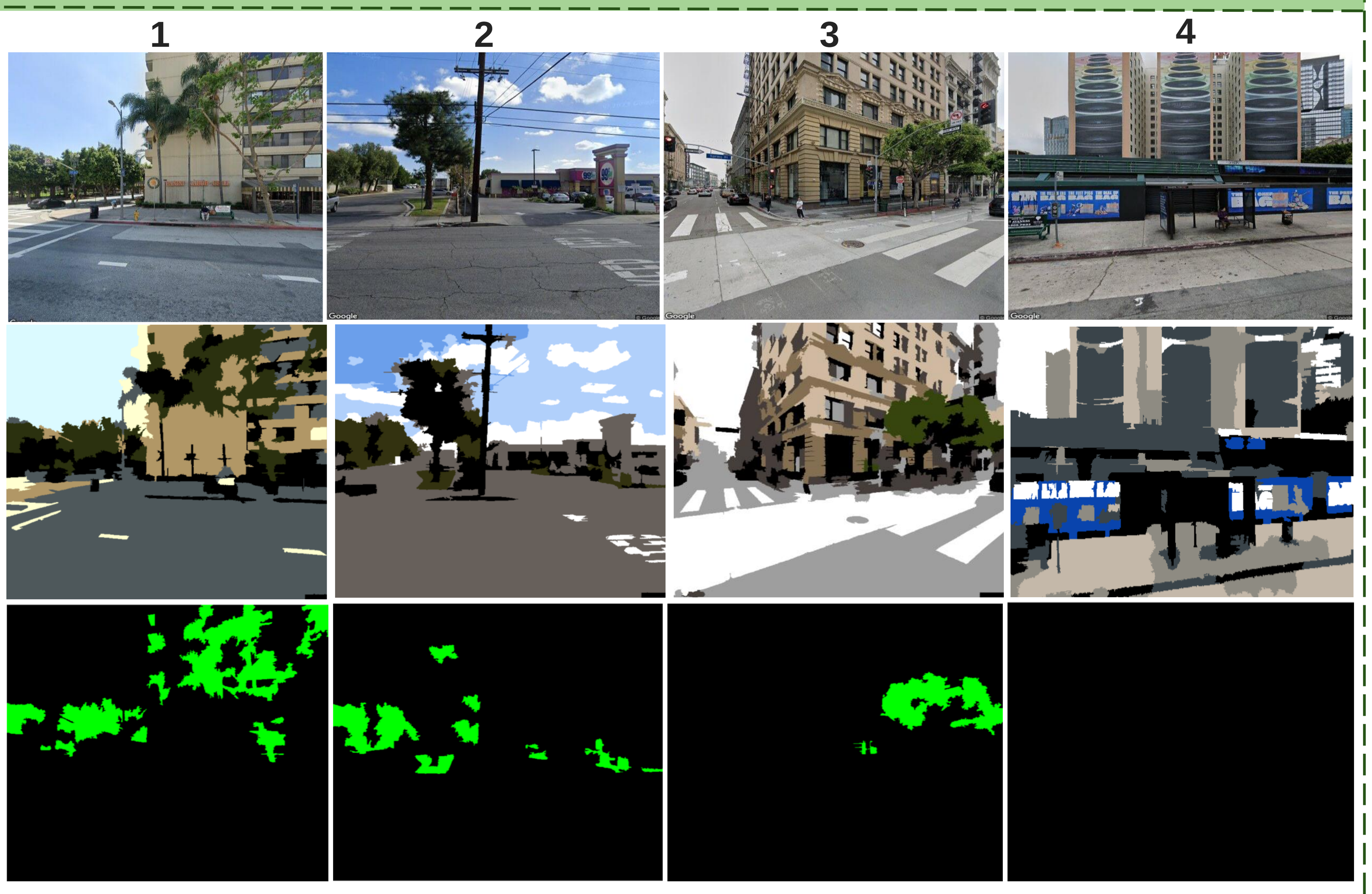

Figure 4. The output of the GSV analysis of a neighborhood location involving an incident of **<u>child neglect</u>** using **<u>image segmentation</u>**.

4. Understanding CAN Risk

<u>A.</u> Detect features in each image and calculate scores representing confidence that the feature is in the image

1 (*Left square in Figure 3*)

```
Building, Confidence Score: 0.95
Sky, Confidence Score: 0.90
Tree, Confidence Score: 0.89
Window, Confidence Score: 0.86
Road surface, Confidence Score: 0.85
Asphalt, Confidence Score: 0.84
Urban design, Confidence Score: 0.83
Condominium, Confidence Score: 0.80
Arecales, Confidence Score: 0.78
Sidewalk, Confidence Score: 0.78
```

4 (*Right square in Figure 3*)

```
Building, Confidence Score: 0.97
Skyscraper, Confidence Score: 0.92
Tire, Confidence Score: 0.90
Sky, Confidence Score: 0.88
Tower block, Confidence Score: 0.85
Urban design, Confidence Score: 0.84
Rolling, Confidence Score: 0.82
Asphalt, Confidence Score: 0.82
Road surface, Confidence Score: 0.78
Facade, Confidence Score: 0.77
```

Figure 5. Example features (e.g., '*Sky*', '*Tire*') detected for two images associated with CAN in *Figure 3* and the confidence score (e.g., .90).

<u>B.</u> Merge results with publicly available environmental data including tree equity, park equity, walkability, and the social vulnerability index (SVI)

<u>C.</u> Spatial Poisson Regression Results (Table 1)

Table 1. Poisson spatial regression results showing significant associations between CAN and the built environment

Characteristic	IRR[1]	95% CI[1]	p-value
Tree Equity Score	0.976	0.976, 0.998	0.003
Land Surface Temperature	1.181	1.056, 1.319	0.003
Physical Activity Rank	1.147	1.025, 1.283	0.016
Financial Strength	0.993	0.986, 1.001	0.096
PM2.5 concentration	1.318	1.174, 1.480	0.000
Proximity to park > 0.25 mile	0.834	0.712, 0.977	0.025

Notes. IRR = Incidence Risk Ratio. IRR > 1 means CAN risk increases; < 1 means CAN risk decreases. PM2.5 refers to 'fine particular matter.'
The model was run using the *fitme* package in R.
Only significant model results are shown.

<u>D.</u> Interpret CAN risk conditioned by built, physical, social environment

Future Directions

We contribute to the innovative use of machine learning models in social work evaluation and research by applying novel analyses to semantically categorize, analyze, and map quantitative and qualitative information for the prevention of CAN. By leveraging geospatial analysis in machine learning models, we improve the predictive capacity of such models to ensure sensitive and effective preventive interventions.

Gia (Left), Sharefa (Right)

The Ohio State University
College of Social Work

Acknowledgements

We would like to thank Olivia McLucas; Simran Grissom; Aisha Rabbani; Sharvari Dhile and the Big Data for Social Justice Research (BDSJR) Team! Come join our Team!