UTRECHT UNIVERSITY
Faculty of Science
Department of Information and Computing Sciences
MSc Artificial Intelligence

USING CALL DETAIL RECORDS DATA TO PREDICT POST-EARTHQUAKE EVACUATION WITH A MACHINE LEARNING APPROACH

A THESIS BY
Jingkang Hu
5454565

Project supervisor Prof. dr. Albert Ali Salah
Daily supervisor Bilgeçağ Aydoğdu
Second examiner Prof. dr. Pınar Yolum Birbil

Utrecht
University

Abstract

The devastating February 2023 Turkey-Syria earthquakes, with magnitudes of $7.8$and 7.5, resulted in over $55,000$ death. While current technology cannot predict earthquakes, efficient evacuation management can significantly reduce secondary casualties and optimize resource allocation. This study explores the application of Call Detail Records (CDR) in times of crisis, with a particular focus on the consequences of earthquakes. This study focuses on two issues, namely, the prediction of population movements after earthquakes, and the features influence post-earthquake evacuation behavior. We use machine learning models with gravity transformed features to predict population movements immediately after earthquake. The experiments show our model have good ability to predict evacuation flow between different district. Our main findings are that population distribution and earthquake intensity are the primary factors of evacuation patterns. The comparative analysis between Turkish population and Syrian population shows the same feature importance rankings but distinct pattern distributions. These results provide valuable insights for emergency management authorities in resource allocation and evacuation planning, such as the effect of social connectedness.

Table of Contents

1 Introduction **4**

1.1 Problem statement 4

1.2 Research questions 5

1.3 Contribution of the thesis 5

2 Related work **7**

2.1 Mobile phone location data for migration and mobility 7

2.1.1 Mobile phone call detail records 7

2.1.2 The applications of CDR in disaster management 8

2.1.3 Public datasets 9

2.2 Processing CDR 9

2.2.1 Home and work location detection 9

2.2.2 Using CDR to analyse mobility 10

2.3 Gravity model 10

2.4 Ethical considerations 11

3 The Data **13**

3.1 CDR data 13

3.2 Data for features 14

3.2.1 Census data 14

3.2.2 Shake Intensity 14

3.2.3 Damaged buildings ratio 16

3.2.4 Relative Wealth Index 17

3.2.5 Facebook connections data 18

4 Methodology **19**

4.1 Transformed features 19

4.2 Machine Learning models 20

4.3 Modeling 21

4.4 Impact of features on prediction 22

5 Experimental Setup **23**

5.1 Processing CDR data 23

5.1.1 Home location 23

5.1.2 Identification of evacuation flows 24

5.2 District-level model 29
5.2.1 Normalization 29
5.2.2 Hyperparameter 31
5.2.3 Baseline model 32
5.3 City-level model 33

6 Experimental Results **35**
6.1 City-level evacuation flows 35
6.1.1 Results of classification model 35
6.1.2 Interpretation of the features 37
6.2 District-level evacuation flows 38
6.2.1 Results of regression model 38
6.2.2 Interpretation of the features 40

7 Discussion **43**
7.1 Model implications 43
7.2 Limitations in long-term migration prediction 43

8 Conclusions **45**

1. Introduction

1.1 Problem statement

Natural disasters such as earthquakes can have a huge impact on people's lives and lead to large-scale population movements. In recent years, call detail records (CDR) have been increasingly used as a unique data source for analysing post-disaster population movement (Blumenstock, 2011). In this thesis, we will use CDRs to construct a migration model to predict the evacuation flow aftermath of an earthquake.

Traditional population survey methods such as census and face-to-face interviews show obvious limitations during natural disasters. The limitations mainly are the timeliness of data collection, data flexibility, and feasibility in emergencies.

Firstly, traditional population surveys are usually conducted every ten years. In emergencies, policymakers and rescue teams need real-time data to develop effective evacuation and rescue plans. Second, traditional survey methods tend to be relatively rigid in their design, collecting mainly specific types of data. In emergencies such as earthquakes, the type of information needed can change rapidly. For example, changes in population density in the affected area after an earthquake are more important than information on education levels. In addition, the damage caused by earthquakes and the potential for secondary hazards such as aftershocks and fires increase the risk of conducting traditional population surveys. These methods are hardly feasible in emergencies.

CDR provide a new perspective for monitoring and analysing population movement patterns after earthquakes (Salah et al., 2019). CDR data contains information on the time and duration of calls and the geographic location of senders and receivers, and their collection does not depend on the infrastructure in the affected area. Mobile networks usually maintain some functionality even if part of the infrastructure is damaged.

A major advantage of CDR data is that it can provide information on population movements in almost real-time. This is critical for guiding emergency evacuation plans, optimising the allocation of relief resources and mitigating the impact of disasters. In addition, the flexibility of CDR data allows researchers and policymakers to conduct a variety of analyses based on actual needs, whether it is analysing population movement patterns over a specific period or tracking the movement of specific groups.

Another advantage is the wide coverage of CDR. The popularization of mobile phones provides a wide and deep sample of the population for research. This ubiquity makes CDR data a powerful tool for analysing and understanding population behaviour in disaster situations such as earthquakes.

There are some challenges and limitations in using these data. It is important to ensure that individual privacy is protected when processing CDR data, which requires strict security measures at every stage of data collection, storage and analysis.

1.2 Research questions

RQ1: How can we use mobile phone data to supplement other data sources to predict the movement of people inside the country after an earthquake?

The first research question explores how mobile phone data can be used to predict post-earthquake movement. These data sources help to understand how populations move in the immediate aftermath of an earthquake. And then this thesis will use machine learning model with gravity transformed features to model the post-earthquake evacuation.

RQ2: Which features will influence post-earthquake evacuation behavior? Will Turkish people and Syrian people respond to similar od different features in their evacuation decisions?

The second research question explores various factors, including earthquake intensity, geographical distance, building damage, and social connections. Understanding the relative importance of these features is important for both theoretical disaster response modeling and practical emergency management. By analyzing Turkish people and Syrian people separately, we investigate whether factors such as social connections, distance preferences, and destination choices vary between these groups.

1.3 Contribution of the thesis

This study helps to improve our understanding of post-earthquake evacuation patterns. We analyzed post-earthquake evacuation patterns and found that approximately half of all displacement flows occurred within the same city. Moreover, the vast majority of evacuation destinations remained within earthquake-affected areas. Our city-level analysis of Turkish and Syrian populations revealed distinct evacuation behaviors: Turkish evacuees showed more dispersed destination choices, while Syrian evacuees demonstrated more concentrated destination preferences.

Our supervised machine learning models demonstrated meaningful predictive capability for post-earthquake evacuation flows. Though these models cannot predict evacuation flows before disasters occur. Understanding how different features influence model predictions allowed us to quantify the impact of various factors (such as earthquake intensity, population, and social connectedness) on evacuation decisions.

Our experiments validated the applicability of gravity-transformed features in predicting post-earthquake population movements, providing methodological guidance for future research in this field.

2. Related work

2.1 Mobile phone location data for migration and mobility

2.1.1 Mobile phone call detail records

Call Detail Records (CDR) are data collected by mobile network operators when providing the service. These records include call initiation time, duration, phone numbers of both parties, call type, and possible device location information, etc Letouzé et al. (2015). CDR plays a large role in counteracting and responding to the adverse effects of disasters (Sarker et al., 2020). The information contained in the original CDR can be broadly categorised into four groups: population dynamics, mobility, social networks, and socio-demographics (Dujardin et al., 2020). The spatial trajectory of each user can be obtained based on the original CDR. Furthermore, the night-time activity range can be used to infer the user's home location. User's social networks may be built from the call and text message logs that they send and receive.

Compared to traditional population surveys, CDR has advantages in dealing with human mobility patterns for emergencies. A post-earthquake study in Nepal used CDR data to rapidly assess mobility patterns after the earthquake, particularly the mass exodus from the Kathmandu Valley (Wilson et al., 2016). This study shows how CDR can be used for quick assessment, which is essential for responding to natural disasters. Users of social media, email, and the web may not be economically representative of the general public (Tufekci, 2014). In comparison to CDR data , GPS data can potentially incur a higher selection bias (Frias-Martinez and Virseda, 2013). Therefore, CDR data is more representative of the general population than social media data and GPS data.

While CDR data has advantages, there still are some challenges and limitations. The biggest limitation is the access to CDR. CDR is stored by telecommunications companies and is not publicly available. Special arrangements and legal agreements are required for governmental or non-governmental organisations to access anonymised and aggregated versions of these valuable datasets. Privacy protection is a major challenge, with the need to ensure anonymity and confidentiality of personal information during data analysis. In addition, CDR also has possible bias. Especially in low-income area, only people who own and use a SIM card are included in the dataset (Li et al., 2019).

[1] `https://reliefweb.int/report/turkiye/undac-deployment-turkiye-06-march-2023`, Accessed 15 October 2024

2.1.2 The applications of CDR in disaster management

Studies using mobile phone location data to study disasters are categorized into three main categories (Yabe et al., 2022).

Population displacement and evacuation modeling

Lu et al. (2012) found that the destinations of people displaced after the earthquake were highly correlated with their prior mobility patterns. This work revealed the potential to predict post-disaster mobility patterns. Kargel et al. (2016) quickly analyzed the behavioral patterns of 12 million mobile phone users following the earthquake. This was the first time that large-scale mobile location data significantly aided disaster relief.

Long-term recovery analyses

CDR can track user activity at high frequencies over long periods. This feature is very useful for studying the long-term effects of disasters. Yabe et al. (2020) used CDR data of over 1.9 million users before and after five natural disasters for this approach. The study showed that although the areas affected and the types of disasters varied, the population recovery trends followed a similar pattern. After all five disasters, the majority of users returned quickly within a few weeks of the disaster, with the remainder returning gradually over a longer period. Therefore, data from the weeks following a disaster is most important in a study to understand the pattern of population recovery after a disaster. This study will investigate the population recovery patterns in the affected cities by analysing CDR data for one and a half months after the 2023 Turkey earthquake.

Inverse inferences about damage to the built environment

Detecting anomalies in human behaviour and movement patterns allows for a rapid assessment of damage to the built environment in the presence of data scarcity, with applications to a variety of downstream tasks, including mapping real-time flood inundation and identifying the location of dysfunctional mobile phone towers (Yabe et al., 2022). While the application potential of these studies is promising, there are not many current applications.

By analysing measures like call duration and mobility, CDR can also describe the social response to a disaster (Moumni et al., 2013). Phithakkitnukoon et al. (2011) found changes in relationship strength before and after the move, particularly six months before and one month after the move, and found that it took an average of seven to eight months to rebuild social networks in the new place of residence after the move. Moumni et al. (2013) found the social response was characterised by an increase in the number of calls, a reduction in the duration of calls, a moderate increase in the number of well-connected citizens, and a moderate increase in mobility.

Different classes have different scopes of activity, phone top-up habits, and communication

costs. Therefore, mobile phone data can also be used to explore the impact of disasters on different social classes (Myers et al., 2008; Steele et al., 2017, 2021). People from the rich class will pay more for each top-up, while the poor class will pay the opposite. The rich will also have higher monthly bills.

2.1.3 Public datasets

The use of CDR for research has been increasingly popular within the last 10 years. CDR datasets provide a rich data source for understanding human patterns at different scales. However, due to commercial sensitivity and privacy considerations, the availability of public CDR datasets is restricted. Despite this challenge, several initiatives have successfully made CDR datasets available to the research community, facilitating a wide range of studies.

The Orange "Data for Development" (D4D) challenge is one of the pioneering efforts to make CDR data available for research. The first D4D Challenge in 2012 provided anonymized CDR data from Orange's operations in Côte d'Ivoire. This initiative has sparked a wealth of research on topics ranging from epidemic modeling to urban planning and socioeconomic analysis (Blondel et al., 2015). Telecom Italia released a CDR dataset as part of its 2014 Big Data Challenge. The dataset covers Milan and the Expo 2015 area. It offers insights into mobility patterns, urban dynamics, and the effects of events on services and infrastructure in urban areas (Barlacchi et al., 2015) Turk Telekom established the Data for Refugees (D4R) initiative. D4R offers CDR data to support studies aiming at enhancing the living conditions of Syrian refugees in Turkey. This dataset has been used to study integration patterns, mobility, and social cohesion among refugee populations (Salah et al., 2019).

2.2 Processing CDR

2.2.1 Home and work location detection

Research on mobile data migration frequently focuses on fixed points in trajectories. Changes in fixed point positions can well reflect the migration of the observed object. Common fixed points include home address and work address. As highlighted by Ahas et al. (2010), early efforts in this field mainly used frequency-based method for home location assignment. This method determine users' home location based on the most common locations within a certain time range. And it often combined with rules related to time of day or week to increase accuracy. But it is limited by reliance on fixed time frames and the potential for inaccuracies due to the dynamic nature of human mobility. In contrast, the segmentation-based method proposed by Chi et al. (2020) marks a major advance in this field. This method does not rely on equally divided time windows but rather

identifies dynamic "time periods" in which users are likely to reside in the same location. This approach not only improves the accuracy of home detection but also filters out noise in the data. So it provides a more insightful understanding of migration events, including confidence scores for identified migrations. A more sophisticated approach described by Yabe et al. (2021) involves using the concept of information entropy to identify a user's schedule to accurately determine when the user is likely to be at home or at work.

2.2.2 Using CDR to analyse mobility

Descriptive analysis focuses on the integration of data and the overview of attributes and involves the statistical summarisation and graphical presentation of data. This type of analysis presents the geographic and temporal distribution of population movements before and after a disaster through visual tools such as graphs and charts. While this approach provides a foundation for understanding the impacts of disasters, it is often insufficient to explain the reasons behind them in depth.

Causal analysis aims to explore direct links between disasters and population movements. This type of analysis tends to use advanced statistical or econometric techniques, such as the use of multi-stage Markov chains to predict the future movement paths of individuals based on past locations (Ghurye et al., 2016), which provide deeper causal reasoning that can assist policymakers and disaster response teams to better understand and anticipate the impacts of hazards on migration trends to formulate more robust responses.

The ex-ante analysis includes forecasting and simulations. Forecasting means anticipating the population movements or destination of displaced people before the event. Forecasting will help the Government and related NGOs to prepare enough shelter, food, and health care for displaced people. Many methods have been used in forecasting, such as statistical models (Martineau, 2010), agent-based models (Suleimenova et al., 2017), and machine learning models (Huynh and Basu, 2020; Nair et al., 2019). Simulations refer to simulating the emergency evacuation after a natural disaster or other emergency. Agent-based modelling, which represents the interactions between individuals and between individuals and their environments, has become the main approach to simulate evacuation (Yin et al., 2020).

2.3 Gravity model

After Ravenstein (1885) demonstrated that the amount of migration is inversely proportional to distance, Stewart (1950) formally proposed to apply the law of gravity to population movement between two locations. The Gravity model assumes that the number of flow between two area decreases with distance, while the opportunity intervention model

assumes that the interval between two area depends on the number of opportunities. The basic equation is as follows

$$T_{ij} \propto \frac{m_i \cdot m_j}{rij} \tag{2.1}$$

where T_{ij} is the amount of flow from area i to area j, m_i and m_j are the population of the two area respectively, r_{ij} is the distance between the two area.

This model assumes that the number of flow from area i is proportional to its population, and the attractiveness of area j is also proportional to m_j. And there is a cost effect with respect to the distance travelled. These concepts can be summarised as the relationship between

$$T_{ij} = K \cdot m_i \cdot m_j \cdot f(rij) \tag{2.2}$$

where K is a constant, m_i and m_j are population of the two area, and $f(r_{ij})$ is known as the deterrence function, a decreasing function of distance. The distance function $f(r_{ij})$ is usually modelled as a power law or exponential (Barbosa et al., 2018)

$$f(r_{ij}) = \alpha \cdot r_{ij}^{-\beta} \cdot e^{-r_{ij}/r_c} \tag{2.3}$$

where α and β are adjustable exponents. Though it looks simple, the gravity model fits internal migration data well (Poot et al., 2016).

The Gravity model is well suited to exploring long-term migration patterns during or after a disaster. Isaacman et al. (2018) simulated migration during the severe drought in Guajira, Colombia, in 2014 using anonymised and aggregated CDR via Gravity model and Radiation model. The results showed that the prediction had a success rate of about 60 percent for the total number of people who migrated and the locations where they migrated.

2.4 Ethical considerations

There are several things that researchers need to be aware of when using mobile data from mobile phones for research. Firstly, CDR data are extremely sensitive because they contain detailed personal information that can be used for a variety of purposes. Most of the literature is based on carefully "anonymised" data, but this is insufficient to alleviate the full range of privacy concerns. The possibility of de-anonymising and re-identifying previously anonymised or aggregated datasets is well known. Narayanan and Shmatikov (2008) applies de-anonymisation methods to the Netflix Price dataset. They used only a little bit of information about a single subscriber to easily identify other records for that subscriber in the dataset and to reveal the user's political preferences and other sensitive information. The researchers tried to inject "noise" into the CDR to make re-identification

more difficult, (Becker et al., 2013). But even that only meant that more data points were needed to pick and choose which individuals to identify. In the future, the potential risks associated with CDR analyses remain, and researchers will need to keep working on this.

In addition, data biases and limitations need to be taken into account. Talking lists typically reflect the structural inequalities of any country:mobile phone ownership is strongly correlated with economic status. Wesolowski et al. (2013) examined the daily movements of nearly 15 million people in Kenya over a year by CDR and combined this analysis with socio-economic status, mobile phone ownership, and usage patterns across the country. Comparing these two data sources reveals that not only does mobile phone ownership tend to favour wealthier individuals, but there is significant geographic heterogeneity in levels of mobile phone ownership across the country. These biases hinder the validity of CDR-based findings and could potentially exacerbate structural inequality. Therefore, this study will analyse CDR in the context of local realities to reduce the bias introduced by the data.

3. The Data

3.1 CDR data

On 6th February 2023, an earthquake with the moment magnitude (Mw) 7.8 and 7.7 struck Kahramanmaraş, Turkey, causing massive damages (Erdik et al., 2023). There are more than 84,000 buildings collapsed or severely damaged (Wang et al., 2023), and more than $55,000$ lives lost during the earthquake (Hussain et al., 2023).

This study will use the fine grained dataset for a total of three months between 01/01/23-31/03/23. The data are coming from the Hummingbird EU project [1]. There are between $325,000$ customers with varying data frequency. As shown in Table. 1, the dataset will include timestamp of calls, base station site information, customer IDs and segment information. Along the guidelines established by ethics committees in past endeavours, the same anonymisation and aggregation methods are employed to remove all personal information (Blondel et al., 2013; Salah et al., 2019). Customers will be sampled with new random identifiers, so it is not possible to correlate customer records from different time periods. Random identifiers may also match in two different periods for two different customers. No personal information will be stored (Aydogdu et al., 2021).

Table 1. Fine Grained Dataset Structure

Fields	Explanations
TIMESTAMP	Day and Hour considered in format DD-MM-YYYY HH (24 h format)
Custmer ID	Randomly assigned ID of the customer generated for 15 days period
Segment Callee	A value denoting the segment of the customer receiving the call
Segment Caller	A value denoting the segment of the customer giving the call
SITE ID Caller	The ID of the base station of the customer receiving the call
SITE ID Callee	The ID of the base station of the customer giving the call

This study will use the site ID with the highest call frequency for each consumer from 19:00 to 7:00 as the home location of that day. The home location with the highest frequency each week will be regarded as the home location for that week, and the same applies to each month. Each site id has a corresponding city and city district, so the city of each month can be obtained from the home location of each month. We can see in Table. 2 that the home location of most consumers in the data set has not changed in three months, so

[1] `https://hummingbird-h2020.eu/publications`, Accessed 15 October 2024

they can be regarded as not migrating. Approximately $140,750$ people stayed in the same city, and $15,817$ people displaced to another one during the three months.

Table 2. CDR Monthly O-D matrices

Month 1	Month 2	Month 3	Count in pattern
City A	City A	City A	140,750
City A	City B	City A	3404
City A	City A	City B	4540
City A	City B	City B	6885
City A	City B	City C	988

3.2 Data for features

3.2.1 Census data

Census data comes from TurkStat [2]. Population data is obtained from the records in the National Adress Database of the General directorate of civil registration. This dataset includes the total population of each city and each district in Turkey. This study will only use the total population of the city and district. According to the statistical results, the total population is $85,372,377$.

Syrian populations in each city comes from the data released by the General Directorate of Migration Management in 2023 [3]. The Syrian population in Turkey is $3,234,000$, and the second most populous city of Syrian, Gaziantep, has $434,045$. The center of this earthquake was in Gaziantep, so it had a particularly large impact on Syrians in Turkey. There is no data on the Turkish population of each city in 2023 on the official website, so the Turkish population of each city will be estimated by subtracting the Syrian population of each city from the total population of each city.

3.2.2 Shake Intensity

The earthquake data were extracted from the USGS Earthquake Hazard Program [4]. The data coverage encompasses central and western Turkey, areas significantly affected by the earthquake. The dataset includes parameters such as Well-Known Text(WKT), intensity, Peak Ground Acceleration, Peak Ground Velocity, and Peak Spectral Acceleration.

[2] `https://data.tuik.gov.tr/Kategori/GetKategori?p=nufus-ve-demografi-109`, Accessed 15 October 2024

[3] `https://multeciler.org.tr/eng/number-of-syrians-in-turkey/`, Accessed 15 October 2024

[4] `https://earthquake.usgs.gov/earthquakes/eventpage/us6000jllz/shakemap/intensity`, Accessed 15 October 2024

The dataset delineates shake intensity in increments of 0.2, with the lowest level being 3.2. In this study, the WKT of Turkey's district map was compared with the WKT representing intensity to determine the intensity range for each district. For districts spanning multiple intensity ranges, the predominant intensity level based on areal coverage was assigned. Eastern Turkey, which experienced less seismic impact with intensities below 3.2, lacked data in the dataset; consequently, these areas were assigned an intensity value of 0. In Figure 1, the distribution of seismic intensities becomes apparent. Considering the actual damage in severely affected cities, areas with intensities above 4.5 were classified as severely impacted zones, while those below 4.5 were designated as less severely impacted zones.

Table 3. Population of different Turkish cities and the sample sizes in the CDR dataset

City	Population	Sample size in the CDR dataset	Percentage %
GAZIANTEP	2,164,131	25,427	1.175
SANLIURFA	2,213,964	19,144	0.865
HATAY	1,544,640	11,787	0.763
ADANA	2,270,298	9,570	0.422
KAHRAMANMARAS	1,116,618	4,256	0.381
KAYSERI	1,445,683	151	0.0105
KILIS	155,179	4,979	3.209
OSMANIYE	557,666	1,324	0.238
MALATYA	742,725	4,044	0.545
ADIYAMAN	604,978	966	0.160

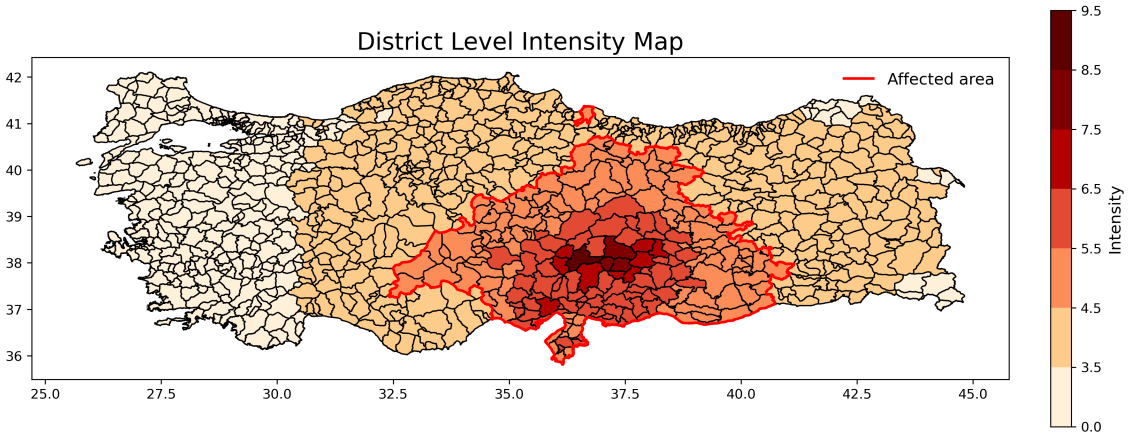

Figure 1. District-level intensity map of the earthquake

The intensity for each city was derived by calculating the mean intensity of its constituent

districts. The resulting city-level intensities are illustrated in the Figure. 2.

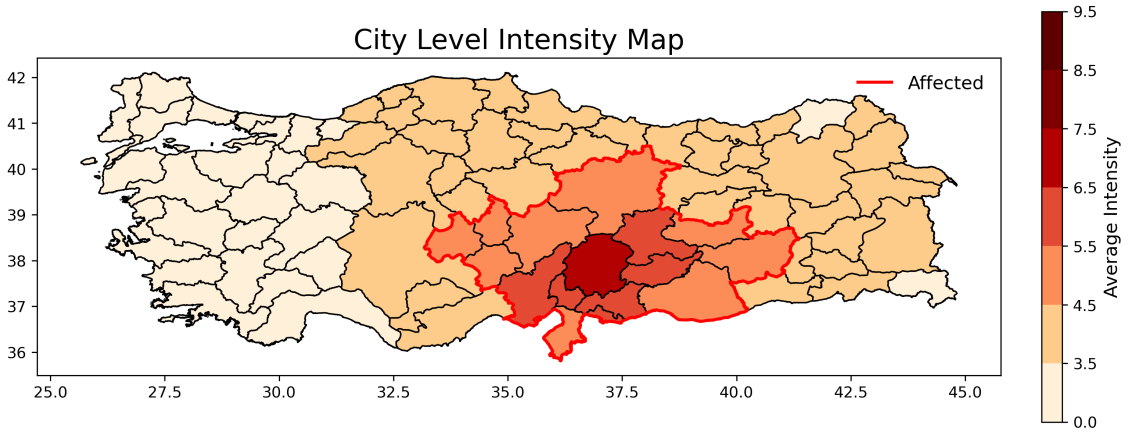

Figure 2. City-level intensity map of the earthquake

3.2.3 Damaged buildings ratio

This study will use data on buildings damaged by earthquakes to assess the varying impacts across different regions [5]. This dataset includes the number of various types of buildings in all districts (such as healthcare, industrial), the total number of buildings, and the number of destroyed buildings. Damage to buildings is categorized into four distinct levels, ranging from severe to minor. These categories are as follows: collapsed buildings, which have suffered the most severe damage; heavily damaged buildings; buildings that need to be demolished; and lightly damaged buildings. In order to reflect different degrees of damage, the following formula will be used for weighting.

$$\begin{aligned} \textit{weighted total damaged buildings} = {} & \textit{collapsed buildings} \\ & + 0.75 \times \textit{heavily damaged buildings} \\ & + 0.5 \times \textit{needs demolished buildings} \\ & + 0.25 \times \textit{slightly damaged buildings} \end{aligned} \tag{3.1}$$

$$\textit{damaged buildings ratio} = \frac{\textit{weighted total damaged buildings}}{\textit{all buildings}} \tag{3.2}$$

From Figure. 3, buildings damaged by the earthquake are predominantly located in major cities near the earthquake's epicenter. In contrast, cities situated farther from the epicenter primarily report slight damage to buildings, with fewer instances of buildings experiencing more severe levels of damage.

[5] https://sheltercluster.org/turkiye-earthquake-2023/pages/damaged-buildings, Accessed 15 October 2024

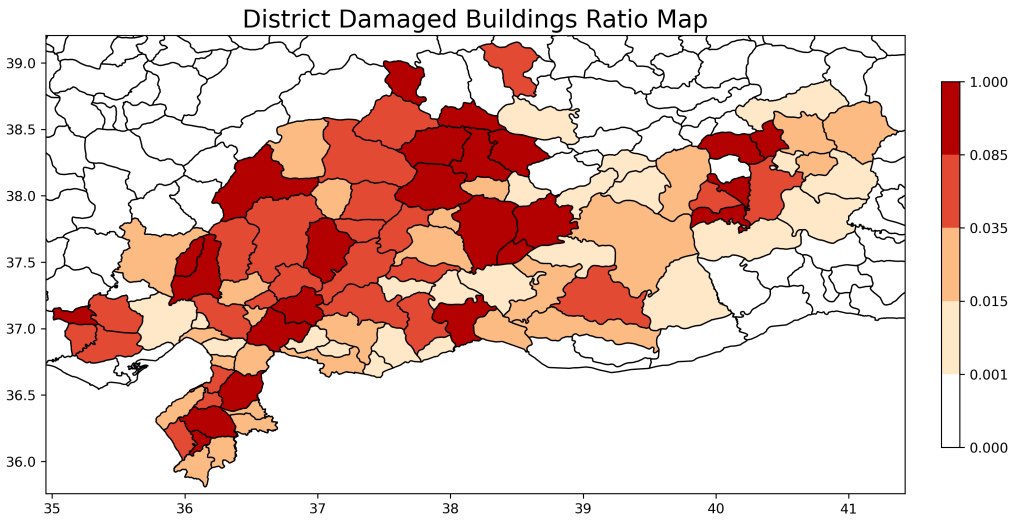

Figure 3. The map of district-level damaged buildings ratio

3.2.4 Relative Wealth Index

The Relative Wealth Index (RWI) was developed by the Center for Effective Global Action at the University of California, Berkeley, in collaboration with Facebook's Data for Good, to accurately assess the economic status of low- and middle-income countries [6]. The RWI is frequently used by nonprofit organizations to inform decisions on allocating cash assistance. This study uses the RWI to evaluate the impact of each district's economic status on population mobility following the earthquake. As shown in Figure. 4, the RWI combines publicly available survey data with non-traditional predictive data. Non-traditional predictive data includes high-resolution satellite imagery, topographic maps, mobile network data, and connectivity data. A variety of geospatially tagged data can aid in predicting relative wealth. For instance, satellite imagery can indicate population density relative to roads, suggesting an increase or decrease in relative wealth based on proximity. These two data sets were integrated to train a machine learning model capable of predicting absolute and relative wealth for each 2.4 square kilometer grid cell across 135 countries (approximately 19.1 million grid cells).

[6] `https://dataforgood.facebook.com/dfg/tools/relative-wealth-index`, Accessed 15 October 2024

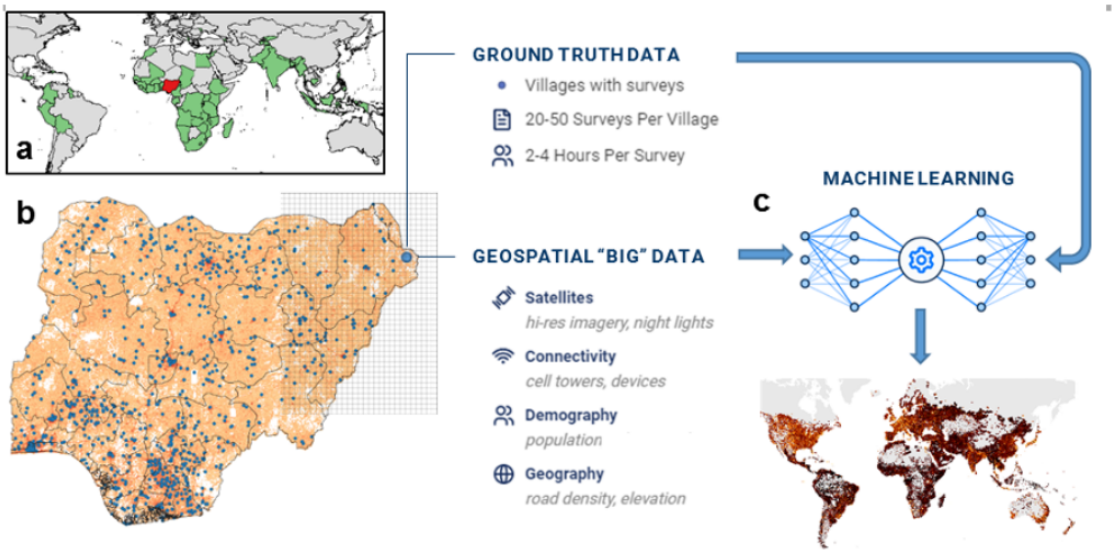

Figure 4. Framework for Deriving the Relative Wealth Index [7]

3.2.5 Facebook connections data

This study will use Facebook connections data to assess the Social Connectedness Index (SCI) between cities [8]. $SocialConnectedness_{i,j}$ measures the relative likelihood that a Facebook user in location i and a user in location j will add each other as Facebook friends (Bailey et al., 2018). The minimum value of $SocialConnectedness_{i,j}$ is 1. The higher the value, the higher the possibility of people in the two places adding friends.

$$Social\ Connectedness_{i,j} = \frac{FB\ Connectedness_{i,j}}{FB\ Users_i \times FB\ Users_j} \tag{3.3}$$

where $FBUsers_i$ and $FBUsers_j$ means the total amount of users from city i and city j; $FBConnectedness_{i,j}$ is the number of times users in city i and city j add each other as Facebook friends.

[7] https://dataforgood.facebook.com/dfg/tools/relative-wealth-index, Accessed 15 October 2024

[8] https://dataforgood.facebook.com/dfg/docs/methodology-social-connectedness-index, Accessed 15 October 2024

4. Methodology

4.1 Transformed features

In this study, we use machine learning model to predict the evacuation flows one week after an earthquake. We were inspired by (Hong and Frias-Martinez, 2020)'s approach. They used gravity model-transformed features to predict post-hurricane evacuation flows. We apply similar methodologies to predict post-earthquake evacuation patterns while also exploring radiation-transformed features. All the features described in the previous section, namely population, damaged buildings ratio, and RWI, are first processed through a formula conversion. Compared to directly using the characteristics of any given two cities to predict the evacuation flow between them, the formula in the form of the gravity model takes into account the influence of the distance between the two cities, although the radiation model does not directly use the distance.

Given a district D_i, we represent the set of features of D_i as,

$$D_i = \{d_{i,1}, d_{i,2}, ..., d_{i,N})\}, \tag{4.1}$$

where i indexes the district. For districts' i and j , the joint feature using a gravity-inspired transformation as follows:

$$g_{ij,k} = \frac{d_{i,k} * d_{j,k}}{dist_{i,j}} \tag{4.2}$$

where $d_{i,k}$ is the k-th feature of district i, $d_{j,k}$ is the k-th feature of district j, $dist_{i,j}$ is the distance between district i and district j, and $g_{ij,k}$ is the k-th joint feature of district i and district j. The reason why we use gravity transformed features is that we believe that the flow between districts depends on both the relationship between their features and the distance between them.

In addition to the gravity model, for the same reason, we also use the following radiation-inspired transformation to calculate the joint features. The joint feature using a radiation-inspired transformation as

$$r_{ij,k} = \frac{d_{i,k}U_{j,k}}{(d_{i,k} + v_{ij,k})(d_{i,k} + d_{j,k} + v_{ij,k})} \tag{4.3}$$

where $d_{i,k}$ is the k-th feature of district i, $d_{j,k}$ is the k-th feature of district j, $v_{ij,k}$ is the sum of between district i and district j, and $r_{ij,k}$ is the k-th joint feature of district i and district j. Radiation-transformed features also consider the mutual influence between the features

of the two districts. However, unlike gravity-transformed features, radiation-transformed features uses the sum of the features of other districts between the two districts to reflect the influence of other districts.

We do the same process for city-level features, and get city-level gravity-transformed features and city-level radiation-transformed features for city pairs. Given a city i, we represent the set of its features as:

$$C_i = \{c_{i,1}, c_{i,2}, ..., c_{i,N})\}, \tag{4.4}$$

where i indexes cities in Turkey. For each pair of cities, the gravity-transformed joint feature is given as follows:

$$G_{ij,k} = \frac{c_{i,k} * c_{j,k}}{dist_{i,j}} \tag{4.5}$$

where $d_{i,k}$ is the k-th feature of city i, $d_{j,k}$ is the k-th feature of city j, $dist_{i,j}$ is the distance between city i and city j, and $G_{ij,k}$ is the k-th joint feature set of city i and city j.

Similarly, the radiation-transformed joint feature for cities is defined as

$$R_{ij,k} = \frac{c_{i,k}c_{j,k}}{(c_{i,k} + V_{ij,k})(c_{i,k} + c_{j,k} + V_{ij,k})} \tag{4.6}$$

where $d_{i,k}$ is the k-th feature of district i, $d_{j,k}$ is the k-th feature of district j, $V_{ij,k}$ is the sum of between district i and district j, and $R_{ij,k}$ is the k-th joint feature of district i and district j.

4.2 Machine Learning models

We use two supervised machine learning models: Random Forest and XGBoost. Random Forest is an ensemble learning method that constructs multiple decision trees and aggregates their results through voting or averaging (Breiman, 2001). XGBoost is an advanced gradient boosting decision tree algorithm. It extends traditional Gradient Boosting Decision Trees (GBDT) by incorporating second-order derivative information and introducing regularization terms to control model complexity (Chen and Guestrin, 2016). Given that our district-level dependent variable contains numerous zero values and exhibits right-skewed distribution, we adopt Tweedie regression as the loss function. The Tweedie distribution is a special case of the exponential dispersion family (Tweedie et al., 1984). And it is particularly suitable for modeling non-negative continuous data with zeros. Anyidoho et al. (2023) successfully applied Tweedie regression loss function in XGBoost for hurricane evacuee prediction, achieving promising results. Here, we distinguish between two XGBoost variants: the standard XGBoost with mean squared error loss

function, denoted as XGBoost-MSE, and XGBoost with Tweedie regression loss function, denoted as Tweedie XGBoost.

4.3 Modeling

Through the above method, we will get the corresponding joint features for any two cities and districts in Turkey. At the district-level, this study aims to predict the evacuation flow after an earthquake. We use joint features as the input variable of the predictive model, and evacuation flow as the output variable to establish a supervised machine learning model. We will use Support Vector Regression, Random Forest and XGBoost to train the regression model. For the city-level, this study hopes to predict which city pairs will have higher evacuation flow. We will use Support Vector Classification, Random Forest and XGBoost to train the classification model.

We also build two baseline models, using original features at the city-level and district-level, respectively. These models will serve as reference points for evaluating the effect of feature transformation. The baseline models are built as follows:

- City-level baseline model: trained using the original features of the origin city and destination city.
- District-level baseline model: trained using the original features of the origin district and destination district.

These baseline models will be trained using the same machine learning algorithms as the transformed feature models (such as support vector machines, random forests, and XGBoost).

There are many city pairs that do not appear in the dataset. We should not ignore these cases. Therefore, this study will use sub-sampling method to randomly extract some city pairs that are not in the dataset and set the evacuation between them to 0. After sub-sampling, there will be a large number of samples with evacuation flow of 0. In order to avoid poor prediction results caused by sample imbalance, this study will use Stratified K-fold Cross-Validation. According to the number of evacuation flows, the samples will be divided into different layers. During cross-validation, samples are independently, randomly, and in a specific proportion from different layers. This approach prevents an over representation of a particular type of sample in any fold during K-fold Cross-Validation.

4.4 Impact of features on prediction

We will use SHAP values to understand how the population, shake intensity, damaged building ratio, RWI and CSI affect the prediction.

$$\phi_i(f) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \left(f(S \cup \{i\}) - f(S)\right) \tag{4.7}$$

where $\phi_i(f)$ is the contribution of this feature to the prediction result of model f; S represents the feature subset, and $S \subseteq N \setminus \{i\}$ is the feature subset that does not contain feature i; N is the set of all features; $|S|$ is the number of features in feature subset S; $|N|$ is the total number of features; $f(S)$ represents the prediction result of model f on feature subset S; $f(S \cup \{i\})$ is the prediction result of model f after adding feature i to feature subset S.

5. Experimental Setup

5.1 Processing CDR data

5.1.1 Home location

In the fine-grained dataset, we already have a site id caller for each data point. Table. 4 presents a synthetic example of CDR data structure used in this study. For privacy protection, all data shown are artificially generated and do not represent actual customer records. And there are latitude and longitude of site id in clusters Voronoi dataset. In clusters Voronoi dataset, the cell towers were clustered based on the proximity of their locations using hierarchical clustering method called Ward's method, and the distance threshold was chosed as 5 km. We can locate the approximate location of the caller when they making this call, and thus know the district where the call was made.

Table 4. An example call detail record (CDR)

date	hour	customer id	segment caller	segment callee	site id caller	site id callee
12/01/2023	15	178938302	1	NaN	17940	11644
10/01/2023	14	138832718	2	1	5956	7497
12/01/2023	22	116489436	1	1	13694	16653
12/01/2023	22	116489436	1	NaN	13694	16653

In this study, we want to predict the evacuation flow after an earthquake through regional features. To know the evacuation flow, we must first know the district where the user's home address is located. The most commonly used method to detect home location is to infer the location of night calls. We will select calls between 19:00 and 7:00, and the location of the site id caller with the highest frequency during this period will be regarded as the user's red location. In order to be more accurate, we will use night calls during a period of time to infer instead of every day. In the case shown in the Figure 5, we select night calls throughout January to infer the user's residence in January (to protect user's privacy, this case is adapted from a real case, not a real user trajectory.)

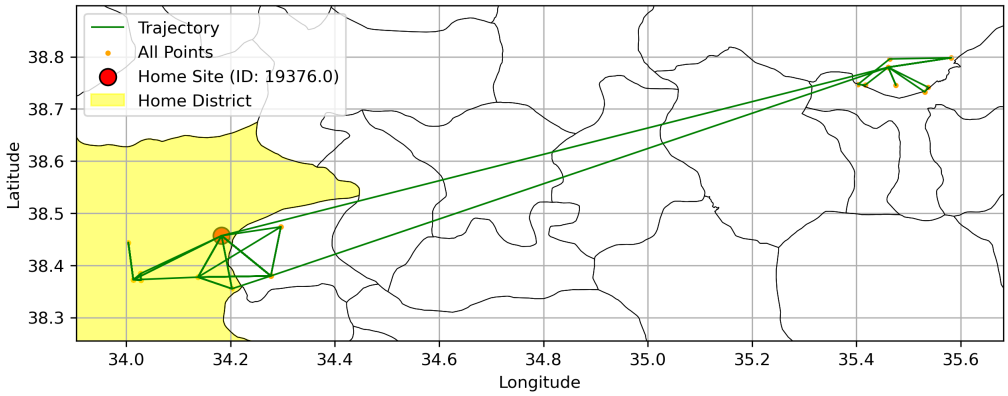

Figure 5. Sankey diagram of flow from Hatay, Sanliurfa, Malatya and Adiyaman

Using the home location between November 2022 and January 2023, we assume $149,579$ users from our CDR dataset lived in affected districts. While census data tells us, the real population of the affected districts is $16,660,931$. So the sample we analyzed accounts for about 0.90% of the total.

5.1.2 Identification of evacuation flows

The next question is how to define evacuation. This study will use the users' trip within one week to infer their mobility. If the first call and the last call of that week were fund to be made in different districts, we will assume that is a trip. We will assume. For example if the first call of user k was made on district i, and the last call was made on district j, then we infer that user k move from district i to district j in that week. Mobility flow before earthquake will be recognized as normal flow, and mobility flow after earthquake will be recognized evacuation flow.

Through the above method, we get the flow count of movement two weeks before the earthquake (from 23-01-2023 to 05-02-2023) and the evacuation flow one week after the earthquake (from 06-02-2023 to 13-02-2023). In order to highlight the impact of the earthquake on the evacuation flow, we should eliminate the flow count of daily movement. Therefore, we take the average of the flow counts two weeks before the earthquake as the baseline flow. If the evacuation flow and the baseline flow have the same district pair (that is, the same origin district and the same destination district). Then the final evacuation flow will be the evacuation flow minus the baseline flow. If the evacuation flow and the baseline flow do not have the same district pair, then the final evacuation flow of those city pairs is the evacuation flow.

District-level evacuation flow

This study focuses on evacuation after an earthquake, so we will select the flow whose origin district is the affected area (district with shake intensity greater than 4.5) from the

final evacuation flow.

First, we show the evacuation flow of some districts, which will help us understand the general situation of evacuation after the earthquake. After processing, there are 10,210 data in our data set, including $42,118$ flow counts.

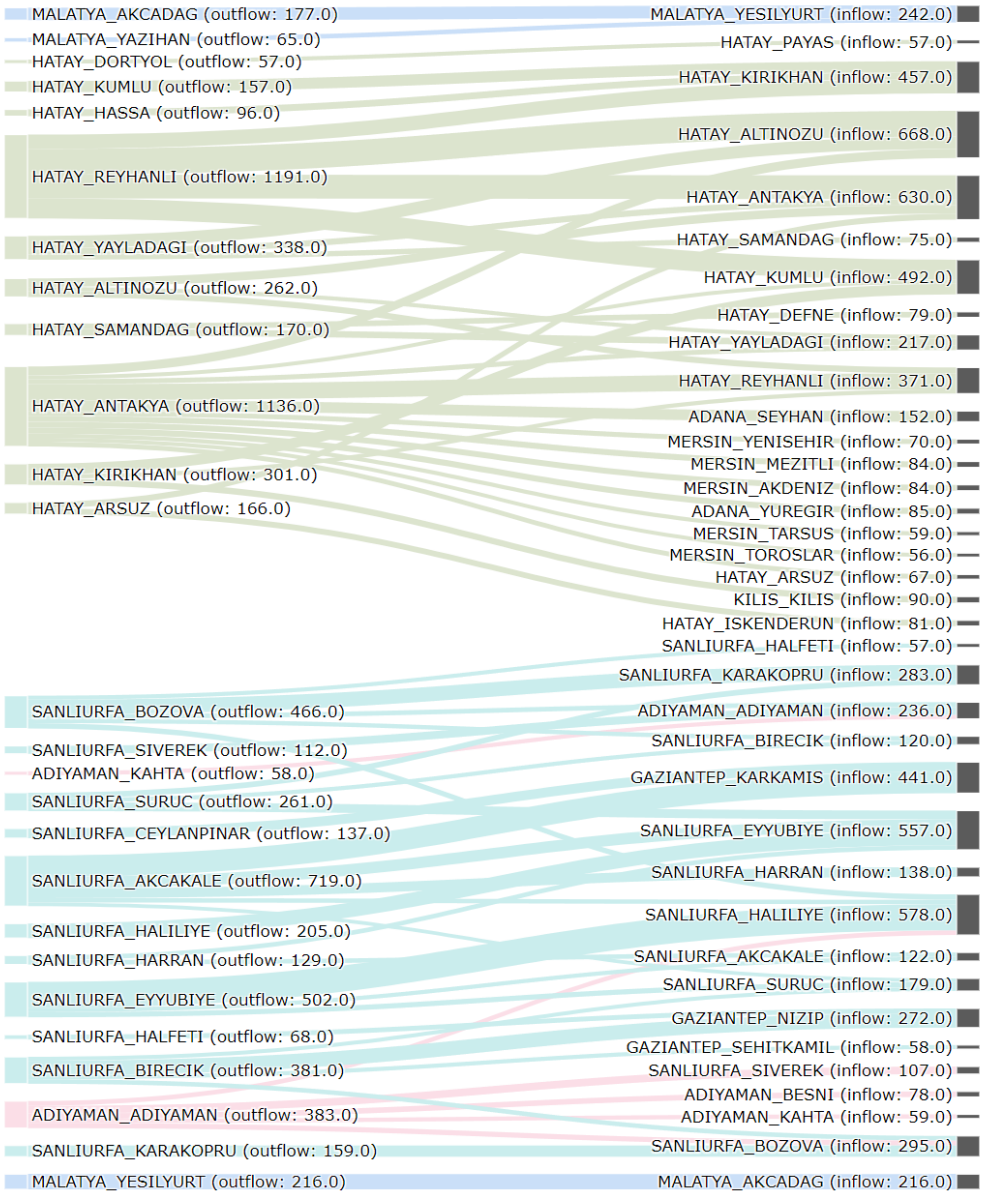

Figure 6. Sankey diagram of flow from Hatay, Sanliurfa, Malatya and Adiyaman

Most origin districts are located in Hatay, Gaziantep, Kilis, Adana Kahramanmaras and Malatya. The flow counts from Hatay and Gaziantep are the highest, reaching 21.6% and 21.1% respectively. We use the data that the origin district belongs to Hatay, Sanliurfa, Malatya and Adiyaman, and the flow count is greater than 50 as an example, to draw a Sankey diagram to analyze the pattern of evacuation flow. In Figure 6, the flows of the

same city in the origin district have the same color. From the flow from Hatay, we can see that even if people leave their district, they often stay in the same city. The data shows that flows within the same city account for 40.74% of all flows. For Hatay city, flows within the city even reached 48.3%. This shows that after the earthquake, most people just evacuated in the same city.

Flow from the Antakya district of Hatay is the largest among the origin districts, with a total of 2152, accounting for 5.1% of the total. The Figure 6 only shows flows with an origin district of Antakya exceeding 50, with a total of 1136. Antakya's destination districts are very diverse, flowing to multiple districts under cities such as Adana, Mersin, Hatay and Kilis. Many of these districts also belong to the affected area (districts with intensity greater than 4.5). Overall, flows from evacuation to affected areas account for 77.76% of all flows. In other words, most people did not leave the affected area, but may have gone to relatively less affected areas.

It can also be seen from Figure 6 that people tend to move to adjacent cities in addition to moving within the city. In Figure 6, the flows from Sanliurfa not only flow to other districts in the city, but also flow to the districts under the adjacent city of Gaziantep.

Among the destination districts that are not part of the affected area, the Mezitli, Erfemli, Toroslar districts of Mersin, the Selcuklu district of Konya, and the Cankaya district of Ankara occupy the top five flow count. The total flow to them is between 600 and 200. These districts are also very close to the affected area. Therefore, distance plays a big role in influencing population evacuation after the earthquake. Most people who choose to move only move within the same city or move to adjacent cities.

City-level evacuation flow

In analyzing the distinct evacuation patterns between Turkish and Syrian populations, we used segment identification in the original data, where segment 1 represents Turkish population and segment 2 represents Syrian population. We established baselines using flow counts from two weeks prior to the earthquake for both populations. The earthquake-induced flow was calculated by subtracting these baselines from the flow counts during the first week post-earthquake. Flows were then aggregated based on origin and destination districts' corresponding cities. Cities were classified as "affected" if their average district-level intensity exceeded 4.5.

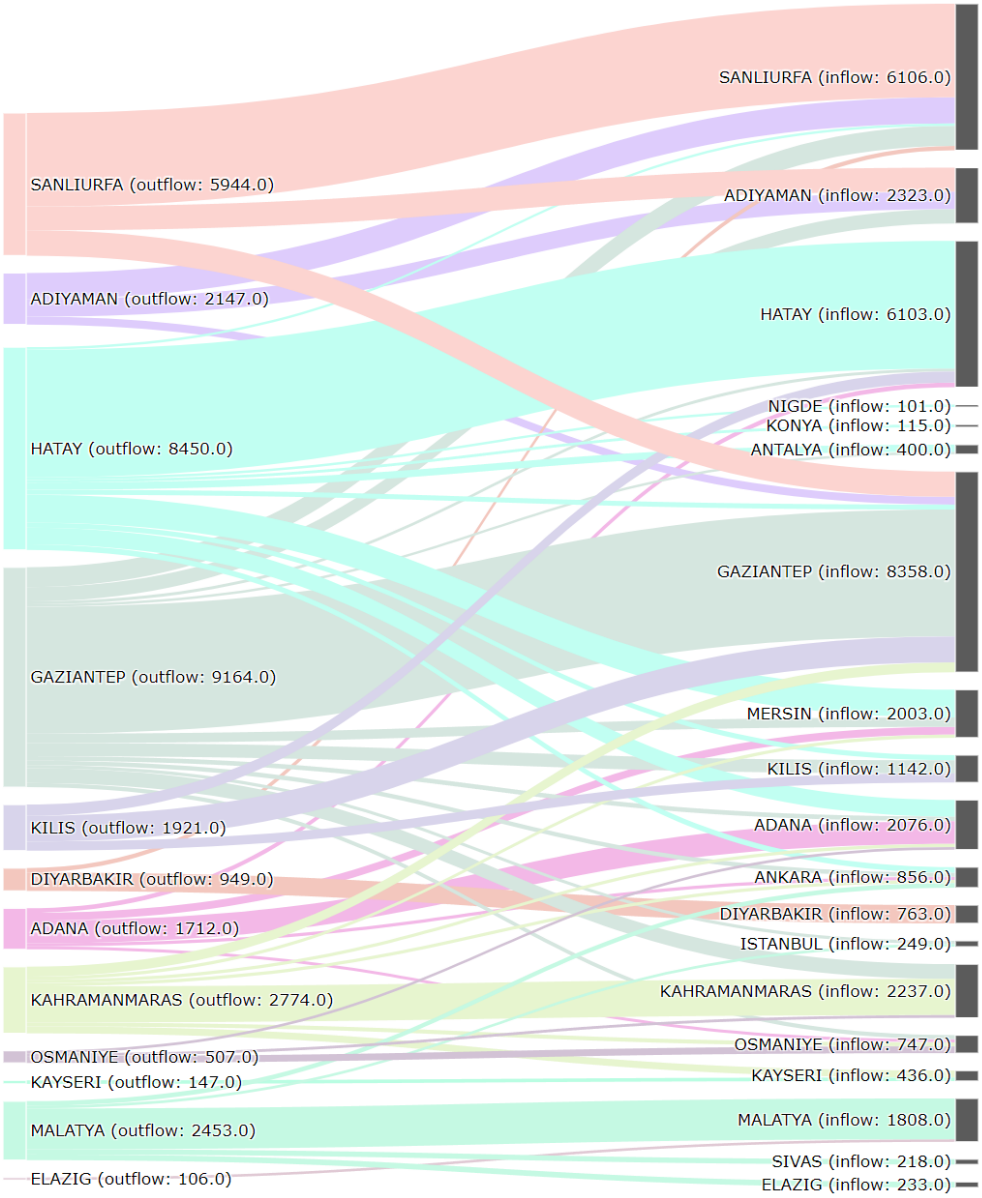

Figure 7. Sankey diagram of Turkish evacuation flow

Post-processing revealed Syrian flow counts of $28,879$, slightly lower than Turkish flow counts of $42,011$. We visualize of flows exceeding 100 count revealed distinct evacuation patterns in Figure 8 and Figure 7. Both populations show similar primary origin cities, with Hatay, Gaziantep, and Sanliurfa being the top three. However, notable differences emerged in destination city patterns. Mersin city, despite having some districts with intensity above 4.5, was classified as unaffected due to its average intensity being below 4.5. Its proximity to Adana and Hatay made it the most popular unaffected destination for both populations, with flow counts around $2,000$. Among unaffected cities, Turkish citizens showed a preference for Ankara (flow count:$1,106$) as their second choice, while Syrians favored Istanbul (flow count: 976). Antalya ranked third for Turkish population (flow count: 725) but did not appear among the top five unaffected destinations for Syrians.

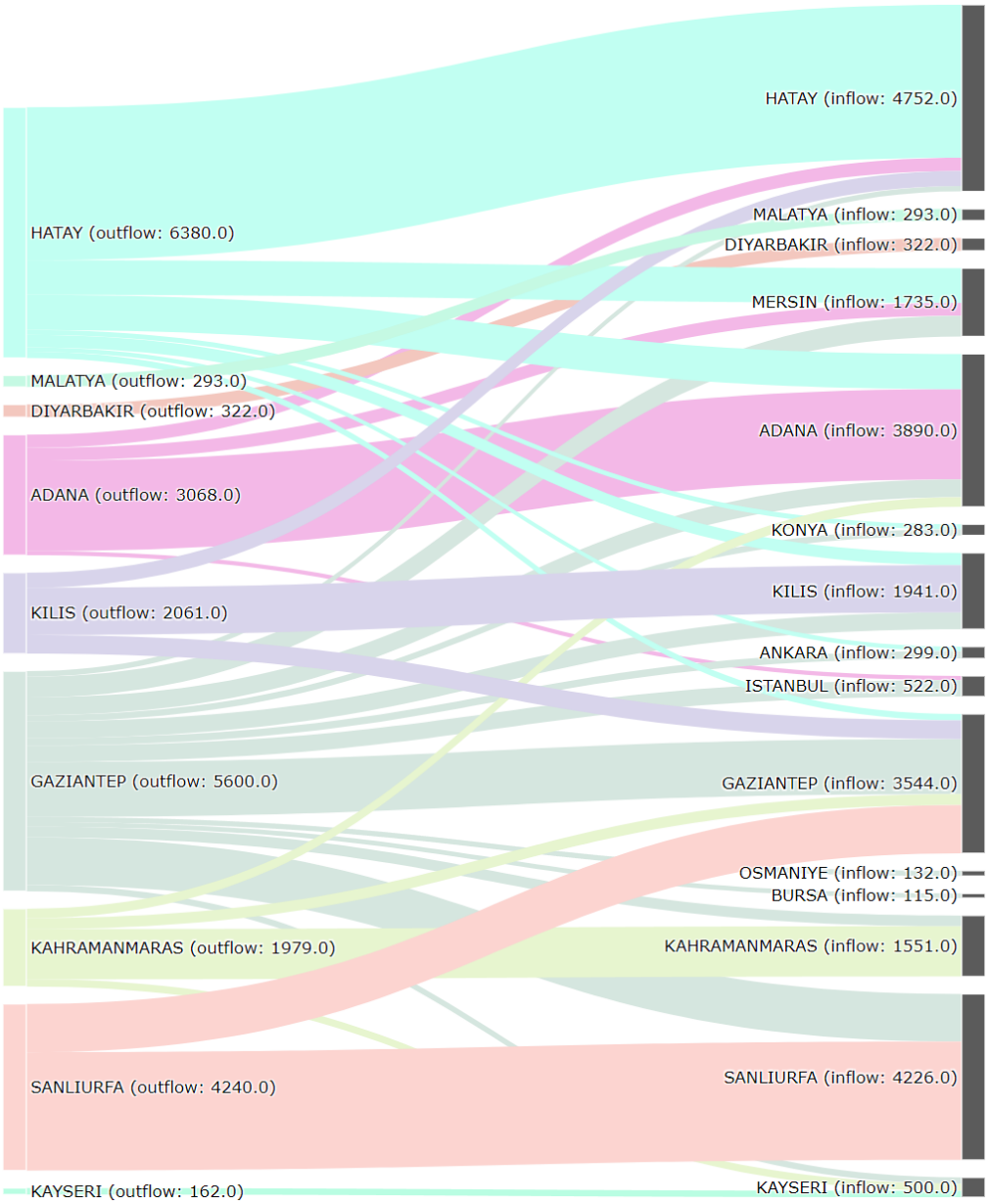

Figure 8. The Sankey Diagram of Syrian evacuation flows

Both populations generally showed significant intra-city movement across different districts within their origin cities, with Gaziantep's Syrian population being a notable exception, where only one-fifth remained within the city. Syrian population demonstrated a more concentrated flow pattern, primarily focusing on major cities like Adana, Hatay, and Mersin. In contrast, Turkish citizens exhibited a more dispersed pattern with a more complex network involving multiple cities. The more concentrated pattern observed among Syrian population suggests potential constraints in their mobility or preferences for areas with established Syrian communities.

5.2 District-level model

In this subsection, we elaborate on the experimental setup of our study. First, we select appropriate normalization methods based on feature distributions. We then explore different approaches using either actual distances or categorized distances for processing. Finally, we discuss the configuration of crucial parameters, such as the sampling size in sub-sampling.

5.2.1 Normalization

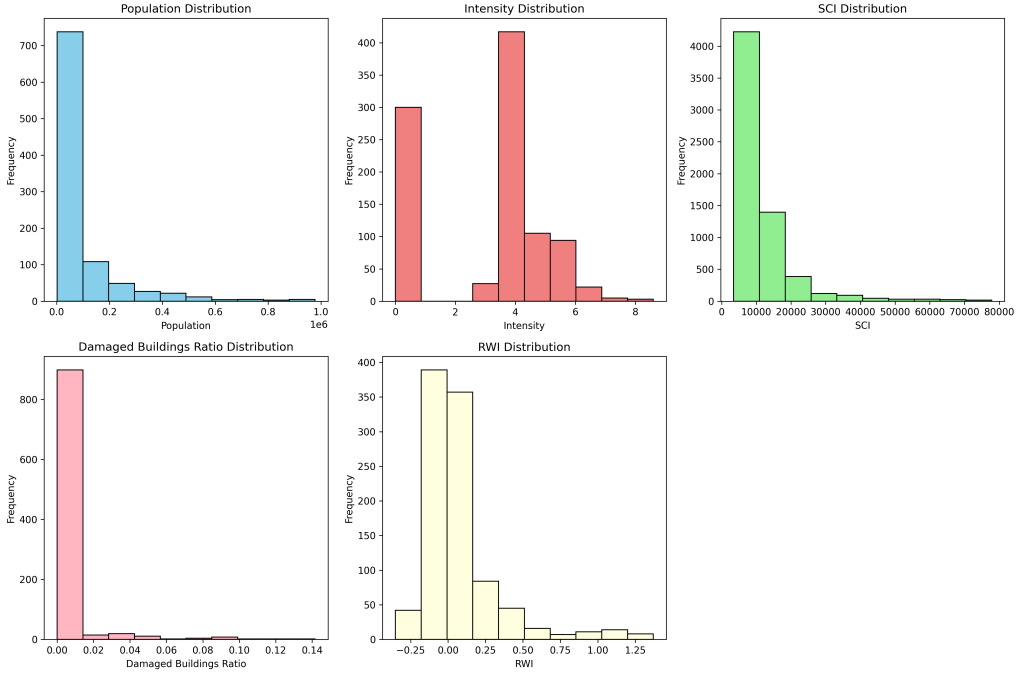

Figure 9. Features Distribution Histogram

As introduced in Section 4.1, we utilize five features for prediction. These features exhibit varying scales, with population typically exceeding tens of thousands and damaged building ratio ranging from 0 to 1. Therefore, feature normalization is necessary to facilitate meaningful comparisons. Rather than arbitrarily selecting a normalization method, we base our choice on the distributional characteristics of the features. As illustrated in Figure 9, these features demonstrate right-skewed and long-tailed distributions. Consequently, logistic z-score normalization becomes the most suitable approach. The formula of logistic z-score is:

$$\begin{aligned} z_i &= \frac{x_i - \mu}{\sigma} \\ \text{logistic z-score}(x_i) &= \frac{1}{1 + e^{-z_i}} \end{aligned} \tag{5.1}$$

where z_i is the Z-score of the feature value, x_i is the original feature value, μ is the mean of the feature values, and σ is the standard deviation of the feature values. In logistic z-score normalization, feature values are initially transformed to their z-scores, then passed through the standard logistic function to yield values within the [0,1] interval.

We proceed to examine the specifics of constructing gravity-transformed features. While we hypothesize that inter-location distances influence feature interactions between locations, the precise nature of this influence remains uncertain. Therefore, we investigate multiple scenarios:

- district-level features : Gravity transformed district-level features without distance.
- district-level features km 1: Gravity transformed district-level features based on raw physical distance.
- district-level features km 2: Gravity transformed district-level features based on raw physical distance square.
- district-level features category 1: Gravity transformed district-level features based on categorical distance.
- district-level features category 2: Gravity transformed district-level features based on categorical distance square.

The relationship between categorical distance and raw physical distance is illustrated in Table 5. A value of 42 represents the average distance between adjacent districts, while a value of 146 represents the average distance between adjacent cities.

Table 5. Categorical distance of districts

Physical Distance	[0,42)	[42,146)	[146,+∞)
Categorical Distance	1	2	3

Radiation-transformed features do not directly incorporate distance division. Instead, they utilize a circular area centered at district i with radius equal to the distance between districts i and district j. The sum of feature values from other districts within this range represents the distance effect, requiring no additional processing.

Upon implementing these features through Random Forest, XGBoost, and Tweedie XGBoost model, we obtain the results shown in Figure. 10. The results indicate superior performance of Tweedie XGBoost model. Tweedie XGBoost combined with feature km3 achieves optimal performance with a score of 0.563, leading us to select this for further optimization.

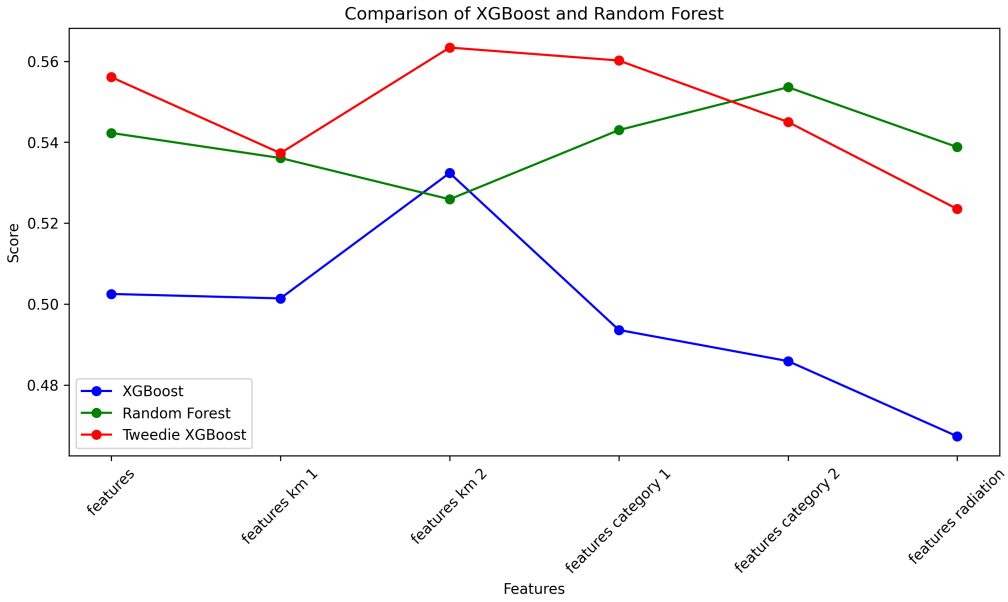

Figure 10. R-Square of Features

5.2.2 Hyperparameter

Sample number of sub-sampling is a very important hyperparameter in this study, because many city pairs are absent in the dataset. We conduct sub-sampling to deal with sample imbalance problem. We consider districts with an intensity greater than 4.5 as origin districts and form city pairs with other districts across Turkey. Excluding the more than $10,000$ city pairs already present in the samples, we have over $36,000$ additional city pairs. The flow counts for these pairs will be set to zero. From this pool, we will randomly select n samples to create a new dataset to evaluate potential improvements in the model. When $n = 0$, it signifies that only data existing in the original dataset is chosen. We select n from [0, 5000, 10000, 15000, 20000, 25000, 30000, 35000]. From Figure 11, the results indicate that when $n = 10000$ the model shows significant improvements.

In addition to varying the sub-sample size, we also optimized the following parameters using grid search with cross validation (GridSearchCV): maximum tree depth $(5, 6, 7, 8, 9, 10)$; number of estimators $(100, 200, 300, 400, 500)$; learning rate $(0, 0.1, 0.2, 0.3, 0.4)$; alpha values of $(0, 0.1, 0.5, 1)$, lambda values of $(0, 0.1, 0.5, 1)$, and gamma values of $(0, 0.1, 0.5, 1)$. Each experiment was conducted using 5-fold stratified cross-validation, with the best results determined by selecting the highest R^2 value. The final optimized parameters were: maximum tree depth of 10; number of estimators of 300; learning rate of 0.2; alpha of 1; lambda of 1; and gamma of 0.1.

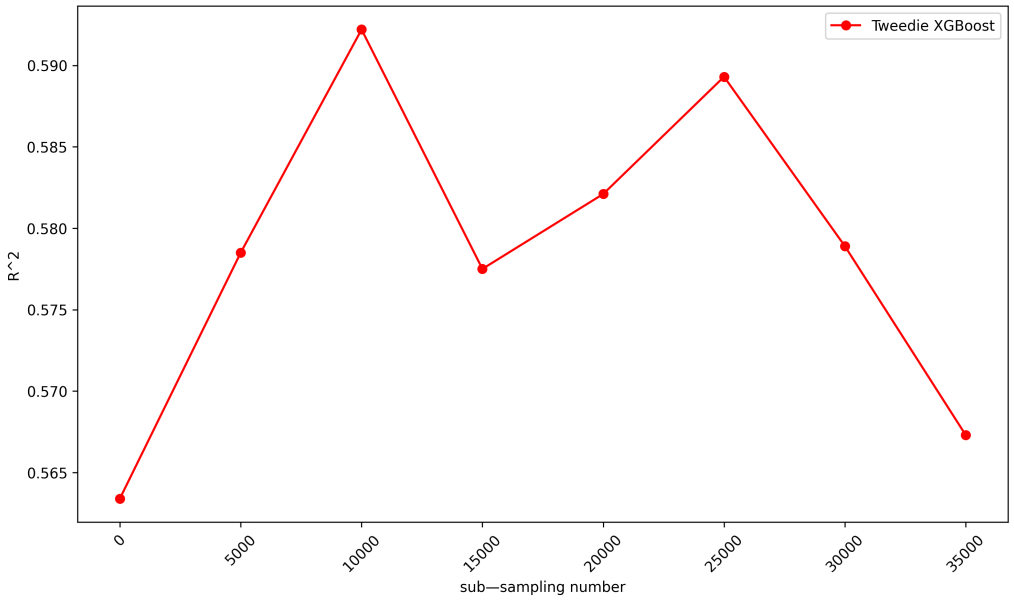

Figure 11. R-Square of different sub-sampling numbers

5.2.3 Baseline model

To compare the results of the machine learning model with transformed feature, we added the following two baseline models.

- Baseline model 1: Gravity model
- Baseline model 2: Machine Learning model using untransformed features

Baseline model 1 is a gravity model that only uses population and distance as features. Baseline model 2 is a machine learning model that directly uses the features and distance of the two cities as features for each city pair.

5.3 City-level model

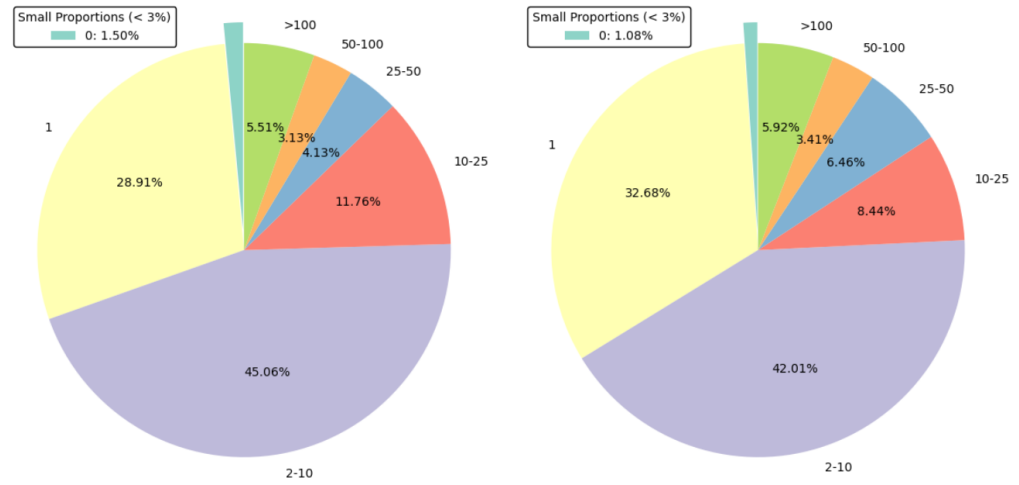

Figure 12. City-level flow counts(Left: Turkish, Right: Syrian)

The distribution of city-level flow count differs significantly from the distribution of district-level flow count. As shown in Fig. 12, approximately 6% of city-level flow counts for both Turks and Syrians exceed 100, with some reaching thousands. Meanwhile, about 75% of flow counts fall between 0 and 10. Consequently, regression models for city-level flow counts would yield very high RMSE and MAE values, failing to achieve the desired results. Therefore, this study decided to develop classification models for city-level flow counts using random forest and XGBoost algorithms. In the classification models, city pairs with relatively high flow counts are considered positive (flow count ≥ 10), while those with lower counts are negative (flow count <10).

We processed city-level features similarly to district-level features, using both actual physical distance and categorized distance to obtain the following features:

- city-level features : Gravity transformed city-level features without distance.
- city-level features km 1: Gravity transformed city-level features based on raw physical distance.
- city-level features km 2: Gravity transformed city-level features based on raw physical distance square.
- city-level features category 1: Gravity transformed city-level features based on categorical distance.
- city-level features category 2: Gravity transformed city-level features based on categorical distance square.

In this model, false negatives represent city pairs with actually high flow counts being

classified as low flow count pairs. This may lead to cities needing more assistance being incorrectly identified as not requiring help. Therefore, considering accuracy, precision, and recall, we chose the random forest model with city-level features category 1 for further optimization.

We considered using GridSearchCV to select these hyperparameters: n estimators: $(50, 100, 200)$, max depth: $(None, 10, 20, 30)$, min samples split: $(2, 5, 10)$, min samples leaf: $(1, 2, 4)$, bootstrap: (True, False). The results showed that for Turks, these hyperparameters were optimal: bootstrap: True, max depth: 10, min samples leaf: 1, min samples split: 2, n estimators: 50. While for Syrian, these hyperparameters were best: bootstrap: True, max depth: None, min samples leaf: 2, min samples split: 10, n estimators: 50.

In feature selection, all five features were important for Turks. For Syrians, removing the damaged building ratio had no significant impact on prediction results, so only the remaining four features were used.

6. Experimental Results

6.1 City-level evacuation flows

6.1.1 Results of classification model

In this section, we present the results of classification machine learning models for predicting city-level evacuation flow. Given the extreme distribution of city-level evacuation flow, with approximately three-quarters ranging from 0 to 9, one-quarter exceeding 10, and a small portion above 100, we opted for classification models to predict cities with higher flow counts. Flow counts greater than or equal to 10 were classified as 'high flow count' (positive).

We used two models, Random Forest and XGBoost, with five feature sets. Based on precision, accuracy, and recall metrics, the Random Forest model with feature category 1 shows superior performance, leading to parameter optimization and feature selection. For both Turkish and Syrian populations, SCI shows minimal impact on results. Consequently, the final model incorporated only four features: RWI, population of Turks or Syrians, damaged building ratio, and intensity.

Table 6. Results of City-level Model

	Accuracy	**Precision**	**Recall**	**F1 Score**	**Confusion Matrix**	
Turks	0.866	0.817	0.610	0.721	562	74
					35	144
Syrian	0.883	0.818	0.748	0.779	389	41
					26	117

As shown in Table. **??**, both Turkish and Syrian models exhibited high and similar accuracy and precision scores. Regarding recall, the Syrian model (0.748) significantly outperformed the Turkish model (0.610). Recall represents the proportion of actual high flow counts successfully identified by the model. The Syrian model's higher recall rate indicates superior detection of actual evacuation behavior with fewer instances of unidentified evacuation cases. This disparity might suggest that Syrian evacuation patterns are more predictable.

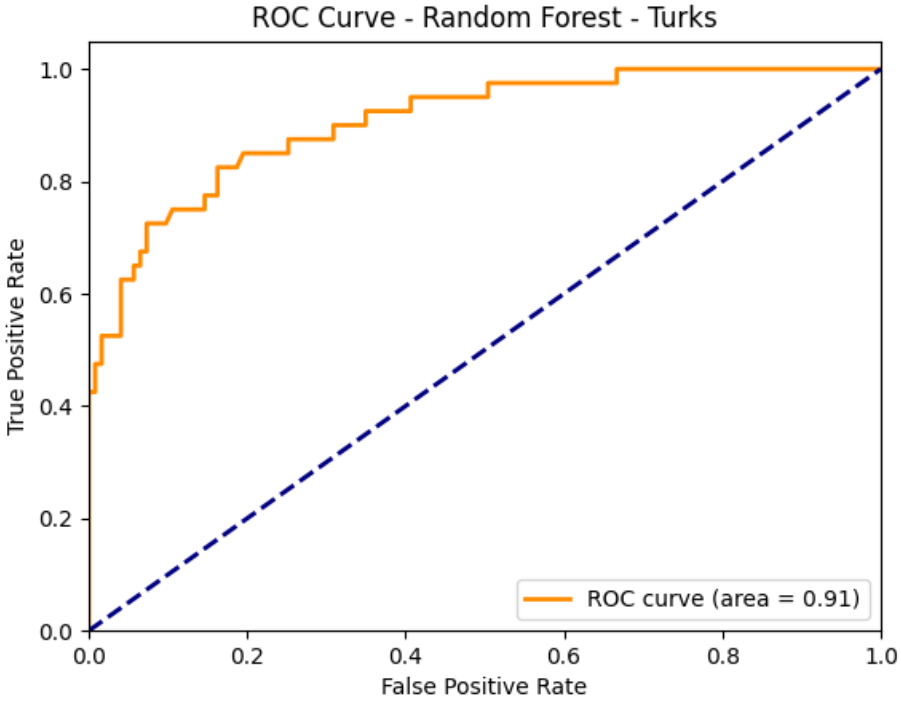

Figure 13. Receiver Operating Characteristic for Turkish population

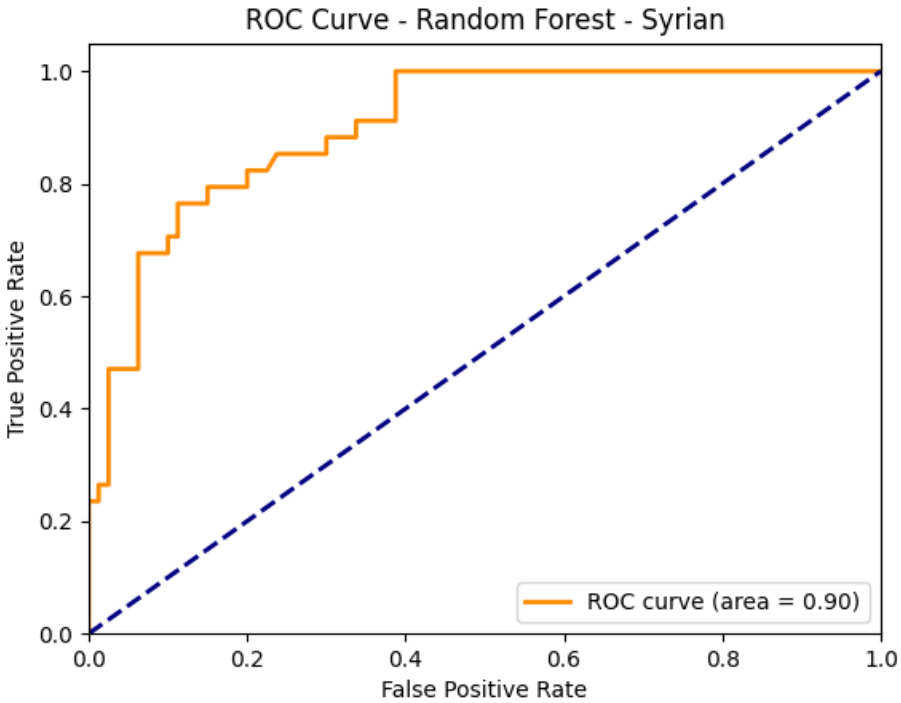

Figure 14. Receiver Operating Characteristic for Syrian population

From Figure 14 and Figure. 13, ROC curve analysis demonstrated excellent classification

capabilities for both models, with AUC values of 0.91 for the Turkish model and 0.90 for the Syrian model.

6.1.2 Interpretation of the features

After demonstrated the effective predictive performance of classification models for both Turkish and Syrian populations, we now use SHAP metrics to examine the differential impacts of various features on these two populations.

Figure. 16 and 15 illustrate the SHAP values (impact magnitude) of different features on evacuation prediction models for Turkish and Syrian populations. Positive SHAP values means the feature tends to increase the prediction value. The analysis reveals that for most features, red points typically correspond to positive SHAP values, indicating that higher feature values between city pairs are more likely to correlate with high flow counts.

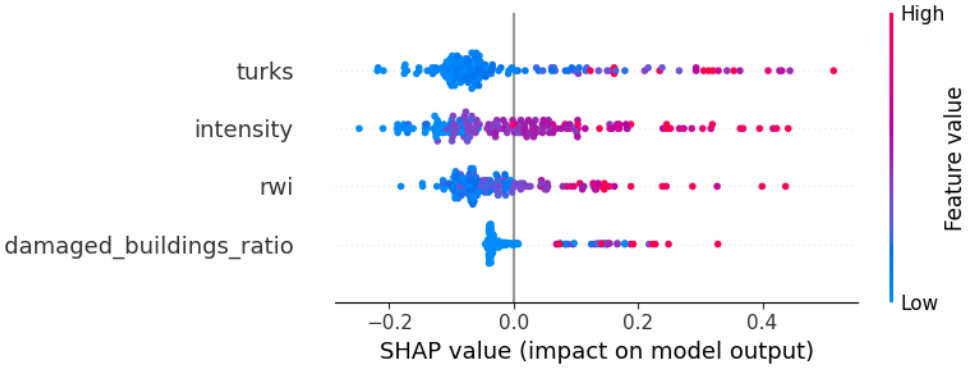

Figure 15. SHAP values of features for Turkish Model

For both populations, their respective population sizes emerge as the most significant predictive factors, suggesting that evacuation behavior largely depends on existing population distribution. Intensity ranks as the second most important factor for both groups, which is logical as higher intensity areas would likely generate more evacuees. However, the impact pattern is more dispersed for the Turkish population. This potentially indicates more diverse responses to earthquake intensity. RWI shows a slightly stronger influence on the Turkish population compared to Syrians. This possibly reflects that urban economic conditions have a greater impact on Turkish evacuation decisions.

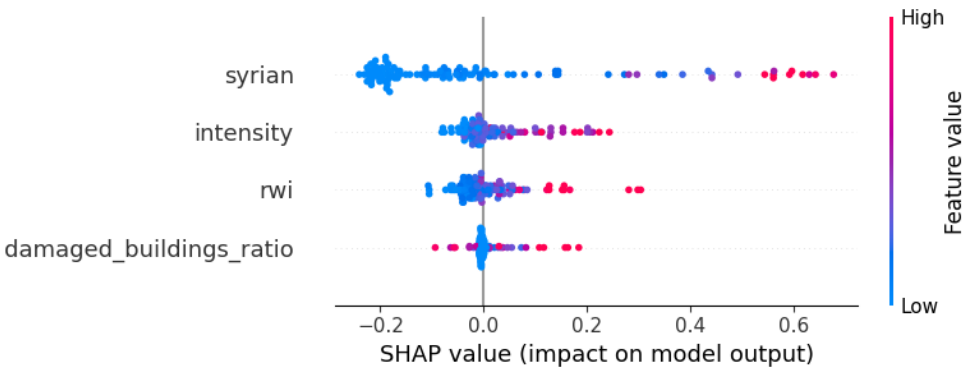

Figure 16. SHAP vaules of features for Syrian Model

The damaged buildings ratio shows relatively minor influence on both groups. Because it provides similar information to intensity, but intensity offers more comprehensive information. Additionally, many cities report a damaged buildings ratio of zero, resulting in numerous zero values when applying gravity transformation to damaged buildings ratios between different city pairs.

6.2 District-level evacuation flows

6.2.1 Results of regression model

In this section, we present prediction results for district-level flows. As discussed in Section 5.2, the final results were obtained using a Tweedie XGBoost model with feature km 2, selecting population, SCI, intensity, and RWI from the five features.

Additionally, we established two baseline models to evaluate whether machine learning with gravity transformation improves post-earthquake evacuation flow prediction. Baseline model 1 represents the traditional gravity model, utilizing only distance and population features. After estimation, baseline model 1 is:

$$T_{ij} = \frac{4.616 \cdot m_i^{0.180} \cdot m_j^{1.135}}{r_{ij}^{0.300}} \tag{6.1}$$

where m_i is the population of district i, m_j is the population of district j, r_{ij} is the distance bewteen district i and district j.

Baseline model 2 is a machine learning model with untransformed features. Specifically, for district i as the origin district and district j as the destination district, it uses the distance between districts i and j, along with their five characteristics, totaling 11 features for prediction.

Table 7. Prediction results for models across evaluation metrics

Model	R2	RMSE	MAE
Final model	0.601	6.921	1.257
Baseline model 1	0.148	15.950	4.487
Baseline model 2	0.063	16.487	3.662

The evaluation results for the final and baseline models are presented in the table. Results indicate that the machine learning model with gravity transformed features (final model) significantly outperforms both baseline models 1 and 2. The final model achieves an R-square of 0.601, explaining approximately 60.1% of data variance. Both RMSE and MAE metrics are relatively small for the final model. The baseline models' R-square values of 0.148 and 0.063 indicate weak predictive power. So machine learning with gravity transformed features substantially improves predictive capability.

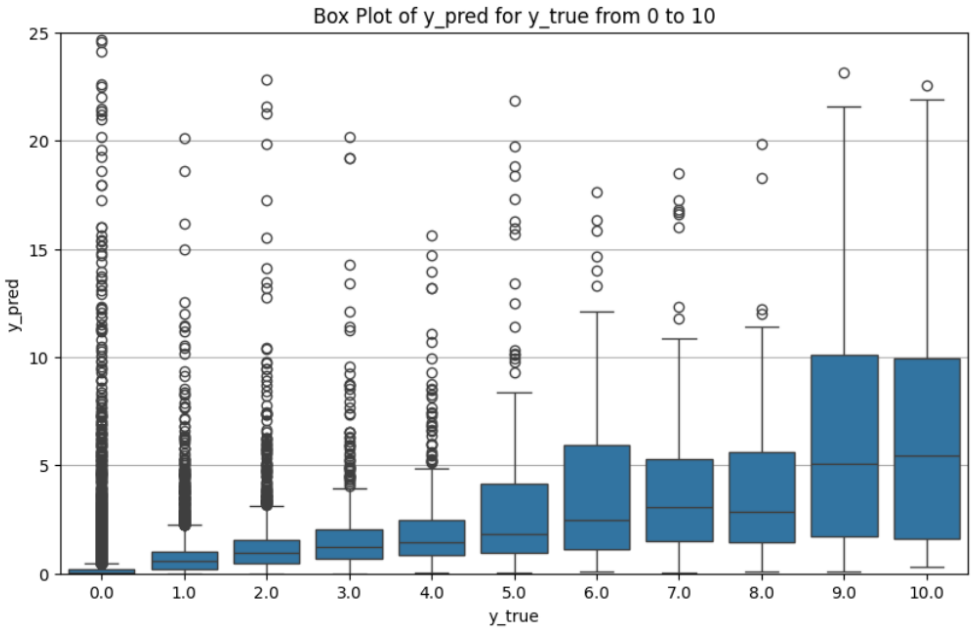

Figure 17. Boxplot of Predicted Values for True Value from 0 to 10

The box plot illustrates the final model's predictive performance for actual values from 0 to 10. Figure. 17 shows increasing median predicted values as actual values increase from left to right, indicating the model captures the basic data trend. For smaller actual values (0 to 4), the boxes are compact, suggesting concentrated predictions.

Overall, while the model accurately reflects the general trend of increasing flow, the median values falling below actual values suggest potential underestimation. The model performs more accurately for small flows, possibly due to their higher representation in

the dataset enabling better feature learning. Conversely, predictions for larger flows show higher uncertainty. For flow counts of 0, despite boxes being very close to 0, numerous outliers exist, potentially due to the sub-sampling process introducing 1.5 times the original dataset's zero flow count city pairs.

6.2.2 Interpretation of the features

The R-square value indicates that the final model demonstrates reasonable predictive capability for post-earthquake evacuation flows. We now examine the selected features in detail. Figure. 18 presents the correlation matrix of the five processed features: RWI, SCI, population, intensity, and damaged building ratio. The highest correlation coefficient (0.64) occurs between intensity and SCI, which is likely coincidental. The second highest correlation (0.58) is between intensity and damaged building ratio, which is logical as areas with higher intensity typically experience greater building damage. All other correlation coefficients are below 0.5.

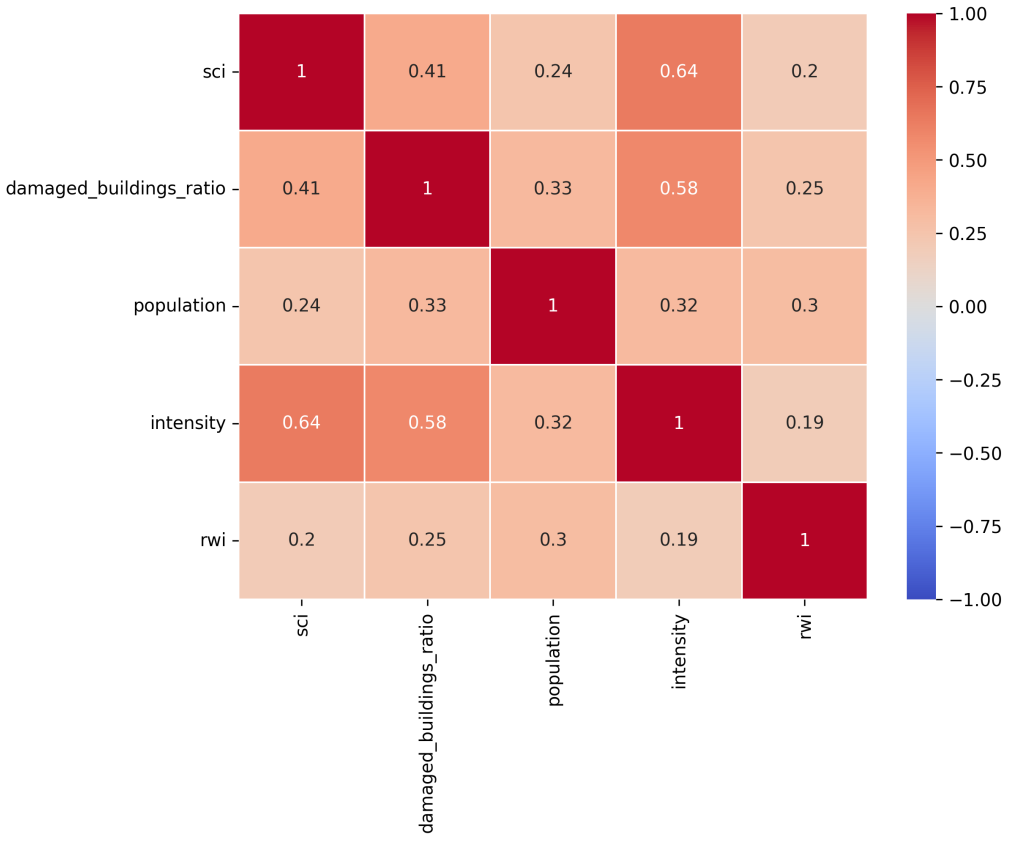

Figure 18. Correlation matrix of the transformed variables used for modeling

During feature selection, we observed that removing the damaged building ratio had minimal impact on prediction metrics (R-square, RMSE, MAE), possibly because it conveyed similar information to intensity. Consequently, the final model incorporated the remaining four features. We used SHAP values to demonstrate these features' impact on the

model. Additionally, all features discussed in this section are gravity transformed features. So the influence of geographical distance has been incorporated into their calculations. This transformation applies the gravity model concept, where the impact of each feature is weighted by the distance between districts. For example, when we discuss population's strong positive influence, we are actually examining the combined effect of population and distance rather than only population.

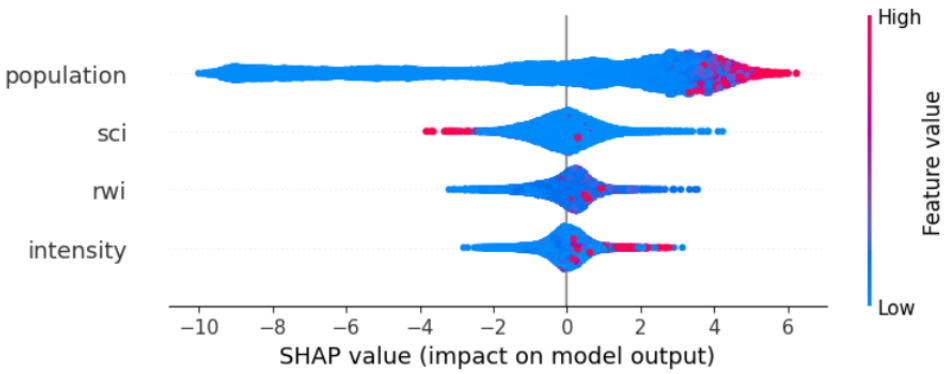

Figure 19. SHAP value for final model

Figure 19 shows the features' influence in descending order: population, SCI, RWI, and intensity. Population and intensity display high values (red dots) concentrated on the right, indicating positive effects on model predictions. Conversely, high SCI values cluster on the left, suggesting that higher SCI values actually decrease model predictions.

Population demonstrates the strongest influence, with SHAP values ranging approximately from -10 to 6. This aligns with traditional gravity models, where population and distance serve as the principal feature.

SCI (Social Connectedness Index) is calculated from Facebook friendship frequencies. Counter to expectations, higher SCI values negatively impact model predictions. This might be attributed to other features being at the district-level, while SCI is initially at the city-level. By assigning identical values to flows with the same origin and destination cities, SCI may fail to capture district-level flow nuances. Although inter-city SCI values are high, these might primarily reflect intra-district user connections, whereas our model analyzes inter-district flows, potentially rendering high SCI values less informative.

RWI (Relative Wealth Index), is calculated from survey data and satellite image through machine learning method. The high values showing positive but less pronounced effects compared to population and intensity. This suggests that district wealth variations moderately influence post-earthquake evacuation patterns.

Intensity with high values clustered on the right side. This aligns with our expectations: high-intensity areas suffer greater damage, motivating stronger evacuation responses.

7. Discussion

7.1 Model implications

Our models are supervised machine learning models requiring labels. Therefore, these models cannot be used to predict evacuation flows before or immediately after an earthquake. Nevertheless, this research provides valuable insights for future studies in this field.

First, we analyzed how features influence post-earthquake population movements. They can guide future researchers in designing pre-earthquake evacuation flow models. For instance, our district-level model indicates that the damaged building ratio conveys similar information to intensity measures. Removing the damaged building ratio has minimal impact on model predictions. Intensity data can be obtained more conveniently than damaged building ratio data. To our surprise, higher SCI values decreased predictive accuracy. This counter-intuitive finding may stem from data granularity issues. City-level SCI data might not accurately reflect district-level social connection patterns, particularly regarding connections between different districts within the same city. While SCI provided the second most significant influence in the district-level model for the overall population, it offered minimal predictive value in the city-level models analyzing Turkish and Syrian populations separately. This might be because Turkish and Syrian communities have distinct social networks that cannot be captured by aggregate SCI measures.

7.2 Limitations in long-term migration prediction

In the beginning, this study also planned to analyze long-term post-earthquake migration patterns. We first estimated users' home locations using CDR data from the three months preceding the earthquake, then compared these with their home locations during June, July, and August after the earthquake. A change in home location between these periods was classified as long-term migration.

Initially, we planned to follow an approach similar to our short-term migration analysis by first developing reliable regression models using our features, then analyzing how different features influenced post-earthquake long-term migration. However the models yielded R-square values lower than zero, indicating worse performance even than mean-value predictions. The poor performance might due to several factors. The first reason may be the methodological differences in processing long-term versus short-term migration data.

Short-term migration was determined by the earliest and latest phone signals within a week, while long-term migration was identified through changes in home locations. These distinct processing methods may cause models suitable for short-term migration not be applicable to long-term patterns. Secondly, mobility may be influenced by more complex set of factors as time progresses. The five datasets employed in this study may not provide sufficient information to capture long-term migration characteristics.

8. Conclusions

To conclude, in this study we have investigated post-earthquake evacuation patterns and their predictive factors. Through the implementation of machine learning gravity-transformed features, we have developed models capable of predicting evacuation flows.

Several features have been identified as important features for post-earthquake evacuation analysis. At the district level, population distribution is the primary factors, followed by earthquake intensity. Social connectedness index and relative wealth index showed moderate influence. The gravity-transformed features demonstrated strong predictive power, while our exploration of radiation-transformed features, although not optimal in the final model, provided valuable insights for future research.

Our comparative analysis of Turkish population and Syrian population revealed identical feature importance rankings but distinct pattern distributions. This suggests varying degrees of factor influence on evacuation decisions. Unfortunately, we were constrained by the limited availability of district-level population data for both Turkish and Syrian populations. This limitation lead us to classification approach rather than regression. The classification model focus on identifying features that predict higher flow counts. Despite these limitations, we have identified patterns in evacuation behaviors and established the effectiveness of gravity-transformed features in predicting population movements. The strong predictive power demonstrated by our classification models for both population groups suggests the viability of our approach for practical applications.

Therefore, our models validated the applicability of gravity-transformed features in studying post-earthquake evacuation flows, providing useful tools in the field of post-earthquake analysis. And the important features we fund provide valuable reference for other researchers studying post-disaster migration.

Bibliography

Ahas, R., Silm, S., Järv, O., Saluveer, E., and Tiru, M. (2010). Using mobile positioning data to model locations meaningful to users of mobile phones. *Journal of urban technology*, 17(1):3–27.

Anyidoho, P. K., Ju, X., Davidson, R. A., and Nozick, L. K. (2023). A machine learning approach for predicting hurricane evacuee destination location using smartphone location data. *Computational Urban Science*, 3(1):30.

Aydogdu, B., Salah, A. A., Ones, O., and Gurbuz, B. (2021). Description of the mobile cdr database. *HumMingBird Project Deliverable*, 6(1).

Bailey, M., Cao, R., Kuchler, T., Stroebel, J., and Wong, A. (2018). Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3):259–280.

Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J. J., Simini, F., and Tomasini, M. (2018). Human mobility: Models and applications. *Physics Reports*, 734:1–74.

Barlacchi, G., De Nadai, M., Larcher, R., Casella, A., Chitic, C., Torrisi, G., Antonelli, F., Vespignani, A., Pentland, A., and Lepri, B. (2015). A multi-source dataset of urban life in the city of milan and the province of trentino. *Scientific data*, 2(1):1–15.

Becker, R., Cáceres, R., Hanson, K., Isaacman, S., Loh, J. M., Martonosi, M., Rowland, J., Urbanek, S., Varshavsky, A., and Volinsky, C. (2013). Human mobility characterization from cellular network data. *Communications of the ACM*, 56(1):74–82.

Blondel, V., De Cordes, N., Decuyper, A., Deville, P., Raguenez, J., and Smoreda, Z. (2013). Mobile phone data for development-analysis of mobile phone datasets for the development of ivory coast. *Orange D4D challenge*.

Blondel, V. D., Decuyper, A., and Krings, G. (2015). A survey of results on mobile phone datasets analysis. *EPJ data science*, 4:1–55.

Blumenstock, J. E. (2011). Using mobile phone data to measure the ties between nations. In *Proceedings of the 2011 IConference*, pages 195–202.

Breiman, L. (2001). Random forests. *Machine learning*, 45:5–32.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Chi, G., Lin, F., Chi, G., and Blumenstock, J. (2020). A general approach to detecting migration events in digital trace data. *PloS one*, 15(10):e0239408.

Dujardin, S., Jacques, D., Steele, J., and Linard, C. (2020). Mobile phone data for urban climate change adaptation: Reviewing applications, opportunities and key challenges. *Sustainability*, 12(4):1501.

Erdik, M., Tümsa, M., Pınar, A., Altunel, E., and Zülfikar, A. (2023). A preliminary report on the February 6, 2023 earthquakes in Türkiye. *Research Briefs*.

Frias-Martinez, V. and Virseda, J. (2013). Cell phone analytics: Scaling human behavior studies into the millions. *Information Technologies & International Development*, 9(2):pp–35.

Ghurye, J., Krings, G., and Frias-Martinez, V. (2016). A framework to model human behavior at large scale during natural disasters. In *2016 17th IEEE International Conference on Mobile Data Management (MDM)*, volume 1, pages 18–27. IEEE.

Hong, L. and Frias-Martinez, V. (2020). Modeling and predicting evacuation flows during hurricane Irma. *EPJ Data Science*, 9(1):29.

Hussain, E., Kalaycıoğlu, S., Milliner, C. W., and Çakir, Z. (2023). Preconditioning the 2023 Kahramanmaraş (Türkiye) earthquake disaster. *Nature Reviews Earth & Environment*, pages 1–3.

Huynh, B. Q. and Basu, S. (2020). Forecasting internally displaced population migration patterns in Syria and Yemen. *Disaster medicine and public health preparedness*, 14(3):302–307.

Isaacman, S., Frias-Martinez, V., and Frias-Martinez, E. (2018). Modeling human migration patterns during drought conditions in La Guajira, Colombia. In *Proceedings of the 1st ACM SIGCAS conference on computing and sustainable societies*, pages 1–9.

Kargel, J. S., Leonard, G. J., Shugar, D. H., Haritashya, U. K., Bevington, A., Fielding, E. J., Fujita, K., Geertsema, M., Miles, E., Steiner, J., et al. (2016). Geomorphic and geologic controls of geohazards induced by Nepal's 2015 Gorkha earthquake. *Science*, 351(6269):aac8353.

Letouzé, E., Vinck, P., and Kammourieh, L. (2015). The law, politics and ethics of cell phone data analytics. *Data-Pop Alliance*.

Li, T., Dejby, J., Albert, M., Bengtsson, L., and Lefebvre, V. (2019). Detecting individual internal displacements following a sudden-onset disaster using time series analysis of call detail records. *arXiv preprint arXiv:1908.02377*.

Lu, X., Bengtsson, L., and Holme, P. (2012). Predictability of population displacement after the 2010 Haiti earthquake. *Proceedings of the National Academy of Sciences*, 109(29):11576–11581.

Martineau, J. S. (2010). Red flags: A model for the early warning of refugee outflows. *Journal of Immigrant & Refugee Studies*, 8(2):135–157.

Moumni, B., Frias-Martinez, V., and Frias-Martinez, E. (2013). Characterizing social response to urban earthquakes using cell-phone network data: the 2012 Oaxaca earthquake. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, pages 1199–1208.

Myers, C. A., Slack, T., and Singelmann, J. (2008). Social vulnerability and migration in the wake of disaster: the case of hurricanes Katrina and Rita. *Population and Environment*, 29:271–291.

Nair, R., Madsen, B. S., Lassen, H., Baduk, S., Nagarajan, S., Mogensen, L. H., Novack, R., Curzon, R., Paraszczak, J., and Urbak, S. (2019). A machine learning approach to scenario analysis and forecasting of mixed migration. *IBM Journal of Research and Development*, 64(1/2):7–1.

Narayanan, A. and Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*, pages 111–125. IEEE.

Phithakkitnukoon, S., Calabrese, F., Smoreda, Z., and Ratti, C. (2011). Out of sight out of mind–how our mobile social network changes during migration. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*, pages 515–520. IEEE.

Poot, J., Alimi, O., Cameron, M. P., and Maré, D. C. (2016). The gravity model of migration: the successful comeback of an ageing superstar in regional science.

Ravenstein, E. G. (1885). *The laws of migration*. Royal Statistical Society.

Salah, A. A., Pentland, A., Lepri, B., Letouzé, E., De Montjoye, Y., Dong, X., and Vinck, P. (2019). *Guide to mobile data analytics in refugee scenarios*. Springer.

Sarker, M. N. I., Peng, Y., Yiran, C., and Shouse, R. C. (2020). Disaster resilience through big data: Way to environmental sustainability. *International Journal of Disaster Risk Reduction*, 51:101769.

Steele, J. E., Pezzulo, C., Albert, M., Brooks, C. J., zu Erbach-Schoenberg, E., O'Connor, S. B., Sundsøy, P. R., Engø-Monsen, K., Nilsen, K., Graupe, B., et al. (2021). Mobility and phone call behavior explain patterns in poverty at high-resolution across multiple settings. *Humanities and Social Sciences Communications*, 8(1):1–12.

Steele, J. E., Sundsøy, P. R., Pezzulo, C., Alegana, V. A., Bird, T. J., Blumenstock, J., Bjelland, J., Engø-Monsen, K., De Montjoye, Y.-A., Iqbal, A. M., et al. (2017). Mapping poverty using mobile phone and satellite data. *Journal of The Royal Society Interface*, 14(127):20160690.

Stewart, J. Q. (1950). The development of social physics. *American Journal of Physics*, 18(5):239–253.

Suleimenova, D., Bell, D., and Groen, D. (2017). A generalized simulation development approach for predicting refugee destinations. *Scientific reports*, 7(1):13377.

Tufekci, Z. (2014). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 505–514.

Tweedie, M. C. et al. (1984). An index which distinguishes between some important exponential families. In *Statistics: Applications and new directions: Proc. Indian statistical institute golden Jubilee International conference*, volume 579, pages 579–604.

Wang, T., Chen, J., Zhou, Y., Wang, X., Lin, X., Wang, X., and Shang, Q. (2023). Preliminary investigation of building damage in Hatay under February 6, 2023 Turkey earthquakes. *Earthquake Engineering and Engineering Vibration*, 22(4):853–866.

Wesolowski, A., Eagle, N., Noor, A. M., Snow, R. W., and Buckee, C. O. (2013). The impact of biases in mobile phone ownership on estimates of human mobility. *Journal of the Royal Society Interface*, 10(81):20120986.

Wilson, R., zu Erbach-Schoenberg, E., Albert, M., Power, D., Tudge, S., Gonzalez, M., Guthrie, S., Chamberlain, H., Brooks, C., Hughes, C., et al. (2016). Rapid and near real-time assessments of population displacement using mobile phone data following disasters: The 2015 Nepal earthquake. *PLoS currents*, 8.

Yabe, T., Jones, N. K., Lozano-Gracia, N., Khan, M. F., Ukkusuri, S. V., Fraiberger, S., and Montfort, A. (2021). Location data reveals disproportionate disaster impact amongst the poor: A case study of the 2017 Puebla earthquake using Mobilkit. *arXiv preprint arXiv:2107.13590*.

Yabe, T., Jones, N. K., Rao, P. S. C., Gonzalez, M. C., and Ukkusuri, S. V. (2022). Mobile phone location data for disasters: A review from natural hazards and epidemics. *Computers, Environment and Urban Systems*, 94:101777.

Yabe, T., Tsubouchi, K., Fujiwara, N., Sekimoto, Y., and Ukkusuri, S. V. (2020). Understanding post-disaster population recovery patterns. *Journal of the Royal Society Interface*, 17(163):20190532.

Yin, L., Chen, J., Zhang, H., Yang, Z., Wan, Q., Ning, L., Hu, J., and Yu, Q. (2020). Improving emergency evacuation planning with mobile phone location data. *Environment and Planning B: Urban Analytics and City Science*, 47(6):964–980.