WARNER: Weakly-Supervised Neural Network to Identify Eviction Filing Hotspots in the Absence of Court Records

Maryam Tabar
Penn State University, USA
mfg5544@psu.edu

Wooyong Jung
Penn State University, USA
wjung@psu.edu

Amulya Yadav
Penn State University, USA
amulya@psu.edu

Owen Wilson Chavez
Child Poverty Action Lab, USA
owen@childpovertyactionlab.org

Ashley Flores
Child Poverty Action Lab, USA
ashley@childpovertyactionlab.org

Dongwon Lee
Penn State University, USA
dongwon@psu.edu

ABSTRACT

The widespread eviction of tenants across the United States has metamorphosed into a challenging public-policy problem. In particular, eviction exacerbates several income-based, educational, and health inequities in society, e.g., eviction disproportionately affects low-income renting families, many of whom belong to underrepresented minority groups. Despite growing interest in understanding and mitigating the eviction crisis, there are several legal and infrastructural obstacles to data acquisition at scale that limit our understanding of the distribution of eviction across the United States. To circumvent existing challenges in data acquisition, we propose WARNER, a novel Machine Learning (ML) framework that predicts eviction filing hotspots in US counties from unlabeled satellite imagery dataset. We account for the lack of labeled training data in this domain by leveraging sociological insights to propose a novel approach to generate probabilistic labels for a subset of an unlabeled dataset of satellite imagery, which is then used to train a neural network model to identify eviction filing hotspots. Our experimental results show that WARNER acheives a higher predictive performance than several strong baselines. Further, the superiority of WARNER can be generalized to different counties across the United States. Our proposed framework has the potential to assist NGOs and policymakers in designing well-informed (data-driven) resource allocation plans to improve the nationwide housing stability. This work is conducted in collaboration with The Child Poverty Action Lab (a leading non-profit leveraging data-driven approaches to inform actions for relieving poverty and relevant problems in Dallas County, TX). The code can be accessed via https://github.com/maryam-tabar/WARNER.

CCS CONCEPTS

• **Applied computing → Sociology**; • **Computing methodologies → Neural networks**.

CIKM '22, October 17–21, 2022, Atlanta, GA, USA

ACM ISBN 978-1-4503-9236-5/22/10…$15.00
https://doi.org/10.1145/3511808.3557128

KEYWORDS

Social Good, The Eviction Crisis, Machine Learning, Weakly Supervised Learning

ACM Reference Format:
Maryam Tabar, Wooyong Jung, Amulya Yadav, Owen Wilson Chavez, Ashley Flores, and Dongwon Lee. 2022. WARNER: Weakly-Supervised Neural Network to Identify Eviction Filing Hotspots in the Absence of Court Records. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM '22), October 17–21, 2022, Atlanta, GA, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3511808.3557128

1 INTRODUCTION

Numerous low-income renting families across the USA are at a high risk of eviction, mainly due to a shortage of federal housing assistance, and an ever-increasing gap between income growth and increases in housing cost, e.g., about 70% of the low-income renters devote most of their income towards housing expenses [12, 13]. Further, eviction is an important cause of several societal problems, such as homelessness [6, 44], and has long-lasting negative effects on individuals' health, and housing stability [8, 12, 17, 21]. As a result, mitigating the eviction crisis is of the utmost importance in order to enhance the well-being of this community.

To tackle this crisis, NGOs and policymakers have been implementing multiple programs at the pre-filing and post-filing stages [5, 33, 43]; especially, they assign several types of resources to increase housing stability and affordability. Efficient resource allocation for these eviction prevention programs is possible only with a data-driven understanding of eviction filing hotspots across the USA. Unfortunately, some state/local policies and infrastructure limitations restrict access to ground-truth eviction filing records for many regions of USA, which in turn, limits our understanding of eviction filing hotspots in those regions [11]. Thus, accurately predicting eviction filing hotspots at a high spatial resolution (without accessing the ground-truth eviction filing records) would enable policymakers to design more well-informed eviction prevention and mitigation programs, and hence, more effectively address the eviction crisis across the USA.

There have been a few studies, at the intersection of sociology and Machine Learning (ML), that use supervised ML models to predict the eviction (or eviction filing) [41, 42] and landlord harassment [47]. However, they are not well-suited for predicting the occurrence of eviction (i.e., the target variable) at a national scale because (i) they rely on input features (such as historical eviction filing data [41] or the outreach activities of specialists at Tenant Support Unit (TSU) [47]) that are only available for certain cities in the USA (which limits their usability to these specific cities), and (ii) they assume that sufficient training data with ground-truth labels is available for training ML models. Unfortunately, this assumption does not hold for the task of eviction filing hotspot prediction because of two reasons: (i) some state/local laws and policies limit access to eviction filing records, e.g., in Illinois, bulk data retrieval is not allowed or in California, tenants may block public access to their eviction records [11], and (ii) the high cost of data collection in some other regions makes it infeasible to collect eviction filing records at scale, e.g., due to some infrastructure limitations, obtaining eviction filing data requires collecting eviction filing records in person from many different judicial courts [11].

To fill this gap, we propose WARNER (**W**eakly-supervised **A**id to **R**elieve **N**ationwide **E**viction **R**ate), a weakly-supervised ML model that leverages publicly available satellite imagery as well as sociological insights (instead of ground-truth labels) to predict eviction filing hotspots across the USA. In particular, we make the following contributions: (i) to account for the lack of sufficient labeled training data in this domain, we propose a novel label generation approach that leverages the findings of past literature in sociology to produce probabilistic labels for a subset of an unlabeled satellite image based training set, (ii) we develop a neural network model to predict eviction filing hotspots from satellite imagery of different shapes, and (iii) we do several experiments to assess the accuracy of WARNER using a real-world dataset with eviction filing records in Dallas County, TX.

Our empirical evaluation shows the high quality of the labels generated by our proposed label generation approach. Furthermore, the results show that WARNER outperforms multiple strong baseline models by obtaining about 36.0% and 1.4% higher F1 and AUC (respectively), which illustrates its suitability for this problem domain. Additionally, the superior accuracy of WARNER can be generalized to different counties across the United States. This work is conducted in collaboration with The Child Poverty Action Lab (CPAL)[1] and its usability in the field is being assessed by domain experts.

2 RELATED WORK

In the following subsections, we survey past literature in the areas of sociology and Machine Learning.

2.1 Sociological Research

Prior work in sociology mostly focuses on understanding the factors associated with eviction using statistical and descriptive analysis. In fact, prior work has studied the association between the risk of eviction and various individual and neighborhood-level characteristics. e.g., they found that job loss, and crime rates in a neighborhood tend to increase the risk of eviction [10]. Further, prior research found that low-income single mothers who have young children tend to be at a high risk of getting evicted [9]. Additionally, according to their findings, eviction could lead to long-lasting health problems (e.g., depression) [12]. Even though this line of work gleaned unique insights about the eviction crisis, they did not address the problem of predicting eviction (or eviction filing) hotspots across the USA (which is the focus of our work).

Some other prior work focuses on finding eviction hotspots in certain geographic regions by counting the total number of eviction filings in their sub-regions. As a result, they find that a large number of evictions in a region can be attributed to a small number of subregions [21, 39]. However, these works require the actual number of evictions (or eviction filing records), which is inaccessible (or highly expensive to obtain) for many regions due to restrictive state/local policies and infrastructure limitations [11]. Therefore, their methodology is not generalizable to all regions within USA. In contrast, this paper proposes a highly generalizable ML-based framework that relies on satellite imagery and sociological insights (rather than the actual number of eviction filings) to predict eviction filing hotspots within US counties in the absence of court records.

2.2 Machine Learning Research

There has been a large number of research on applying ML techniques to tackle societal problems. One line of research developed predictive ML models using a tabular dataset consisting of several factors with potential impacts on the dependent variable [42, 47]. For example, Ye et al. [47] relied on some classical ML models to predict the risk of landlord harassment using a tabular dataset. However, these models have limited real-world usability, as their predictive performance is highly dependent on data sources that are either (i) unavailable for many regions across the USA, or (ii) highly expensive to obtain as they need to be gathered by conducting surveys. Additionally, a recent study [41] developed a multi-view time-series neural network model to forecast the number of eviction filings for each census tract. However, it assumes that the historical eviction filing data is available for the target region, but this assumption does not necessarily hold at the national scale. In contrast, in this paper, we rely on publicly available datasets that cover all census tracts across USA (namely, satellite imagery and the American Community Survey data[2]).

Another line of research takes advantage of imagery data and variants of Convolutional Neural Network (CNN) models [22] to predict factors related to poverty and human development. In particular, some prior work focused on predicting poverty from satellite imagery in the face of sparsely labeled data [19, 46], and to tackle this issue, they proposed to incorporate night-time light intensity as a proxy for poverty during training. However, a subsequent study [16] showed that this methodology does not necessarily generalize to predicting some other human development factors (such as access to water and average child weight-to-height percentile). Additionally, some studies used computer vision approaches (such as object detection techniques [2, 3], panoptic image segmentation [23], and CNN-based neural networks [18, 48]) for predicting poverty and/or other development indicators from imagery data. However, relying on a supervised learning paradigm, all these studies trained their models on a dataset consisting of ground-truth labels. In contrast, this paper proposes a framework for predicting eviction filing hotspots without access to ground-truth labels; instead, it addresses the lack of labeled training data by leveraging

[1]https://childpovertyactionlab.org

[2]https://www.census.gov/programs-surveys/acs/data.html

insights of prior work in sociology to develop a weak supervision approach to generating labels.

3 A PROBLEM STATEMENT

In this section, we provide a formal definition of the problem of identifying eviction filing hotspots in US counties. Intuitively, **eviction filing hotspots** of a county c over a period of m years refer to the census tracts[3] (in that county c) which "*consistently have high contributions*" to the total eviction filings in c during a period of m years.

More formally, we define **top-k% eviction filing hotspots** of a county c over a period of m years as follows. Suppose that c has n census tracts, and E_t^i denotes the total number of eviction filings in the i^{th} census tract of c in year t ($t \in \{1, ..., m\}$) after sorting census tracts of c in the descending order of their number of eviction filings in year t. Additionally, let S_t^d refer to the largest set of census tracts (in descending order) whose combined number of eviction filings is less than or equal to d% of the total number of filings in c in year t (i.e., $\sum_{tract_i \in S_t^d} E_t^i \leq \frac{d}{100} \times \sum_{i=1}^{n} E_t^i$). Please note that we add census tracts to S_t^d in decreasing order of the number of eviction filings. Then, the *top-k% eviction filing hotspots of c over a period of m years* are defined as the set $\cap_{t=1}^{m} S_t^d$ such that $|\cap_{t=1}^{m} S_t^d| \approx \lceil \frac{k \times n}{100} \rceil$. In this definition, k and m are considered to be fixed (defined by stakeholders) and d is chosen to be the largest number such that $|\cap_{t=1}^{m} S_t^d| \leq \lceil \frac{k \times n}{100} \rceil$. Note that these conditions can be satisfied with fractional values of d and k.

Finally, we formulate the problem of identifying top-k% eviction filing hotspots as a *binary classification problem*, in which the ultimate goal is to predict if a census tract belongs to the top-k% eviction filing hotspots of its county (i.e., positive label) or not (i.e., negative label). In this paper, we propose an ML model that takes the satellite images of a census tract ($\{x_1^{tract}, x_2^{tract}, ..., x_m^{tract}\}$) and its county ($\{x_1^{county}, x_2^{county}, ..., x_m^{county}\}$) as input and outputs a prediction for the binary variable of interest. To assess the effectiveness of the proposed model for this problem domain, we experiment with different values of k in Section 6.

4 DATASETS

In this study, we use three datasets: (i) American Community Survey data, (ii) Satellite imagery, and (iii) Eviction filing records.

4.1 American Community Survey (ACS)

ACS data[2] contains various pieces of information on demographic characteristics, housing characteristics, work status, and poverty status in the past 12 months. This dataset is published by the U.S. Census Bureau annually, but with delay of about two years. We use ACS 5-Year Experimental Estimates in this work as it provides annual statistics for all census tracts in the U.S. *Note that we only use the ACS data to generate weakly supervised labels for our satellite imagery training datasets, which we describe next.*

[3]A census tract refers to a sub-division of a county and is usually used for national surveys.

4.2 Satellite Imagery

We use Sentinel imagery[4], which provides a bird's eye view of the environment with spatial and temporal resolutions of 10 meters and 10 days, respectively. For each census tract, we crawl one image corresponding to the bounding box of that tract (i.e., the minimum rectangle surrounding the polygon of that census tract). Further, for each image, we generate a mask matrix to be able to distinguish the pixels that fall inside that census tract (i.e., valid pixels) from the other ones (i.e., invalid pixels).

4.3 Eviction Filing Records

Through our collaboration with CPAL, we also have access to all historical *ground-truth* eviction filing records collected from Dallas County, TX since 2017. As part of its operations, CPAL gets daily updates regarding new eviction filing records across Dallas County on business days. Each record contains information about the eviction filing time, tenant's address (i.e., latitude and longitude), names of both parties (i.e., landlord and tenant), etc. However, this data does not have the court's decision about each eviction case. *Please note that while this dataset on eviction filing records is available for Dallas County (through our collaboration with CPAL), getting similar datasets from other US counties is very challenging, if not impossible. Since we want a generalizable ML model that can predict eviction filing hotspots across all US counties (not just Dallas County), we do not use this dataset to train our ML model. Instead, we only use this source of data for evaluating the performance of WARNER.*

5 THE PROPOSED FRAMEWORK: WARNER

We propose a weakly-supervised framework to address the problem of predicting eviction filing hotspots from satellite imagery in the face of a lack of ground-truth eviction filing data. Figure 2 illustrates the architecture of WARNER, which is composed of two components: (i) a *label generation model* that generates probabilistic labels for a subset of an unlabeled satellite imagery dataset by leveraging insights from prior work in sociology as well as the ACS data, and (ii) a *hotspot prediction model* that predicts eviction filing hotspots from satellite imagery (along with the generated labels) using a neural network model. In the following subsections, we elaborate on the architecture of each component.

5.1 The Label Generation Model

As the first step toward predicting eviction filing hotspots, we build an ML model that uses sociological insights along with the ACS data (as input) to generate labels for our satellite imagery training dataset. In fact, the following steps are taken: (i) we survey past literature in sociology to find several factors that are highly associated with eviction. (ii) Then, we define one labeling function (LF)[5] for each associated factor; i.e., a labeling function labels each data point based on the value of the underlying associated factor. (iii) Finally, we use a weak supervision framework via Snorkel [38] to combine the results of various labeling functions; i.e., since each data point might be labeled by several labeling functions, we use Snorkel to

[4]https://sentinel.esa.int/web/sentinel/missions/sentinel-2

[5]A labeling function is a piece of code that takes a data point (i.e., census tract) as input and assigns a label (positive, negative, or abstain for binary classification) using some rules (or heuristics, etc.).

Table 1: The definition of ACS factors underlying our labeling functions ([41]).

LF#	Explanation of the Underlying ACS Feature	Polarity
1	# of housing units occupied by renters	Positive
2	# of renter-occupied units whose householder has zero or negative earnings in the previous 12 months	Positive
3	# of renter-occupied units, with monthly housing costs $\geq$ (0.3$\times$income)	Positive
4	# of renter-occupied units whose householder's level of education is less than high school	Positive
5	% of renter-occupied units whose householder is a high school graduate (or equivalent)	Negative
6	% of renter-occupied units whose householder has a college or associate's degree	Negative
7	% of renter-occupied units whose householder has at least a bachelor degree	Negative
8	% of families below poverty line who get paid SSI or cash public assistance income	Negative
9	% of full-time workers with some earnings	Negative

convert that set of *potentially noisy* labels into one probabilistic label. The following paragraphs provide further details regarding each step.

Sociological Insights. To mitigate the eviction crisis, sociologists and social work scientists have been studying various aspects of the eviction crisis and housing instability. As a result, they have discovered several data-driven insights; for example, they found a high level of association between the risk of eviction and some demographic and financial characteristics of renters (and neighborhoods) [10, 36, 40]. However, these studies rely on datasets collected through conducting in-person surveys from a relatively small population, and thus, their datasets and studied features are not available as-is for all U.S. census tracts. To tackle this challenge, we use the ACS dataset, which consists of various demographic, financial, and housing characteristics at the census tract level (across the entire U.S.). Next, we review prior work in sociology to find a set of factors associated with eviction and housing instability [4, 10, 13, 36, 40]. For each associated factor, we try to locate that factor among the set of features present in the ACS dataset. If an associated factor is not found as-is in the ACS dataset, an ACS feature that is semantically close to that factor is selected, instead. Note that in spite of their association with eviction, some neighborhood-level characteristics (such as past eviction rate) [10] and social network properties (such as network disadvantage) [10] are not considered in this study because they are not gathered in the nationwide ACS dataset. As a result of taking these steps, we successfully locate nine associated factors (as reported in prior sociology literature) in the ACS dataset (similar to [41]), and these nine factors form the basis of our label generation model.

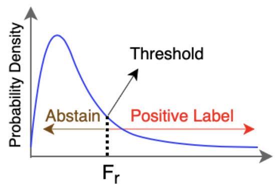
Table 1 provides the definition of these nine associated factors that were located in the ACS dataset. Each of the selected factors is shown to have some sort of association with eviction. For example, while job loss, and hence, zero income (LF#2) tend to increase the risk of eviction [10, 40], being employed (LF#9) has shown to be a protective factor[6] for housing instability [36]. Additionally, most low-income renting families reportedly spend a considerable amount of their income on housing expenses; in fact, about 70% of them devote most of their earnings on housing expenses [13]. Accordingly, we include LF#3 in our feature set. Furthermore, individuals with educational attainment of less than high school (LF#4) tend to be at higher risk of eviction [40], whereas having higher level of education (LFs #5, #6, and #7) tends to be associated with a lower likelihood of housing instability [4]. Intuitively, census tracts with higher numbers of renter-inhabited housing units (LF#1) tend to have high contributions to the county's total evictions. Finally, recipients of public assistance (LF#8) are found to be at lower risk of housing instability [4, 36]. *Please note that, due to the potential ethical implications of labeling data points only based on some protected characteristics (such as gender, race, and age), we did not utilize such factors for defining labeling functions.* Next, we describe how we use this set of associated factors to generate probabilistic labels for an unlabeled dataset.

[6]Protective (risk) factors refer to factors that are associated with a lower (higher) chance of a negative outcome.

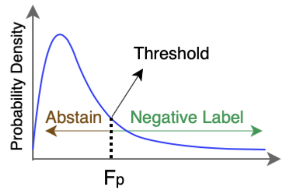
Design of Labeling Functions. Although prior work distinguishes between protective and risk factors for eviction, no rule has been defined for identifying a concerning level of eviction (or eviction filing) risk from the value of an associated factor. In this paper, we propose a novel approach to design such rules that mainly relies on (i) the shape of the probability distribution of the selected ACS features, (ii) whether the underlying factor is found to be a risk factor or a protective factor, (iii) the value of k (i.e., the desired percentage of hotspots in a county), and (iv) the characteristics of the county of each census tract. In the following paragraphs, we elaborate on the role of the aforementioned criteria in the design of our labeling functions.

We define a labeling function for each of our nine associated factors (as shown in Table 1) separately. Each of these labeling functions can abstain from providing labels for a data point if it is highly uncertain about the label of that data point. To this end, we analyze the probability distributions of our nine factors in each county separately, and find out that all distributions are right-skewed (similar to Figure 1), where the distribution's right tail is longer than its tail on the left side. Therefore, the data points that fall on the left hand side of the probability distribution look somewhat similar to each other with respect to that selected factor. This piece of evidence has motivated us to design labeling functions that abstain from labeling the data points that fall on the left hand side.

Next, we need to decide on the **polarity** of each labeling function, which refers to the type of labels that it can assign (e.g., in a binary classification problem, the polarity can be any of the following:



(a) **Risk factor F_r**



(b) **Protective factor F_p**

Figure 1: The proposed approach for defining labeling functions.

Positive, Negative, or {Positive, Negative}). The polarity of each labeling function is defined as follows: A labeling function corresponding to a risk factor only assigns positive labels, and similarly, the one corresponding to a protective factor only assigns negative labels. We made this decision because when a factor is known to be a risk factor (resp. protective factor) for the prevalence of eviction, it is positively correlated with the higher (resp. lower) number of eviction. Thus, a larger (resp. smaller) value of this factor provides a signal on a larger (resp. smaller) number of evictions and eviction filings in that region. The polarity of our labeling functions are given in Table 1.

Furthermore, we need to specify the exact value of the threshold (shown in Figure 1) to complete the definition of our labeling functions. Since we want to find the top-$k\%$ hotspots of each county and both high precision and recall are equally important in this domain, the threshold for factor f and county c is defined as the $\lceil \frac{k \times n}{100} \rceil^{th}$ largest value of that factor among census tracts in county c, where n refers to the total number of census tracts in county c. As a result, each labeling function labels about $k\%$ of data points. Figure 1 summarizes our schema for defining labeling functions. Next, we explain how to integrate these noisy signals to assign (at most) one probabilistic label to each data point.

Probabilistic Label Generation. Each labeling function provides a signal, with unknown accuracy, regarding the label of each data point. Now, we need to integrate those signals to generate (at most) one label per data point. One simple approach is to take the majority vote, however, due to some potential correlations between selected factors, majority voting might result in the "double counting" issue [37]. Therefore, we use Snorkel [38] to integrate the outputs of our labeling functions. To produce a probabilistic label, Snorkel learns a generative model (over labeling functions) that (i) models the correlations between labeling functions, and (ii) estimates their accuracy (through examining the overlaps/conflicts in their output) during learning [38]. *Please note that this step is done in an unsupervised manner and Snorkel does not utilize any ground-truth labels for integrating the outputs of labeling functions*. In section 6.3, we evaluate the gain of employing Snorkel rather than the majority voting approach.

Although the proposed labeling approach is suitable (i.e., fast, easy to compute, and inexpensive) for creating a labeled training set from a large unlabeled dataset, it has two weaknesses that limit its usability for identifying eviction filing hotspots directly: (i) It does not necessarily label all data points (i.e., census tracts) because the underlying labeling functions refrain from labeling a data point if they are highly uncertain (i.e., the total coverage is associated with the value of k). (ii) The algorithm relies on the ACS data, which is released with a delay of about two years, and hence, it cannot be used for monitoring the most recent situation as the input of the model is not available for the past two years. Next, we propose a neural network model that can label all data points using satellite imagery, which is available at a high temporal resolution.

5.2 The Hotspot Prediction Model

In this section, first, we explain our rationale for choosing satellite imagery as input. Then, we describe the architecture of our proposed neural network model that aims at identifying top-$k\%$ eviction filing hotposts from satellite imagery.

Rationale for the Use of Satellite Imagery. We choose satellite imagery as the input of our model mainly because of three reasons: (i) Past literature [24, 25] has shown that urban poverty can be identified using satellite imagery (e.g., urban trees provide useful signals for identifying income inequalities, and distinguishing poor neighborhoods from the rich ones [30, 49]), and given the strong association between eviction and poverty, we believe that satellite imagery would be a suitable source of data for identifying eviction filing hotspots at the census tract level as well. (ii) We hope that our neural network model can identify signs of gentrification, which has been shown to be associated with eviction, from satellite imagery [7, 29, 31]. (iii) Satellite imagery is available at high spatial (~10 meters) and temporal (~10 days) resolutions, which makes it an appropriate source of data for monitoring eviction filing hotspots in a timely manner.

The Neural Network Model. We now describe the architecture of our hotspot prediction model (the component on the left-hand side of Figure 2). The model takes the satellite images of a census tract and its county (as well as the mask matrices) as input and predicts whether that census tract is among the top-$k\%$ hotposts of that county or not. This neural network model extends the idea behind the ResNet model [15], while considering the challenges and characteristics of this problem domain. In the following paragraphs, we elaborate on these challenges and how they are addressed in this model.

The first challenge is that different census tracts have various shapes and sizes, and hence, input images can have various sizes. To address this challenge, we take the following steps: (i) we set the width (and height) of each satellite image to the third quartile of the width (and height) of all satellite images, (ii) we build a mask matrix for each image to distinguish valid pixels from invalid ones, and (iii) we incorporate *Partial Convolutional Layer* [26, 27] into our instance of ResNet (*Partial ResNet)* to make sure that the result of convolution in each layer only depends on the valid pixels. *Partial ResNet* mainly consists of three residual blocks whose parameters are the same as the first three residual blocks in ResNet-18. Please note that the kernel initializer for all partial convolutional layers is set to *he_normal* [14].

Furthermore, as the hotspots are defined with respect to each county, the neural network should consider the characteristics of

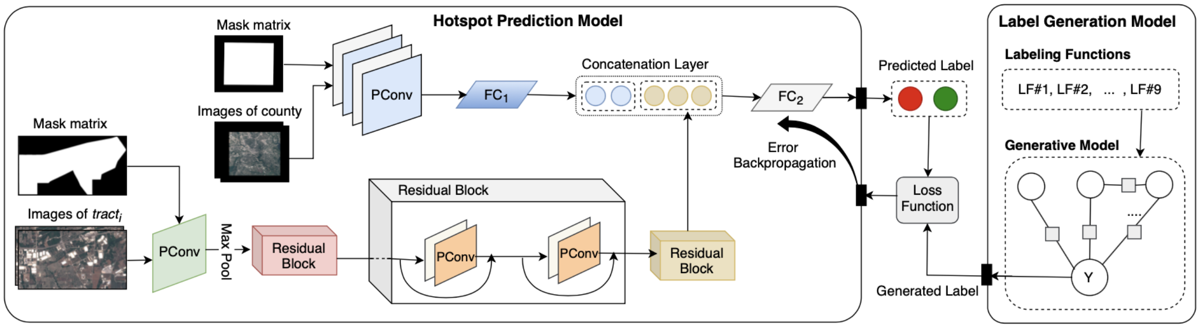


Figure 2: The architecture of WARNER.

that county in the prediction process. To this end, we employ the feature concatenation approach [1]; i.e., we apply a CNN model (i.e., CNN_{county}) on the satellite images of a county and concatenate extracted features (i.e., the output of FC_1) with the output of our *Partial ResNet* model. CNN_{county} applies a partial convolutional layer with 4 filters on each input image of a county, concatenates their outputs, and then, employs three partial convolutional layers with 16, 32, and 64 filters (respectively). The kernel size for all partial convolutional layers in CNN_{county} is set to (7×7). Also, FC_1 and FC_2 (represented in Figure 2) are fully-connected layers with 64 and 128 neurons (respectively) and the *ReLU* activation function.

Finally, since the generated probabilistic labels could be noisy, we consider two loss functions in our experiments: (i) the binary cross-entropy loss function (Equation 1), which is commonly used under a small noise rate, and (ii) the *Active Passive Loss (APL)* (Equation 2), which has been shown to be highly effective under a large noise rate [28]. *APL* is defined as the sum of Normalized Cross Entropy (the left term in Equation 2) and Reverse Cross Entropy [45] (the right term in Equation 2). In the following equations, p refers to the output probability of the neural network classifier and q denotes the ground-truth.

$$-\sum_{K=0}^{1}(q(k|x) \times \log p(k|x)) \quad (1)$$

$$\frac{-\sum_{K=0}^{1} q(k|x)\log p(k|x)}{-\sum_{j=0}^{1}\sum_{K=0}^{1} q(y=j|x)\log p(k|x)} - \sum_{K=0}^{1}(p(k|x) \times \log q(k|x)) \quad (2)$$

6 EXPERIMENTS

In this section, we first describe the set-up and data preparation process. Then, we provide an empirical evaluation of the performance of our label generation approach. Finally, we conduct a comparison between the accuracy of WARNER and various baselines and assess the contributions of various components of WARNER on its performance.

6.1 Experimental Set-up

We implemented our codes in Python and used the following packages/libraries: keras (v. 2.7.0), tensorflow (v. 2.7.0), pandas (v. 1.1.5), numpy (v. 1.19.5), and scikit-learn (v. 1.0.2).

In our experiments, we utilize Adam [20] with a learning rate of 2×10^{-4}, β_1 of 0.9, and β_2 of 0.999 as the optimizer for training the neural network models. Also, the maximum number of epochs is 100 and the early stopping technique [35] is used to stop the training process once the loss value on the validation set does not degrade after ten epochs. Our code and further hyper-parameters are available at https://github.com/maryam-tabar/WARNER.

6.2 Data Preparation

We now explain our data preparation process. For each census tract (and county), we consider the median of the three least cloudy satellite images collected from the beginning of June to the end of July of a year[7] as the satellite image of that census tract (and county) in that year. Then, we convert the value of each pixel into the range of $[0, 1]$.

Additionally, to prepare the eviction filing records of Dallas county (for which we have ground-truth labels), we take three main steps. First, following [11, 41], we exclude the eviction cases filed against business defendants and remove duplicates. Then, for each census tract, we calculate the number of eviction records in each year (Dallas has 529 census tracts). Finally, the data is split into the ratio of 60:20:20 while keeping the class distribution among training, validation, and test sets.

6.3 Evaluation of Generated Labels

In this section, we evaluate the accuracy of the generated labels (under various conditions) by comparing them to the ground-truth labels available for Dallas county, TX. Table 2 compares the performance of our label generation approach against the majority voting technique with different choices of k ($k \in \{5, 10, 15\}$) and training regions. We make the best performance bold and report the percentage of increase (in the predictive performance) achieved by employing WARNER (in the best case) compared to the majority voting approach in the last row (i.e., Gain). Please note that all performance metrics are computed on the subset of test set (i.e., the testing portion of Dallas data) labeled by all models. In our experiments, the set of data points labeled by our model is a super-set of the set of data points labeled by majority voting.

[7] We use the satellite images taken in summer because, over that period of time, the climate condition seems to be suitable across the USA for taking clear images.

Table 2: A comparison between the performance of our label generation model and majority voting.

Model	Training Region	$k=5$ F1	$k=5$ AUC	$k=10$ F1	$k=10$ AUC	$k=15$ F1	$k=15$ AUC	Avg. ($k \in \{5,10,15\}$) F1	Avg. ($k \in \{5,10,15\}$) AUC
Majority Vote	–	0.307	0.690	0.538	0.809	0.400	0.634	0.415	0.711
WARNER	Dallas County, TX	0.307	0.851	**0.666**	0.914	0.500	0.766	0.491	0.843
WARNER	Other counties in TX	**0.666**	**0.886**	0.533	**0.943**	**0.533**	**0.771**	**0.577**	**0.866**
	Gain (%)	116.9%	28.4%	23.7%	16.5%	33.2%	21.6%	39.0%	21.8%

According to the results, on average, the majority voting approach (which does not involve learning an ML-based model) achieves an AUC of 0.711, which could be an indicator of the good quality of our labeling functions. Further, in general, employing Snorkel leads to a considerable improvement against majority voting. In fact, on average, our model outperforms the majority voting approach by 21.8% in terms of AUC, which shows the value of employing Snorkel (compared to taking the majority vote) for integrating outputs of our labeling functions.

Additionally, we do a cross-region test to investigate the generalizability of our label generation approach; i.e., we train our model on the unlabeled data of other counties in TX (i.e., all Texas counties except Dallas County), and then, evaluate its performance on the testing portion of the Dallas data. As a result, it achieves an AUC of 0.866 (on average), which is higher than the average AUC of the model being trained on the training portion of Dallas data (i.e., 0.843). This shows that our label generation approach could be generalized to different counties within USA.

Finally, in spite of the high accuracy with different choices of k, the coverage of this label generation approach can change with the value of k (as the coverage of the underlying labeling functions changes with k); e.g., in total, about 77%, 59%, and 35% of data points are labeled by at least one labeling function when k is equal to 15, 10, and 5, respectively. However, a low coverage does not result in a serious issue in our problem domain because unlabeled data can be collected easily.

6.4 Evaluation of the Hotspot Prediction Model

We conduct three sets of experiments to assess the effectiveness of WARNER for the task of top-k% hotspot prediction. First, we conduct a comparison between the accuracy of WARNER and several strong deep learning-based baseline models. Then, we investigate the impact of WARNER's components on the value of different performance metrics. Finally, we evaluate the potential of WARNER trained for a specific k (e.g., $k = 10$) to be generalized (easily) to other values of k (e.g., $k \in \{5, 15\}$).

Comparison with Baseline Models. In this set of experiments, we consider the following three baseline models: (i) A Convolutional Neural Network (CNN) model with four convolutional layers, (ii) Partial-CNN that incorporates partial convolutional layers [26] into the CNN model, instead of the standard convolutional layer, and (iii) ResNet-18 [15] which is a residual network with 18 layers. While CNN and ResNet-18 take the masked satellite images as input, Partial-CNN takes satellite images and mask matrices as separate inputs as it can distinguish valid and invalid pixels. Further, the binary cross-entropy loss function is utilized for training all neural models and evaluated on the testing portion of the Dallas data (in the next section, we compare the effectiveness of cross-entropy with that of APL in our problem domain).

Table 3 shows the performance of WARNER and the aforementioned baselines for $k \in \{5, 10, 15\}$. The first three rows show the performance of baseline models being trained on the ground-truth data of Dallas County in a fully-supervised manner. In addition, the fourth and fifth rows represent the performance of Partial-CNN and WARNER (respectively) being trained on the labels that our label generation approach produced for the data of other counties in Texas. According to the results, WARNER outperforms the best-performing fully-supervised model (i.e., Partial-CNN) by 36.0% and 1.4% (on average) in terms of F1 and AUC, respectively. Therefore, although WARNER has not seen any data from Dallas County during the training phase, it works better than the best-performing fully-supervised baseline model trained on the training portion of the Dallas data, which has ground-truth labels.

Further, we observe higher improvements when training WARNER and the best-performing baseline model (i.e., Partial-CNN) on the same training dataset. In fact, the results of training both WARNER and Partial-CNN on the data of other counties in TX (with generated labels) show that WARNER outperform Partial-CNN by 86.8% and 8.7% (on average) in terms of F1 and AUC, respectively. Further, the performance of Partial-CNN decreases significantly (by 27.2% and 6.7% in F1 and AUC, respectively) when being trained on the data of other counties and evaluated on the testing portion of Dallas data. This observation could show the impact of considering the satellite imagery of a county in the decision-making process because different characteristics of various counties could mislead the network.

Additionally, comparing the accuracy of fully-supervised baselines, we see that incorporating partial convolutional layers into CNN improves F1 and AUC by 37.3% and 7.9% (on average), respectively. Also, employing residual learning with a deeper network (i.e., ResNet-18) leads to 2.0% and 6.9% improvement (on average) in terms of F1 and AUC, respectively.

Finally, we note that almost all deep learning-based models have an AUC of over 0.6 in the task of predicting top-5% hotspots (with significantly imbalanced dataset), which could show the models' capability in distinguishing positive samples from the negative ones. However, achieving a high F1 (which is calculated using the threshold of 0.5 on the predicted probability of belonging to the

Table 3: An evaluation of the performance of WARNER and baseline models.

Model	Training Region	$k=5$ F1	$k=5$ AUC	$k=10$ F1	$k=10$ AUC	$k=15$ F1	$k=15$ AUC	Avg. ($k \in \{5, 10, 15\}$) F1	Avg. ($k \in \{5, 10, 15\}$) AUC
CNN	Dallas County, TX	0.000	0.631	0.162	0.544	0.136	0.560	0.099	0.578
ResNet-18	Dallas County, TX	0.000	0.632	0.166	0.588	0.138	0.634	0.101	0.618
Partial-CNN	Dallas County, TX	0.000	0.639	0.208	0.596	0.200	**0.639**	0.136	0.624
Partial-CNN	Other counties in TX	0.000	0.580	0.117	0.528	0.181	0.638	0.099	0.582
WARNER	Other counties in TX	**0.083**	**0.650**	**0.222**	**0.644**	**0.250**	0.607	**0.185**	**0.633**

positive class) is an extremely difficult task in this case and we plan to address that in our future research.

Ablation Study. We now investigate the effect of various components on the overall accuracy of WARNER. Table 4 shows the outcome of ablation study on WARNER while assuming $k = 10$. According to the results, replacing partial convolutional layers with the standard convolutional layers (i.e., `WARNER-w-Conv`) results in 20.2% and 9.0% decrease in F1 and AUC, respectively. This observation suggests that simply masking the invalid pixels could pose significant challenges to the learning process of our neural network.

Further, we compare the suitability of feature concatenation approach with that of a recent condition approach called *Feature-wise Linear Modulation (FiLM)* method [34]. In fact, three common approaches have been usually used for incorporating multiple signals into a model [1]: input concatenation, feature concatenation, and conditioning layer. As the size of county images differs a lot from the size of census tract images, input concatenation is not an appropriate choice in our case. However, we tried a conditioning layer approach (i.e., `WARNER-w-FiLMLayer`) called *Feature-wise Linear Modulation (FiLM)* [34]. `WARNER-w-FiLMLayer` takes the following steps: (i) for the i^{th} census tract, it extracts two sets of features (i.e., $a^j_{i,m}$ and $b^j_{i,m}$) from the output of FC_1 through applying linear layers, and then (ii) use them to influence the m^{th} feature map of the j^{th} convolution layer ($f^j_{i,m}$) in the Partial ResNet via the feature-wise affine transformation given in Equation 3 [34]. According to the results, in our problem domain, employing FiLM leads to 22.0% and 14.5% decrease in F1 and AUC, respectively. Thus, the use of a concatenation layer seems to be a more appropriate choice.

$$FiLM(f^j_{i,m}|a^j_{i,m}, b^j_{i,m}) = a^j_{i,m} \times f^j_{i,m} + b^j_{i,m} \tag{3}$$

Additionally, comparing the performance of `WARNER` with that of `WARNER-w-APLLoss` shows that incorporating either of the two mentioned loss functions, i.e., cross-entropy and APL, results in a similar predictive performance. Finally, to evaluate the value of our label generation approach in the WARNER framework, we replace it with a naive label generation approach and train our neural network model on this newly labeled data. This naive approach works as follows: A census tract is among top-k% hotspot if it shows up among the top-k% census tracts of its county in terms of the total population. The experimental results show that, the use of this naive label generation approach, i.e., `WARNER-w-NaiveLabel`, leads to 31.0% and 14.1% decrease in F1 and AUC, respectively. This observation shows the key role of our label generation approach in the WARNER's architecture for building a more accurate eviction filing hotspot prediction model.

Table 4: The results of ablation study when $k = 10$.

Model	F1	AUC	Drop in AUC (%)
WARNER	0.222	0.644	–
WARNER-w-Conv	0.177	0.586	-9.006%
WARNER-w-FiLMLayer	0.173	0.550	-14.596%
WARNER-w-APLLoss	0.226	0.643	-0.001%
WARNER-w-NaiveLabel	0.153	0.553	-14.130%

Generalizability of WARNER to Various Values of k. In our previous experiments, we trained a separate model for each value of k because the task changes with the value of k; i.e., `if` $k_1 \neq k_2 \rightarrow p(y_{k=k_1}|x) \neq p(y_{k=k_2}|x)$ (in this formula, y_k shows the binary label of interest for different values of k). However, training a separate model for each k of interest could be time-consuming. To tackle this challenge, we propose to use a transfer learning approach [50] to be able to easily transfer knowledge from a single pre-trained WARNER to the target task. The transfer learning algorithm works as follows: First, we train a single model for a specific value of k. Then, we freeze all weight matrices, except the parameters of the last two layers (as a result of freezing, the knowledge can be transferred to the target task). Finally, we fine-tune the parameters of the last two layers using the training data of the target task.

Table 5 represents the results of fine-tuning a WARNER model trained with $k = 10$ (i.e., source task) to the target tasks of top-5% hotspot prediction and top-15% hotspot prediction (please note that the results of training a separate WARNER for each target task are given in parenthesis). This table shows that employing the aforementioned transfer learning approach leads to comparable results. Thus, we can easily transfer knowledge from a pre-trained WARNER model (trained with $k = 10$, for example) to the task of top-k′ hotspot prediction with various values of k′ (e.g., $k' \in \{5, 15\}$ in our experiments).

7 REAL-WORLD USAGE OF WARNER

One possible use case for WARNER is to help NGOs, policymakers, and/or federal agencies improve their resource allocation plans to enhance housing stability. In fact, various rental assistance programs (like Emergency Rental Assistance Program (ERAP) [33]) are being implemented to assist high-need renters. While being

Table 5: An evaluation of the generalizability of a pre-trained WARNER (with $k = 10$) to the task of top-k′ hotspot prediction ($k' \in \{5, 15\}$).

k'	F1 (original)	AUC (original)
5	0.000 (0.083)	0.676 (0.650)
15	0.181 (0.250)	0.590 (0.607)

available nationwide, a big variability has been observed in the utilization of those financial resources across the USA; e.g., some regions have sent back part of the funding to the government as it was too much money for those regions [32]. One important reason for this issue could be the limited understanding of the distribution of eviction filings at the local level, which, in turn, results from some infrastructure/resource limitations that hinder the collection of eviction filing records across the USA. As a result, in the absence of ground-truth eviction filing data, WARNER could serve as an ML-based assistant for informing actions across the USA. In other words, WARNER's predictions can be visualized as a map of hotspots similar to Figure 3. Figure 3 represents WARNER's prediction of the top 10% hotspots (shown in red) across Texas over a period of three years (from 2017 to 2019). Such a heatmap could assist NGOs and policymakers in (i) monitoring eviction filing hotspots over a period of time, (ii) identifying high-need areas, particularly hidden hotspots (i.e., actual hotspots from where no/limited reports have been received due to the aforementioned obstacles to data acquisition), and hence, (iii) distribute resources and funding more efficiently.

As an step toward making a broader impact in the real world, we also reached out to domain experts working for mitigating housing problems in southeast Texas, where robust court records are not accessible at the census tract level due to the existing obstacles to data acquisition. More specifically, we reached out to subject matter experts at Texas Housers[8], which is a non-profit organization attempting to advance low-income housing policies and tackling housing issues. Regarding the potential benefits of eviction data tools for mitigating the eviction crisis, Ben Martin, who is a senior research analyst (at Texas Housers), mentioned that:

> "Eviction data tools are like the smoke that help us find the fire, and once we find the fire we can figure out what tools and resources to use to mitigate the problem."

He also described the importance of monitoring the eviction-related situation when eviction data is out of reach. Further, he provided more details on the contributions of such tools to improving the existing eviction diversion programs and related policies as follows.

[8]https://texashousers.org

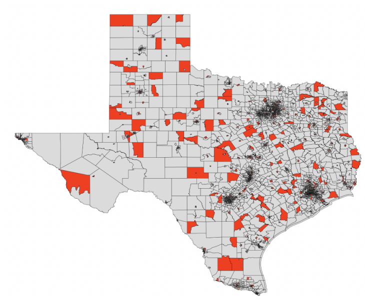


Figure 3: The WARNER's prediction regarding the top-10% eviction filing hotspots over a period of three years (from 2017 to 2019) across Texas. Hotspots and non-hotspots are shown with the red and gray colors, respectively.

> "Eviction data tools can contribute to these programs by, for instance, setting a baseline of need. Or, for a statewide ERA program, eviction data tools could help administrators identify areas of high need for targeted outreach."

The results of this discussion confirms (1) the significance such ML-based tools for improving the existing eviction mitigation plans in the absence of eviction filing records (across a large region), and hence, (2) their high potential for making significant social impacts in the field. Moreover, through this network of domain experts, we hope that we can assess the effectiveness of the proposed methodology in the field in the near future.

8 CONCLUSION

This paper proposed WARNER, which is a weakly-supervised ML-based framework for identifying eviction filing hotspots in US counties from satellite imagery in the absence of court records. In fact, first, we proposed a label generation approach that leverages sociological insights on the eviction crisis to label an unlabeled training dataset of satellite imagery. Then, relying on those generated labels, we built a neural network model for predicting eviction filing hospots from satellite imagery. To assess the performance of WARNER, we conducted various experiments using eviction filing data of Dallas County, TX. The experimental results show the suitability of the proposed label generation approach for this problem domain. Furthermore, WARNER outperforms multiple strong (fully-supervised) baseline models and its superior accuracy could be generalized to various counties within the US. In the absence of eviction filing records, the data-driven insights produced by WARNER could assist policymakers in distributing resources more efficiently and improving eviction mitigation programs. In this work, we collaborated with CPAL and it is being assessed by domain experts.

ACKNOWLEDGMENTS

This work was partly supported by NSF awards #1940076 and #1934782 and the Bill and Melinda Gates Foundation Grant #141840.

REFERENCES

[1] Badour AlBahar and Jia-Bin Huang. 2019. Guided image-to-image translation with bi-directional feature transformation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 9016–9025.

[2] Kumar Ayush, Burak Uzkent, Marshall Burke, David Lobell, and Stefano Ermon. 2020. Generating Interpretable Poverty Maps using Object Detection in Satellite Images. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*. 4410–4416.

[3] Kumar Ayush, Burak Uzkent, Kumar Tanmay, Marshall Burke, David Lobell, and Stefano Ermon. 2021. Efficient Poverty Mapping from High Resolution Remote Sensing Images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 12–20.

[4] Ellen L Bassuk, John C Buckner, Linda F Weinreb, Angela Browne, Shari S Bassuk, Ree Dawson, and Jennifer N Perloff. 1997. Homelessness in female-headed families: childhood and adult risk and protective factors. *American journal of public health* 87, 2 (1997), 241–248.

[5] Emily A Benfer, Solomon J Greene, and Margaret Hagan. 2020. Approaches to eviction prevention. *Available at SSRN 3662736* (2020).

[6] Brian Bieretz, Kimberly Burrowes, and Emily Bramhall. 2020. *Getting Landlords and Tenants to Talk: The Use of Mediation in Eviction.* Retrieved November 2021 from https://www.urban.org/sites/default/files/publication/101991/getting-landlords-and-tenants-to-talk_3.pdf

[7] Antony Chum. 2015. The impact of gentrification on residential evictions. *Urban Geography* 36, 7 (2015), 1083–1098.

[8] Matthew Desmond. 2012. Eviction and the reproduction of urban poverty. *American journal of sociology* 118, 1 (2012), 88–133.

[9] Matthew Desmond. 2015. Unaffordable America: Poverty, housing, and eviction. *Fast Focus: Institute for Research on Poverty* 22, 22 (2015), 1–6.

[10] Matthew Desmond and Carl Gershenson. 2017. Who gets evicted? Assessing individual, neighborhood, and network factors. *Social science research* 62 (2017), 362–377.

[11] Matthew Desmond, Ashley Gromis, Lavar Edmonds, James Hendrickson, Katie Krywokulski, Lillian Leung, and Adam Porton. 2018. *Eviction Lab Methodology Report: Version 1.0.* Technical Report. Princeton: Princeton University, Princeton, NJ. www.evictionlab.org/methods.

[12] Matthew Desmond and Rachel Kimbro. 2015. Eviction's Fallout: Housing, Hardship, and Health. *Social Forces* 94 (02 2015).

[13] Frederick J. Eggers, Fouad Moumen, and Inc. Econometrica. 2010. Investigating Very High Rent Burdens Among Renters in the American Housing Survey. *U.S. Department of Housing and Urban Development, Washington, DC* 94 (2010).

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*. 1026–1034.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[16] Andrew Head, Mélanie Manguin, Nhat Tran, and Joshua E Blumenstock. 2017. Can Human Development be Measured with Satellite Imagery?. In *Proceedings of the Ninth International Conference on Information and Communication Technologies and Development*. 1–11.

[17] Gracie Himmelstein and Matthew Desmond. 2021. Association of eviction with adverse birth outcomes among women in Georgia, 2000 to 2016. *JAMA pediatrics* 175, 5 (2021), 494–500.

[18] Wenjie Hu, Jay Harshadbhai Patel, Zoe-Alanah Robert, Paul Novosad, Samuel Asher, Zhongyi Tang, Marshall Burke, David Lobell, and Stefano Ermon. 2019. Mapping missing population in rural India: A deep learning approach with satellite imagery. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. 353–359.

[19] Neal Jean, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell, and Stefano Ermon. 2016. Combining satellite imagery and machine learning to predict poverty. *Science* 353, 6301 (2016), 790–794.

[20] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[21] The Eviction Lab. 2018. *The Eviction Lab.* Retrieved November 2021 from https://evictionlab.org/

[22] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. 1989. Backpropagation Applied to Handwritten Zip Code Recognition. 1, 4 (1989), 541–551.

[23] Jihyeon Lee, Dylan Grosz, Burak Uzkent, Sicheng Zeng, Marshall Burke, David Lobell, and Stefano Ermon. 2021. Predicting Livelihood Indicators from Community-Generated Street-Level Imagery. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 268–276.

[24] Guie Li, Zhongliang Cai, Yun Qian, and Fei Chen. 2021. Identifying Urban Poverty Using High-Resolution Satellite Imagery and Machine Learning Approaches: Implications for Housing Inequality. *Land* 10, 6 (2021), 648.

[25] Li Lin, Liping Di, Chen Zhang, Liying Guo, and Yahui Di. 2021. Remote Sensing of Urban Poverty and Gentrification. *Remote Sensing* 13, 20 (2021), 4022.

[26] Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. 2018. Image inpainting for irregular holes using partial convolutions. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 85–100.

[27] Guilin Liu, Kevin J. Shih, Ting-Chun Wang, Fitsum A. Reda, Karan Sapra, Zhiding Yu, Andrew Tao, and Bryan Catanzaro. 2018. Partial Convolution based Padding. In *arXiv preprint arXiv:1811.11718*.

[28] Xingjun Ma, Hanxun Huang, Yisen Wang, Simone Romano, Sarah Erfani, and James Bailey. 2020. Normalized loss functions for deep learning with noisy labels. In *International Conference on Machine Learning*. PMLR, 6543–6553.

[29] Peter Marcuse. 2013. Abandonment, gentrification, and displacement: the linkages in New York City. In *Gentrification of the City*. Routledge, 169–193.

[30] Per Square Mile. 2012. *Urban trees reveal income inequality.* Retrieved November 2021 from https://persquaremile.com/2012/05/17/urban-trees-reveal-income-inequality/

[31] Kathe Newman and Elvin K Wyly. 2006. The right to stay put, revisited: Gentrification and resistance to displacement in New York City. *Urban studies* 43, 1 (2006), 23–57.

[32] PBS NewsHour. 2022. *States clash over rental assistance as the federal government reallocates funds.* Retrieved August 2022 from https://www.pbs.org/newshour/nation/states-clash-over-rental-assistance-as-the-federal-government-reallocates-funds

[33] U.S. Department of the Treasury. 2021. *Emergency Rental Assistance Program.* Retrieved November 2021 from https://home.treasury.gov/policy-issues/coronavirus/assistance-for-state-local-and-tribal-governments/emergency-rental-assistance-program

[34] Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. 2018. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.

[35] Lutz Prechelt. 1998. Early stopping-but when? In *Neural Networks: Tricks of the trade*. Springer, 55–69.

[36] Alan Puckett, Lynette M. Renner, and Kristen Shook Slack. 2002. *Trends in homelessness & housing insecurity.* Technical Report. Northwestern University.

[37] Alexander Ratner, Stephen H Bach, Henry Ehrenberg, Jason Fries, Sen Wu, and Christopher Ré. 2020. Snorkel: Rapid training data creation with weak supervision. *The VLDB Journal* 29, 2 (2020), 709–730.

[38] Alexander J Ratner, Stephen H Bach, Henry R Ehrenberg, and Chris Ré. 2017. Snorkel: Fast training set generation for information extraction. In *Proceedings of the 2017 ACM international conference on management of data*. 1683–1686.

[39] Devin Q Rutan and Matthew Desmond. 2021. The concentrated geography of eviction. *The ANNALS of the American Academy of Political and Social Science* 693, 1 (2021), 64–81.

[40] Sten-Åke Stenberg, Lars Brännström, Christine Lindberg, and Ylva B Almquist. 2020. Risk Factors for Housing Evictions: Evidence from Panel Data. *European Journal of Homelessness _ Volume* 14, 2_ (2020).

[41] Maryam Tabar, Wooyong Jung, Amulya Yadav, Owen Wilson Chavez, Ashley Flores, and Dongwon Lee. 2022. Forecasting the Number of Tenants At-Risk of Formal Eviction: A Machine Learning Approach to Inform Public Policy. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*.

[42] Jialu Tan. 2020. *Using machine learning to identify populations at high risk for eviction as an indicator of homelessness.* Master's thesis. Massachusetts Institute of Technology, Cambridge, MA.

[43] Mark Treskon, Solomon Greene, Olivia Fiol, and Anne Junod. 2021. Eviction Prevention and Diversion Programs. (2021).

[44] Igor R van Laere, Matty A de Wit, and Niek S Klazinga. 2009. Pathways into homelessness: recently homeless adults problems and service use before and after becoming homeless in Amsterdam. *BMC public health* 9, 1 (2009), 1–9.

[45] Yisen Wang, Xingjun Ma, Zaiyi Chen, Yuan Luo, Jinfeng Yi, and James Bailey. 2019. Symmetric cross entropy for robust learning with noisy labels. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 322–330.

[46] Michael Xie, Neal Jean, Marshall Burke, David Lobell, and Stefano Ermon. 2016. Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*. 3929–3935.

[47] Teng Ye, Rebecca Johnson, Samantha Fu, Jerica Copeny, Bridgit Donnelly, Alex Freeman, Mirian Lima, Joe Walsh, and Rayid Ghani. 2019. Using machine learning to help vulnerable tenants in new york city. In *Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies*. 248–258.

[48] Christopher Yeh, Anthony Perez, Anne Driscoll, George Azzari, Zhongyi Tang, David Lobell, Stefano Ermon, and Marshall Burke. 2020. Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature communications* 11, 1 (2020), 1–11.

[49] Pengyu Zhu and Yaoqi Zhang. 2008. Demand for urban forests in United States cities. *Landscape and urban planning* 84, 3-4 (2008), 293–300.

[50] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. 2020. A comprehensive survey on transfer learning. *Proc. IEEE* 109, 1 (2020), 43–76.