Passive Health Monitoring Using Large Scale Mobility Data

YUNKE ZHANG, Beijing National Research Center for Information Science and Technology, Tsinghua University
FENGLI XU*, CSE, Hong Kong University of Science and Technology
TONG LI, CSE, Hong Kong University of Science and Technology
VASSILIS KOSTAKOS, School of Computing and Information Systems, University of Melbourne, Australia
PAN HUI, CSE, Hong Kong University of Science and Technology
YONG LI, Beijing National Research Center for Information Science and Technology, Tsinghua University

In this paper, we investigate the feasibility of using mobility patterns and demographic data to predict hospital visits. We collect mobility traces from two thousand users for around two months. We extract 16 mobility features from these passively collected mobility traces and train an XGBoost model to predict users' hospital visits. We demonstrate that the designed mobility features can significantly improve prediction accuracy ($p < 0.01$, AUC = 0.79). We further analyze how these mobility features affect the prediction results and measure their importance by using Shapley additive explanation values. We discover that users with less mobility activity, less visit diversity, and few sports facilities, bountiful entertainment around their visited locations are more likely to visit hospitals. Moreover, we conduct predictions on the populations with different demographic features, which achieves meaningful and insightful results, i.e. maintaining a high mobility activity is crucial for older people's health, while fast food store more substantially affects younger people's health; visit patterns can indicate females' health, while the neighborhood environment is more indicative of males, etc. These results shed light on how to use and understand large scale mobility data in health monitoring and other health-related applications in practice.

CCS Concepts: • **Human-centered computing** → **Ubiquitous and mobile computing**; • **Information systems** → *Data mining*;

Additional Key Words and Phrases: Health monitoring, human mobility, mobile computing

ACM Reference Format:
Yunke Zhang, Fengli Xu, Tong Li, Vassilis Kostakos, Pan Hui, and Yong Li. 2021. Passive Health Monitoring Using Large Scale Mobility Data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 1, Article 50 (March 2021), 23 pages. https://doi.org/10.1145/3448078

*This is the corresponding author: fenglixu2020@hotmail.com

Authors' addresses: Yunke Zhang, zhangyk17@mails.tsinghua.edu.cn, Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing; Fengli Xu, CSE, Hong Kong University of Science and Technology, Hong Kong; Tong Li, CSE, Hong Kong University of Science and Technology, Hong Kong; Vassilis Kostakos, School of Computing and Information Systems, University of Melbourne, Australia; Pan Hui, CSE, Hong Kong University of Science and Technology, Hong Kong; Yong Li, Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2021 Association for Computing Machinery.
2474-9567/2021/3-ART50 $15.00
https://doi.org/10.1145/3448078

1 INTRODUCTION

People's interest in health monitoring has consistently grown in the past decade, with more than half of mobile phone users now have health applications to track their personal health status [18]. Currently, health monitoring mainly relies on questionnaires to regularly log self-assessments and custom-designed hardware to collect physiological signals [24, 40]. However, these methods are difficult to scale due to intensive human labor and high expense [6, 22, 24, 35]. Alternatively, mobility data, ubiquitously collected by pervasive mobile devices, offers a novel opportunity for passive health status monitoring. Compared with other data sources, mobility data not only has the advantages of passively collecting and monitoring but also provide a proxy to measure the level of user's physical activities and their interactions with various locations (e.g. park, gym, fast food restaurant, etc), which has been proven informative to infer health status [1, 8, 15]. In this work, we aim to explore the correlations between mobility data and users' health status and design a predictive model for large-scale health monitoring.

Extensive research efforts have been invested in designing efficient and effective health monitoring mechanisms [6, 10, 19, 24, 38]. However, most of them suffer from the following three important limitations. First, survey-based methods relies on recruiting a small number of participants [6, 24], which is considerably expensive to be applied to a large population and geographical coverage. This limitation in scale also cast doubts on the representativeness of the recruited participants. Second, previous works on health prediction by mobility data only use simple and limited mobility features as indicators of health status [6, 38], and did not consider the semantic meanings of the visited locations. Since the environment of individuals' visitations varies, confining the sights on mobility metrics is insufficient and comprehensive considerations need to be taken. Third, most works focus on several specific health issues, like depression, drug craving, and cardiac surgery readmission [10, 19, 38]. The understanding of the overall health status monitoring is vital in many scenarios yet woefully inadequate.

Specifically, there are three key challenges to conduct large scale health monitoring based on mobility data. First, collecting an appropriate mobility data source for health status monitoring is essential but challenging. Wearable sensors and survey-based data collection need to leverage dedicated devices and recruit participants, which limit the exploitation in large scale passive health status monitoring. As we aforementioned, mobility data is a potential option. Thus, we need to find out a method for large-scale passive collection of both mobility and health data. Second, modeling the complicated correlation between individuals' health status and mobility is difficult as well. Previous studies only worked on the surface, predicting specific health metrics, such as depression, bipolar disorder, and re-hospitalization after cardiac surgery [6, 28, 38]. However, in order to analyze overall health, deeper mining and understandings of mobility patterns are required. Third, the mechanism of some factors' influence on health status is still unclear and needs to be investigated. For example, the effect of the environment around the visited locations on human health is still unknown. Therefore, we need to delve into more profound mechanisms.

In this paper, we address the above challenges with three key designs. First, we collect a large scale mobility dataset covering users for nearly two months. It is collected from the mobility networks by recording the base station accessing of mobile users with high temporal and spatial resolution, which provides sufficient and fine-grained data for us to monitor human health status. Second, to seek more in-depth insights into mobility's predictability on health, we propose different features extracted from the mobility data. These features characterize mobility patterns from diverse aspects, including mobility willingness, mobility diversity, neighborhood environment, visit characteristics, etc. Thirdly, we conduct in-depth analyses by these features to analyze both the direct influence of each factor on health and the interaction between features. To summarize, the contribution of our work is threefold.

- We investigate the correlations between large scale passively collected mobility dataset and overall health status. It is feasible to extend our results to a larger scale of population.

- We define comprehensive classes of features that can be drawn from human mobility traces and train a model to predict users' visits to the hospital by using these features. The model reaches an AUC(Area Under Curve) of 0.79, which is significantly higher($p < 0.01$) than baseline models(AUC = 0.70).
- Our analyses provide deeper insights into the way of how mobility features influence health prediction. These findings offer implications on real-time large scale overall health monitoring.

2 DATA COLLECTION AND OVERVIEW

2.1 Data Collection

We collect the research data from a health survey conducted in 13 major hospitals in Beijing. From this survey, we gather the demographic features of age, gender, and income level from 4,321 outpatients who paid at least one visit to one of these hospitals, and we also collect the same information from 10,399 randomly sampled healthy persons as a control group. To study the correlation between a hospital visit and mobility data, we ask the participants to donate 2 months' worth of mobility traces extracted from the cellular network records, which cover the period from 5th July to 31st August 2017. The traces capture users' visits across the metropolitan area. Specifically, they contain geographic coordinates, arrival time, and stay time for each visit. The cellular network records the ID of mobile devices, the location of the base stations, and the data connection periods. It provides high spatial and temporal resolution in real-time. Previous work [45] has shown that data collected with this method offers comprehensive records of mobile users and is highly credible.

Considering further processing and calculation on mobility trace, we select a mobility record balanced population as experiment data. It consists of 1,056 outpatients and 1,056 healthy persons. We first filter out the traces with less than 20 records per day. Then we match the distributions of records on outpatients and healthy persons. Thus, the experiment population contains sufficient and unbiased trajectory information.

2.2 Ethical Considerations

Given the sensitivity of the data, we enforce the following protocols to address the privacy and ethical risks in data sharing and analysis. First, all survey participants are informed of the purpose of data collection and they explicitly authorize the usage for research purposes. The dataset is properly anonymized by the data collector before they are shared with researchers. The real user IDs are never made available to or utilized by the research project. Second, all the researchers that have been authorized to access the datasets are bounded by strict non-disclosure agreements, and our research protocols are approved by the local institutional board. Third, we store all the data in a secure off-line server, and only the authorized core researchers can access the data.

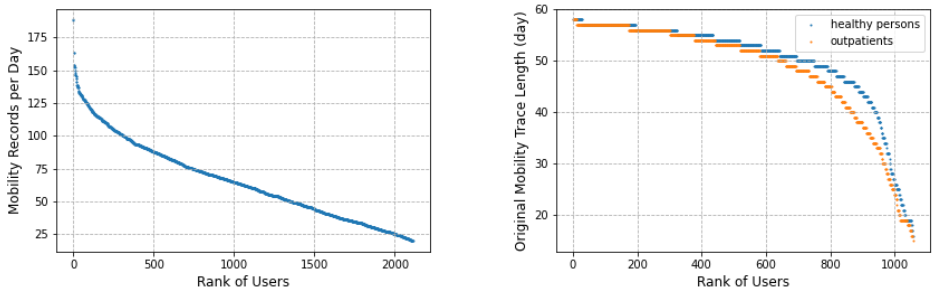
2.3 Data Statistics

We first provide an overview of population characteristics in our experiment. The 1,056 outpatients have visited various departments, including cardiology, internal medicine, respiratory, gastroenterology, neurology, etc. To study the relation between mobility pattern and overall health status, which is represented as an overall hospital visit, we do not distinguish chronic disease patients from other outpatients but treat all outpatients equally to investigate how to make predictions of future hospital visits risk by mobility features. The demographic distributions on both outpatients and healthy persons are showcased in Table 1. The analysis of how demographic features influence hospital visit prediction is discussed in section 4.2.1. We can observe some distribution bias in our dataset, such as female users account for 69% of the population. To handle this bias, we leverage SHAP value analysis, which will be introduced in section 3.5, which unbiasedly extracts each feature's additive influence on prediction. We also conduct experiments on female and male population separately in section 4.3, which exclude the noise brought by gender bias.

Table 1. The distribution of user demographic features and mobility trace length.

Feature	Healthy persons	Outpatients
Gender	female(76.4%), male(23.6%)	female(62.0%), male(38.0%)
Age	0~30(34.1%), 30~50(55.2%), 50~99(10.7%)	0~30(18.1%), 30~50(47.8%), 50~99(34.1%)
Income	Low(30.0%), Medium(38.7%), High(31.3%)	Low(28.3%), Medium(30.2%), High(41.5%)
Mobility Trace Length (days)	≤20(4.2%), 20~30(4.8%), 30~40(9.6%), 40~50(21.3%), >50(60.1%)	≤20(2.6%), 20~30(4.4%), 30~40(4.5%), 40~50(22.8%), >50(65.7%)

To provide a basic understanding of our collected mobility trace dataset, we showcase the distribution of mobility records per day and mobility trace length in Fig. 1. We draw the distribution of mobility records per day, which has identical distribution on two groups, on all users in Fig. 1(a). It measures the temporal resolution of the data. We can inspect that the mobility records per day follow a power-law distribution, with 63.7% of users (1,346) having at least 50 records per day. Besides, the trace length distributions of both outpatients and healthy persons are plotted in Fig. 1(b), with the users sorted by the number of days that the mobility trace covers. The stratified distribution of mobility trace length is provided in Table 1. We calculate the relative entropy [20], which measures how one probability distribution is different from another, between two distributions and the result is 0.001. Since a relative entropy close to zero indicates that the two distributions are almost identical, this confirms that trace lengths are almost evenly distributed among two groups of users. In addition, we can observe that 66.8% of users (1,410) have no less than 50 days of mobility records. These properties guarantee the high temporal resolution of the dataset, which concords with our expectations. As for the spatial resolution of the mobility trace data, they have a granularity of 200 to 300 meters.



(a) Mobility records per day (b) Mobility trace length

Fig. 1. The distribution of mobility trace characteristics.

To demonstrate the environment features in the neighborhood and on user's visits, we adopt the Point of Interest(POI) dataset collected from BaiduMap[1]. It contains the distribution of 1,474,504 POIs in Beijing. Geographic location, category, and sub-category are provided for each POI entry. It is feasible to calculate the distribution of specific category POI in an area. There are 18 categories and 357 sub-categories in total. The 18 POI categories are listed in Table 2.

[1] https://maps.baidu.com/

Table 2. POI category taxonomy

#	POI Category	#	POI Category	#	POI Category
1	food	7	sports	13	education
2	corporate business	8	vehicle	14	bank
3	governmental agency and public organization	9	healthcare	15	place name and address
4	shopping	10	hotel	16	infrastructure
5	living service	11	scenic spot	17	real estate
6	entertainment	12	cultural venue	18	indoor facility

3 METHOD

One of the primary goals of our work is to investigate the relationship between users' mobility traces and their health status. Therefore, we first formulate and interpret this problem in a computational and statistical vision. As mentioned in the data collection section, the users are labeled with hospital visitors and non-visitors. We extract various features from mobility trace data. In this way, the research question can be rephrased as to which features are of a higher success probability to classify a user into a visitor or non-visitor group. In this section, we first introduce the formal definition of mobility traces and the details of preprocessing procedures. Next, we define a set of features extracted from mobility traces and cluster them into several groups by using k-Medoids. After identifying all essential mobility features, we adopt a statistical learning model to make the health status prediction.

3.1 Mobility Data Definition

The mobility traces consist of a sequence of points the user has visited, sorted by the user's arrival time. Each visited point is represented as a tuple containing the geographic coordinates, arrival time, and the stay time of the user. We define a visited point as follows,

$$P =< L, t^a, t^s >, \tag{1}$$

where L is the longitude-latitude pair of the visited location; t_a is the arrival time; t_s is the stay time. The time is recorded in minutes. Formally, we defined the mobility trace of a specific user as

$$T = (P_1, P_2, ..., P_{N(T)}), \tag{2}$$

where $N(T)$ is the number of points on this user's mobility trace. We denote the longitude-latitude pair of the ith point on the trace T as L_i. Similarly, we define t_i^a and t_i^s as the arrival time and stay time of the ith location point, respectively. The visited points are sorted in terms of arrival time. Thus, the mobility traces satisfy that $t_i^a + t_i^s \leq t_{i+1}^a, \forall i = 1, ..., N(T) - 1$. Empirically, we consider the time gap between two consecutive visited points as the period when the user moves from the former to the latter point.

As for the POI dataset, each POI entry has the name of POI, geographic coordinate, category, and subcategory information. Specifically, we denote the geographic coordinate of the ith POI as PL_i, the category of the ith POI as C_i, and the subcategory of the ith POI as SC_i. For a specific location, we define a particular (sub)category's POI density around that location as the number of that (sub)category's POI.

3.2 Data Preprocessing

To predict users' likelihood of being as a patient and visiting a hospital, we need to extract well-defined and meaningful features from the collected mobility traces. In this respect, we preprocess traces by using three approaches: trace reduction, calibration, and mean-field filtering.

First, as the problem we focus on is to predict the likelihood of users' hospital visits, the adopted features should be computed based on the traces before the hospital visit. Therefore, we first determine the time when the patients first visit one of the collected 13 hospitals and then remove their subsequent mobility records. More specifically, for each of the 13 hospitals where visitors in our dataset have visited, we fetch the geographic coordinates of its boundary by using BaiduMap[2]. We next regard the union set of the 13 polygons enclosed by fetched points as the hospital region, notated as H. Therefore, in terms of the geographic coordinate of each point a user has visited, we can determine whether it is located in the hospital region. We then define the hospital stay time of a visited point in a visitor's mobility trace as

$$HosStay(P_i) = \begin{cases} t_i^s, & L_i \in H \text{ and } t_i^a \in [8:00, 17:00] \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

This function will determine the time of a user's visit to hospitals during the open hours (8 am to 5 pm). Then, we split the mobility traces with respect to the date of each visited point and compute the user's total visit duration on different days. The divided traces are denoted as $T_1, ...T_{58}$. If the user stays in the hospital for at least 20 minutes on a specific day, that day will be labeled as a visit day.

$$HosVisit(T_i) = \begin{cases} 1, & \sum_{P \in T_i} HosStay(P) > 20 \\ 0, & otherwise \end{cases} \tag{4}$$

Given the hospital visit days derived from 1,056 hospital visitors' mobility traces, we demonstrate the total number of visitors each day in Fig. 2. All users paid their first visit after July 15th, which guarantees a pre-visit period that may contain latent signals of a future hospital visit.

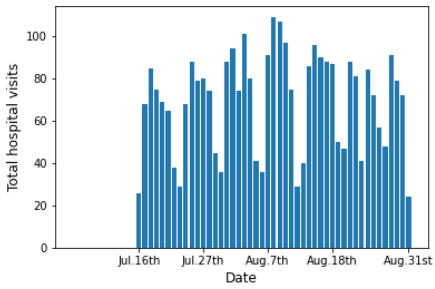


Fig. 2. Number of hospital visits each day.

After determining the visit days of users, we remove the mobility traces after the first visit day, i.e., only the visited points before the first visit day are kept. For visitors, the first visit day is expressed as $FV(T) = min(\{i|HosVisit(T_i) = 1\})$. As for non-visitors, their traces of all 58 days are preserved and the $FV(T)$ equals 59. Thus, the reduced trace is defined as:

$$T_r = (T_1, T_2, ..., T_{FV(T)-1}). \tag{5}$$

[2] https://api.map.baidu.com/lbsapi/getpoint/index.html

As users' location information in our dataset is determined by the associated base stations, sometimes there are noise points on the collected mobility traces. For instance, continuously short-time shifts between two distant base stations have been detected in our dataset. These noise points may influence the accuracy of feature extractions. To mitigate the negative effects, we recognize the abnormal movements in users' mobility traces and remove these noise points.

To identify these noise movements, we computed the velocity of a user moving from one point to the next in the mobility trace. As mentioned in section 3.1, the time gap between two adjacent points is considered as the moving time. Thus, the velocity of the user moving from ith point to the next point is computed as,

$$v_i = \frac{dist(L_i, L_{i+1})}{t^a_{i+1} - t^a_i - t^s_i}, \tag{6}$$

where $dist(L_i, L_{i+1})$ is the distance between two locations L_i and L_{i+1}. In our case, we regard the movement with velocity below 2 kilometers per minute as normal ones. Thus, we remove the points with abnormal movements correspondingly. Hence, the calibrated mobility trace T_c is defined as,

$$T_c = \{P_i | max(v_i, v_{i-1}) < 2000m/min\}. \tag{7}$$

Another type of false movement in mobility traces is the small-scope shift. For example, in some mobility traces, the pattern of continuously moving from two locations no further than 1000 meters at night is detected. This is caused by unstable positioning. We need to exclude these false shifts from mobility traces as well.

Apart from false shifts, users' true movements in a small range may also bring confusion. For example, a user wanders around one visited location is often captured in our dataset. This wander behavior should not be considered as valid movements of users. Thus, in [6], the researchers applied a state machine to determine whether the users are really moving between different locations visited. In our study, we capture the effective movements by leveraging mean-field filtering. In mean-field filtering, the filtered trace T_f is initialized as an empty set. We denote the current point cluster as (P_1) where P_1 is the first point in T_c. Then, for each point in the T_c, it will be appended into the current point cluster, if its distance from the centroid of the current point cluster is no longer than 1 kilometer. Otherwise, if the distance is larger than 1 kilometer, this movement is regarded as invalid. The current point cluster and the arrival time of the first point will be combined as a new point in T_f. At last, we add the centroid of the final current point cluster into the filtered trace. The centroid of the current point cluster is the weighted average of each point's coordinate with the stay times of each point as weights. The algorithm is shown in Algorithm 1.

ALGORITHM 1: Mean-field Filtering for Mobility Traces.

Input: The cleaned trace T_c
Output: The filtered trace T_f
$T_f \leftarrow \{\}$, $CurrentCluster \leftarrow \{P_1\}$, $FirstPointTime \leftarrow t^a_1$;
for *i=2:N(T_c)* **do**
  if $dist(Centroid(CurrentCluster), L_i) \leq 1000$ **then**
    $CurrentCluster.append(L_i)$;
  else
    $T_f.append((Centroid(CurrentCluster)), FirstPointTime)$;
    $CurrentCluster \leftarrow \{P_i\}$;
    $FirstPointTime \leftarrow t^a_i$;
  end
end
$T_f.append((Centroid(CurrentCluster)), FirstPointTime)$;

3.3 Feature Design

Since our study aims to discover the role of the pre-processed mobility data in predicting users' health status, we next extract features from mobility traces to characterize health-related patterns. Previous literature [6, 15] has considered a wide range of mobility features to capture the mobility characteristics of people with poor health status. Based on their results and other works related to mobility patterns, we design the following features, which are assigned to a specific user's filtered mobility trace.

1) **The radius of gyration** ρ: we define this feature to depict the mobility vigor of a user. As discussed in [6, 32], the pattern of reduced mobility is associated with depression. In [4], sedentary behavior can improve the prediction of hospital readmission. It is also common when sick people reduce their movement, resulting in a smaller range of activities. Thus, we pick the radius of the gyration of users' trace to capture the coverage area of users' mobility. This feature is defined as the deviation of visited points in the mobility trace from its centroid L_C, which can be calculated as follows,

$$\rho = \sqrt{\frac{1}{N(T_f)} \sum_{i=1}^{N(T_f)} dist(L_i, L_C)}. \tag{8}$$

where the centroid L_C is the average of geographic coordinate L_i's of all points on a filtered trace.

2) **The standard deviation of displacements** σ: the radius of gyration characterizes users' mobility levels. More generally, [6, 19] discussed that people with a mental disorder might lack the willingness to visit many places. This gives us the insight that people with sub-health status may have more regular movement and following daily routines instead of impromptu activities like watching a movie after work. To characterize such regularity of movement, we introduce the standard deviation of displacements. Displacements are defined as the distances between two consequent points on a trace, $D_i = dist(L_i, L_{i+1})$. The average displacement is notated as $\bar{D}$. Then, we can calculate the standard deviation of displacements as follows,

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N(T_f)-1} (D_i - \bar{D})^2}{N(T_f) - 1}}. \tag{9}$$

3) **Distribution entropy of locations visited** ϵ: besides the regularity of movements, we also extract the feature of distribution entropy of locations visited to capture the diversity of users' movements. This lack of diversity is expressed as having less distribution in impromptu visits in mobility trace. We adopt a quantitative metric, the distribution entropy, to depict this condensation. To illustrate the distribution of locations visited, we segment the city of Beijing into 300 meters by 300 meters region grids. For each visited point in the mobility trace, we assign it to the region it locates in. We then take the distribution of regions visited as the distribution of locations. Let n_i be the number of points in the filtered trace that fall in the ith region. Then the entropy is defined as

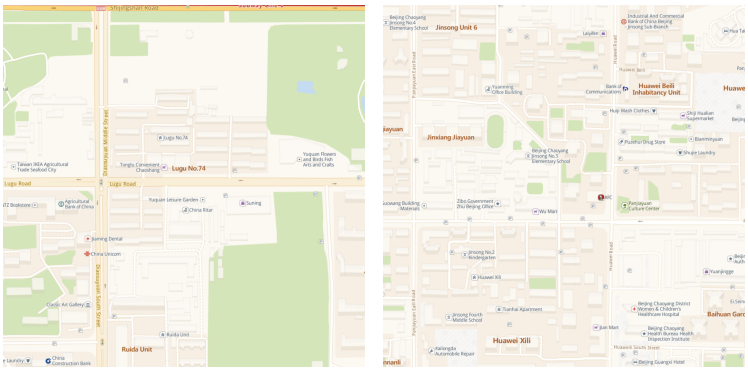
$$\epsilon = 1 - \sum_{i=1}^{\text{number of regions}} (\frac{n_i}{N(T_f)})^2. \tag{10}$$

For the users with locations visited or with a number of visits taking up the majority of locations on mobility trace, they will have smaller ϵ's.

4) **Area of green spaces in neighborhood** $N(green)$: Hasthanasombat *et al.*[15]has discussed the effect of neighborhood environment on public health. The accessibility to nature is considered to be linked with mental and physical health. More areas of green spaces in the neighborhood can trigger more visits to them. Thus, we select the area of green spaces in the neighborhood as one of the features that may help to predict people's health status. We first determine the location of a user's home by figuring out the place the user appears most during the night. We adopt the segmentation regions introduced when calculating E. For each point P_i in T_f, if the arrival

time t_i^a lies at 9 p.m. to 7 a.m, we will increase the count of the region P_i falls in by one. Eventually, we choose the region with the most visited points that appeared as the user's home, denoted as h. We then extract the map around h from BaiduMap, which is 1,000 pixels by 1,000 pixels picture covering an area of 500 meters by 500 meters. An example map is shown in Fig. 3. Then, we count the pixels with the color that denotes the green spaces area and set it as $N(green)$, the area of green spaces in the neighborhood.

It is worthwhile to mention that we also adopt the POI database to count the number of scenic POIs which include parks and green areas in the neighborhood. However, it differs from $N(green)$ in several aspects. $N(green)$, the area of green spaces in the neighborhood, depicts the abundance of green spaces, while the POI density captures the number of green spaces nearby. Besides, some small green spaces are not counted as a POI. To cover comprehensive factors, we consider both of these features.



(a) Neighborhood with large green space area.

(b) Neighborhood with small green space area.

Fig. 3. Examples of neighborhood maps extracted from BaiduMap.

5) **POI density in neighborhood** $N(food), N(ent), N(sport), N(scene), N(fast), N(tob)$: aside from the green spaces in the neighborhood, several other factors are considered to be related to health status in previous literature. For example, [13] discussed that the density of tobacco stores in the neighborhood might affect life expectancy. Also, [15] worked on the effect of sports facilities in the neighborhood on public health. [25, 39] looked over the relationship between food venue density, fast food environment, and health indices like BMI and waist circumference. Thus, we pick 4 POI categories - food, entertainment, sports, and scenic spots, and two subcategories - fast food and tobacco/liquor shop, as potential health-related POIs. It is worth mentioning that food is parallel with other categories, while fast food is a subcategory that is regarded as health-related in literature. We consider both of them in our model and examine the differences between their contributions in hospital visit prediction. We regard the number of POIs in a specific (sub)category distributed in the location as that location's environmental feature. For instance, the food POI density around a user's home is formally defined as follows,

$$N(food) = \#\{i | dist(h, PL_i) < 500m, C_i = food\}. \tag{11}$$

The POIs are counted within 500 meters around users' homes. Specifically, we can obtain the entertainment, sports, scenic spot, fast food store, and tobacco/liquor shop POI density in the neighborhood denoted as $N(ent)$, $N(sport)$, $N(scene)$, $N(fast)$, $N(tob)$, by applying the same method.

6) **POI densities around visited points** $V(food), V(ent), V(sport), V(scene), V(fast), V(tob)$: apart from the neighborhood, the places where user visits can also reveal users' health status. Instead of exercising in sports facilities in the neighborhood, users' may visit gyms or stadiums in other regions. Hence, we further treat the

environment around locations visited by users as another essential feature. For each user's mobility trace, we calculate the POI density of each visited point and take the average as trace features. Moreover, we assign each point with a weight of its distance from the user's home when computing the average. In this way, the food POI density around visited points denoted as $V(food)$ can be calculated as follows,

$$V(food) = \frac{\sum_{i=1}^{N(T_f)} dist(L_i, h) Den(L_i, food)}{\sum_{i=1}^{N(T_f)} dist(L_i, h)} \tag{12}$$

where $Den(L_i, food) = \#\{j | dist(L_i, PL_j) < 500m, C_j = food\}$.

Similarly, we further obtain $V(ent), V(sport), V(scene), V(fast)$ and $V(tob)$, which represent the entertainment, sports, scenic spot, fast food store and tobacco/liquor shop POI density around visited points respectively.

We note that visitors' mobility traces are cut off before their first hospital visits, as introduced in section 3.2. There are sufficient pieces of evidence that confirm this pre-process does not influence our designed features. First, the demographic and neighborhood features are not related to trace length. Second, Song *et al.* [36] demonstrate the ultra-slow diffusion property in human mobility modeling, in which the mean square displacement appears to follow a slower than logarithm growth. The radius of gyration and the number of places visited both converge rapidly with respect to time. These guarantees that radius of gyration ρ, the standard deviation of displacement σ, distribution entropy ϵ, and POI densities around visited points are influenced minutely by trace length when it is over 100 hours. Thus, all considered features are not affected by the trace reduction process.

3.4 Feature Clustering

By combing the above designed 16 features extracted from mobility traces and the collected 3 demographic features, i.e., gender, age, and income, we obtain a set of 19 features for each user. To construe the intrinsic relation of these features and analyze their roles in health status prediction, we adopt the k-Medoids algorithm [29], which is robust to noises and outliers, to cluster them. We aim to cluster highly related features in the same group. Thus, we use the absolute correlation coefficient to measure the similarity of features.

As the cluster number k is a hyperparameter, we adopt the Silhouette coefficient [21] to evaluate the clustering performance and determine the optimal k. In practice, we obtain the best Silhouette coefficient at $k = 4$. The clustering result is shown in Table 3. Age, gender, and income form the first group. The second group consists of the radius of gyration, the standard deviation of displacements, and the distribution entropy of visited locations. Moreover, green spaces in neighborhood and POI density in neighborhoods compose the third group. The final group consists of POI densities around visited points.

The clustering result provides several interesting insights. In practice, we assign a semantic label for each group. The first group consists of demographic features of users. The second group is the mobility-related features that capture users' mobility coverage and diversity. Features in the third group depict the environment in users' neighborhood. Thus, we label it as neighborhood environment features. The fourth group is recognized as visitation environment features. In Fig. 4, we depict the absolute correlation heatmap of features. Obviously, we can discern four distinct clusters. Thus, we treat each feature group as a whole to present their physical meaning in the following analyses.

3.5 Health Status Prediction

To classify users into health and un-health, we train a binary XGBoost classifier to classify the users [7]. We adopt 5-fold cross-validation for each experiment and select the ROC AUC, F1-score as the metrics to evaluate model performances. We conduct Welch's t-test [41] to determine statistically significant differences in model performance.

Table 3. Clustering result of k-Medoids algorithm

Demographic Features	Mobility Features	Neighborhood Environment Features	Visitation Environment Features
Age, gender, income	ρ, σ, ϵ	N(green), N(food), N(ent), N(sport), N(scene), N(fast), N(tob)	V(food), V(ent), V(sport), V(scene), V(fast), V(tob)

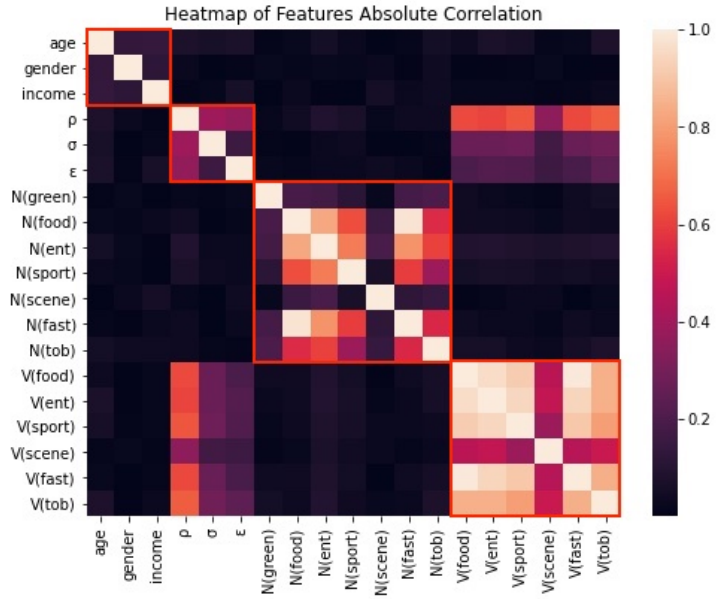


Fig. 4. The absolute correlation heatmap of 19 features.

To interpret the contribution of features in the boosting model, we adopt the analysis framework of SHAP (Shapley additive explanation) values [23]. Like Shapley value, the SHAP value can measure the additive feature's importance for a particular prediction, which helps us analyze and understand how the feature affects the prediction. The Shapley value is defined as the average of model output change when introducing a specific feature on all possible feature orderings, which is computed as follows,

$$\Phi_i(f, x) = \sum_{R \in \mathcal{R}} \frac{1}{M!} [f_x(P_i^R \cup i) - f_x(P_i^R)], \tag{13}$$

where i is the feature, x is a specific input, f is the model, M is the number of input features in the model, $\mathcal{R}$ is all $M!$ possible feature orderings, P_i^R is the set of all features come before i in the ordering R, and $f_x(S)$ is the conditional expectation function of model's output when the features S are conditioned as in input x.

SHAP value provides an opportunity to robustly interpret the boosting model with accuracy and consistency. A feature with a positive SHAP value in a particular user's prediction is considered to lead to a higher probability of predicting the user be a visitor. On another scale, a feature with a higher absolute SHAP value in a specific prediction is considered to have a higher contribution to it than other features.

4 RESULTS

In this section, given the set of demographic and mobility features and users' hospital visit patterns, we conduct extensive experiments to answer the following questions: is it possible to monitor health status by mobility trace? To what extent and in what way the features affect the prediction? How do mobility features and demographic features interact in the prediction?

In order to answer the above questions, we first demonstrate the model performances of different groups of features as input, illustrating the significant assistance of mobility trace. Then we analyze each feature's impact on the prediction based on a quantitative method. We investigate how different features contribute to the outcome of the model. With similar means, we examine how feature importance varies in different demographic groups. Deeper insights can be drawn from the analysis. Finally, we inspect how model performance varies with the length of mobility trace, which is instructive to further application.

4.1 Prediction Performance

To answer the essential question of whether possible to monitor health status by mobility traces, we set up a predictive analysis with collected features. To quantify the performance, we adopt the widely used metrics of AUC and F1-score [31], where the higher the value, the better the performance.

We conduct six experiments with different combinations of feature sets as input. First four experiments with four feature groups respectively. The fifth model's input is all features collected from mobility traces. Finally, we conduct the prediction with both demographic features and mobility trace features. We tune the model's hyperparameters with cross-validation to achieve the best performance. The results are shown in Table 4. We can observe that mobility features and visitation environment features are with the lowest performances. Demographic features result in the best performance among four single group predictions. Its AUC, 0.7006, is significantly higher than that of the other three groups ($p < 0.01$). This indicates that demographic features alone provide enough information for health monitoring. Also, only using mobility trace features cannot predict hospital visits better than demographic features. This matches the challenge of the complexity of modeling the relationship between health status and mobility. Therefore, we have to combine different features in multiple semantics.

More importantly, we can observe that the combination of mobility trace features(AUC=0.7271, F1-Score=0.6747) outperforms the prediction made with demographic features(AUC=0.7006, F1-Score=0.6467). This provides strong evidence that passively collected mobility data can be used to predict users' health status, and even better than demographic features. To understand how the mobility data interact with demographic features, we investigate the performance of combing all the features. We recognize that prediction with all features has the best performance among all experiments with 0.7957 AUC and 0.7206 F1-Score. The significance level p is lower than 0.001. By combining all of our findings, we can conclude that it is possible to use mobility traces to monitor health status. Moreover, mobility data can improve the performance of demographic features significantly.

Table 4. Predicting hospital visit with different feature groups.

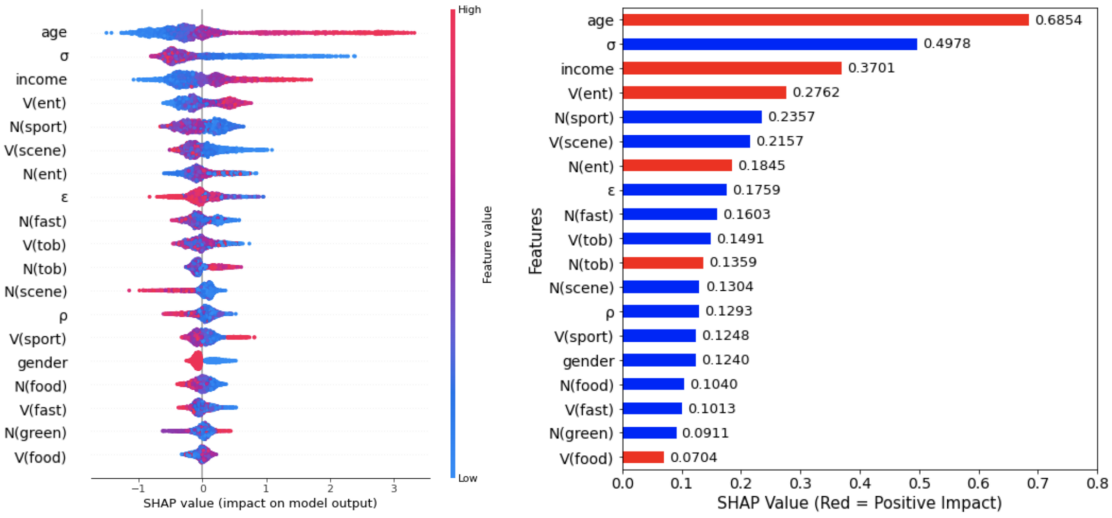
Features	AUC	F1-Score
Demographic Features	0.7006	0.6467
Mobility Features	0.6080	0.5710
Neighborhood Environment Features	0.6476	0.5947
Visitation Environment Features	0.6105	0.5819
Mobility Trace Features	0.7271	0.6747
All Features	**0.7957**	**0.7206**

4.2 Feature Analysis

Although we have illustrated that mobility data can benefit health monitoring, the prediction mechanism is still unclear. To understand the extent and how the features affect the prediction, we have to take a deeper analysis. We adopt SHAP value to interpret the features' impact in the XGBoost model.

As introduced in section 3, SHAP value provides us a chance to seize precise and deep insights from the prediction results. Generally speaking, features with higher absolute SHAP values are more critical in the model, since they impact the prediction result most. In another aspect, the feature's impact on the prediction can be analyzed by the correlation coefficient of that feature's value and SHAP value. If the correlation is positive, a higher feature value results in a positive SHAP value. Correspondingly, the prediction will be more likely to be 1, which is the label of visitors in our mode.

The SHAP values of different features are depicted in Fig. 5(a). For each feature, the SHAP values of 2,112 users are presented in a horizontal line, and the color denotes feature values. Red points refer to higher values, while blue points represent lower values. Specifically, for the feature of gender, female users are denoted as red points. We then compute the average absolute SHAP value of each feature and the correlation of feature value and SHAP value, as shown in Fig. 5(b). Features with red bars have a positive correlation coefficient. The higher value in these features, the more likely the model predicting a hospital visit. As for features with blue bars, a higher feature value is related to a lower likelihood of hospital visit prediction. From Fig. 5(a), the four most important features are age, the standard deviation of displacements σ, income, and entertainment category POI density around visited points. Their average absolute SHAP values are higher than 0.25. Nevertheless, the four features with the least importance are food and fast food category POI around visited points, food category POI in neighborhoods, and green space area in neighborhoods. Their average absolute SHAP values are lower than 0.11. We will analyze the features by groups after investigating the importance of these features in the prediction test.



(a) A summary of features' SHAP values

(b) Feature's average absolute SHAP values, where color denotes positive/negative impact on prediction

Fig. 5. SHAP value plots of the XGBoost model

To demonstrate the importance of these features, we conduct two ablation experiments. In the first experiment, we only use the four most important features as inputs to predict hospital visits. As for the second one, we adopt all features without the four features with the least importance. The results, compared with the scenario taking all features as input, are shown in Table 5. The AUC of using the four most important features to predict hospital visitation is 0.7316, which is significantly lower than using all features($p<0.01$). However, it outperforms the model using demographic features and the model using all mobility trace features. This implies that introducing one important mobility feature and one important visitation environment feature can enhance the prediction. Still, more features are needed to enhance it substantially. Also, we can observe that removing the four least important features do not decrease the performance obviously($p=0.9$). This finding indicates that it is reasonable to treat these features as unimportant features in the following analysis.

Table 5. Ablation study considering the different feature importance.

Features	AUC	F1-Score
All Features	**0.7957**	**0.7206**
4 Most Important Features	0.7316	0.6704
W/o 4 Least Important Features	0.7922	0.7149

4.2.1 Demographic Features. As shown in Fig. 5(a), age and income are positively correlated to hospital visits and are among the four most important features. The result demonstrates that older people tend to visit the hospital, which fits our empirical experience. Older people are more likely to fall ill than robust young people. As for gender, male users have positive SHAP values in Fig. 5(a), which corresponds to the truth that the life expectancy of women in China is five years more than men. Besides, our model indicates that people with higher income levels are more likely to be predicted of visiting hospitals. This finding contradicts our intuition that wealthy people are healthier and the study displays that rich people have longer life expectancy [9]. However, we have to mention that the hospitals in our experiment are for pay, outpatient registration fees cannot be covered by medical insurance and medical service fees are only partially covered. This situation should be considered when analyzing the relation between income and hospital visits. Studies [47–49] have shown the healthcare inequity related to socioeconomic status in China. Individuals who worked in low-income occupations are at a disadvantage of outpatient expenditure. This is consistent with our model, where wealthy people are more likely to seek hospital treatment.

4.2.2 Mobility Features. Previous papers have shown that the reduction in mobility is associated with mental health problems like depression [6]. In our analysis, the radius of gyration ρ's value is positively correlated to its SHAP value. Users with a higher coverage area of movements are less likely to be predicted as hospital visitors, while people demonstrating inactive mobility are having a higher possibility of being sick. The mobility range's impact on general health status is consistent with it on mental health. However, some studies [30, 43] state that larger coverage area is correlated with higher socioeconomic status, which contradicts with our model where income and ρ have different signs in SHAP value-feature value correlation. In fact, a recent study [46] has discovered both positive and negative correlation between socioeconomic status and gyration radius. In typical American cities like Boston where rich people live in the suburb, the conventional wisdom holds. While in Singapore, a polycentric city, wealthy people have lower work-residence distance and travel less. Thus, the result in Beijing, another typical Asian city similar to Singapore, concord with the situation in Singapore. σ and ϵ are designed to illustrate the regularity and diversity of movements. Both features are negatively correlated to their SHAP value and are highly ranked in terms of feature importance. This finding corresponds to the presupposition that a decrease in users' mobility diversity can reveal their unhealthy situation. A possible explanation is that

people in worse health status are less likely to make impromptu visits like dating in a far restaurant or go shopping. Instead, they follow their daily routine and visit limited places, thus having lower σ and ϵ. To conclude, mobility features can indicate users' vigor, therefore implying users' health status. The symbolic features of visitors are the low radius of gyration, low standard deviation of displacements, and low distribution entropy of locations visited.

4.2.3 Neighborhood Environment Features. This feature set contains the POI density and the green space area in users' neighborhood. From 5(b), we observe that the importance ranking within this group is sports POI, entertain POI, fast food POI, tobacco/liquor POI, scenic POI, food POI, and green space area. Among three groups of mobility trace features, the neighborhood environment features' prediction on hospital visitation has the best performance. Moreover, we observe that the density of sports facilities in the neighborhood is negatively correlated to its SHAP value in predicting hospital visits. Similarly, more scenic sights and green areas in the neighborhood is positively correlated to users' healthy status. By contrast, people living in a neighborhood with abundant entertainment, fast food stores, or tobacco and liquor shops are more likely to be in worse health status. These findings indicate how the neighborhood environment affects our health. More accessibility to traditional considered healthy activities such as exercising in sports clubs, jogging in a delightful park indeed enhance our health status. Also, empirically considered unhealthy places, such as noisy recreational entertainment places, junk food, and tobacco, are indeed correlated to harmed health status.

4.2.4 Visitation Environment Features. This class of features characterizes the environment of locations users visit, which differs from the neighborhood environment. This group of features does not perform as well as neighborhood environment features in predicting hospital visits. However, interesting insights can be drawn from the feature importance analysis.

Similar to the result of home environment features, densities of scenic sights and sports facilities around users' visitations are negatively related to hospital visits. However, the number of tobacco and liquor store around places users visited is negatively correlated with their SHAP values. This contradicts empirical cognition. By observing Fig. 5(a), there are red points with positive but lower SHAP values for $V(tob)$. Then we check the correlation between this $V(tob)$'s values and SHAP values, which is -0.02, an almost irrelevant relationship. Thus, we regard this feature as an irrelevant feature.

The most important feature in this group, by contrast with $N(sport)$ in the neighborhood environment features group, is $V(ent)$. This variation indicates the status difference of the user's home and visits. It is common to go to entertainment places far from home, which often happens more frequently than going to these in the neighborhood. Conversely, we tend to go to gyms near our home instead of one far away from the neighborhood. These differences result in the order of feature importance. But in any case, entertainment places have been negatively correlated with health status.

To sum up, we investigate how features affect hospital visit prediction. By the analysis, a typical hospital visitor is likely to have the following features: inactive moving level (low ρ, low σ and low ϵ), few sports facilities, green spaces, and scenic sights in the neighborhood, bountiful entertainment places, and tobacco/liquor shop near home, few sports facilities and scenic sights around places visited and many entertainment visits.

4.3 Demographic and Mobility Interaction

Our results so far show that demographic features and mobility features have no substantial difference in prediction performance, and significant improvements only occur when we combine them together. In the full model, demographic features such as age and income score high, and their correlation analysis corresponds with our empirical understanding. Yet, based on the analysis of demographic features, more questions can be proposed. We already know that elders are more likely to go to a hospital, but what if we only consider them in

the analysis? Does the feature importance order change? For example, is sport POI in the neighborhood still important? To answer these questions, we carry out an analysis focusing on specific demographic groups.

Given that user age is the most important feature in our analysis, we divide the population into three groups: people younger than 30, people ages between 31 and 50, and those over 50. In addition, we apply our analysis to male users and female users separately. For each group, we record the AUC score and the six most important features in the prediction. The results are shown in Table 6. The signs denote the directions of features' correlation with hospital visits.

Table 6. Top 6 important features in different demographic groups.

Demographic groups	**AUC**	**Top 6 important features**
>50 years	0.7164	$\sigma(-), V(ent)(+), N(scene)(-), N(green)(-), \rho(-), V(tob)(-)$
31~50 years	0.7293	$\sigma(-), income(+), V(scene)(-), N(sport)(-), V(ent)(+), N(ent)(+)$
≤30 years	0.7225	$\sigma(-), income(+), N(fast)(+), \epsilon(-), N(green)(-), gender(-)$
Male	0.7545	$age(+), \sigma(-), income(+), N(sport)(-), N(food)(-), N(ent)(+)$
Female	0.7723	$age(+), \sigma(-), income(+), V(ent)(+), \epsilon(-), V(scene)(+)$

The results show that certain patterns remain consistent with the full model. For each age group, the most important feature is the standard deviation of displacements σ, which is the second most important feature in the full model. This suggests that the regularity of users' mobility is consistently a good health indicator for people in all age groups. For the 30-50 age group, the six most important features are exactly the second to the seventh important features in the full model. This consistency suggests that the age group serves as a good representative of the whole population.

More insights can be drawn by considering the divergence from the full model. For people over 50, income is no longer among the most important features. This can be explained in the sense that income level does not determine hospital visits as it only for younger people. Older people, who are more likely to become sick, are less likely to consider their financial status when deciding to visit a hospital's treatment, or alternatively, their illness may be severe and a hospital visit becomes absolutely necessary. $N(sport)$ is also less important for people over 50, while $N(green)$ and ρ are listed in the most important features. This indicates that green space and mobile coverage are more important for elder people's health, while sports facilities become less important. This represents a preference shift from the 30-50 age group: for older people, green spaces are more important than sports clubs. Furthermore, older people need to maintain a certain level of mobility to keep healthy, thus a low mobility coverage is more likely to be a signal of unhealthy status in elders than in younger users. For young people less than 30, $N(fast)$'s importance is substantially higher and it is positively correlated to hospital visits. This finding indicates that as the main target of fast food stores, young people are indeed affected by them. Given that our analysis shows that young males are more likely to visit hospitals than young females, the results indicate the need for young males to manage their visits to fast food venues. The distribution entropy of locations visited is also important in this group, which indicates that younger people's health status depends on the diversity of their visits a lot.

Aside from age analysis, we can observe interesting insights from our gender analysis. According to Table 6, both men's and women's hospital visits are mostly influenced by age, the standard deviation of displacements, and income. However, the comparison between the following three features indicates certain differences between the two group's health. $V(ent)$, the distribution entropy of locations visited and $V(scene)$ are important in predicting women's hospital visits. These features are all related to the diversity of mobility traces. Hence, it is essential for women to extend their willingness to visit various places, however, they also have to keep away from too many

entertainments and stay with scenic sights to enhance their health status. In contrast, the important features for men are all related to the neighborhood environment. Men living in a neighborhood with sports facilities, restaurants, and fewer entertainment places are more likely to avoid hospital visits. This comparison provides interesting insights into gender's interaction with mobility trace features.

4.4 Impact of Mobility Trace Length

Aside from the feature analysis, the data attributes can also influence the performance of our prediction. To understand this influence and discover the best data that can be applied in further application, we divide the users into four groups by their mobility traces' length (in days) and calculate the prediction performance of each group. As shown in Fig. 1(b), users' original mobility trace lengths are distributed from less than 20 days to 58 days. Thus, users have different trace lengths that we can divide them into clusters with respect to trace length where users in the same cluster have similar trace lengths. We analyze four clusters: users with trace no longer than 20 days, between 21 and 30 days, between 31 and 40 days, and longer than 40 days. Similar to former experiments, we calculate the average AUC score over 5-fold cross validation for each group. The results are shown in Fig. 6. We notice that the group with a trace length of 21 to 30 days achieves the best AUC score. This provides insight into the appropriate trace length to consider when applying our findings. Shorter traces are likely to be affected by anomalous feature value, which is brought by non-health-related actions. On the other hand, patterns that can be used to monitor unhealthy status are easily be diluted, thus be neglected in longer traces. Thus, a trace lasting about 20 to 30 days is an appropriate choice to adopted to predict health status due to our findings.

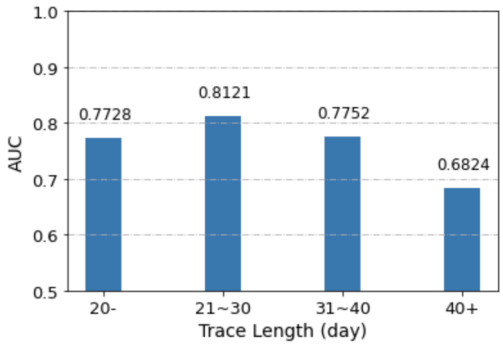


Fig. 6. Model performance on groups with similar trace length.

5 DISCUSSION

5.1 Related Work

5.1.1 Mobility Modeling. As mentioned in [2], smartphones are carried globally, with almost 70% of adults in developed countries adopted and almost half in developing countries. With the prevalence of mobile phones, access to large scale data of human mobility trace is constantly growing. In recent years, a growing number of examinations have explored the potential of using mobility patterns in revealing multiple aspects of information [1, 14, 37, 42, 44]. González *et al.* [14] demonstrated that human moves follow simple duplicable patterns with high spatial and temporal regularity. By evaluating these patterns from a macroscopic perspective, studies are conducted on regional and demographic pattern analysis by mobility data [1, 42, 44]. Xu *et al.* [44] reported the feasibility of detecting popular temporal modes and socioeconomic status from trajectory data. In [1], large scale mobile data are used to study activity levels in more than 100 regions in the world. Correlation between walkability

and activity is found. The impact of mobility data on the prevalence of infectious disease is studied in [42], which shows how mobility data help in the study of malaria epidemiology. Compared with the above studies on coarse-grained data with limited understandings of mobility patterns, we conduct our study with fine-grained mobility features that provide more comprehensive insights.

5.1.2 Health Analysis. A significant proportion of the mobility data applications in previous literature is related to health analysis. Hillebrand *et al.* [16] adopted mobile network data together with app usage data in inferring aggregate objective and subjective well-being, which primarily focused on happiness and life satisfaction. Mental health predictions are studied in [5, 6, 10]. Canzian *et al.* [6] studied the correlation of metrics extracted from mobility trajectories and user's depressive moods and adopted these metrics to predict depressed mood. Bauer *et al.* [5] detected stress periods by studying the difference of the user's mobility pattern in the stressed and non-stressed mood. Epstein *et al.*[10] predicted patients' stress and drug craving by therapeutic features given by mobility data.

In [4, 11, 38], researchers evaluated and predicted hospital readmission by patient's mobility patterns. Sedentary behavior is considered to be predictive of readmission in [4]. Takahashi *et al.* [38] studied steps' influence on predicting readmission after cardiac surgery. Instead of adopting mobility data, abundant studies work on health analysis from other perspectives. Nie *et al.* [27] processed the community-generated health data to link the vocabulary gaps between health seekers and healthcare knowledge. Hasthanasombat *et al.* [3, 15] provided insights on how neighborhood environments such as fast food, tobacco store, and sports facilities influence public health. Kelly *et al.* [17] discussed that user profiles generated by sensors could help predict health status automatically. Big data tools were adopted in [26, 34] to make health status prediction with voluminous healthcare data.

We predict people's overall health status by features extracted from large scale passively collected mobility data and demonstrate deeper insights into the mechanism of mobility features influencing health prediction. Our work differentiates from previous studies in several aspects. First, we propose an approach to predict general health status, which is presented by users' hospital visits, by using mobility traces, while most of the existing works [6, 16, 22, 35] focus on mental health detection and well-being inference. Second, we consider the neighborhood built environment and the characteristics of visitations on mobility trace for health prediction. By contrast, most studies [3, 15] only find the effect of the neighborhood. Third, unlike studies [6, 12, 35] that require user interactions such as surveys quantifying mental status, mobility features, and hospital visits in our study are all obtained passively. Fourth, [5, 22] mainly focus on the detection of mental health status, [6, 16] merely analyzes the absolute correlation between mobility metrics and predicted feature. By contrast, we interpret the features' effect on health prediction to provide typical "symptoms" of hospital visitors. Besides, we study the interaction of demographic features and mobility trace and find deeper insights that offer implications on large scale overall health monitoring.

5.2 Research Implications and Applications

To investigate the feasibility of using passively collected mobility trace in health monitoring, we conduct experiments on predicting user's hospital visit by demographic and mobility trace features. Specifically, we compare the prediction performance by a different set of features and demonstrate the effectiveness of mobility trace in the prediction. Furthermore, through feature importance analysis, we provide possible explanations of the impact of features on health status and explore how mobility features' indicative varies in distinct demographic groups. These findings have broad implications for research communities and application developers, which can be summarized as follows.

Our performance analysis reveals that mobility features achieve slightly better performance in predicting hospital visits than demographic characteristics and enhance the performance significantly when combined with

demographic features. These findings demonstrate that mobility is indicative of health status, which validates and strengthens previous understandings. In the feature importance analysis, mobility trace features such as standard deviation of displacements and entertainment POI density around points visited have a high impact on the prediction. Specifically, the importance of entertainment and scenic POIs, which are seldom considered previously, shed light on future feature design and selection. Also, we adopt an approach to evaluate mobility trace's influences on prediction. The results of the positive contribution of sports facilities and the negative effect of tobacco stores validate prior understandings [13, 15]. Moreover, this approach establishes the exact correlation directions between features and health status, which provides an alternative for feature analysis in future studies.

Our work also contributes to understanding the interaction between demographic features and mobility trace, while previous works mostly focus on one set of mobility features' effectiveness. For example, the comparison of feature importance on different age groups reveals that older people's mobility coverage indicates their health more substantial than younger people. In contrast, people lower than 30 are more likely to be affected by fast food in the neighborhood. On the other hand, the analysis of two gender groups provides other insights that have not been discussed. Visitation environment and diversity are more indicative for women in health prediction, while men's health status is mainly influenced by neighborhood. These findings have important implications for most health prediction projects that deeper insights can be drawn from analyzing different demographic groups.

Our findings provide implications for health-related application developers in terms of the app's efficiency in health monitoring. We have demonstrated the effectiveness of mobility trace in health status prediction and provide an ideal trace length for evaluation. Users can assign health-related apps the permission to the passively collected mobility data when running the app under strict protocols that enforce data anonymization and non-disclosure agreements. Health-related apps can utilize the data to accurately predict the likelihood of users requiring hospital treatment, then raise timely warnings to the user. Besides, under the privacy protection protocol, users can select to actively provide their demographic features, which helps the application give health tips for app users. These procedures include none or merely user interaction. Thus the app can propose precise recommendations to users with less interference.

Before elaborating on the beneficial applications that can be drawn from our study, we have to reiterate the meticulous user privacy protection protocols we enforce in the experiments. First, when the data collectors gather the demographic data from survey participants, they are informed of the purpose of the data collection and clearly authorized data usage for research purposes. Before the dataset is shared with researchers, it is properly anonymized by the data collectors. The real user IDs and phone numbers are never provided to or used by the researchers. Aside from data anonymization, all researchers authorized to access the dataset are constrained by strict confidentiality non-disclosure agreements. To eliminate the risk of privacy leakage by an on-line attack, all the data are stored in an off-line server. Only the authorized core researchers can get access to the data. These protocols ensure that the experiments are conducted within ethical limits where the privacy of users can be protected. Only under such consideration, we are able to discuss the implications and applications of our work. Furthermore, all following applications should only be taken under ethical considerations similar to or stricter than our privacy protection protocols.

From an individual's perspective, the results of the feature effect analysis can be applied to form a healthier lifestyle. The pattern that patients visit entertainment POI more and sports facilities less suggests that even our living environment may not be health-supportive, we can exercise more and avoid visiting entertainment places often to improve our health. Second, from the medical institution's perspective, if they can collect mobility patterns of the examinees with their permissions as additional data for medical examination, it can enhance the efficiency and reduce the examiners' workload. Furthermore, applying our analysis of feature importance in specific demographic groups to local health propaganda can potentially benefit its effect on the exact proportion of the population. Third, from the whole society's perspective, our findings can be applied in constructing a healthier society. If mobility data on a larger population, for instance, the population in a district, can be collected

under strict privacy protection, it can be used to evaluate public health status and raise corresponding policies to improve it.

5.3 Limitations and Future Works

In this study, we principally obtain our research findings from one dataset from the city of Beijing. Therefore, they might be biased due to the city's traditional lifestyle and cultural background. However, the mechanism established in designing mobility features is based on insights of researches conducted on different cultural contexts. This provides the possibility of a ubiquitous pattern in this study. Admittedly, the gender distribution among users is biased, 69 percent are female. Inevitable noise may be brought in our experiment. In fact, we leverage several approaches to minimize the influence of the limitations and provide sufficiently reliable results. Albeit the bias of demographic features may affect prediction, the SHAP value analysis grants an additive and unbiased explanation of mobility features, where the contributions of features are averaged over all possible feature sequences. We also demonstrate the sufficient and significant improvement on performance provided by mobility features by Welch's t-test. As for the bias brought by the data collection period, we have balanced distributions of trace length and trace records for two groups of users.

Although our prediction and feature analysis validate the effectiveness of using mobility trace in predicting hospital visits and revealing health status, the causal relation between mobility features and hospital visits is under the potential influence of confounding variables. Deriving the exact causality between hospital visits and mobility behavior can have an instructive impact on health monitoring. Possible solutions to the causality test can be hinted at from related works. In [15], the authors assess the causality test between neighborhood environment and resident health based on the propensity score matching method. The findings in our experiments can help in designing a causality test with a similar framework. It is demonstrated in section 4.3 that demographic features are both correlated with hospital visits and mobility patterns. This property is enlightening in designing the causal diagram and identifying confounder variables. Features with similar attributes should be treated as confounder variables when testing the causality between hospital visits and mobility patterns. Then a propensity score can be obtained from the confounder variables. We can match users with propensity scores alike but a gap in specific mobility features to inspect the difference between their health status outcomes. Specific method and experiment settings require further consideration. Aside from testing the causality relation, there are other urban features with the potential of improving health prediction. For example, the image data of neighborhood and locations which the user has visited. Authors in [33] adopt thermal sensor images to assess the urban heat island effect and its influence on health. It is possible to extract other health-related factors, such as PM 2.5 and green roof, from real-time urban image data with specific image processing methods. These features may be indicative of local residents' health status, which can help to improve the effectiveness of health prediction. Furthermore, we mainly focus on the interaction between age or gender and mobility pattern in section 4.3. An interesting research topic is to analyze health prediction within a group of users with similar neighborhood environments or mobility patterns [16], which may provide more insights into precise health monitoring. We are also actively seeking opportunities for performing this work on a larger population in future work.

6 CONCLUSION

The availability of large scale mobility data provides an exclusive opportunity to study how mobility patterns influence the overall health status. In this work, we demonstrate an approach to predict hospital visits by using users' mobility data. We reveal that the features extracted from mobility traces can significantly enhance health status prediction ($p < 0.01$). We also show how mobility features affect the prediction. For example, users with less coverage of mobility and less visitation to sports facilities are more likely to visit hospitals. Furthermore, we investigate feature importance in demographic groups and obtain deeper understandings of the features.

Our results can be adopted in various applications. The predictive pattern of mobility traces can be used in health-related mobile apps that can easily collect large scale mobility data. Besides, our findings on different demographic groups can help in customized health monitoring. As for future works, one promising direction is to conduct the causality test to understand the causal relationship of extracted features and health status. Besides, extending the current study on a larger population with an extended feature set is another exciting topic.

REFERENCES

[1] Tim Althoff, Rok Sosič, Jennifer L. Hicks, Abby C. King, Scott L. Delp, and Jure Leskovec. 2017. Large-scale physical activity data reveal worldwide activity inequality. *Nature* 547, 7663 (jul 2017), 336–339.

[2] Emily Anthes. 2016. Mental health: There's an app for that. *Nature* 532, 7597 (apr 2016), 20–23.

[3] Mariana C. Arcaya, Reginald D. Tucker-Seeley, Rockli Kim, Alina Schnake-Mahl, Marvin So, and S.V. Subramanian. 2016. Research on neighborhood effects on health in the United States: A systematic review of study characteristics. *Social Science & Medicine* 168 (nov 2016), 16–29.

[4] Sangwon Bae, Anind K. Dey, and Carissa A. Low. 2016. Using passively collected sedentary behavior to predict hospital readmission. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM.

[5] Gerald Bauer and Paul Lukowicz. 2012. Can smartphones detect stress-related changes in the behaviour of individuals?. In *2012 IEEE International Conference on Pervasive Computing and Communications Workshops*. IEEE.

[6] Luca Canzian and Mirco Musolesi. 2015. Trajectories of depression. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*. ACM Press.

[7] Tianqi Chen and Carlos Guestrin. 2016. XGBoost. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM.

[8] Xinlei Chen, Zheqi Zhu, Min Chen, and Yong Li. 2018. Large-scale mobile fitness app usage analysis for smart health. *IEEE Communications Magazine* 56, 4 (2018), 46–52.

[9] Raj Chetty, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler. 2016. The Association Between Income and Life Expectancy in the United States, 2001-2014. *JAMA* 315, 16 (apr 2016), 1750.

[10] David H. Epstein, Matthew Tyburski, William J. Kowalczyk, Albert J. Burgess-Hull, Karran A. Phillips, Brenda L. Curtis, and Kenzie L. Preston. 2020. Prediction of stress and drug craving ninety minutes in the future with passively collected GPS data. *npj Digital Medicine* 3, 1 (mar 2020).

[11] S. R. Fisher, Y.-F. Kuo, G. Sharma, M. A. Raji, A. Kumar, J. S. Goodwin, G. V. Ostir, and K. J. Ottenbacher. 2012. Mobility After Hospital Discharge as a Marker for 30-Day Readmission. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 68, 7 (dec 2012), 805–810.

[12] Mads Frost, Afsaneh Doryab, Maria Faurholt-Jepsen, Lars Vedel Kessing, and Jakob E. Bardram. 2013. Supporting disease insight through data analysis. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing - UbiComp '13*. ACM Press.

[13] Panagis Galiatsatos, Cynthia Kineza, Seungyoun Hwang, Juliana Pietri, Emily Brigham, Nirupama Putcha, Cynthia S Rand, Meredith McCormack, and Nadia N Hansel. 2017. Neighbourhood characteristics and health outcomes: evaluating the association between socioeconomic status, tobacco store density and health outcomes in Baltimore City. *Tobacco Control* 27, e1 (nov 2017), e19–e24.

[14] Marta C. González, César A. Hidalgo, and Albert-László Barabási. 2008. Understanding individual human mobility patterns. *Nature* 453, 7196 (jun 2008), 779–782.

[15] Apinan Hasthanasombat and Cecilia Mascolo. 2019. Understanding the Effects of the Neighbourhood Built Environment on Public Health with Open Data. In *The World Wide Web Conference on - WWW '19*. ACM Press.

[16] Martin Hillebrand, Imran Khan, Filipa Peleja, and Nuria Oliver. 2020. MobiSenseUs: Inferring Aggregate Objective and Subjective Well-Being from Mobile Data. In *ECAI 2020 - 24th European Conference on Artificial Intelligence (Frontiers in Artificial Intelligence and Applications)*, Vol. 325. IOS Press, 1818–1825.

[17] Daniel Kelly, Kevin Curran, and Brian Caulfield. 2017. Automatic Prediction of Health Status Using Smartphone-Derived Behavior Profiles. *IEEE Journal of Biomedical and Health Informatics* 21, 6 (nov 2017), 1750–1760.

[18] Paul Krebs and Dustin T Duncan. 2015. Health App Use Among US Mobile Phone Owners: A National Survey. *JMIR mHealth and uHealth* 3, 4 (nov 2015), e101.

[19] Kurt Kroenke, Robert L. Spitzer, Janet B.W. Williams, and Bernd Löwe. 2010. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. *General Hospital Psychiatry* 32, 4 (jul 2010), 345–359.

[20] S. Kullback and R. A. Leibler. 1951. On Information and Sufficiency. *The Annals of Mathematical Statistics* 22, 1 (mar 1951), 79–86.

[21] R. Lletı, M.C. Ortiz, L.A. Sarabia, and M.S. Sánchez. 2004. Selecting variables for k-means cluster analysis by using a genetic algorithm that optimises the silhouettes. *Analytica Chimica Acta* 515, 1 (jul 2004), 87–100.

[22] Hong Lu, Denise Frauendorfer, Mashfiqui Rabbi, Marianne Schmid Mast, Gokul T. Chittaranjan, Andrew T. Campbell, Daniel Gatica-Perez, and Tanzeem Choudhury. 2012. StressSense. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*. ACM Press.

[23] Scott M. Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. 2020. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence* 2, 1 (jan 2020), 56–67.

[24] Anmol Madan, Manuel Cebrian, David Lazer, and Alex Pentland. 2010. Social sensing for epidemiological behavior change. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*. ACM.

[25] Kate E Mason, Neil Pearce, and Steven Cummins. 2018. Associations between fast food and physical activity environments and adiposity in mid-life: cross-sectional, observational evidence from UK Biobank. *The Lancet Public Health* 3, 1 (jan 2018), e24–e33.

[26] Lekha R. Nair, Sujala D. Shetty, and Siddhanth D. Shetty. 2018. Applying spark based machine learning model on streaming big data for health status prediction. *Computers & Electrical Engineering* 65 (jan 2018), 393–399.

[27] Liqiang Nie, Yi-Liang Zhao, Mohammad Akbari, Jialie Shen, and Tat-Seng Chua. 2015. Bridging the Vocabulary Gap between Health Seekers and Healthcare Knowledge. *IEEE Transactions on Knowledge and Data Engineering* 27, 2 (feb 2015), 396–409.

[28] Venet Osmani, Alban Maxhuni, Agnes Grünerbl, Paul Lukowicz, Christian Haring, and Oscar Mayora. 2013. Monitoring activity of patients with bipolar disorder using smart phones. In *Proceedings of International Conference on Advances in Mobile Computing & Multimedia - MoMM '13*. ACM Press.

[29] Hae-Sang Park and Chi-Hyuck Jun. 2009. A simple and fast algorithm for K-medoids clustering. *Expert systems with applications* 36, 2 (2009), 3336–3341.

[30] Dudley L. Poston. 1972. Socioeconomic Status and Work-Residence Separation in Metropolitan America. *The Pacific Sociological Review* 15, 3 (jul 1972), 367–380.

[31] Sebastian Raschka. 2014. An overview of general performance metrics of binary classifier systems. *arXiv preprint arXiv:1410.5330* (2014).

[32] Babak Roshanaei-Moghaddam, Wayne J. Katon, and Joan Russo. 2009. The longitudinal effects of depression on physical activity. *General Hospital Psychiatry* 31, 4 (jul 2009), 306–315.

[33] Valentina Sagris and Mait Sepp. 2017. Landsat-8 TIRS Data for Assessing Urban Heat Island Effect and Its Impact on Human Health. *IEEE Geoscience and Remote Sensing Letters* 14, 12 (dec 2017), 2385–2389.

[34] Prasan Kumar Sahoo, Suvendu Kumar Mohapatra, and Shih-Lin Wu. 2016. Analyzing Healthcare Big Data With Prediction for Future Health Condition. *IEEE Access* 4 (2016), 9786–9799.

[35] Akane Sano and Rosalind W. Picard. 2013. Stress Recognition Using Wearable Sensors and Mobile Phones. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE.

[36] Chaoming Song, Tal Koren, Pu Wang, and Albert-László Barabási. 2010. Modelling the scaling properties of human mobility. *Nature Physics* 6, 10 (sep 2010), 818–823.

[37] C. Song, Z. Qu, N. Blumm, and A.-L. Barabasi. 2010. Limits of Predictability in Human Mobility. *Science* 327, 5968 (feb 2010), 1018–1021.

[38] Tetsuya Takahashi, Megumi Kumamaru, Sue Jenkins, Masakazu Saitoh, Tomoyuki Morisawa, and Hikaru Matsuda. 2015. In-patient step count predicts re-hospitalization after cardiac surgery. *Journal of Cardiology* 66, 4 (oct 2015), 286–291.

[39] M. C Wang, S. Kim, A. A Gonzalez, K. E MacLeod, and M. A Winkleby. 2007. Socioeconomic and food-related physical characteristics of the neighbourhood environment are associated with body mass index. *Journal of Epidemiology & Community Health* 61, 6 (jun 2007), 491–498.

[40] Rui Wang, Weichen Wang, Alex daSilva, Jeremy F. Huckins, William M. Kelley, Todd F. Heatherton, and Andrew T. Campbell. 2018. Tracking Depression Dynamics in College Students Using Mobile Phone and Wearable Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (mar 2018), 1–26.

[41] B.L. Welch. 1947. The generalization of Student's problem when several different population variances are involved. *Biometrika* 34 (1947), 28–35.

[42] A. Wesolowski, N. Eagle, A. J. Tatem, D. L. Smith, A. M. Noor, R. W. Snow, and C. O. Buckee. 2012. Quantifying the Impact of Human Mobility on Malaria. *Science* 338, 6104 (oct 2012), 267–270.

[43] James O. Wheeler. 1967. Occupational Status and Work-Trips: A Minimum Distance Approach. *Social Forces* 45, 4 (jun 1967), 508.

[44] Fengli Xu, Tong Xia, Hancheng Cao, Yong Li, Funing Sun, and Fanchao Meng. 2018. Detecting Popular Temporal Modes in Population-scale Unlabelled Trajectory Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (mar 2018), 1–25.

[45] Fengli Xu, Pengyu Zhang, and Yong Li. 2016. Context-aware real-time population estimation for metropolis. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM.

[46] Yang Xu, Alexander Belyi, Iva Bojic, and Carlo Ratti. 2018. Human mobility and socioeconomic status: Analysis of Singapore and Boston. *Computers, Environment and Urban Systems* 72 (nov 2018), 51–67.

[47] Xin Zhang, Qunhong Wu, Yongxiang Shao, Wenqi Fu, Guoxiang Liu, and Peter C. Coyte. 2015. Socioeconomic Inequities in Health Care Utilization in China. *Asia Pacific Journal of Public Health* 27, 4 (jan 2015), 429–438.

[48] Yuyang Zhang and Peter C. Coyte. 2020. Inequality of opportunity in healthcare expenditures: evidence from China. *BMC Health Services Research* 20, 1 (may 2020).

[49] Dawei Zhu, Na Guo, Jian Wang, Stephen Nicholas, and Li Chen. 2017. Socioeconomic inequalities of outpatient and inpatient service utilization in China: personal and regional perspectives. *International Journal for Equity in Health* 16, 1 (dec 2017).