2024 IEEE 6th International Conference on Cognitive Machine Intelligence (CogMI) | 979-8-3503-8672-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/CogMI62246.2024.00020

Data-Driven Vulnerable Community Identification During Compound Disasters

Jainil Anilkumar Patel*, Mohammadreza Akbari Lor*, Shu-Ching Chen†, Mei-Ling Shyu*, Steven Luis‡
*School of Science and Engineering, University of Missouri-Kansas City, Kansas City, MO 64110, USA
†Data Science and Analytics Innovation Center (dSAIC), University of Missouri-Kansas City, Kansas City, MO 64110, USA
‡Knight Foundation School of Computing and Information Sciences, Florida International University, Miami, FL 33199, USA
Email: {japmyy, ma7fy, s.chen, shyum}@umkc.edu; luiss@fiu.edu

***Abstract*—This paper aims to introduce research directions to identify vulnerable communities during compound disasters, such as the intersection of hurricanes and pandemics like COVID-19. We suggest integrating mobility, socioeconomic, and geographic factors to develop a comprehensive vulnerability index. By analyzing mobility data, U.S. Census socioeconomic data, disaster data, pandemic data, and geographic data, we extract information like hospital visits, business closures, increased travel distances, decreased inbound movement to commercial census block groups (CBGs), decreased outbound from home CBGs, and changes in mobility trends to various categories of stores, to develop a vulnerability index that integrates all of these factors. This index aims to capture and analyze the unique impacts of natural disasters on communities, especially those exacerbated by simultaneous disaster scenarios. We also propose a model design that can capture the spatial and temporal nature of the data and can be trained to perform the dual task of vulnerability index calculation in a disaster scenario along with predicting pandemic growth features in case of a compound disaster.**

***Keywords*— Machine Learning, Disaster Response, SafeGraph, COVID-19, Compound Disasters, Community Support.**

I. Introduction

Accurately identifying vulnerable communities during compound disasters is critical for effective disaster response and resource allocation. Traditional approaches often need to catch up, relying heavily on news data, calls to emergency services, and hospital admits data, which can lead to delays and inaccuracies in identifying those who need more help. We used real-time mobility data, particularly from SafeGraph [1], along with socioeconomic indicators, to pinpoint communities at more risk.

Our aim is to integrate mobility patterns and socioeconomic factors to identify real-time changes in community behavior and needs. For instance, during disasters such as the COVID-19 pandemic compounded with hurricanes, we observe that communities forced to travel further for essential supplies due to local store closures or gathered in shelters for safety, faced increased risks of exposure to the virus.

The primary objective of this research is to develop methodologies to identify vulnerable communities during compound disaster events, such as the concurrent occurrence of a pandemic and natural disasters. In this paper, we discuss movement pattern shifts that can suggest vulnerability changes after a disaster to create a comprehensive vulnerability index, leveraging mobility data and socioeconomic indicators along with disaster and pandemic data to analyze the compounded effects.

We propose a framework that identifies vulnerable Point Of Interest (POIs) based on the people's demographics and the number of people visiting the commercial CBGs. The framework is designed to both predict the vulnerability index of areas during a disaster and forecast the pandemic spread throughout the region due to its effect. It is meant to be capable of detecting heavy fluctuations in movement within specific areas by employing a Graph Attention Network (GATs) [2], [3] along with a Long Term Short Memory network (LSTM) [4] for the dual purpose of detection and forecasting. We show that a model designed for and trained on data of disaster timeframes can be effectively used to give indication of areas that would be effected the most by a disaster along with forecasting spread of viruses for a better preparation.

This identification enables targeted interventions and efficient allocation of resources, helping authorities to mitigate the effects of disasters more effectively [5]–[9]. Additionally, since the framework seeks to forecast how pandemics might spread through these vulnerable areas, it provides valuable insights for better preparation and response strategies. By predicting potential transmission pathways, public health officials can implement preventive measures to control the spread of diseases. Ultimately, this approach enhances disaster preparedness and public health responses by offering actionable insights, contributing to the development of more resilient and responsive communities.

The rest of this paper is organized as follows. Section 2 reviews the related work in vulnerability prediction, impact analysis for disasters, pandemic spread, and disaster preparedness. The data sources we studied and/or used, including mobility data, disaster data, socioeconomic data, COVID-19 data, and geographic data are discussed in Section 3. Section 4 delves into our methodologies for detecting venerability signals from changes in mobility patterns and how to combine them into a unified vulnerability index. Section 5 presents our model design for using the combined mobility, pandemic, and disaster data for the dual task of vulnerability prediction and pandemic growth forecasting. Finally, Section 6 provides the conclusion of this paper.

979-8-3503-8672-1/24/$31.00 ©2024 IEEE
DOI 10.1109/CogMI62246.2024.00020

II. Literature Review

The CDC/ATSDR Social Vulnerability Index (SVI) provides a comprehensive dataset that captures these socioeconomic variables across the United States, as detailed in [10]. The SVI is instrumental in identifying communities more susceptible to harm during and after disasters, allowing for more targeted and effective interventions. Further, the methodology for assembling this socioeconomic dataset is elaborated in [11]. The research outlines the processes and criteria used to select and integrate various socioeconomic indicators, ensuring that the index accurately reflects the vulnerability of different communities. Though this vulnerability index is measured per year, it is not as reliable for disasters such as hurricanes, where vulnerability is determined foremost by the path of the hurricane. However, it can be used as a feature in detecting the degrees of vulnerability in a disaster scenario.

PolicyMap [12] offers a robust mapping and visualization tool that compiles and presents various layers of socioeconomic data. These layers include demographic information, income levels, housing conditions, quality of life, education, and health indicators. Through PolicyMap, users can generate detailed dashboards that aid policymakers and researchers in identifying communities at risk. The platform's visualizations highlight areas with varying degrees of vulnerability, such as regions with high levels of homelessness, enabling targeted disaster preparedness and response strategies. However, while PolicyMap effectively identifies vulnerable communities based on static socioeconomic factors like income, age, and housing, it does not account for effects of a disaster as when a disaster strikes, communities exhibit specific shifts in mobility patterns—changes that can serve as critical indicators of their vulnerability.

The impact of dual crises, such as a volcanic eruption and the COVID-19 pandemic, on food production and security in a Small Island Developing State was systematically examined in a recent study [13]. The research focused on how these concurrent disasters affected local food systems, considering variables such as food production methods, unemployment, food insecurity, transportation, and work hours.

Another recent study [14] proposed a framework for real-time nowcasting of human mobility during disasters, aiming to enhance emergency response and resource allocation. The researchers leveraged Twitter data with geolocations to capture real-time sentiments and relate them to mobility of individuals during crises.

Location intelligence offers valuable insights into hurricane preparedness extent, timing, and spatial variation, as demonstrated in a recent study [15]. The researchers utilized SafeGraph mobility data to monitor the behavior of CBGs leading up to hurricanes, focusing on visits to essential establishments such as gas stations, grocery stores, pharmacies, and home improvement stores.

With the COVID-19 pandemic bringing in a new aspect to be considered for disaster study and management,the authors in [16] examined the impact of Hurricane Laura on COVID-19 spread, particularly in lower-income CBGs. By utilizing mobility data, hospital locations, COVID-19 case counts, and mobility to hospitals, the research demonstrated that the hurricane disproportionately affected these lower-income CBGs, leading to a higher spread of COVID-19.

Similarly, researchers investigated the intersection of hurricane preparation and COVID-19 transmission in Harris County, Texas, during Hurricane Harvey [17]. The authors combined the population movement data, COVID-19 case counts, and social vulnerability indices to model the impact of hurricane-induced actions on virus spread. The study utilized digital trace data and network analysis, offering a framework for future risk assessments in multi-hazard scenarios and providing valuable insights for disaster planning and urban resilience.

In [18], how economic policies influenced consumer spending patterns was examined. It uses data from multiple sources, including government reports, consumer surveys, and spending data from financial institutions, to look at changes in spending on various categories such as essentials (like groceries and healthcare) and non-essentials (such as entertainment and travel). [19] analyses the emergency transportation systems during disasters, which involves examining the factors that influence the extent of disruption and recovery. The study emphasizes three core components: preparedness, impact factor, and recovery period.

The compounded effects of economic and environmental disasters on mental health were studied in [20]. The research reveals that the exposure to Hurricane Harvey, combined with the income loss due to the COVID-19 pandemic, significantly increases the likelihood of higher anxiety levels among individuals. The analysis indicates a cumulative effect, where the combined impact of these events exacerbates mental health issues.

A detailed analysis of COVID-19 spread in ten major U.S. metropolitan areas using dynamic mobility networks is presented in [21]. The study reveals that a small fraction of "superspreader" POIs contributes disproportionately to the infection spread. It also highlights that reopening strategies, particularly in high-density POIs like restaurants and gyms, could significantly increase infection rates if not coupled with additional precautions. Furthermore, it underscores racial and socioeconomic disparities, predicting higher infection rates in disadvantaged neighborhoods due to higher mobility and more crowded POIs.

The concept of systemic impact, initially defined by Vugrin et al. [22], describes the difference between a targeted system performance level and the actual system performance following a disruptive event. In this study, systemic impact values were determined as the lowest percentage change value observed in each category following the landfall of Hurricane Harvey.

III. Data Collection

In our research, we utilize a variety of datasets to gather comprehensive information on mobility, socioeconomic fac-

tors, and vulnerability indices to identify vulnerable communities along with disaster and pandemic datasets. Each dataset provides valuable insights for detecting vulnerable communities and assessing the potential impact of disasters. By combining mobility data with socioeconomic indicators and vulnerability indices, we aim to create a comprehensive vulnerability index that can better predict and prepare for disasters.

A. Mobility Data

1) SafeGraph Mobility Data: The detailed mobility data from SafeGraph [1] provides insights into population density, movement patterns, and average distance traveled. This dataset includes information at the Points of Interest (POI) and Census Block Group (CBG) levels, offering granular details down to specific tracts and block groups. The granularity of this data is crucial for assessing localized movement patterns and identifying potential vulnerabilities.

2) Google Mobility Data: Google offers county-level mobility data that, while not as granular as the SafeGraph data, still offers valuable insights into regional movement trends and behaviors. This data helps provide a broader picture of mobility patterns across different counties [23].

3) U.S. Department of Transportation Mobility Data: Additional mobility data can be sourced from the U.S. Department of Transportation, which provides county-based data on changes in mobility, particularly during the COVID-19 pandemic. While lacking tract-level detail, this dataset offers another perspective on mobility changes [24].

B. Disaster Data

1) Hurricane Path Visualizer Data: The National Oceanic and Atmospheric Administration (NOAA) provides an interactive map and API access for visualizing hurricane paths. This tool is critical for understanding the historical and projected paths of hurricanes, which can impact vulnerable communities [25].

2) Hurricane Datasets from the U.S. Government: The U.S. Government Data Catalog provides comprehensive hurricane datasets compiled from various federal agencies, including NOAA, FEMA, and NASA. These datasets include historical records of hurricanes and tropical storms, detailing their paths, wind speeds, pressure, duration, and impact assessments. The data encompasses information on storm surges, rainfall, flooding events, and damage reports. Researchers and policymakers can utilize this information to analyze patterns, improve predictive models, assess risk areas, and develop mitigation strategies for hurricane-prone regions. The datasets are available in various formats such as GIS shapefiles, CSV files, and KML, facilitating the integration with mapping tools and statistical software [26].

3) Disaster Data by State: The Emergency Events Database (EM-DAT), managed by the Centre for Research on the Epidemiology of Disasters (CRED), offers detailed data on various natural and technological disasters affecting different states in the U.S [27]. This dataset includes comprehensive information on disaster types, dates, affected locations, human impacts such as fatalities and injuries, and economic losses. It is a valuable resource for analyzing historical disaster trends, assessing state-level vulnerabilities, and informing disaster preparedness and mitigation strategies.

C. Socioeconomic Data

1) FEMA Household Preparedness Data: The FEMA National Household Survey (NHS) provides valuable insights into the preparedness behaviors and attitudes of households across the United States [28]. Conducted annually, the survey collects data on topics such as emergency planning, disaster supply kits, communication plans, evacuation readiness, and awareness of local hazards. The dataset includes demographic variables like age, income, education, and region, enabling analysis of preparedness levels across different population groups and geographic areas. This information helps emergency management officials, policymakers, and community organizations identify gaps in preparedness, tailor public outreach efforts, and enhance community resilience.

2) Social Vulnerability Index (SVI): The Social Vulnerability Index (SVI) dataset [29], available from the Socioeconomic Data and Applications Center (SEDAC), provides an index that assesses the vulnerability of communities based on factors such as age, income, and marital status. The dataset includes maps in TIFF format showing the socioeconomic status of people in different regions for the years 2000, 2010, 2014, 2016, 2018, and 2020. This index is crucial for predicting the vulnerability of populations in specific areas.

3) Census Quick Facts: : The U.S. Census Bureau's QuickFacts [30] provides essential demographic, economic, and social statistics for all states, counties, cities, and towns with populations of 5,000 or more. This tool offers quick access to data on population size, age distribution, racial and ethnic composition, housing characteristics, income levels, education attainment, and business and industry statistics. These insights are crucial for understanding community profiles, identifying vulnerable populations, and planning resource allocation..

D. COVID-19 Data

1) CovidTracking.com: This source provides detailed data on hospitalizations and vaccinations during the COVID-19 pandemic, including state-by-state breakdowns and historical trends. The dataset encompasses testing numbers, confirmed cases, deaths, and recovery rates, which are essential for epidemiological studies and public health planning [31].

2) Google COVID-19 Data: Google's COVID-19 Open Data repository [32] aggregates a wide array of datasets from authoritative sources, providing up-to-date statistics on COVID-19 cases, deaths, testing, and vaccination rates worldwide. The dataset also includes mobility reports, healthcare infrastructure data, and demographic information. This comprehensive resource is valuable for researchers and policymakers aiming to understand the pandemic's progression, identify hotspots, and develop data-driven strategies for mitigation and recovery.

3) Trips During COVID-19: The Bureau of Transportation Statistics (BTS) provides detailed data on daily travel patterns during the COVID-19 pandemic through its "Daily Travel" dataset [33]. This includes information on the number of trips taken by different transportation modes such as personal vehicles, public transit, airlines, and rail—across various regions. The dataset helps analyze how travel behaviors changed in response to pandemic-related restrictions and health concerns, which is essential for understanding the pandemic's impact on mobility and planning for future public health emergencies.

E. Geographic Data

Elevation and Weather Data: The Integrated Surface Dataset (ISD) [34], accessible via the U.S. government's data catalog, provides comprehensive weather and elevation data from the weather stations across the globe. The dataset includes hourly observations of weather elements such as temperature, precipitation, wind speed and direction, and atmospheric pressure, along with elevation information for each station. This data is critical for analyzing how geographical and climatic factors influence community vulnerability to disasters, weather-related hazards, and climate change impacts. This dataset is helpful in understanding geographical factors that may influence vulnerability.

IV. Determining Vulnerability

Determining vulnerability after a disaster is a challenging task due to many factors that need to be taken into account and many different definitions and interpretations of vulnerability that can be considered. Vulnerability can be affected by or even defined as poor health, healthcare access, economic state, etc.

Here, we define vulnerability by considering what we can derive from mobility data first and foremost. Human movement can be viewed as a fundamental factor in determining the well being of groups, in this case, people in the same CGBs. We assume that as a disaster occurs, a very telling factor that can be traced is movement - whether there is a change (increase or decrease) in the movement pattern. In this section, we explain the major ways in which the movement pattern's immediate change in a disaster scenario can be used to derive a vulnerability index specific to that disaster.

A. High Visits to the Hospitals

According to [16], there is a notable correlation between high hospital visits and increased COVID-19 spread. The research indicates that the areas with higher rates of hospital visits are more likely to experience rapid COVID-19 transmission. This observation suggests that census block groups (CBGs) with frequent visits to hospitals and pharmacies are more vulnerable compared to others.

To quantify this vulnerability for the detection of CBGs that are affected the most by disasters, we use a metric that incorporates the following factors:

- **Hospital Visits**: the frequency of visits to hospitals and pharmacies by residential CBGs;
- **Baseline Visits**: the rate of hospital visits before a disaster as a baseline measure;
- **Population Density**: the population density of the area, which provides context for the volume of hospital visits relative to the population.

An index is calculated based on these factors, with CBGs exhibiting higher rates of hospital visits deemed more vulnerable. This metric helps in understanding and predicting vulnerability by linking hospital visit patterns to the affectedness by a disaster, providing a basis for targeted public health responses and resource allocation in the aftermath of a disaster.

$$\mathrm{VI} = \frac{(\mathrm{NHV}_{\mathrm{post}} - \mathrm{BHV}_{\mathrm{pre}})}{\mathrm{PD}_{\mathrm{CBG}}}$$

where:

$$\begin{aligned} \mathrm{VI} &= \text{Vulnerability Index} \\ \mathrm{NHV}_{\mathrm{post}} &= \text{New Hospital Visits per Week After a Disaster} \\ \mathrm{BHV}_{\mathrm{pre}} &= \text{Baseline Hospital Visits per Week} \\ \mathrm{PD}_{\mathrm{CBG}} &= \text{Population Density of Residential CBG} \end{aligned}$$

B. Closure of business

Inspired by the work in [35], the following methodology is proposed as a factor in determining community vulnerability:

1) Calculate the Total Number of Businesses: Determine the total number of businesses within the commercial area of interest. This provides a baseline for assessing the extent of business activity in the area.
2) Calculate the Number of Closed Businesses: Identify and count the number of businesses that have been closed within the same commercial area. This figure reflects the impact of disruptions such as economic downturns or pandemics.
3) Assess Vulnerability Based on Business Closure Percentage: Calculate the percentage of closed businesses relative to the total number of businesses. The vulnerability of the commercial area is directly proportional to the percentage of closures. A higher percentage indicates greater economic distress and, consequently, higher vulnerability.
4) Determine Vulnerability of Nearby Residential Areas: Extend the assessment to nearby residential areas (home CBGs) surrounding the commercial area. These areas are also considered vulnerable if the commercial area experiences a high rate of business closures. This is based on the assumption that economic downturns in commercial areas have ripple effects on adjacent residential communities.

This methodology (as shown in Figure 1) helps in identifying vulnerable areas by evaluating the impact of business closures and assessing the broader implications for neighboring communities.

$$\mathrm{VI}_{\mathrm{closure}} = \frac{\mathrm{BC}_{\mathrm{post}}}{\mathrm{TB}_{\mathrm{region}}}$$

where:

$VI_{closure}$ = Vulnerability Index by business closure
BC_{post} = Total number of businesses closed after disaster
TB_{region} = Total number of businesses in the region

Note: Business closure is determined if the business has no visits after the disaster for the next 4 weeks. A total number of businesses in the region includes only those businesses which have a significantly large amount of visits. This is also visualized in Figure 1.

Fig. 1: Vulnerability by Business Closures

C. Increased Travel Distance to Stores

The increased travel to stores, particularly after a disaster suggests the destruction of or inaccessibility nearby POIs. The methodology involves identifying home CBGs that must travel farther to access essential services, making them more vulnerable to compound disasters such as hurricanes and COVID-19. The methodology is visualized in Figure 2 and described as follows.

1) **Pre-disaster Data Collection:** We collect data for one month before the anticipated disaster, excluding the last week as communities prepare for the disaster. This leaves three weeks of data reflecting standard travel patterns unaffected by the disaster. Next, we analyze the pre-disaster data to identify POIs that are frequently visited by individuals traveling from more distant locations. We then designate these POIs as distant commercial POIs, indicating that they attract visitors from further away, possibly due to a lack of nearby alternatives. For each distant commercial POI identified, we track the home CBGs of the visitors. Finally, we count the occurrences to identify which common home CBGs are visiting these distant POIs.
2) **Post-disaster Data Collection and Analysis:** For after the disaster, we collect data to identify changes in travel patterns to distant commercial POIs for 4 weeks. Then the identification (the same as the last step) is repeated to determine which home CBGs are now traveling to distant commercial POIs.
3) **Vulnerability Index:** At last, we subtract the list of vulnerable home CBGs identified before the disaster from those identified after the disaster. This comparison highlights the additional home CBGs that have become vulnerable post-disaster due to increased travel distances for essential services, indicating a reduction in nearby commercial POIs.

Fig. 2: Vulnerability by Travel Distance

D. Decreased Inbound to Commercial CBGs and Decreased Outbound from Home CBGs

Similarly, the vulnerability of communities is assessed by analyzing changes in the inbound movement to commercial CBGs and outbound movement from home CBGs before and after a natural disaster, such as a hurricane. The focus is identifying CBGs that experience significant changes in mobility patterns, which may indicate issues such as road blockages or evacuations. The methodology explained below is designed to find these changes as they are shown in Figure 3.

1) **Data Collection:** First, data on the number of inbound and outbound movements for each CBG over a specified disaster period is collected. The data should cover at least four weeks before the disaster and four weeks after the disaster.
2) **Calculation of Inbound and Outbound Movements:**
   a) **Inbound to Commercial CBGs:** Sum the total weekly inbound movements to each commercial CBG. Calculate the median inbound movements for the period before the disaster (pre-disaster) and after the disaster (post-disaster).

b) **Outbound from Home CBGs:** Sum the total number of outbound weekly movements from each home CBG. Calculate the median outbound movements for the pre-disaster and post-disaster periods.

3) **Determination of Vulnerability:**

a) **Inbound Decrease to Commercial CBGs:** If the median inbound movements to a commercial CBG significantly decrease after the disaster compared to before the disaster, it may indicate road blockages or reduced accessibility to that CBG. This CBG is labeled as *vulnerable* due to potential disruptions in access to essential services.

b) **Outbound Decrease from Home CBGs:** If the median outbound movements from home CBG significantly decrease after the disaster, it may suggest limitations to residents' mobility. This home CBG is labeled as *vulnerable* due to reduced connectivity.

4) **Vulnerability Index Calculation:** The vulnerability index for each CBG is calculated based on the magnitude of changes in inbound and outbound movements.

a) The more significant decrease in inbound movements to commercial CBGs, the higher the vulnerability index, indicating more severe road blockages or access issues.

b) Similarly, the more significant decrease in outbound movements from home CBGs, the higher the vulnerability index, indicating more significant impacts from evacuations or mobility restrictions.

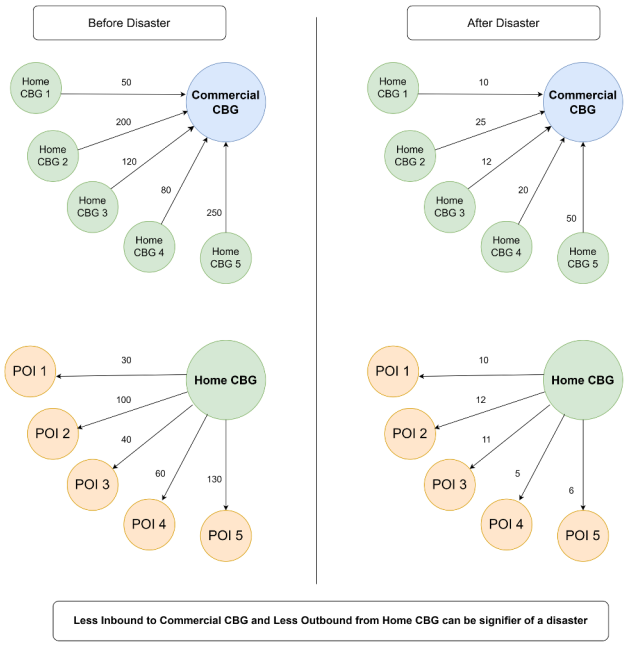

Fig. 3: vulnerability by changes in inbound and outbound

E. Change in Mobility Trends to Various Places

The change in the trend to visit various categories of stores signals about the vulnerability of the community.

- **Data Collection:** Data on mobility patterns from different CBGs and store categories is collected. This includes visit patterns to other locations before and after a compound disaster event.
- **Trend Analysis:** The outbound mobility trends to different place categories are analyzed, noting how the visit patterns change over time. Specifically, the weekly trends leading up to and following the disaster can be observed to change.
- **Matrix Construction:** The Matrix consists of the outbound from each CBG to the categories of POIs, where the POI categories are taken from the SafeGraph dataset.

$$D_{ij} = \frac{\sum_{i \in \text{POI}_j} V_i}{P_i}$$

where:

- D_{ij} is the normalized outbound visit frequency to POI category j from CBG i.
- $\sum_{i \in \text{POI}_j} V_i$ is the total visit frequency from CBG i to all points of interest (POIs) in category j.
- P_i is the population density of CBG i.

- **Distance Measurement:** We first take the median of all the weeks before the disaster and after the disaster. Then we compute the Euclidean distance between each location's pre-disaster and post-disaster vectors. This distance helps quantify the magnitude of changes in mobility patterns. The Euclidean distance d_i for each CBG i is calculated as:

$$d_i = \sqrt{\sum_j \left(D_{ij}^{\text{post}} - D_{ij}^{\text{pre}}\right)^2}$$

where:

- D_{ij}^{pre} is the normalized outbound visit frequency to POI category j from CBG i before the disaster.
- D_{ij}^{post} is the normalized outbound visit frequency to POI category j from CBG i after the disaster.

We can determine how mobility patterns shift in response to compound disasters by analyzing the distance matrix and the Euclidean distances. Large changes in the matrix values indicate significant deviations in mobility, potentially reflecting areas of increased vulnerability. The results can be used to identify those regions that are more affected by the compound disasters, providing valuable insights for emergency response and resource allocation.

F. Combining Signals Into A Unified Index

In each of the previous subsections, we calculated a measure of changes in movements before and after a disaster. For each of these, if we consider the set of results for all CBGs, we can assume that they follow an almost normal distribution. While we cannot take each of these alone as a measure of vulnerability necessarily (unless only considering the most extreme

outliers), we can consider their combination. By standardizing all of these measures, we can combine them into a unified index. This index describes an aggregated standardized change in mobility of each CBG. Therefore in a disaster scenario, the outliers can be considered as having been impacted the most by the aforementioned disaster. The process can be described as follows.

$$Z_k = \frac{x_k - \mu_k}{\sigma_k}$$

$$Z_{\text{avg}} = \frac{1}{n}\sum_{k=1}^{n} Z_k$$

where:

- Z_k is the standardized distribution of mobility changes for the kth measure. x_k is the actual distribution, μ_k and σ_k are the mean and standard deviation.

This can be used as a priority tier list. The most extreme tier includes those with the highest thresholds, are experiencing the largest change in movements across all categories, and will therefore need the most immediate attention. The threshold can be dynamically chosen to accommodate each scenario and region.

$$|Z_{\text{avg}}| > T$$

where:

- Z_{avg} is the aggregated standardized vulnerability index and T is the actual the threshold.

While this index is calculated by incorporating the data from aftermath of a disaster as well as before the disaster, we need a way to predict this index for future disasters by noticing the first changes of movements that happen in the week that the disaster hits a region. Therefore in the next section, we propose to train a model on the past and during the disaster data, in order to recognize the shift in the mobility pattern and predict the CBG level vulnerability index.

V. Model Design

The proposed model (as shown in Figure 4) is designed to perform dual tasks on spatio-temporal data of the CBGs. Given the complexity of the data, which consists of both spatial relationships among CBGs and temporal dependencies over time, the model is designed using a combination of a Graph Attention Network (GAT) and a Long Short-Term Memory (LSTM) component tailored to temporal dynamics. This architecture leverages the inherent structure of the data by processing the spatial aspect of the data first, then the temporal aspect of the processed features at each node, and finally using the acquired state to perform the dual tasks of predicting the vulnerability index and forecasting pandemic growth in the next few weeks.

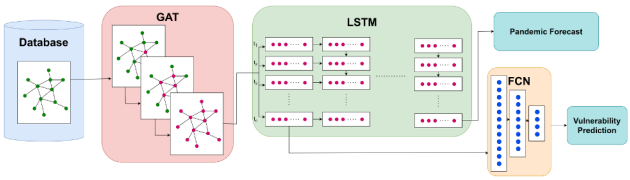

Fig. 4: Model Design

A. Spatial Component: Deep Graph Attention Network

The spatial component of the model uses a graph-based structure where each CBG is represented as a node containing the feature vector of that CBG, and the edges connecting nodes that are geologically connected, such as proximity or shared characteristics. To capture the complex spatial dependencies, we use a GAT network. The attention mechanism in a GAT allows the model to weigh the importance of neighboring CBGs dynamically, enabling it to learn which neighbors contribute most to the classification decision for a given CBG. This adaptability makes the GAT suitable for handling varying degrees of influence across CBGs, making the spatial dependencies non-uniform and context-dependent.

Formally, let the graph be represented as G = (V, E), where V is the set of nodes (CBGs) and E the set of edges between them. Each node v_i in V contains features representing the characteristics of a CBG. The GAT layer computes a node embedding h_i for each v_i, based on both its own features and the features of its neighboring nodes, weighted by the learned attention coefficients a_{ij} that reflect the importance of node v_j's contribution to node v_i.

B. Temporal Component: Long Short-Term Memory

For capturing the temporal dependencies, the model leverages an LSTM network which is well-suited for modeling sequential data and learning long-range dependencies in time series. Each CBG evolves over time, and the LSTM is tasked with learning the patterns of changes across different time steps. By utilizing the memory cells within the LSTM architecture, the model is capable of retaining important temporal information while filtering out irrelevant data, ensuring that the classification decision is informed by both short-term fluctuations and long-term trends. In our scenario, we want to capture the short term fluctuation while considering the impact and contrast it has with the previous small fluctuations happening in a larger period.

Given a sequence of temporal data X = x_1, x_2, ..., x_T, where each x_t represents the features of CBGs at time t, the LSTM processes the sequence to generate a hidden representation at each time step. The final hidden state of the LSTM captures the aggregated temporal information, representing the temporal evolution of a CBG's features over time. By using multiple layers, the model can learn the longer range dependencies better and perform the forecasting task more efficiently. Therefore, the last hidden state of the first layer's output will be used for the vulnerability index prediction and the final layer's output will be used to forecast the pandemic spread features.

C. Dual-Purpose Output

To address the dual objectives of predicting a continuous float-type index and forecasting an input feature for each CBG, the model employs separate output layers that both utilize the final hidden state from the LSTM. This approach allows the model to leverage the shared spatio-temporal representations while specializing in each task.

The prediction of the continuous float-type index $\hat{Z}_{\text{avg}}$ for each node v_i is performed by passing the final hidden state of the first layer of the LSTM network to a Fully Connected Network (FCN) with ReLU activation function between layers, so that the entirety of the feature set of the hidden state can be leveraged and distilled into a single number. The way the hidden state is passed to each input node of the FCN is described as follows.

$$\mathbf{C}_l = \mathbf{w}_y^\top \mathbf{h}_{i,T}^{(1)} + b_y \tag{1}$$

where: $\mathbf{C}_l$ is the value passed to the l-th FCN input node. $\mathbf{w}_y \in \mathbf{R}^H$ is a learnable weight vector. $b_y \in \mathbf{R}$ is a bias term. H is the dimensionality of the LSTM hidden state. $\mathbf{h}_{i,T}^{(1)}$ is the hidden state itself. This network with non-linear activations effectively maps the spatio-temporal features captured in the LSTM's hidden state to the continuous index prediction.

As mentioned before, to enhance the model's forecasting capability, we employ multiple layers in the LSTM network and utilize the final hidden state of this layer that has richer temporal information, rather than relying on the first layer's final hidden state. This allows the model to capture stronger temporal dynamics over the entire sequence, which is particularly suitable for forecasting tasks.

Let the LSTM consist of L layers. The output of the l-th LSTM layer at time step t for node v_i is denoted as $\mathbf{h}_{i,t}^{(l)}$. The final (last) LSTM layer produces a sequence of hidden states for each time step:

$$\mathbf{H}_i^{(L)} = \left[\mathbf{h}_{i,1}^{(L)}, \mathbf{h}_{i,2}^{(L)}, \ldots, \mathbf{h}_{i,T}^{(L)}\right] \tag{2}$$

For forecasting the input feature $\hat{f}_{i,T+1}$ at the next time step $T+1$, the model again leverages the final hidden state $\mathbf{h}_{i,T}^{(L)}$:

$$\hat{f}_{i,T+1} = \mathbf{w}f^\top \mathbf{h}_{i,T}^{(L)} + b_f \tag{3}$$

where:

$\mathbf{w}_f \in \mathbf{R}^H$ is a learnable weight vector. $b_f \in \mathbf{R}$ is a bias term. If multiple features are to be forecasted simultaneously, the equation generalizes to:

$$\hat{\mathbf{f}}_{i,T+1} = \mathbf{W}_f \mathbf{h}_{i,T}^{(L)} + \mathbf{b}_f \tag{4}$$

where:

$\hat{\mathbf{f}}_{i,T+1} \in \mathbf{R}^F$ is the vector of forecasted features. $\mathbf{W}_f \in \mathbf{R}^{F\times H}$ is a learnable weight matrix. $\mathbf{b}_f \in \mathbf{R}^F$ is a bias vector. F is the number of features to forecast. This output layer enables the model to predict future feature values based on learned spatio-temporal patterns.

D. Loss Function and Optimization

To train the model for both tasks simultaneously, a combined loss function is defined, capturing the errors in both the index prediction and feature forecasting.

1) Mean Squared Error (MSE) for Index Prediction: The loss for predicting the float-type index is calculated using the Mean Squared Error (MSE):

$$\mathcal{L}_{\text{index}} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{5}$$

where:

y_i is the true float-type index for node v_i. $\hat{y}_i$ is the predicted index. N is the number of nodes.

2) MSE for Feature Forecasting: Similarly, the loss for feature forecasting is computed as:

$$\mathcal{L}_{\text{forecast}} = \frac{1}{N}\sum_{i=1}^{N}\left(f_{i,T+1} - \hat{f}_{i,T+1}\right)^2 \tag{6}$$

where:

$f_{i,T+1}$ is the true feature value at time $T+1$ for node v_i. $\hat{f}_{i,T+1}$ is the forecasted feature value.

3) Total Loss Function: The total loss function combines the two losses with weighting factors λ_{index} and $\lambda_{\text{forecast}}$ to balance the importance of each task:

$$\mathcal{L} = \lambda_{\text{index}}\mathcal{L}_{\text{index}} + \lambda_{\text{forecast}}\mathcal{L}_{\text{forecast}} \tag{7}$$

where:

λ_{index} and $\lambda_{\text{forecast}}$ are hyperparameters that control the trade-off between the two tasks.

The model parameters are optimized by minimizing the total loss $\mathcal{L}$ using gradient-based optimization methods such as the Adam optimizer. During training, errors are backpropagated through the output layers, the FCN network, the LSTM network, and the GAT layers to update all learnable parameters.

E. Capabilities and Limitations

The model design suggested in this section offers several advantages that make it well-suited for the disaster effect prediction tasks:

- The unique combination of GAT and LSTM networks allows the model to effectively handle the dual tasks in a single framework by capturing complex spatio-temporal dependencies.
- Utilizing the final LSTM hidden state of different layers for dual tasks leverages different temporal features, enhancing learning efficiency and performance.
- Separate output layers for index prediction and feature forecasting enables the model to specialize in each task without mutual interference.
- The combined loss function ensures simultaneous optimization of both tasks, potentially improving overall model performance due to shared learning. Handling two tasks within a single model reduces computational

resources and training time compared to training separate models for each task.
- The ability to adjust the hyperparameters λ_{index} and $\lambda_{\text{forecast}}$ allows practitioners to control the focus on each task based on specific application needs or priorities.
- The model can be extended to forecast multiple features or adapted to different types of spatio-temporal data, demonstrating its scalability and flexibility.

By combining these strengths, the proposed model provides a robust framework for spatio-temporal prediction and forecasting, making it a valuable tool for analyzing complex datasets with spatio-temporal nature for various tasks.

However there are possible considerations and limitations to this design as well. While the model's design offers a substantial opportunity for dual-purpose training, there is a risk that performance may decline compared to a model solely focused on each task individually. The reasoning is that both tasks benefit from shared spatio-temporal features, with dedicated output layers fine-tuned for their specific goals. However, one could argue that the early layers might be better optimized for one task over the other. To fully understand the effectiveness of this dual-purpose design, an in-depth analysis and ablation study are necessary, testing various hyperparameters λ_{index} and $\lambda_{\text{forecast}}$ to assess its performance.

Another consideration is the hyperparameters of each component of the model, which could severely effect the training results and can also be a factor to consider for the effectiveness of the dual purpose output layers of the model. Further study and experimentation are required to determine to what extent the model can be pushed in terms of the size in either direction and its effect of accuracy and adaptability to larger and more diverse datasets for different disasters.

VI. Conclusion

In this paper, we introduced a comprehensive framework for identifying vulnerable communities during compound disasters by integrating mobility data, socioeconomic indicators, and advanced machine learning techniques. By leveraging real-time mobility patterns from SafeGraph data and socioeconomic factors from various sources, we developed a vulnerability index that captures the multifaceted impacts of disasters on different communities.

Our proposed model combines GATs and LSTM networks to effectively capture the spatial and temporal dependencies inherent in the data. This dual-purpose model not only predicts a continuous vulnerability index but also forecasts pandemic spread, enabling timely and informed decision-making during emergencies.

The methodology emphasizes the importance of understanding the changes in mobility trends, hospital visits, business closures, and other critical factors that contribute to community vulnerability. By standardizing and aggregating these signals into a unified index, we provide a robust tool for prioritizing resource allocation and emergency response efforts.

Future work includes validating the model with real-world data from past disasters, refining the model's predictive capabilities, and exploring the integration of additional data sources such as real-time spending data or social media analytics. The framework has the potential to significantly enhance disaster preparedness and resilience by providing actionable insights to policymakers, emergency responders, and community organizations.

Acknowledgment

This work is partially supported by the National Science Foundation (NSF) Grant Number CNS-2301552.

References

[1] M. Lissner. (2016, mar) Safegraph website. [Online]. Available: https://www.safegraph.com

[2] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph Attention Networks," *International Conference on Learning Representations*, 2018. [Online]. Available: https://openreview.net/forum?id=rJXMpikCZ

[3] S. Brody, U. Alon, and E. Yahav, "How attentive are graph attention networks?" *arXiv preprint arXiv:2105.14491*, 2021.

[4] S. Hochreiter, "Long short-term memory," *Neural Computation MIT-Press*, 1997.

[5] S.-C. Chen, M. Chen, N. Zhao, S. Hamid, K. Chatterjee, and M. Armella, "Florida public hurricane loss model: Research in multi-disciplinary system integration assisting government policy making," *Government Information Quarterly*, vol. 26, no. 2, pp. 285–294, 2009.

[6] S. Hamid, B. Golam Kibria, S. Gulati, M. Powell, B. Annane, S. Cocke, J.-P. Pinelli, K. Gurley, and S.-C. Chen, "Predicting losses of residential structures in the state of florida by the public hurricane loss evaluation model," *Statistical Methodology*, vol. 7, no. 5, pp. 552–573, 2010.

[7] K. Saleem, S. Luis, Y. Deng, S.-C. Chen, V. Hristidis, and T. Li, "Towards a business continuity information network for rapid disaster recovery," in *Proceedings of the 2008 International Conference on Digital Government Research*. Digital Government Society of North America, 2008, p. 107–116.

[8] Y. Tao, T. Wang, A. Sun, S. S. Hamid, S.-C. Chen, and M.-L. Shyu, "Florida public hurricane loss model: Software system for insurance loss projection," *Software: Practice and Experience*, vol. 52, no. 7, pp. 1736–1755, 2022.

[9] L. Zheng, C. Shen, L. Tang, C. Zeng, T. Li, S. Luis, and S.-C. Chen, "Data mining meets the needs of disaster information management," *IEEE Transactions on Human-Machine Systems*, vol. 43, no. 5, pp. 451–464, 2013.

[10] C. for Disease Control, A. f. T. S. Prevention (CDC), and D. R. (ATSDR), "Cdc/atsdr social vulnerability index (svi): Data and documentation download," https://www.atsdr.cdc.gov/placeandhealth/svi/data_documentation_download.html, 2022, accessed: 2024-08-18.

[11] B. Flanagan, E. Gregory, E. Hallisey, J. Heitgerd, and B. Lewis, "A social vulnerability index for disaster management," *Journal of Homeland Security and Emergency Management*, vol. 8, 01 2011.

[12] "Policymap: Mapping and data visualization tool," https://www.policymap.com/newmaps/#/, accessed: August 2, 2024.

[13] E. Augustus, M. M. Murphy, C. Guell, K. Morrissey, D. Ramdath, M. Woodward, S. G. Anderson, and N. Unwin, "The double burden of covid-19 and a major volcanic eruption on local food production and food security in a small island developing state," *Frontiers in Sustainable Food Systems*, vol. 7, 2023. [Online]. Available: https://www.frontiersin.org/journals/sustainable-food-systems/articles/10.3389/fsufs.2023.1268330

[14] R. Jiang, Z. Wang, Y. Tao, C. Yang, X. Song, R. Shibasaki, S.-C. Chen, and M.-L. Shyu, "Learning social meta-knowledge for nowcasting human mobility in disaster," in *Proceedings of the ACM Web Conference 2023*, ser. WWW '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 2655–2665. [Online]. Available: https://doi.org/10.1145/3543507.3583991

[15] B. Li and A. Mostafavi, "Location intelligence reveals the extent, timing, and spatial variation of hurricane preparedness," 2022. [Online]. Available: https://arxiv.org/abs/2203.06567

[16] M. Tracey, A. Plemmons, and A. Belasen, "Throwing caution to the wind: How hurricanes affect covid-19 spread," *Health Economics*, vol. 31, 03 2022.

[17] J. S. Dargin, Q. Li, G. Jawer, X. Xiao, and A. Mostafavi, "Compound hazards: An examination of how hurricane protective actions could increase transmission risk of covid-19," *International Journal of Disaster Risk Reduction*, vol. 65, p. 102560, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212420921005215
[18] Z. Yang, Y. Choe, and M. Martell, "Covid-19 economic policy effects on consumer spending and foot traffic in the u.s." *Journal of Safety Science and Resilience*, vol. 2, no. 4, pp. 230–237, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666449621000475
[19] Z. Ren and J. Zhu, "Disaster Resilience through Diverse Evacuation and Emergency Transportation Systems," *https://cammse.charlotte.edu/wp-content/uploads/sites/191/2023/06/CAMMSE-UNCC-2022-UTC-Project-Report-10-Zhu-Final.pdf*, no. 2022 Project 10, 2022, not publicly available.
[20] R. Callender, J. M. Canales, C. Avendano, E. Craft, K. B. Ensor, and M. L. Miranda, "Economic and mental health impacts of multiple adverse events: Hurricane harvey, other flooding events, and the covid-19 pandemic," *Environmental Research*, vol. 214, p. 114020, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0013935122013470
[21] S. Chang, E. Pierson, P. Koh, J. Gerardin, B. Redbird, D. Grusky, and J. Leskovec, "Mobility network models of covid-19 explain inequities and inform reopening," *Nature*, vol. 589, no. 7840, pp. 82–87, Jan. 2021, publisher Copyright: © 2020, The Author(s), under exclusive licence to Springer Nature Limited.
[22] C. Podesta, N. Coleman, A. Esmalian, F. Yuan, and A. Mostafavi, "Quantifying community resilience based on fluctuations in visits to points-of- interest derived from digital trace data," *Journal of The Royal Society Interface*, vol. 18, 04 2021.
[23] Google. (2016, mar) Google covid 19 community mobility reports. [Online]. Available: https://www.google.com/covid19/mobility/
[24] B. of Transportation and statistics. (2016, mar) Changes in mobility by state. [Online]. Available: https://www.bts.gov/browse-statistical-products-and-data/covid-related/changes-mobility-state-0
[25] NOAA. (2016, mar) Historical hurricane tracks. [Online]. Available: https://coast.noaa.gov/hurricanes/#map=4.06/26.98/-70.3
[26] ——. (2016, mar) Us hurricane datasets. [Online]. Available: https://catalog.data.gov/dataset/?q=hurricane
[27] I. D. Database. (2016, mar) Emergency event database. [Online]. Available: https://public.emdat.be/data
[28] FEMA. (2016, mar) National household survey. [Online]. Available: https://www.fema.gov/about/openfema/data-sets/national-household-survey
[29] N. S. Data and A. Center. (2016, mar) National household survey. [Online]. Available: https://sedac.ciesin.columbia.edu/data/set/usgrid-us-social-vulnerability-index-rev01
[30] U. S. C. Bereau. (2016, mar) Quick facts. [Online]. Available: https://www.census.gov/quickfacts/
[31] CovidTracking. (2016, mar) Covid tracking dataset. [Online]. Available: https://covidtracking.com/data/download
[32] Google. (2016, mar) Google covid 19 open data. [Online]. Available: https://health.google.com/covid-19/open-data/
[33] T. department. (2016, mar) Daily travel dataset. [Online]. Available: https://www.bts.gov/daily-travel
[34] NOAA. (2016, mar) Global hourly - integrated surface database (isd). [Online]. Available: https://www.ncei.noaa.gov/products/land-based-station/integrated-surface-database
[35] L. D. Crane, R. A. Decker, A. Flaaen, A. Hamins-Puertolas, and C. Kurz, "Business exit during the covid-19 pandemic: Non-traditional measures in historical context," *Journal of Macroeconomics*, vol. 72, p. 103419, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0164070422000210