

Article

Assessing Climate Disaster Vulnerability in Peru and Colombia Using Street View Imagery: A Pilot Study

Chaofeng Wang ^{1,2,*} , Sarah E. Antos ³, Jessica G. Gosling-Goldsmith ³, Luis M. Triveno ³, Chunwu Zhu ⁴, Jason von Meding ^{1,5} and Xinyue Ye ⁴ 

- ¹ M.E. Rinker, Sr. School of Construction Management, College of Design, Construction and Planning, University of Florida, 573 Newell Dr., Gainesville, FL 32611, USA
- ² Department of Civil and Coastal Engineering, Herbert Wertheim College of Engineering, University of Florida, 1949 Stadium Rd., Gainesville, FL 32611, USA
- ³ World Bank Group, 818 H St NW, Washington, DC 20433, USA
- ⁴ Department of Landscape Architecture & Urban Planning, Center for Geospatial Sciences, Applications and Technology, Texas A&M University, 789 Ross Street, College Station, TX 77843, USA
- ⁵ Florida Institute for Built Environment Resilience, University of Florida, 720 SW 2nd Ave, Gainesville, FL 32601, USA
- * Correspondence: chaofeng.wang@ufl.edu

Abstract: Community and household vulnerability to natural hazards, e.g., earthquakes, hurricanes, and floods, is a concern that transcends geographic and economic boundaries. Despite the abundance of research in this field, most existing methods remain inefficient and face the challenge of data scarcity. By formulating and investigating the correlation between the household vulnerability and street view images of buildings, this research seeks to bridge the knowledge gap to enable an efficient assessment. Especially in developing countries, the widespread prevalence of outdated or inadequately enforced building codes poses a significant challenge. Consequently, a considerable portion of the housing stock in these regions fails to meet acceptable standards, rendering it highly vulnerable to natural hazards and climate-related events. Evaluating housing quality is crucial for informing public policies and private investments. However, current assessment methods are often time-consuming and costly. To address this issue, we propose the development of a rapid and reliable evaluation framework that is also cost-efficient. The framework employs a low-cost street view imagery procedure combined with deep learning to automatically extract building information to assist in identifying housing characteristics. We then test its potential for scalability and higher-level reliability. More importantly, we aim to quantify household vulnerability based on street view imagery. Household vulnerability is typically assessed through traditional means like surveys or census data; however, these sources can be costly and may not reflect the most current information. We have developed an index that effectively captures the most detailed data available at both the housing unit and household level. This index serves as a comprehensive representation, enabling us to evaluate the feasibility of utilizing our model's predictions to estimate vulnerability conditions in specific areas while optimizing costs. Through latent class clustering and ANOVA analysis, we have discovered a strong correlation between the predictions derived from the images and the household vulnerability index. This correlation will potentially enable large-scale, cost-effective evaluation of household vulnerability using only street view images.



Citation: Wang, C.; Antos, S.E.; Gosling-Goldsmith, J.G.; Triveno, L.M.; Zhu, C.; von Meding, J.; Ye, X. Assessing Climate Disaster Vulnerability in Peru and Colombia Using Street View Imagery: A Pilot Study. *Buildings* **2024**, *14*, 14. <https://doi.org/10.3390/buildings14010014>

Academic Editor: Yung Yau

Received: 13 October 2023

Revised: 26 November 2023

Accepted: 28 November 2023

Published: 20 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: street view image; building vulnerability; household vulnerability; deep learning

1. Introduction

Decades of climate change projections are now being observed worldwide as significant impacts are becoming increasingly evident [1]. Disastrous consequences of natural hazards may significantly impact a region on several levels [2]. Community vulnerability is a critical variable that helps shape strategies in disaster preparedness, urban planning,

and socio-economic studies. Understanding the level of community vulnerability is vital for mitigating risks and optimizing resource allocation during natural disaster events [3,4].

However, the process of assessing this vulnerability traditionally hinges on the collection of extensive demographic and socio-economic data. Such information is typically gathered through survey work, in-person interviews, and the collection of other statistical data. These conventional methods, while thorough, can be resource-intensive, necessitating significant time, financial investment, and personnel. Furthermore, the collection of these data often requires direct interaction with the community, which can sometimes be seen as invasive or burdensome by the very communities under study.

To address these challenges, there is a growing interest in developing innovative methodologies that can provide valuable insights about community and, specifically, household vulnerability, while being more accessible, time-efficient, and non-invasive. The ubiquity and increasing quality of visual data, particularly street view imagery, offers an interesting opportunity in this regard. This type of imagery, made widely available through platforms like Google Street View, contains a wealth of visual details about houses and neighborhoods that could potentially be leveraged to infer aspects of household vulnerability.

Inspired by the conventional visual screening methodology, such as FEMA 154 [5], our research applies cutting-edge technology to advance the field of household vulnerability assessment. We propose a novel method, using deep learning (DL) techniques applied to street view imagery, to identify potentially vulnerable households based on visible details of houses. This approach aims to reduce the need for traditional, more invasive data collection methods, thereby making the process more efficient, less costly, and respectful of household privacy.

The main objective of this study is to establish a correlation between street view imagery and socio-economic conditions (e.g., household vulnerability). Demonstrating this correlation could enable predictions of socio-economic status based on these images alone. Firstly, we employ DL models to extract relevant features from street view images, including construction quality, materials used, overall condition, and usage type. Following that, we define household vulnerability using a new metric, the 'K3 index', derived from census data. Specifically, K represents the K-means method used to classify Gower's similarity of census data into three clusters. Finally, we examine the correlation between the features identified in the images and the socio-economic data encapsulated by the K3 index. This correlation demonstrates the potential for automatic quantification of household vulnerability at scale, with improved reliability. The innovation here lies not just in the automated visual inspection of physical structures, but also in the correlation we establish between these building details and household vulnerability. By linking these two data sources, our research offers an innovative way to approximate vulnerability conditions more efficiently, which is particularly beneficial for constrained budgets and targeted areas. In this way, our work could revolutionize the approach to vulnerability assessments, setting a new standard for both scalability and respect for household privacy.

2. Methodology: Vision-Based Vulnerability Evaluation

This research is inspired by the conventional visual screening methodology, which is used for assessing building performance by employing a scoring scheme mainly based on the visual clues manifested at the exterior of buildings, without the need to access the inner space of buildings. An example of this type of method is FEMA 154 [5–7]. The rationale under this type of visual-based assessment method is that the structure's performance largely depends on its construction type, material, maintenance, etc. This paper focuses on household vulnerability, which is different from building performance. However, they share a similar rationale.

Although the visual screening method has been broadly adopted, such screening can be expensive and prone to errors because it is extremely labor-intensive when gathering a vast amount of data (i.e., images) of the buildings being investigated. Additionally, the subjectiveness in human decisions can potentially lead to diverging interpretations and erroneous results. To address this issue, this research presents an alternative procedure, which first collects street images from the region of interest using a ground vehicle and

then employs a DL approach to recognize structures from these images. The recognized properties can then be used for determining the household vulnerability.

Street view imagery is an inexpensive data source. The collection of street view images requires minimum equipment—images can be captured using cameras mounted on moving objects, such as vehicles and pedestrians. Yet, street view imagery can provide rich visual information of the road and buildings. Recent works have proved that street view images are applicable to a variety of different studies. For example, the degree of urban public security can be calculated from visual cues perceived from numerous street view images [8,9]. Ref. [8] found the machine-extracted visual presence of the urban environment can be used to measure the life quality of inhabitants in a neighborhood. Similarly, Ref. [10] found the correlation between the street visual cues of cars and the demographic makeup by analyzing millions of street view images collected from multiple cities in the United States. In another study, Refs. [11–13] discovered strong links between the housing price and the visual cues of building facades from the street. Refs. [14–18] employed street view imagery to identify specific building features that may lead to severe damage under major natural hazard events, such as earthquakes and hurricanes. A wide range of applications can also be found in the literature: neighborhood environment auditing [19], urban greenery assessment [20,21], and many other aspects in urban planning [22–24]. Street view images can also be combined with data to provide more information about the built environment; for example, Ref. [25] utilized satellite and street view images to infer the function and occupancy type of buildings.

All the aforementioned examples use DL-based methods to extract information from images. This paper develops a DL-based system in which the risk-related building attributes are automatically perceived from street view images and then, more importantly, are correlated with household vulnerability (e.g., two-stages, as shown in Figure 1, firstly infers from the image then uses the inferred information for a vulnerability assessment). The proposed system possesses many advantages when compared to traditional screening approaches in terms of the consistency in evaluation, cost efficiency, and scalability. Consistency means one model for the prediction of the whole investigated region, compared to many evaluators hired for different streets in traditional in-person assessments; Cost efficiency and scalability mean that during the deployment period, it takes only several thousand dollars to collect data in several hours and run the model to make the prediction, comparing to expensive in-person and invasive data collection, which would take months or years.

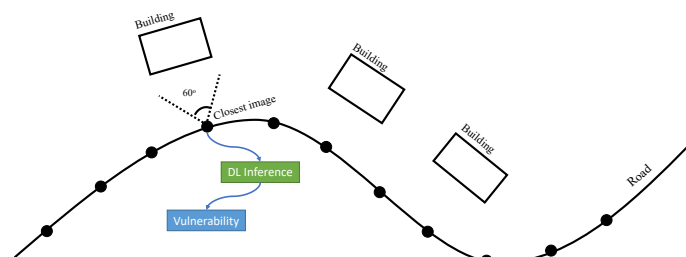


Figure 1. Rapid vulnerability evaluation workflow.

3. Explicit Visual Inference from Images

The objective of this research is to develop and validate an integrated workflow for automatic large-scale screening of building and household vulnerabilities at the city level. The first part of this workflow is inferring information from the images, which is based on DL instance segmentation. The procedure can be divided into 1. Collection of Data Street level images are first collected. Each image has its camera parameters and GPS coordinates recorded. 2. Image Annotation For training the model, we created an annotated dataset, where the building objects in the collected street view images are annotated with bounding boxes. Each annotated building object is also labelled with the attributes, such as the use type, facade material, construction type, and the condition. In Section 3.2, we describe the details of the image collection and annotation procedures. 3. Model Training Once the

images are annotated, the next step is to train multiple models for instance segmentation. The details of the model can be found in Section 3.1. 4. Inference Once models are trained, they are then used for detecting buildings from other images collected from the streets and predict the attributes of each detected building. 5. Geocoding In this step, the detected building attributes will be linked with building footprints. This step is not the focus of this paper and is detailed in [16]. The core of this procedure is to successfully train instance segmentation models that are capable of identifying building instances and attributes from street view images. It is a challenging task because of the complex contexts in street view images. The model needs to be trained on a dataset that captures the variations in building facades so that the model can learn to identify building objects from images and further classify the objects according to their different attributes.

3.1. Inference Model

An image classification model takes a given picture as input and returns a classification determining whether the object of a specific class is displayed in the picture or not. In contrast, instance segmentation is about categorizing and grouping pixels in an image into predefined classes. In this study, we implement the instance segmentation technique in the workflow for information extraction from street view images because it classifies the images at the pixel level and identifies the locations of building objects of specific categories.

There are various DL-based instance segmentation models that have been developed in recent years. It should be noted that the proposed vulnerability evaluation workflow can use information inferred by any similar model. The choice of models is out of the scope of this paper. We only demonstrate the workflow using one model. The exact model we used for this task is based on the Mask R-CNN approach originally developed by [26]. This approach is an extension of the Faster R-CNN algorithm [27]. Besides the original classification branch with bounding box regression, a new branch that can predict the segmentation masks is added to the head. Details of the architecture can be found in its original paper.

The backbone of the neural network is a ResNet, which is pretrained on the COCO dataset. A street view image is first fed into the backbone to generate a feature map, based on which a Region Proposal Network (RPN) can propose regions of interest (RoIs). Each RoI will enter head two branches for mask generation and bounding box prediction/classification. The branch responsible for mask generation is a fully convolutional network that can predict the segmentation mask for each RoI at the pixel level. The other branch responsible for the bounding box consists of a set of fully-connected layers, which performs the bounding box regression and softmax classification. The loss will be back-propagated to the network to update the weights. For each RoI, the loss can be calculated like this: $L = L_{cls} + L_{box} + L_{mask}$, in which L represents the total loss; L_{cls} represents the classification loss and L_{box} represents the bounding box loss; and L_{mask} represents the segmentation loss. The detailed definitions of L_{cls} , L_{box} , and L_{mask} are referred to in [26,27].

3.2. Data Preparation

During 2018–2021, we collected all street view images in eight cities in both Latin America and Asia (as shown in Figure 2) using a 360° camera mounted on a ground vehicle. As illustrated in Figure 1, the vehicle drives through the streets with the cameras taking images at a constant speed, resulting in spaces of 5–10 m between two images. The GPS coordinates and camera parameters are recorded for each image at the moment the image was taken. The images selected for this research are those perpendicular to the travel direction.

From all the street view images, a random subset (about 100,000 images) is selected for annotation. The cities are selected because they are populated cities in developing countries, and they are subject to natural hazards including earthquakes, hurricanes, and floods. The same buildings appear at different angles in different images, which are annotated by different annotators who are trained in advance. Building objects are annotated on the selected images using bounding boxes. The bounding boxes are labelled with four tags: construction type, material, use, and condition. For each tag, the values are listed in Table 1, where we also show

the number of annotated objects for each label class. We use the Computer Vision Annotation Tool (CVAT) for annotation. Regarding the accuracy of the annotation, we implemented quality control to ensure the reliability of the data. Firstly, the annotation was performed by trained annotators who were familiar with building construction and material types. They were given a comprehensive annotation guide to ensure the labels applied were consistent across different annotators and images. We also performed regular inter-rater reliability checks to assess the level of agreement among different annotators. A subset of 80% of these annotations are used for training the model, the rest for validation.



Juchitan	(Mexico)	Salina Cruz	(Mexico)
Cartagena	(Colombia)	Neiva	(Colombia)
St Maarten	(Caribbean)	Lima	(Peru)
Asuncion	(Paraguay)	Padang	(Indonesia)

Figure 2. Data source distribution.

Table 1. Annotation tags.

Tag Name	Value	Counts
Construction type	confined	8839
	unconfined	1642
Material	plaster	76,584
	mix_other_unclear	19,446
	brick_or_concrete_block	16,613
	wood_crude_plank	2303
	wood_polished	1097
	corrugated_metal	566
	adobe	207
	stone_with_mud_ashlar_with_lime_or_cement	164
	container_trailer	72
	plant_material	46
Use	residential	87,872
	non_residential	16,765
	mixed	12,461
Condition	fair	71,927
	poor	28,768
	good	16,403

3.3. Inference Performance

The aforementioned annotation dataset is used for training the Mask R-CNN model. This section shows the performance of the trained model. A subset of annotated images

that was not seen by the model during the training is used for calculating the performance. We estimate the performance based on Intersection over Union (IoU) greater than 75%. IoU is calculated as the intersection of the prediction region and ground truth region divided by the union of prediction and ground truth regions. For all $\text{IoU} > 75\%$ predictions, we show the accuracy and F1-score (a measure of a test's accuracy, which is calculated from the precision and recall of the test) in Table 2.

Table 2. Accuracy and F1-score.

Attribute	Accuracy (%)	F1-Score (%)
Construction type	98.97	98.97
Material	97.08	93.30
Use	94.62	88.31
Condition	78.85	69.68

A few prediction examples are presented in Figure 3, which shows the segmentation results compared with ground truth annotations. Figure 3a,c,e,g are the original images annotated by humans. Correspondingly, Figure 3b,d,f,h are the same images segmented by the model. In these examples, the model can accurately predict the construction type, material, use, and condition. It should be noted that these images are collected from different locations in the world: Latin America and Southeast Asia. These two regions share similar geographical and social–economic characteristics. To demonstrate this, we show in Figure 4 some examples of street view images collected from these two regions.

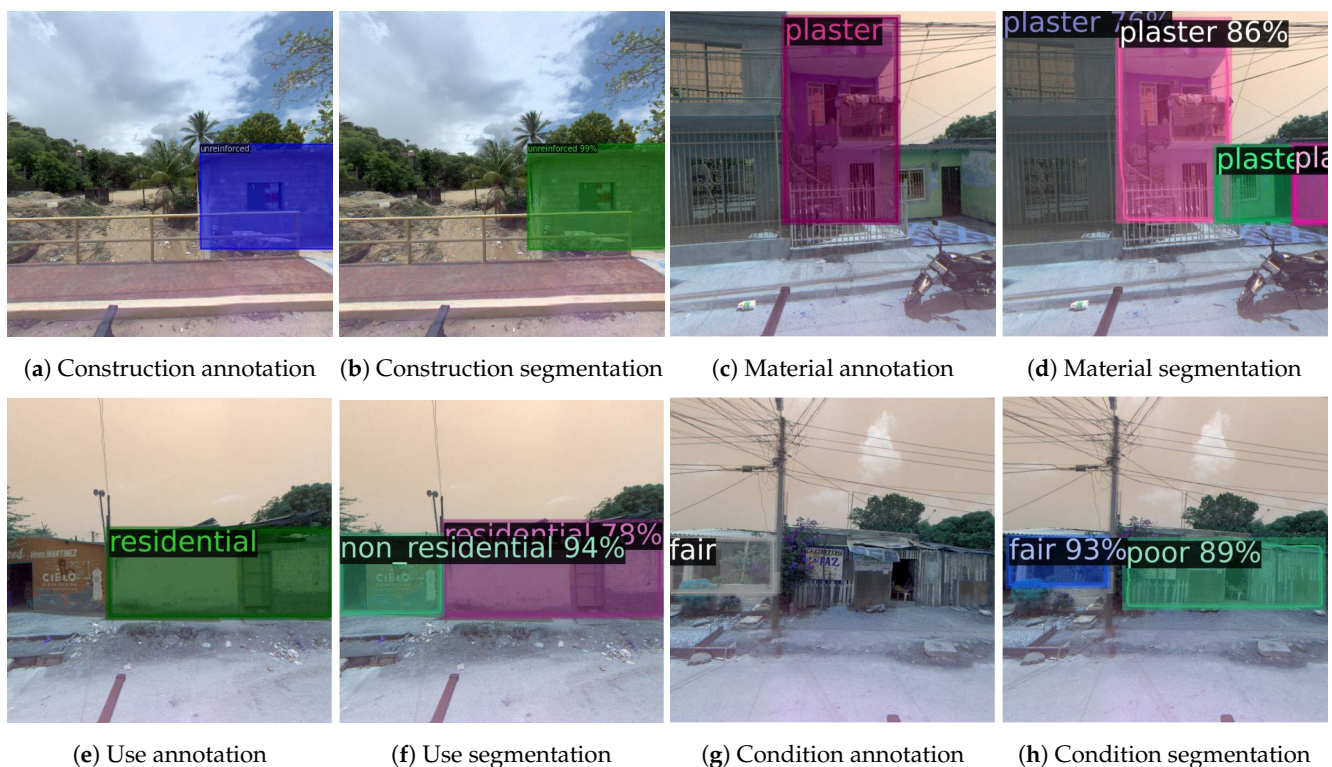


Figure 3. Segmentation results compared with ground truth annotations: (a) construction annotation; (b) construction segmentation; (c) material annotation; (d) material segmentation; (e) use annotation; (f) use segmentation; (g) condition annotation; (h) condition segmentation.

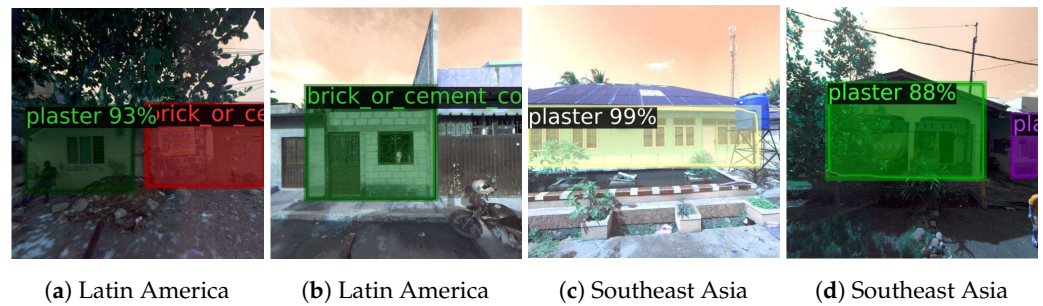


Figure 4. Images from different regions: (a) Latin America; (b) Latin America; (c) Southeast Asia; (d) Southeast Asia.

Some examples of false predictions are shown in Figure 5. It should be noted that a limitation of street view images is that some buildings are not visually accessible from streets, for example, in slums or in regions with heavy vegetation. This is a common issue. Potential solutions would be capturing the facade images from multiple angles or even from drones, or using other methods to predict the occluded features based on other information, such as the method stated in this paper [17], which uses statistical models [28].

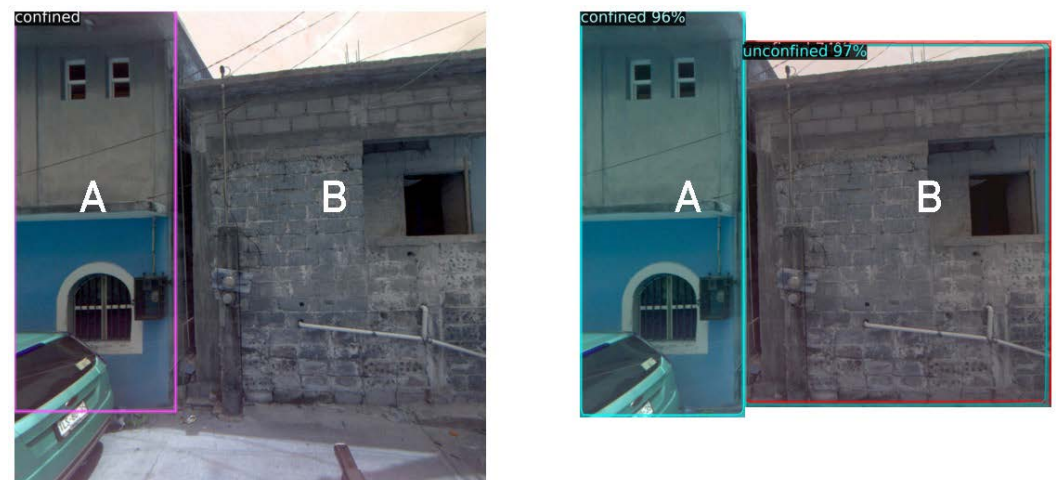


Figure 5. Examples of false predictions (a) Left is annotation (building B is missing in the annotation); right figure shows building B is detected but the classification is false. (b) Left is annotation; right figure shows that the model failed to detect the building

4. Household Vulnerability

A household is typically defined as an individual or group of people (e.g., a family) who live together in one residence and share living arrangements. Vulnerability refers to the degree to which a system, individual, or group is likely to experience harm due to exposure to hazards, stresses, and risks. It also encompasses the inability to withstand or recover from these adverse effects. Vulnerability is a function of the character, magnitude, and rate of climate variation to which a system is exposed, its sensitivity, and its adaptive capacity. In the context of social and economic systems, vulnerability is often a measure of how external stresses (such as economic downturns, natural disasters, or health crises) affect individuals or groups, often influenced by factors like economic stability, health status, social networks, and access to resources or services.

Household vulnerability refers to the susceptibility of a household to potential losses or adversities due to a combination of factors, such as socioeconomic status, physical health, access to resources, and social networks. It typically considers various aspects including the household's capacity to anticipate, cope with, resist and recover from the impact of a natural or anthropogenic hazard. Importantly, household vulnerability is not a fixed characteristic but is dynamic and can change over time due to shifting circumstances and interventions.

In the second part of this paper, we develop the K3 index for quantifying the household vulnerability. K3 is a proxy for a ground-truth testing. It incorporates granular data that are generally unavailable. The index is calculated based on 27 variables that we obtained from census data (Table 3). The dataset is provided by the local governments. The data contain two categories: housing unit data and household members data. The housing unit data are categorical data, which contain questions with Yes/No answers. The household members data are quantitative data, which contain questions with numerical answers. Based on the data, the calculation of K3 is shown in Figure 6. First, for each household, all 27 variables are normalized in values from 0 to 1. The normalization follows a Gower style. For Yes/No questions (e.g., walls made of industrial materials), 1 represents Yes and 0 for No. For numerical variables (e.g., household members between the ages of 15 and 64 that are working), a value from 0–1 is used to indicate the percentage of members with that characteristic. For each household, the values are added up: $k = \sum_{i=1}^{27} v_i$, where v is the normalized value for each variable. Then, Gower's similarity index is calculated, based on which we performed the K-means clustering analysis to group the households into three classes. (The details of Gower's similarity can be found in [29].) So, every household is classified and assigned a value k' according to its group (1, 2, or 3). For each neighborhood block, the averaged value is then calculated $K = \frac{\sum_{j=1}^n k'_j}{n}$, where n is the total counts of households in this block. To ensure that all the clusters are actually distinctive, an ANOVA analysis is conducted. The final value for each block is called the K3 index, ranging from 1 to 3. A lower K3 value means the household is more vulnerable from the socio-economic point of view.

The K3 index is based fully on census data that are heavily validated in the field by the local governments who collaborate with us. To cross validate, we compared the K3 Index geographic data layer that identifies the highly vulnerable areas with another layer that the city planning department has been using to prioritize and target highly vulnerable areas. This government layer was made independently and unpublished. The result of our prediction was employed by the governments to cross validate—the results show they agree with each other. It should be noted that because the built environment is a manifestation of geography and social–economic situation of a region, the K3 should be calibrated when applying the method to a different region or distinctive built environment.

Single building attribute–household vulnerability relationship

To explore the relationship between the visual inference of housing characteristics and the socio-economic status of those living in them, two neighborhoods were viewed with distinct qualities. The first is El Pozón (2.4 sq km), an informally established neighborhood of low-income residents living on the periphery of Cartagena, Colombia. In contrast, the second is Breña, a densely-populated neighborhood near the historical city center of Lima, Peru. Breña (3.2 sq km) is a middle- to low-income neighborhood that was formally established in

the mid-twentieth century. Their K3 map and the histogram are plotted in Figure 7. From the figures, we can see most households in El Pozón have a K3 value less than 1.5 (more vulnerable), while most households in Breña are greater than 1.5 (less vulnerable). This means our K3 index is capable of capturing the distinctions between households.

Table 3. Household variables for calculating K3.

Category	Variable
Housing unit (Yes/No)	More than one household per housing unit? Walls made of industrial materials? Floor made of industrial materials? Access to electricity? Access to water? Access to sewerage? Access to natural gas? Access to waste collection services? Waste collection more than 3 times a week? Access to internet connection (fixed or mobile)? WC connected to sewage network? At least one room in the house is a bedroom? Independent room for the kitchen? Kitchen connected to the water network?
Household Members (Percentages)	Less than 3 persons per bedroom? Members that are men. Members under the age of 20. Members older than 64. Households without members of indigenous origin. Members that live in the municipality they were born. Members that did not get sick. Members without any disability. Members that are not illiterate. Members above 24 years old with college or higher education. Members between the ages of 15 and 64 that are working. Members that are married or in another partnership arrangement. Women that have children.

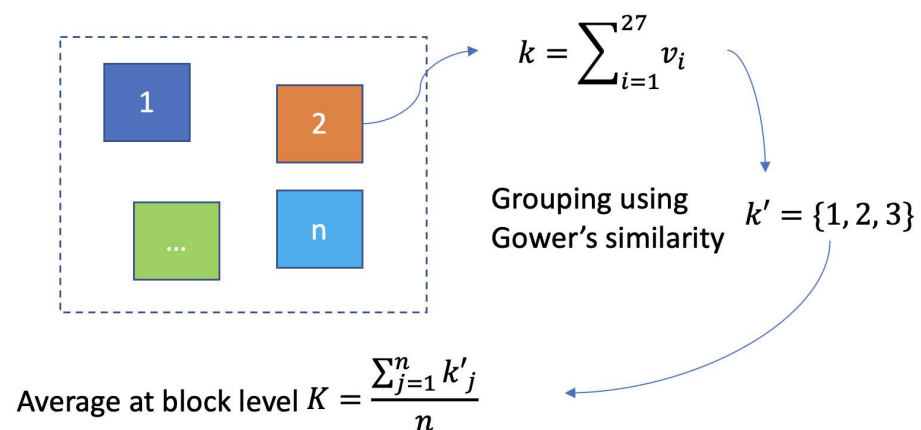
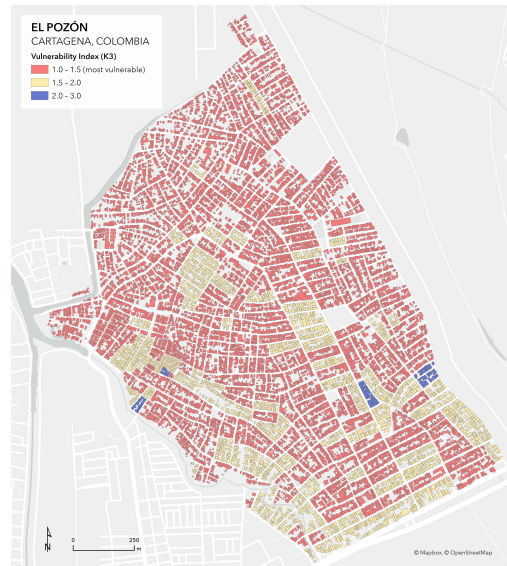


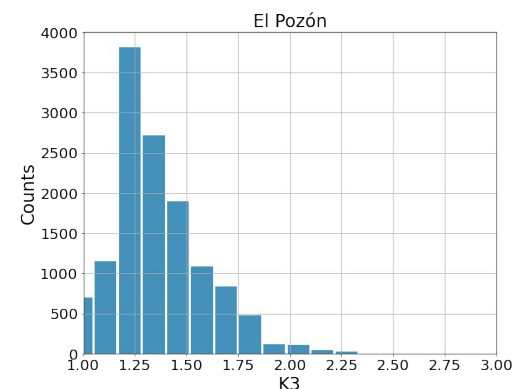
Figure 6. Derivation of K from census data.

We then combined the K3 data for two neighborhoods. The histogram of the combined K3 data is presented in Figure 8. The results show that most of the households in these two areas are considered to be vulnerable since more households have a K3 value less than 2. For each household, we use the segmentation model to infer the building's attributes (construction type, material, use, and condition) from the street view images. We inferred 16,725 street view images of houses using the combined dataset of Breña and El Pozón. Table 4 shows the results of the building attributes inferred from our DL-based model. Our results show that most of

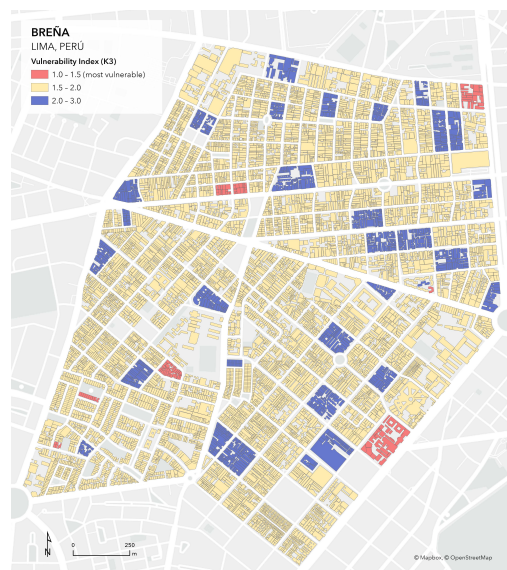
the buildings (95.7%) in these two areas are confined structures; the majority of the house conditions are predicted to be fair (83.0%), while poor-condition houses (14%) outweighed good-condition houses. Most of the houses are made by plaster (66.3%); mix, other, or unclear material (21.9%); and brick or cement concrete block (11.2%). The majority of the houses are residential (89.9%), followed by mixed-use houses (5.9%) and non-residential houses (4.1%). The results indicate houses in these areas are rather consistent in construction type, conditions, and use, while they show more variation in materials.



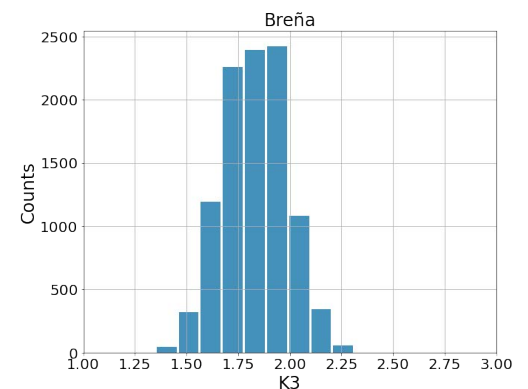
(a) El Pozón K3 map



(b) El Pozón K3 histogram



(c) Breña K3 map



(d) Breña K3 histogram

Figure 7. K3 distributions: (a) El Pozón K3 map; (b) El Pozón K3 histogram; (c) Breña K3 map; (d) Breña K3 histogram.

In order to examine the relationship between each building attribute and their corresponding K3 values, we drew scatter plots accompanied by trend lines to visually represent their correlations. Figure 9 shows that the predictions are correlated with the K3 index. Regarding the construction type, it shows that unconfined buildings have lower K3 values, indicating more vulnerable. For building materials, mix, plaster, brick or concrete block have the highest K3 values, while wood and corrugated metal buildings are found with

lower K3 values. Regarding the use of the building, non-residential buildings have higher K3 values than residential, while mixed building type has the highest. For the condition prediction, the correlation is also very clear and reasonable: poor < fair < good. Based on these observations, we believe that the prediction from street view images and the K3 derived from census data agree with each other very well. Thus, it could be possible to quantify the vulnerability of the household using deep learning and street view images.

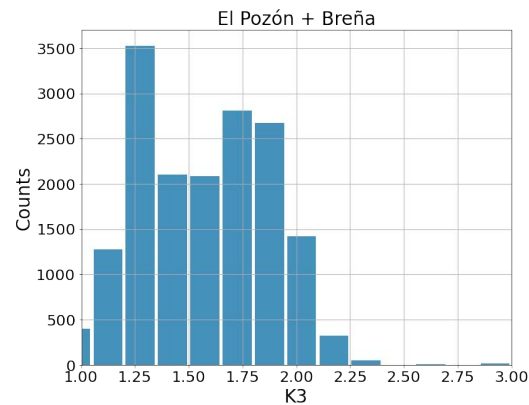


Figure 8. Histogram of K3 (El Pozon and Brena).

Table 4. Descriptive information of building attributes (n = 16,725).

Category	Attributes	Label	Counts (Percentage)
Construction type	confined	1	16,014 (95.7%)
	unconfined	2	711 (4.3%)
Condition	good	1	490 (2.9%)
	fair	2	13,886 (83.0%)
	poor	3	2349 (14.0%)
Material	corrugated_metal	1	38 (0.2%)
	brick_or_cement_concrete_block	2	1869 (11.2%)
	plaster	3	11,082 (66.3%)
	mix_other_unclear	4	3661 (21.9%)
	wood_crude_plank or wood_polished	5	75 (0.4%)
Use	residential	1	15,044 (89.9%)
	mixed	2	987 (5.9%)
	non_residential	3	694 (4.1%)

Mapping visual inference to household vulnerability

Visual inference can tell the susceptibility of the physical built environment, as influenced by factors like construction materials and design. On the other hand, social vulnerability is shaped by the socio-economic and demographic characteristics of the households residing in those buildings. Previous literature has indeed shown that physical susceptibility and social vulnerability often interact to exacerbate the impacts of hazards on communities, e.g., households in physically vulnerable buildings might also have fewer resources to recover from disasters due to their socio-economic conditions. There is plenty of research that has demonstrated the likelihood of a relationship between the two aspects. The article entitled “Social vulnerability to environmental hazards” discussed how social vulnerability interacts with the physical characteristics of locations to enhance the risk posed by environmental hazards [3]. The study entitled “Vulnerability” discussed how household vulnerability to environmental hazards has the contribution from the physical attributes of the built environment [30].

As shown in the previous section, the correlation between single household attributes and household vulnerability provides a foundation for predicting household vulnerability based on their visual attributes. Furthermore, the overall character of one household can be identified as a combination of several visual attributes generated from our DL-based

model. To identify the potential types of households, we applied a Latent Class Clustering (LCA) method to categorize the household according to the visual attributes, including construction type, condition, material, and use. LCA is a probability-based clustering approach that can divide the households into homogenous groups that maximizes the similarity within each group and minimizes the similarity across groups. The number of clusters is determined by the value of Bayesian Information Criteria (BIC). A lower BIC value indicates the LCA model has better goodness-of-fit.

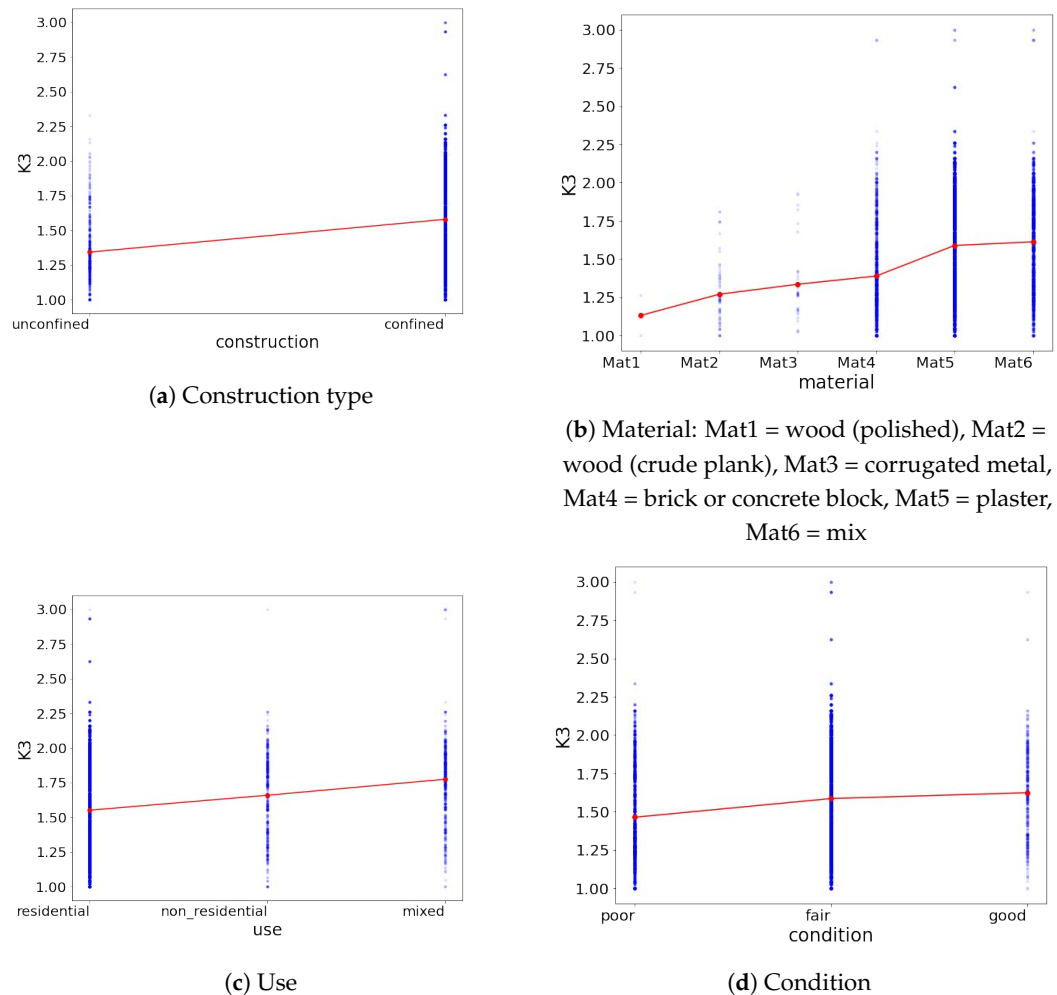


Figure 9. Predictions versus K3 Index: (a) construction type; (b) material: Mat1 = wood (polished), Mat2 = wood (crude plank), Mat3 = corrugated metal, Mat4 = brick or concrete block, Mat5 = plaster, Mat6 = mix; (c) use; (d) condition.

According to Figure 10, when the number of clusters is three, the LCA model achieves the lowest BIC value. The LCA results for three clusters are shown in Table 5. According to the probability of each building attribute in each household type, type 1 households can be characterized as living in “confined, fair-condition residential or mixed-use houses made of plaster or mixed material”; type 2 households can be identified as living in “confined, fair-condition residential houses made of plaster, brick, cement concrete block, or mixed material”; type 3 households can be labeled as living in “unconfined, poor-condition residential houses made of brick, cement concrete block, or plaster”. Type 2 households constitute more than half of all households (67.7%). Type 1 households represent 25.4% of all the households and differ from type 2 households with more mixed-use and mixed material. Type 3 households make up 6.9% of all the households and show larger differences, for they contain most of the unconfined houses with poor conditions.

To further compare the household vulnerability of the three household types, we drew a boxplot and ran a Welch's ANOVA test, as illustrated in Figure 11. The p -value of the ANOVA test is less than 0.01 ($F_{\text{Welch}}(23, 129.87) = 611.41, p \leq 0.01$), which indicates there are significant differences between the three household types in terms of their K3 value. The average K3 values of type 1, 2, and 3 are 1.64, 1.56, and 1.33, respectively, and they are significantly different from each other (p -value less than 0.01 for each pair in Figure 11). This means type 1 households are the least vulnerable, while type 3 households are the most vulnerable. The results demonstrated the potential for predicting household vulnerability based on visual building attributes, as the visually unconfined and poor-conditioned type 3 households have a lower average K3 value, thus are more vulnerable, and the mixed-use and mixed-material-made type 2 households have a higher average K3 value, thus are less vulnerable.

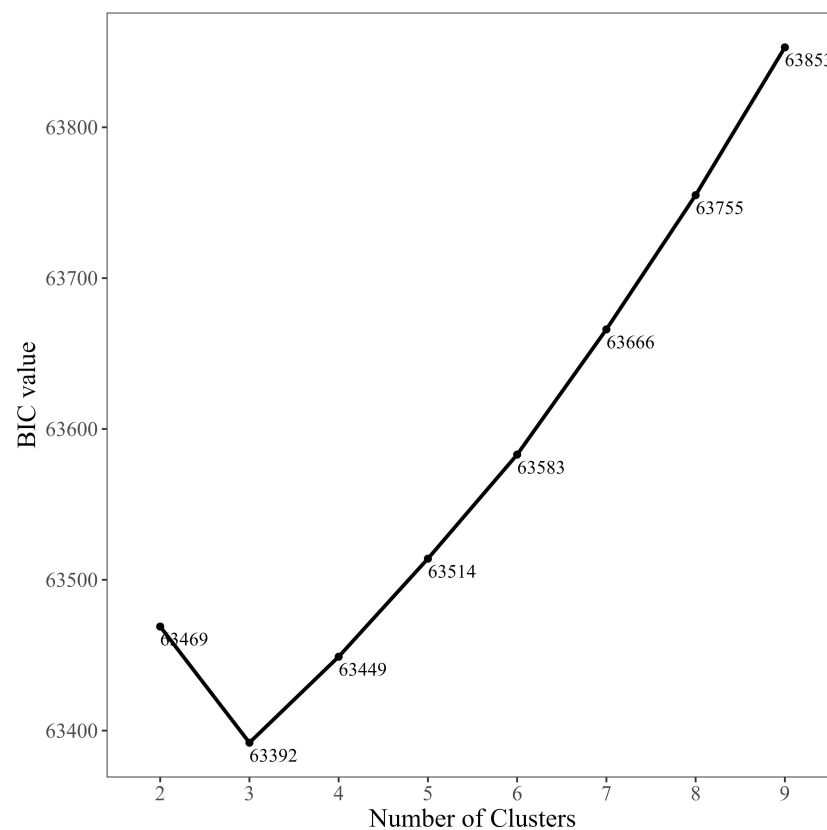


Figure 10. The number of clusters and corresponding BIC value.

Table 5. Latent class clustering results.

Category	Label	Type 1	Type 2	Type 3
Count (Percentage) of total households		4252 (25.4%)	11,320 (67.7%)	1153 (6.9%)
Construction type	1	100%	98.1%	61.5%
	2	0%	1.9%	38.5%
Condition	1	2.3%	3.7%	0%
	2	90.6%	88.3%	14.4%
	3	7.1%	8.1%	85.6%
Material	1	0%	0%	2.3%
	2	0.6%	9%	68.3%
	3	62.9%	74.4%	19.2%
	4	36.5%	16.5%	4.7%
	5	0%	0%	5.5%
Use	1	77.1%	95.4%	99.6%
	2	18.6%	0%	0.1%
	3	4.3%	4.6%	0.3%

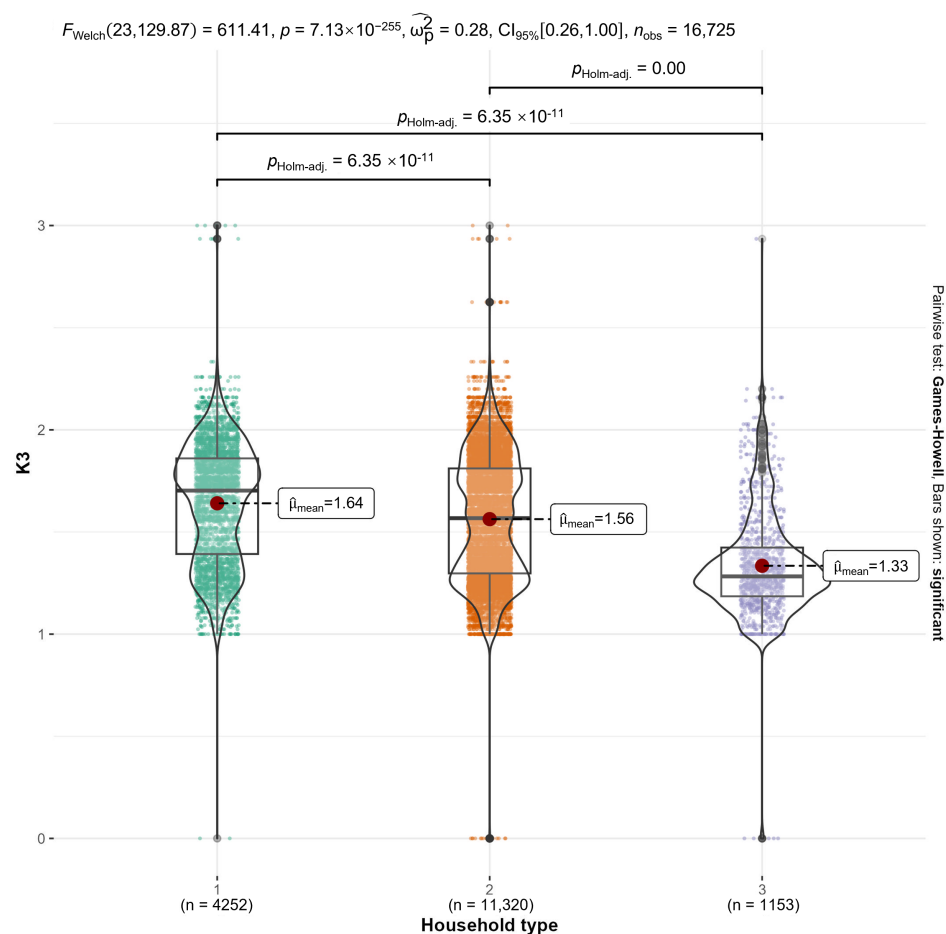


Figure 11. Boxplot with ANOVA test results.

5. Discussion

The increased frequency and severity of natural hazards cause disastrous consequences that significantly impact the built environment. It is essential to evaluate the characteristics of building stock and the vulnerability of households, especially in developing countries, because assessing household vulnerability is key to informing public policies (e.g., housing subsidies, urban upgrading, social cash transfers, etc.) and private investments (e.g., real estate, credit, insurance, education, health and entertainment services, etc.). This study demonstrates that deep learning-based image segmentation can be used to help identify building attributes from street view imagery, which leads to the rapid assessment of household vulnerability. The performance of the trained model is promising: the construction type accuracy is 98.97%, the material type accuracy is 97.08%, the use class accuracy is 94.62%, while the condition classification accuracy is 78.85%.

It is important to note that the annotation dataset used for training is imbalanced. For example, in the material labels, there are three major classes (plaster/mix_other_unclear/brick_or_concrete_block) that are dominant, while the remaining minor classes have much fewer labels ranging from only 46 to 2000. Such imbalances could cause difficulties for the model to recognize the minor classes. It is possible to use re-sampling or similar techniques to enhance the minor classes, but this might not work well for such extreme imbalance, i.e., 46:76,584. Future improvements are possible with more labels in those classes. However, it should also be noted that these minor classes are not common; therefore, they are always hard to find.

Among the four building attributes, the condition prediction has the lowest performance. There are two possible reasons. First, the training dataset is imbalanced. There are far fewer labels for the ‘good’ class. Second, it is hard to distinguish “good” and “fair”,

even for trained labelers. Despite the training and inter-rater reliability checks, when we investigated the annotated images further, we found bias in the labels regarding the two classes. The bias and noise could possibly be one of the causes; therefore, the model might be improved by eliminating those in future studies.

The K3 index has the potential to be used for quantifying the vulnerability of a household. As we found clear correlation between the street view predictions and the K3 index. This implies that it is possible to evaluate the household vulnerability directly from street view images. With the deep learning model, the evaluation procedure has the potential to be scalable.

This study advances the theoretical understanding of household vulnerability by introducing a novel framework that integrates the correlations between the household vulnerability and the street view images. By applying this framework to two selected cities, it not only tests its applicability across diverse geographic regions but also enriches the current discourse with new examples. The findings challenge the conventional approach to household vulnerability assessment, particularly in how a vulnerability index is conceptualized and how it can be predicted by street view images. This contributes to a more nuanced understanding of household vulnerability, offering a platform for future research to build upon. Such theoretical advancements are crucial for developing robust, context-sensitive strategies that resonate with global efforts to mitigate household vulnerability.

There are several limitations or untouched questions in this study, which can be potentially answered in future studies, including the uncertainty in the data and estimations, and the generalization to other regions. For example, one home can have several construction materials, which requires more detailed classification beyond the approach stated in this paper. Another challenge is that occlusions can often cause difficulties to accurately infer the building information. This is a common issue when using street view images, which is also beyond the scope of this paper. While our method helps in identifying physical attributes of buildings that might suggest physical weakness, we acknowledge that it does not directly measure social vulnerability. However, the correlation we observed between the K3 index and the physical attributes derived from the street view images suggests that our method could serve as a valuable preliminary screening tool to identify areas that might be particularly vulnerable to hazards due to both their physical and social characteristics. We perceive the relationship between building characteristics and household vulnerability as intrinsically linked, given that the physical condition of a home can directly impact the livelihood and wellbeing of its inhabitants. However, we also acknowledge that household vulnerability is influenced by a myriad of social, economic, and demographic factors that are not directly tied to the physical state of the housing structure. Though there is a significant correlation between poor building conditions and lower socio-economic status, exceptions do exist, such as in condominium situations, where individual family conditions may not directly reflect the overall building's appearance. Our machine learning models are adept at identifying patterns and trends within large datasets, yet they are not infallible and do not claim absolute certainty. The premise that poor building conditions unequivocally imply poor family conditions would overlook the nuanced reality of urban socio-economic landscapes. As such, while a strong likelihood exists that poor building conditions correlate with lower socio-economic status, this is not a universal rule. In future developments, we aim to refine our models to account for such discrepancies and explore additional data sources that could provide more insight into individual household conditions. This may include cross-referencing building conditions with other socio-economic indicators or integrating more granular data to enhance the accuracy of our socio-economic assessments.

6. Conclusions

This study presents an automated method to assess household vulnerability at a large scale. Building attributes are firstly characterized from street view images with a deep learning-based instance segmentation method. The model can detect four building attributes with high accuracy: construction type (98.97%), material (97.08%), use (94.62%),

and condition (78.85%). The model is broadly applicable to regions that have similar geographical street views, indicating the similar construction types, development, and economy levels.

We then demonstrated that a census data-based index, K3, can be used for quantifying the vulnerability of a household: financially robust households have higher K3 values, while households with lower K3 values are more vulnerable. Applying the developed segmentation model and K3 model to two neighborhoods, Breña, Peru, and El Pozón, Colombia, we found a clear correlation between these two sources. Therefore, we believe it is possible to develop a deep learning-based automatic system to rapidly evaluate household vulnerabilities from street view images.

When applying this framework to regions significantly different from those studied in this paper, calibration is essential. For instance, the DL model may require retraining with new data to accommodate diverse building types and designs, and the K3 parameter should be adjusted based on updated census data.

This work appears to be among the first studies that uses a deep learning-based image analysis for a household vulnerability study. The present approach aims at scalability and higher level reliability—it provides an automated and inexpensive method for large-scale regional examinations of vulnerability at the household level. The method requires minimum interactions, providing flexibility that enables implementations even during a period like COVID-19. It overcomes the difficulties in traditional assessments that are expensive or dependent on either slowly-developed datasets, such as a census, or third-party datasets inaccessible to most users with the level of detail needed to make key policy or business decisions. The major innovation of this study is that we established a correlation between machine-inferred explicit facade characteristics with household vulnerability, which paves the way to a rapid and large-scale assessment.

Author Contributions: Conceptualization, C.W., S.E.A., J.G.G.-G. and L.M.T.; Methodology, C.W.; Software, C.W.; Validation, C.W., J.G.G.-G., C.Z., J.v.M. and X.Y.; Formal analysis, C.W. and J.G.G.-G.; Investigation, C.W.; Resources, C.W., S.E.A. and L.M.T.; Data curation, C.W., S.E.A., J.G.G.-G. and L.M.T.; Writing—original draft, C.W.; Writing—review & editing, C.W., S.E.A., J.G.G.-G., L.M.T., C.Z., J.v.M. and X.Y.; Visualization, C.W. and J.G.G.-G.; Supervision, C.W.; Project administration, C.W.; Funding acquisition, C.W. All authors have read and agreed to the published version of the manuscript.

Funding: This project is partially supported by the NVIDIA Applied Research Accelerator Program, granted to Chaofeng Wang.

Data Availability Statement: Testing data, trained models, and the codes that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The images and labels used in this paper were generated by the team of the World Bank's Global Program for Resilient Housing.

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

References

1. Mori, N.; Takemi, T.; Tachikawa, Y.; Tatano, H.; Shimura, T.; Tanaka, T.; Fujimi, T.; Osakada, Y.; Webb, A.; Nakakita, E. Recent nationwide climate change impact assessments of natural hazards in Japan and East Asia. *Weather. Clim. Extrem.* **2021**, *32*, 100309. [[CrossRef](#)]
2. Wallemacq, P.; Below, R.; McClean, D. *Economic Losses, Poverty & Disasters: 1998–2017*; United Nations Office for Disaster Risk Reduction: Geneva, Switzerland, 2018.
3. Cutter, S.L.; Boruff, B.J.; Shirley, W.L. Social vulnerability to environmental hazards. *Soc. Sci. Q.* **2003**, *84*, 242–261. [[CrossRef](#)]
4. Wisner, B.; Blaikie, P.M.; Blaikie, P.; Cannon, T.; Davis, I. *At Risk: Natural Hazards, People's Vulnerability and Disasters*; Psychology Press: London, UK, 2004.
5. ATC. *Rapid Visual Screening of Buildings for Potential Seismic Hazards: A Handbook, FEMA 154*; Federal Emergency Management Agency: Washington, DC, USA, 1988.
6. ATC. *Rapid Visual Screening of Buildings for Potential Seismic Hazards, a Handbook, FEMA 154 Second Edition*; Applied Technology Council, National Earthquakes Hazards Reduction Program: Redwood City, CA, USA, 2002.

7. ATC. *Rapid Visual Screening of Buildings for Potential Seismic Hazards, a Handbook, FEMA 154 Third Edition*; Federal Emergency Management Agency: Washington, DC, USA, 2015.
8. Naik, N.; Philipoom, J.; Raskar, R.; Hidalgo, C. Streetscore-predicting the perceived safety of one million streetscapes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014.
9. Liu, X.; Chen, Q.; Zhu, L.; Xu, Y.; Lin, L. Place-centric visual urban perception with deep multi-instance regression. In Proceedings of the 25th ACM International Conference on Multimedia, Mountain View, CA, USA, 23–27 October 2017.
10. Gebru, T.; Krause, J.; Wang, Y.; Chen, D.; Deng, J.; Aiden, E.L.; Li, F.-F. Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 13108–13113. [\[CrossRef\]](#)
11. Bency, A.J.; Rallapalli, S.; Ganti, R.K.; Srivatsa, M.; Manjunath, B. Beyond spatial auto-regressive models: Predicting housing prices with satellite imagery. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Santa Rosa, CA, USA, 24–31 March 2017.
12. Kang, Y.; Zhang, F.; Gao, S.; Peng, W.; Ratti, C. Human settlement value assessment from a place perspective: Considering human dynamics and perceptions in house price modeling. *Cities* **2021**, *118*, 103333. [\[CrossRef\]](#)
13. Law, S.; Paige, B.; Russell, C. Take a look around: Using street view and satellite images to estimate house prices. *arXiv* **2018**, arXiv:1807.07155.
14. Lenjani, A.; Dyke, S.J.; Bilionis, I.; Yeum, C.M.; Kamiya, K.; Choi, J.; Liu, X.; Chowdhury, A.G. Towards fully automated post-event data collection and analysis: Pre-event and post-event information fusion. *Eng. Struct.* **2020**, *208*, 109884. [\[CrossRef\]](#)
15. Lenjani, A.; Yeum, C.M.; Dyke, S.; Bilionis, I. Automated building image extraction from 360 panoramas for postdisaster evaluation. *Comput.-Aided Civ. Infrastruct. Eng.* **2020**, *35*, 241–257. [\[CrossRef\]](#)
16. Wang, C.; Antos, S.E.; Triveno, L.M. Automatic detection of unreinforced masonry buildings from street view images using deep learning-based image segmentation. *Autom. Constr.* **2021**, *132*, 103968. [\[CrossRef\]](#)
17. Wang, C.; Yu, Q.; Law, K.H.; McKenna, F.; Stella, X.Y.; Taciroglu, E.; Zsarnóczy, A.; Elhaddad, W.; Cetiner, B. Machine learning-based regional scale intelligent modeling of building information for natural hazard risk management. *Autom. Constr.* **2021**, *122*, 103474. [\[CrossRef\]](#)
18. Yu, Q.; Wang, C.; McKenna, F.; Stella, X.Y.; Taciroglu, E.; Cetiner, B.; Law, K.H. Rapid visual screening of soft-story buildings from street view images using deep learning classification. *Earthq. Eng. Eng. Vib.* **2020**, *19*, 827–838. [\[CrossRef\]](#)
19. Rundle, A.G.; Bader, M.D.; Richards, C.A.; Neckerman, K.M.; Teitler, J.O. Using Google Street View to audit neighborhood environments. *Am. J. Prev. Med.* **2011**, *40*, 94–100. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Xia, Y.; Yabuki, N.; Fukuda, T. Development of a system for assessing the quality of urban street-level greenery using street view images and deep learning. *Urban For. Urban Green.* **2021**, *59*, 126995. [\[CrossRef\]](#)
21. Li, X.; Zhang, C.; Li, W.; Ricard, R.; Meng, Q.; Zhang, W. Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban For. Urban Green.* **2015**, *14*, 675–685. [\[CrossRef\]](#)
22. Ma, X.; Ma, C.; Wu, C.; Xi, Y.; Yang, R.; Peng, N.; Zhang, C.; Ren, F. Measuring human perceptions of streetscapes to better inform urban renewal: A perspective of scene semantic parsing. *Cities* **2021**, *110*, 103086. [\[CrossRef\]](#)
23. Samany, N.N. Automatic landmark extraction from geo-tagged social media photos using deep neural network. *Cities* **2019**, *93*, 1–12. [\[CrossRef\]](#)
24. Taecharungroj, V.; Mathayomchan, B. The big picture of cities: Analysing flickr photos of 222 cities worldwide. *Cities* **2020**, *102*, 102741. [\[CrossRef\]](#)
25. Kang, J.; Körner, M.; Wang, Y.; Taubenböck, H.; Zhu, X.X. Building instance classification using street view images. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 44–59. [\[CrossRef\]](#)
26. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
27. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 91–99. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Wang, C. NHERI-SimCenter/SURF: v0.2.0. *Zenodo*, 2019. Available online: <https://zenodo.org/records/3463676> (accessed on 12 October 2023).
29. Gower, J.C. A general coefficient of similarity and some of its properties. *Biometrics* **1971**, *27*, 857–871. [\[CrossRef\]](#)
30. Adger, W.N. Vulnerability. *Glob. Environ. Chang.* **2006**, *16*, 268–281. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.