HealthWalks: Sensing Fine-grained Individual Health Condition via Mobility Data

ZONGYU LIN and SHIQING LYU, Beijing National Research Center for Information Science and Technology, Tsinghua University
HANCHENG CAO, Department of Computer Science, Stanford University
FENGLI XU, Beijing National Research Center for Information Science and Technology, Tsinghua University
YUQIONG WEI, China Mobile
PAN HUI, CSE, Hong Kong University of Science and Technology
HANAN SAMET, Department of Computer Science, University of Maryland
YONG LI*, Beijing National Research Center for Information Science and Technology, Tsinghua University

Can health conditions be inferred from an individual's mobility pattern? Existing research has discussed the relationship between individual physical activity/mobility and well-being, yet no systematic study has been done to investigate the predictability of fine-grained health conditions from mobility, largely due to the unavailability of data and unsatisfactory modelling techniques. Here, we present a large-scale longitudinal study, where we collect the health conditions of 747 individuals who visit a hospital and tracked their mobility for 2 months in Beijing, China. To facilitate fine-grained individual health condition sensing, we propose *HealthWalks*, an interpretable machine learning model that takes user location traces, the associated points of interest, and user social demographics as input, at the core of which a Deterministic Finite Automaton (DFA) model is proposed to auto-generate explainable features to capture useful signals. We evaluate the effectiveness of our proposed model, which achieves 40.29% in micro-F1 and 31.63% in Macro-F1 for the 8-class disease category prediction, and outperforms the best baseline by 22.84% in Micro-F1 and 31.79% in Macro-F1. In addition, deeper analysis based on the SHapley Additive exPlanations (SHAP) showcases that *HealthWalks* can derive meaningful insights with regard to the correlation between mobility and health conditions, which provide important research insights and design implications for mobile sensing and health informatics.

CCS Concepts: • **Information systems → Data mining**; • **Human-centered computing** → *Empirical studies in ubiquitous and mobile computing*;

Additional Key Words and Phrases: Human Mobility, Health Sensing

*This is the corresponding author.

Authors' addresses: Zongyu Lin; Shiqing Lyu, Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing; Hancheng Cao, Department of Computer Science, Stanford University, California; Fengli Xu, Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing; Yuqiong Wei, China Mobile, Beijing; Pan Hui, CSE, Hong Kong University of Science and Technology, Hong Kong; Hanan Samet, Department of Computer Science, University of Maryland, Maryland, lin-zy17@mails.tsinghua.edu.cn; Yong Li, liyong07@tsinghu.edu.cn, Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing.

2474-9567/2020/12-ART138 $15.00
https://doi.org/10.1145/3432229

ACM Reference Format:
Zongyu Lin, Shiqing Lyu, Hancheng Cao, Fengli Xu, Yuqiong Wei, Pan Hui, Hanan Samet, and Yong Li. 2020. HealthWalks: Sensing Fine-grained Individual Health Condition via Mobility Data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4, Article 138 (December 2020), 26 pages. https://doi.org/10.1145/3432229

1 INTRODUCTION

Can we infer an individual's health condition from his/her mobility patterns? There has been a long line of research exploring the correlation between physical activity and human well-being [1]. As one of the most ubiquitous forms of physical activity, mobility has been shown to be closely linked to user well-being [2, 3]. Meanwhile, with the popularization of the mobile phones and wearable devices, individual mobility trace data have become easily accessible and have been widely used to profile user attributes, e.g. gender, age, achieving considerable accuracy [4, 5]. A natural and intuitive extension would be to investigate whether individual mobility traces, can be leveraged to sense fine-grained health conditions, i.e. potential health risks the individual is prone to, of the users.

Traditional health profiling methods mainly rely on leveraging physical examination and questionnaire-based methods to give a coarse grained classification [6, 7]. However, existing approaches require slow and manual interactions between patients and care providers, which cost a lot of time and money[6, 7]. Thus, the effectiveness of the system for sensing potential health risks is becoming increasingly important, since an efficient and instant system can provide great convenience in patient care as well as controlling the medical costs. To achieve this vision, recent years have seen a surge of research interest to leverage ubiquitous computing techniques, e.g. sleep logs are collected from smart bands and user surveys for user health sensing [8], and the relapse of psychotic disorder has been detected in [9] by patient-generated and patient-contributed digital data from Facebook, yet they quite often require specially designed sensors or frequent feedback from users. Thus if health sensing can be strengthened through mining user mobility signals, the cost of sensing can be even lowered as mobility data are one of the most easily accessible data sources for an individual.

Despite its potential impact, little work has been done on profiling users' fine-grained health condition using mobility data. The most relevant work [10] intuitively defined several mobility metrics to infer individual mood disorder, and [11] utilized the points of interest (POI) data to predict the evolution of chronic disease at the region level. Yet none of the literature has addressed the problem of using mobility trace to sense disease category level individual health condition. The key obstacle not only lies in the lack of a novel dataset, but also the difficulty of modelling to capture useful signals from mobility traces that can be used to infer health risks at a more fine-grained level.

Present Work. Here, we launch a longitudinal study where we collect and explore a novel dataset, which includes mobility traces of 747 patients and the associated medical information. Preliminary analysis confirmed the characteristic difference in the distribution of classical mobility metrics across different disease categories. To enable accurate sensing of individual health conditions, we further propose *HealthWalks*, a machine learning model that takes user location traces, their associated points of interest, and user demographics as input, at the core of which a deterministic finite automaton (DFA) model is used to auto-generate a large number of interpretable features to help capture useful signal.

Furthermore, we present insightful analysis on the interpretability of *HealthWalks*. Many machine learning techniques have been utilized in previous work about user profiling based on mobility data, including RNN-based methods [12, 13], unsupervised clustering [14, 15], and network embedding methods [16, 17]. However, they are unable to explain the signals that contribute to the prediction captured in their model, neither can they offer a reasoning procedure for the prediction similar to human experts. Considering that interpretability plays a vital role in the medical machine learning task, we adopt the latest popular explainable model, the SHapley Additive exPlanations (SHAP) [18] to construct an explainable system for our model, where the influence of each variable on each prediction can be computed by utilizing a game theoretic approach. Within this technique, some interesting insight has been extracted, e.g., the endocrinology group is found most associated with sports and restaurant POI, the mental and neurology is found to be the most aligned with the workplace-related features, and the birth group showcases the most inactive mobility pattern. Finally, we present a reasoning procedure to showcase how our model infers a user's potential health risk through his mobility trace.

To the best of our knowledge, our paper is the first to profile fine-grained health condition using individual mobility traces, which is of great importance to the boosting of effectiveness of the existing system for sensing potential health risks. The major contributions of this work can be summarized as follows:

- We propose a novel question to sense fine-grained health risks using mobility data. To make this study possible, we utilize a questionnaire-based method to collect the outpatient information from hospitals in Beijing. Then, through collaboration with mobile operator, we collect the mobility trace covering two months from relevant users with their permission.
- We propose a method *HealthWalks* based on a deterministic finite automaton (DFA) to generate numerous features which are then selected by regularization method for the final classification. We also aggregate the points of interest (POI) data and demographics to generate features capturing the semantics of the trajectory.
- Then, we conduct empirical experiments on a real world dataset to showcase the significance and robustness of our model to predict severity and disease category. We demonstrate that our model has achieved up to 40.29% in Micro-F1 and 31.63% in Macro-F1 for the 8-class disease category prediction, which significantly outperforms the best baseline by 22.84% in Micro-F1 and 31.79% in Macro-F1.
- Finally, we conduct extensive experiments to cast insight on the correlation between mobility patterns and health conditions. We combine the global feature importance and local explanations to gain a meaningful and comprehensive interpretation of the health signals reflected by the mobility trace data. These results bring us knowledge which is vital to numerous applications, including intelligent medical system for sensing potential health risks, medical-related apps and even the scheduling for the urban medical resources.

2 BACKGROUND AND RELATED WORK

With the prevalence of mobile devices, there has been a soaring number of applications based on the large-scale mobility data, including traffic prediction [19], social relationship inference [16], user profiling [20], and disease evolution prediction [21]. In this paper, we concentrate on leveraging the mobility data to predict user's fine-grained health conditions. We review the most relevant related work, which can be summarized as three major perspectives.

2.1 Physical Activity, Mobility and Health

Previous work from biological and medical fields has studied the relationship between physical activity and health status. Buchner et al. [1] has proved that physical exercise can improve older people's health considering the physiological measures of physical fitness. A conceptual framework is proposed [22] to outline how adolescent physical activity might contribute to adult health. Also, a number of researchers [23, 24] have demonstrated that regular physical activity of moderate intensity can reduce the risk of numerous chronic diseases and extends longevity. Blair et al. [25] combine preventive etiologic associations with the therapeutic effects of physical activity on health and disease to assess how much physical activity is required for health. Evidence provided in [26] demonstrates that even light activities can have health benefits.

Physical mobility, as a specific instance can also have association with the health condition. Actually, there has been some related work that has explored the correlation between health condition and human mobility. Longitudinal analytical techniques have been used in [2] to better understand and identify the relationship between residential mobility and health. Althoff et al. [27] demonstrated the relationship between mobility and obesity rate at the country level. All the background theoretical knowledge mentioned above has strengthened our motivation to study the health condition using individual's mobility traces, and we hypothesize that individual human mobility pattern can help predict user health condition at fine-grained level.

2.2 Health Sensing with Ubiquitous Computing

Unobtrusive sensing and wearable devices monitoring health condition have been discussed in [28, 29]. Sleep quality is predicted in [8] using sleep logs collected from smart bands and user surveys. The monitoring of bipolar disorder by means of smartphones are discussed in [30]. The relapse of psychotic disorder has been detected in [9] by patient-generated and patient-contributed digital data from Facebook. However, the existing approaches cost a lot of money and time on the wearable devices or requires users to give feedback (i.e., fill in questionnaire on a daily basis) to the monitoring devices or related apps, which is unlikely to be put into use on a large scale. Besides, the previous related work has mainly focused on the detection and prediction of the psychological states and mental health conditions. In contrast, our work focus on investigating the possibility of using individual mobility data captured by their smartphones to predict more fine-grained health conditions, e.g. potential health risk at disease category level, for each user.

Some work has been done on sensing health using mobility data. The prediction of the chronic disease using POI check-in data has been studied in [11]. Malaria transmission foci has been identified in [3] using large-scale human mobility data. User locations and social connections are exploited by [31] to detect the stressful periods. However, all of these papers sense people's health condition with mobility data by a coarse granularity, either from region level or at a large population scale. Furthermore, the previous work simply leverage the physical mobility data or just the POI check-in data, which cannot extract a whole picture of the mobility pattern. Therefore, in order to narrow the gap existed in previous work, we are dedicated to predicting the fine-grained health condition at disease category level, where we combine the physical mobility traces and the points of interest (POI) information to enhance the semantics of the mobility traces and better represent a comprehensive mobility pattern linked with health conditions.

2.3 User Profiling with Mobility Data

The availability and proliferation of the large-scale mobility data has aroused extensive studies on exploring the user profiling. Previous work mainly focuses on leveraging the mobility data to infer social relationship based on the co-location behavior [16, 17], or combined the mobility data with other source of ubiquitous data, such as app usage and light sensor, to infer user attributes like age and gender [4, 5]. Other work has investigated the possibility of sensing user living pattern from mobility [32–34]. Also, studies [35, 36] have found that important functional locations (i.e., residence and workplace) can be identified by individual trajectory data. Furthermore, another important related work proposed a temporal pattern based trajectory clustering algorithm, which effectively captured the similarity in time allocation patterns [14]. It showcased the possibility of predicting the user attributes based on mobility data. However, one key obstacle for inferring the complex user attributes solely based on mobility data is the difficulty of dissecting the motivations behind mobility transitions, which prevents the models from capturing the underlying correlation between mobility behavior and user attributes. Besides, the methods to deal with the mobility data utilized in the previous work are mainly unsupervised clustering [14, 35, 37], network embedding methods [16, 17] or RNN-based methods [12, 13], which lack decent interpretability in how a mobility pattern contributes to the inference of the user profiles.

Different from previous work, our goal is to infer the fine-grained health condition (i.e., the potential health risks) based on the mobility records collected by the widely-used smartphones, which has been absent in the ubiquitous computing area. Also, different from the methods mentioned in the related work, we propose a feature-based model together with the state-of-the-art explainable machine learning technique to improve the performance while preserving fine-grained interpretability.

3 DATA COLLECTION, PRELIMINARY ANALYSIS AND TASK FORMULATION

3.1 Data Collection

We collect a real-world mobility trace along with user improvement demographics and medical data, which combine the raw mobility records with the points of interest (POI) data to enrich the semantics of the trajectory.

Table 1. The basic information of the utilized datasets.

City	Time Duration	Records	Users	Locations
Beijing, China	1st July – 31st August, 2017	907,345	747	15,509

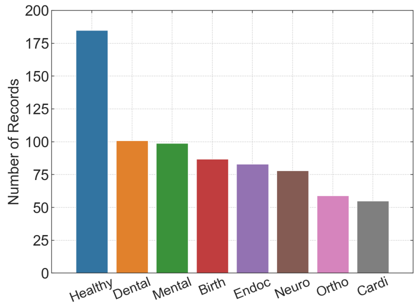
3.1.1 *Mobile Network Data.*

We collect the dataset jointly from the local mobile operator, China Mobile and 11 hospitals in Beijing, which contains over 7 million mobility records covering 747 outpatient users and 15,509 associated base station locations for two months, i.e., from 1 July. 2017 to 31 August. 2017. It should be noted that the eleven hospitals we collaborate with are well-known and representative, including 8 general hospitals and 3 specialized hospitals. The details of how we collect the dataset are as follows.

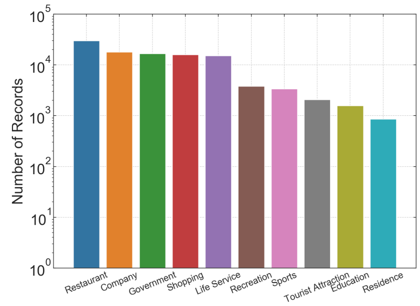
- First, we set digital questionnaires in the collaborated hospitals in Beijing by QR-code, and set several questions about their outpatient information (i.e., including their gender, age, education level, specific department he or she visits, etc.). It should be noted that we regard the department he or she visits as the specific kind of disease category he or she suffers. Several measures have been adopted to verify the raw data. We check the validity by collecting the completion time for each question to verify the quality of the collected data. Also, we set a trap question (i.e., a question that has a known answer) in the questionnaire to check the validity of the questionnaire. We accepted 7,759 questionnaires and only pick out 562 validated users from seven representative disease categories (i.e.,dental, mental, birth, endocrinology, neurology, orthopedics and cardiology) and 185 healthy people.
- Then, for the validated questionnaires, we ask their permission to join our potential research project and permit our analysis of the mobility trace. If they are willing to share this kind of private information, then we can collect the mobility records of them which covers a period of 2 months through close collaboration with the mobile operator. Specifically, the fine-grained mobility trace records the time stamps and the connected base stations with longitude and latitude information whenever the mobile users access cellular network, e.g., sending text messages and making phone calls, consuming mobile data traffic. The period of two months ensures the sufficient number of records after the preprocessing procedure in Section 3.1.4. Furthermore, the mobile operator offers us the user attributes, i.e., age, gender and average revenue per user (ARPU.).

Table 2. The distribution of user demographic in China Mobile dataset.

Demographic	Category
Gender	male (27.98%), female (72.02%)
Age	0∼30 (21.95%), 30∼40 (35.74%), 40∼60 (35.21%), 60∼99 (7.10%)
Education	junior high school (21.02%), senior high school (25.57%), undergraduate (25.97%), postgraduate (27.44%)
Income	very low (10.17%), low (18.34%), high (59.71%), very high (11.78%)



(a) department



(b) POI

Fig. 1. The distribution of disease and POI categories of the dataset.

- To further verify the data, we check the consistency of the basic information (i.e., age and gender) from both hospitals and mobile operator, and we discard non-reliable questionnaires. Furthermore, we adopt several measures to protect the privacy of this sensitive data (e.g., map the telephone information by hash technique for each user). The issue of the privacy and ethical considerations will be discussed in the following section.

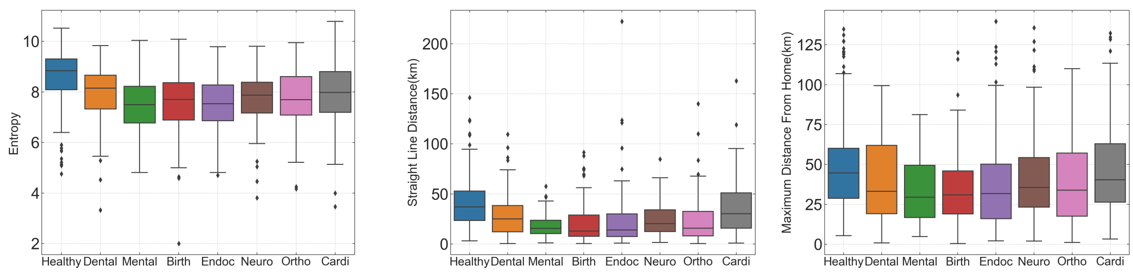
Specifically, we display the distribution of user demographics (i.e., gender, age, education and income) of collected users in Table 2. Most importantly, the hospitals provided the questionnaires information filled out by the users mentioned above, which includes the associated disease information. We collect healthy users as well as patients with 7 representative disease categories, including dental, mental, birth, endocrinology, neurology, orthopedics and cardiology. The distribution of the disease categories, i.e., healthy (185), dental (78), mental (55), birth (83), endocrinology (101), neurology (87), orthopedics (59), and cardiology (99), is shown in Figure 1(a).

3.1.2 POI of Beijing. Through collaboration with Tencent incorporation, we were granted access to the points of interest (POI) information collected by Tencent Maps. In order to link the raw mobility records with the POI data, we first partition Beijing with a grid-based map segmentation. Specifically, we adopt a $100m \times 100m$ disjointed grid, which means the distance between the adjacent minimal units on the grid-based map is 100 meters. Then we compute the distribution of the POI for each grid. It should be noted that we only pick out ten possible related POI categories proposed by [11] for use, i.e., restaurant, company, government, shopping, life service, recreation, sports, tourist attraction, education and residence. The total number of the points of interest for each category is displayed in Figure 1(b).

3.1.3 Privacy and Ethical Considerations. Considering the sensitivity of this kind of data, we have adopted the following protocols to tackle the potential privacy and ethical risks in data analytics. First of all, the whole data is properly anonymized by the data owners before being available to us. Specifically, real user IDs are never made available to the researchers. Moreover, our data analysis procedures are reviewed and authorized by the dataset owners to ensure the compliance with privacy protocols in the Term-of-Use statements. Second, all the researchers that have been authorized to access the datasets are bounded by strict non-disclosure agreements, and our research protocol is approved by the local institutional board. Finally, all the data are stored in a secure off-line server, and only the authorized core researchers can access the data.

3.1.4 Preprocessing. To prepare a cleaner and higher-quality dataset for the downstream tasks, we take several preprocessing procedures to denoise the raw trajectories. First of all, for our problem setting, we only utilize the mobility records before their first visit to the hospital for the patients who visit a specific department. Specifically, we filter out the users who never show up around the hospital to check the validation of the questionnaire information from the spatial perspective. Secondly, we limit the highest speed of the trajectories to 120 kilometers per hour, since the expressway in China limits the maximum speed per hour to 120 kilometers, and the speed limits of most of the means of transportation for daily use, including the bus and subway, are lower than that bound. Finally, to fully explore the mobility patterns, we only preserve the users with more than 50 records.

3.2 Preliminary Analysis



(a) Entropy (b) Straight Line Distance (c) Maximum Distance From Home

Fig. 2. The distribution of basic mobility metrics across different disease categories.

To initially explore the correlations between mobility patterns and disease categories, we design three basic and classical metrics to represent the mobility patterns and showcase their distribution among disease categories. First of all, we compute mobility entropy (i.e., the historical probability that a location i was visited by u) for each user. As is shown clearly in Figure 2(a), the healthy people rank first with 8.84 median of the entropy, followed by dental (8.14), while the mental, endocrinology and birth falls behind, which has only 7.49, 7.53, 7.70 for the median of the mobility entropy, respectively. Secondly, we figure out the sum of distances travelled by an individual per day. As can be observed clearly from Figure 2(b), healthy, dental and cardiology group traveled 37.01, 30.36 and 25.01 kilometers for the median of the distance respectively, while the birth, endocrinology, mental, and orthopedics group has only 12.75, 13.86, 15.50, 15.87 kilometers for the median of the distance respectively. In addition, we detect the "home" by finding the places where the user visited most frequently during nighttime, and then figure out the maximum distance from home [10]. Figure 2(c) indicates that the healthy people go farthest from home, and have 44.69 kilometers for the median of the distance. In contrast, the median distance from home of the mental, birth and orthopedics group is only 25.56, 30.92 and 33.92 kilometers, respectively.

In brief, these three figures showcase the difference of the mobility metrics lying in different disease categories, where the healthy and dental group displays an active mobility pattern, while the mental, birth and orthopedics groups show inactive mobility patterns, which is in agreement with our intuition and common knowledge. Most importantly, this preliminary analysis indicates that we might be able to predict the disease category based on these interpretable mobility metrics that can reflect the behavioral pattern of the user. However, it is difficult for our researchers to pick out all the metrics that can show great variance among different disease categories for downstream task (i.e., prediction task), since the distribution of the mobility metrics across disease categories, to some extent, is subject to the characteristics of the dataset, including the number of the users, the length of the mobility traces, the granularity of the disease categories or even the regions or countries covered in the dataset. Therefore, it is of great importance to design a model that can automatically generate numerous mobility metrics while preserving interpretability.

3.3 Task Formulation

Finally, we formally define our utilized data and problem as follows.

Definition 3.1. (Urban mobility record, UMR) An urban mobility record is a quaternion (u, l, t, d), which denotes that user u visits location l at time t and stays at that location for duration d , where l denotes a unique area with geographical coordinates (i.e., longitude and latitude) and boundary in the urban space.

Definition 3.2. (Point of interest, POI) A point of interest P is defined as a uniquely identifiable venue with specific function C, e.g., residence, workplace, restaurant and park. In our model, each location has a POI distribution vector which indicates the number of different types of POIs around that location.

***HealthWalks* Target.** Given the mobility records of users with the associated POI distribution data, we aim at predicting the potential health risk for each user based on the mobility metrics automatically generated by a deterministic finite automaton. The derived mobility metrics should not only be able to enhance the performance of the prediction task, but also to preserve fine-grained interpretability that can explain the health signals reflected by the mobility trace.

4 METHOD

In this section, we are transforming the raw mobility records into a series of mobility metrics that can be used to predict the disease category for each user, while preserving fine-grained interpretability. There has been a major defect in the most relevant related work [10], where several mobility metrics are intuitively chosen to predict the psychological state. Since it takes a lot of time to specify each feature one by one, this intuitive approach, to some extent, improves the interpretability at the sacrifice of the accuracy of the model. Here, we balance the two key factors, accuracy and interpretability, and propose a model based on deterministic finite automata, to automatically generate numerous interpretable features.

Furthermore, the relevant work [10, 11] either solely relies on raw mobility trace or POI check-in data, which cannot fully extract the behavioral patterns of the raw mobility trace. Therefore, we combine the physical mobility trace with the associated POI data and the demographics as input of the model to better capture the patterns and health signals reflected by the mobility trace. To conclude, we propose a model based on a deterministic finite automaton (DFA) to generate a great number of mobility features automatically. Then, we aggregate POI information to enrich the semantics for the individual trajectory. Finally, we utilize regularization techniques to eliminate redundant features from the model for the final prediction.

4.1 Definitions of Metrics

Before formally introducing the main structure of our model, we should first give the definitions of the notations and mobility metrics utilized in our model.

4.1.1 Mobility Metrics. First of all, we regard the mobility trace as a sequence of mobility records defined in Section 3.3, which is widely used in the research of mobility modeling [38]. Then, we choose some classical mobility features from [39], [40] and [38] to represent an individual's moving behavior and pattern. It should be noted that we identify the location that user i visits most frequently during nighttime as his or her home and the place visits most frequently during daytime as his or her workplace. Previous work [36] has suggested that residence and workplace are two important locations since people spend most of their time around residence and workplace during daytime and nighttime, respectively. Therefore, we identify these two locations and link some mobility metrics with the two vital locations using well-established techniques [36, 41]. The definitions of the mobility metrics utilized in our work are displayed as follows.

- **The number of different locations visited, *Nd*** [10] This metric Nd_u is defined as the total number of unique locations visited by user u.

- **Maximum distance, *Dm*** [42]. The metric is computed as the maximum distance between two locations travelled by an individual.

$$Dm_u = \max_{l_i, l_j \in l_1, \ldots l_N} \mathrm{dist}(l_i, l_j),$$

- **Straight line distance, *Ds***. It is defined as the sum of distances travelled by a user:

$$Ds_u = \sum_{i=1}^{N-1} \mathrm{dist}(l_i, l_{i+1}),$$

  where N denotes the number of locations visited by user u.
- **Radius of gyration, *R* [43]**. This metric is widely used to weigh the area covered by the user, which is computed as:

$$R_u = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \mathrm{dist}(l_i, c)},$$

  where N is the number of locations visited by user i, and c is the coordinates of the center of mass of the locations visited by user u.
- **Maximum distance from home, *Dh* [10]**. We first identify the location that user u visits most frequently during nighttime as the home, and then compute the maximum distance traveled from there:

$$Dh_u = \max_{l_i \in l_1, \ldots l_N} \mathrm{dist}(l_h, l_i),$$

  where N denotes the number of locations visited by user u and l_h denotes the coordinates of the home. Another metric **The maximum distance from workplace, *Dw* [36]** can be computed in a similar way.
- **Mobility entropy** [44] Mobility entropy is inspired by the Shannon entropy to quantify the predictability of an individual's movements based on the historical probability that a location was visited by the user. We incorporate two types of entropy in our method, the **uncorrelated entropy *Eu*** and the **real entropy *Er***:

$$Eu_u = -\sum_{l} p_l \log_2 p_l, \quad Er_u = -\sum_{s} p_s \log_2 p_s,$$

  where l iterates over locations visited by user u, and p_l represents the probability that a location l was visited by user u, i.e. the number of visits to location l divided by the total number of records generated by user u; s iterates over all possible sequences of locations, and p_s denotes the probability that s appears as a contiguous subsequence in the user's trajectory.

4.1.2 POI Metrics. POI features are demonstrated to be closely related to user living patterns [33, 45]. For each location, the POIs in its vicinity are classified into 10 types and counted as per Section 3.1.2. Three types of metrics for a location can thus be defined:

- **The count of surrounding POIs, P_{sum}**.
- **The proportion of a certain type among surrounding POIs, P_i** ($0 \le i \le 9$). Defined to be the number of surrounding POIs of type i divided by P_{sum}. Types $i = 0, 1, \ldots, 9$ correspond to POIs marked with tags (0) restaurant, (1) company, (2) government, (3) shopping, (4) life service, (5) recreation, (6) sports, (7) tourist attraction, (8) education, and (9) residence, respectively.
- **The variety of function, P_{sd}**. Defined to be the standard deviation of $P_0, P_1, \ldots, P_9$.

Now that all the notations and metrics utilized in our model have been defined, we introduce the major framework of *HealthWalks*.

4.2 Deterministic Finite Automaton for Feature Generation

Previous related work on spatio-temporal health analytics [10] has focused on manually picking out a limited number of mobility metrics to capture the behavioral characteristics. The major drawback of this approach is that even for an experienced researcher or medical staff, it is really hard for them to select the decent mobility metrics that can effectively link with people's health condition at a first glance. Therefore, it costs them a lot of time to verify the performance for the permutation and combination of those state-of-art mobility metrics. Inspired by [46] and [47], which prove the effectiveness and interpretability of the deterministic finite automaton (DFA) on sequential data and successfully applies it on poverty and wealth index inference, we are motivated to design a novel structure based on DFA to deal with the mobility traces, which can generate numerous interpretable mobility metrics automatically. The key part for a deterministic finite automaton is the operations and the legal transitions between different states, which are introduced as follows.

- **Overview.** As is shown in Figure 3, from the initial state, the list of raw urban mobility records (UMRs) consisting of a series of fields (user ID, date, time, duration, coordinates of location, etc.) are sent into the DFA and then transferred from one state to another. The final output at the terminating state are numerical values corresponding to each user, being a mobility metric generated according to the traversed path on the automaton. The transitions are defined with clear meanings, so that any mobility metric generated by the DFA have a physical interpretation. We list the operations here and describe the details in Section 4.3.
- **Filter $f(\cdot)$.** Transforms a set of UMRs into a subset according to a certain criterion.
- **Select $s(\cdot)$.** Transforms a set of UMRs into a set of values by selecting a single field from a row.
- **Merge $m(\cdot)$.** Groups a set of selected fields (with their corresponding UMRs) by a certain attribute.
- **Reduce $r(\cdot)$.** Assigns each user the set of field values of the group they belong to, again forming (possibly identical) collections grouped by user IDs (ego).
- **Aggregate $a(\cdot)$.** Calculates a certain metric on the set of fields, transforming it into a single numerical value. Details of metrics used in this step are described in Section 4.1.1.

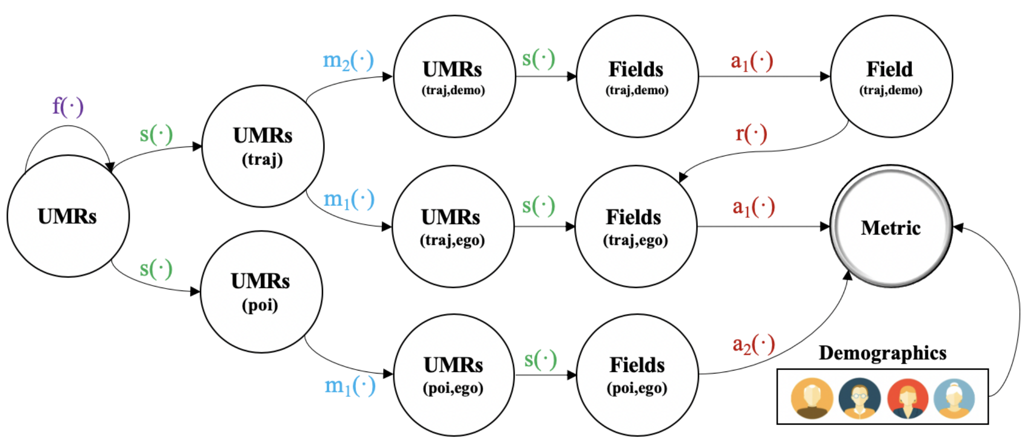


Fig. 3. The workflow of our proposed model *HealthWalks*. Each circle represents a state and each arrow symbolizes a transition. The initial input of the deterministic finite automaton is the sequence of raw mobility records together with the points of interest (POI) distribution vectors, and the final output is the mobility metric for users. Demographics are both externally and internally aggregated into the complete framework.

4.3 Incorporating Semantics for Mobility Traces

Obviously, our model provides a flexible and scalable framework that can easily aggregate additional information. Here, we are going to add the points of interest (POI) information into our model to boost the prediction of the disease category. Meanwhile, the interpretability of the model is improved since a POI represents an identifiable venue with specific function, which might be able to reflect what the user is doing at that location, thus activating the potential to explain the correlation between health condition and mobility patterns from a more fine-grained perspective. After the preprocessing of the points of interest (POI) data in Section 3.1.2, we can now map each location visited by users to a specific $100m \times 100m$ grid of the city with a POI distribution vector. Specifically, we regard the points of interest data as a "semantic mobility trace", which consists of a sequence of points of interest distribution. Then, the sequence of POI distribution vectors is treated in the same way as the sequence of mobility records. Finally, the output of the metric related with POIs aggregates decent semantics into the model, which is of great benefit to the downstream prediction and interpretation tasks.

The overall design and details of the legal transitions in *HealthWalks* are listed in Table 3. An abbreviation is assigned to each transition, so that each path on the automaton (hence each generated feature) can be uniquely identified with the concatenation of all transitions went through. As an instance, “`wd+_age_traj_dailynlocs`” denotes a path that generates for each user the average number of all distinct locations visited during daytime per day by users among their age group.

Table 3. The legal transitions on the automaton and their corresponding abbreviations.

Operation	Notation	Transitions	Abbreviations
Filter	$f(\cdot)$	Keep records on weekdays or weekends	`wd+`, `wd-`
		Keep records during daytime (09:00–21:00) or nighttime	`whour+`, `whour-`
		Keep records around home or workplace	`resi+`, `work+`
Select	$s_1(\cdot)$	Select the coordinates of the location	`traj`
	$s_2(\cdot)$	Select the count of surrounding POIs P_{sum}	`poisum`
		Select the variety of function P_{sd}	`poisd`
		Select the proportion of a type of POIs P_i	`poi0`, ..., `poi9`
Merge	$m_1(\cdot)$	Group by user ID (i.e. no grouping)	`id`
	$m_2(\cdot)$	Group by demographic traits	`age`, `edu`, `income`
Reduce	$r(\cdot)$	Unique transition; operation determined by selected $m(\cdot)$	N/A
Aggregate	$a_1(\cdot)$	Radius of gyration R	`radgyr`
		Average uncorrelated entropy Eu over all days	`dailyuentr`
		Average real entropy Er over all days	`dailyrentr`
		Average sum of straight-line distances Ds over all days	`dailysumdist`
		Average number of distinct locations Nd over all days	`dailynlocs`
	$a_2(\cdot)$	Numeric statistical measures (sum, maximum, average, standard deviation, entropy)	`sum`, `max`, `mean`, `sd`, `entr`

4.4 Regularization for Feature Selection

From the hundreds or thousands of mobility metrics generated by the DFA model, we utilize the supervised technique to pick out a smaller subset of predictors that can serve as the best joint features of the disease category. Specifically, we adopt lasso regularization [48] to penalize the complexity of the model and prevent our model from ‘overfitting’ on the relatively small number of training samples. For the model coefficient of the n−th value, the penalty of Lasso regression is defined as: $\lambda\Sigma_p||\beta_n||$. This norm-1 regularizer typically results in a sparse solution in the feature space, and assigns zero weights to most irrelavant or redundant features, which is the most widely used and classical technique for feature selection. Theoretical analysis [49] has proved that Lasso regression is highly effective when there are only a few training instances with many redundant and irrelevant features. In conclusion, our goal is to reduce the dimensionality of the data and then use another powerful classifier to make the final prediction.

4.5 Downstream Prediction Model (XGB)

The large number of features generated by *HealthWalks* can be used on any prediction model for the final target. In this paper, we adopt the gradient boosting machine trees because previous work found that they achieve superior and robust performance gains over other state-of art models. Gradient boosting machines were introduced by Friedman [50], which utilize an ensemble of weak prediction models to perform classification or regression tasks in an iterative fashion. Specifically, we utilize XGBoost [51], a popular and powerful implementation of gradient boosting machines which is derived from the extreme gradient boosting, and optimizes both the training loss and regularization for the ensemble of the trees generated. In addition to the multiclass XGBoost model, we also adopt two other strategies, one-versus-rest (OVR) and one-versus-one (OVO) for the downstream explainable classification [52]. On one hand, the final decision in OVR strategy selects the class that has the maximum value of binary decision function among all OVR binary classifiers. On the other hand, binary classifiers for all pairs of classes are constructed in OVO strategy, and then the class with the largest number of votes by all classifiers is selected. We only report the results of the multiclass XGBoost due to its higher performance and apply SHAP technique on it to offer global interpretability. The additional two strategies are applied only when we want to offer local explanations for each prediction case.

4.6 Interpretation Technique (SHAP)

SHAP is a recently popular technique aimed at explaining the output of so-called black box machine learning models by utilizing a game theoretic approach. Most of the previous work to explain the tree simply gives the global feature importance, while the SHAP value computes the contribution of each input variable in each prediction of a machine leaning model. It connects optimal credit allocation with local explanations using the classic Shapley values from game theory and their related extensions, and the calculation formula for the SHAP value ϕ_i is computed as follows [18]:

$$\phi_i = \Sigma_{S\in R\setminus i}\frac{1}{N!}\left[f(x_{S\cup i}) - f(x_S)\right] \times \frac{|S|! \times (N-|S|-1)!}{N!}, \tag{1}$$

where R denotes the whole set of the input features, S represents a subset of input features and M stands for the number of input features. $f(x_S)$ denotes the output of the machine learning model when the ith feature is withheld, and $f(x_{S\cup i})$ denotes the output when the ith feature is present. However, although global interpretability can be given by aggregating the SHAP values across all the training examples, it was difficult to consider the influence of all variables on each prediction for the multi-class classification. Therefore, we further adopt other two strategies, OVR and OVO to extract fine-grained local interpretability.

Furthermore, much related work [53, 54] has proved the effectiveness to explain the medical machine learning problems. Therefore, we focus on the tree explainer using SHAP value approach [18] to combine many high-quality local explanations with global ones to give a comprehensive explanation to the model, thus casting insight on the correlation between mobility patterns and health conditions.

5 EXPERIMENT

5.1 Experiment Settings

5.1.1 Baselines. We select a number of baselines to compare against our method. Specifically, the baselines can be divided into three categories: classic models, deep learning models and variants of our model. (i)First, we choose two classic methods to serve as benchmarks: random guess and prediction based on demographic and intuitive mobility features. (ii) Then, we compare our method with neural network based method(i.e., MLP and LSTM). (iii) After that, we present the results of the four variants of *HealthWalks* which represents the results on the partial structure of our whole model. The details of the variants are discussed as follows.

- Random guess. It is the most basic method that sets a naive benchmark for the prediction task. It randomly chooses a disease category for each user.
- Raw Demographic Feature (demo). It is the classic feature based method, where the demographics (i.e., age, gender, income and education level) are leveraged to predict the disease category by XGBoost. It should be noted that this is the traditional questionnaire-based method widely used in hospitals.
- Manual Feature (manual). We select five hand-picked mobility features (i.e., number of records, radius of gyration, entropy, straight line distance and maximum distance from home) and combine them with the points of interest (POI) distribution around users' residence to predict the disease category by XGBoost.
- Demographic & Manual Features (demo+manual) We combine the demographics and the hand-picked mobility features together with POI distribution around a user's residence for prediction by XGBoost.
- Multilayer Perceptron (MLP) [55]. It is a conventional deep neural network which learns the non-linearities from the historical frequency distributions over all locations.
- Long Short-term Memory (LSTM) [56]. It is an artificial recurrent network with feedback connections. LSTM is often used for text classification [57]. Here we input the points of interest sequence as time series data to predict the disease category for each user.
- *HealthWalks* (ego). It is a variant of *HealthWalks*, which only exploits the mobility records grouped by ego network in our framework.
- *HealthWalks* (ego+demo). It is a variant of *HealthWalks*, which aggregates the mobility features grouped by demographic network in our framework.
- *HealthWalks* (ego+poi). It is a variant of *HealthWalks*, which incorporates the POI information into the trajectory.
- *HealthWalks* (ego+demo+poi). It is a variant of *HealthWalks*, which aggregates both the demographics and POI information into our framework.

Except for the random guess and raw demographic feature method, all the evaluated baselines utilize the mobility traces, where the MLP model extracts the historical frequency distributions, and the LSTM makes use of the points of interest information as time series data. Considering that all the users in our evaluation datasets have ground truth information about their disease categories, we feed all the mobility metrics generated by *HealthWalks* into a supervised classification model to predict the disease category. Specifically, we adopt the widely used and highly effective boosting tree system: XGBoost [51] as our classifier, and it can be easily switched to other supervised and semi-supervised models.

5.1.2 Feature Selection. From the 1,032 mobility metrics generated by the *HealthWalks*, we utilize the norm-1 regularizer to remove the most redundant and irrelevant metrics and preserve the ones than can be the best joint predictors. Figure 4 illustrates how model performance depends on the parameter of regularization λ. Here we report both the cross-validated performance (5-fold cross-validation) in terms of Micro-F1 and Macro-F1. For a small regularization parameter λ, a large subset of the features are selected, which might cause 'overfit' on the small number of training examples. When the parameter λ is extremely high, the universal set of the mobility metrics is preserved, and both the Micro-F1 and Macro-F1 levels off. As the parameter λ increases, more redundant or irrelevant features are removed, and both Micro-F1 and Macro-F1 increases. The Macro-F1 reaches its peak with 750 features preserved. As the parameter λ continues to decrease, less features are selected and the performance begins to degrade, which means that the model is so strict that it even drops some useful predictors. When the parameter λ is extremely large, very few features are preserved, so the performance is poor. It should noted that all the baselines related with our *HealthWalks* have conducted feature selection before the final prediction using XGBoost.

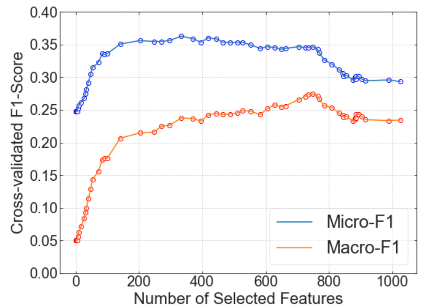


Fig. 4. The influence of regularization on model performance. More features are selected as the regularization parameter λ increases, and the optimal regularized model contains 750 features. Adding more features will lead to overfitting while removing features will lower the model performance.

5.1.3 Evaluation Metric. We adopt Micro-F1 [58] Macro-F1 [59] and Top-k Accuracy [60] metrics to evaluate the prediction performance. They are the most adopted metrics in real-world application [61, 62], and they balanced measure model's performance from different aspects.

$$\text{Macro-F1} = \frac{1}{L}\sum_{l=1}^{L}\frac{2\cdot TP_l}{2\cdot TP_l + FN_l + FP_l},\ \text{Micro-F1} = \frac{\sum_{l=1}^{L} 2\cdot TP_l}{\sum_{l=1}^{L}(2\cdot TP_l + FN_l + FP_l)},\ \text{Accuracy} = \frac{\sum_{i=1}^{M}|T_i \in S_i(k)|}{M},$$

where M stands for the number of samples, L represents the number of disease category, $S_i(k)$ is the set of the top-k attributes the model predicts user i exhibits, and T_i is the ground truth. $|T_i \in S_i(k)|$ equals to 1 if the ground truth attribute is included in $S_i(k)$, otherwise it equals to 0. Without loss of generality, we set $k = 2$ for the Top-k Accuracy metric.

In order to ensure the robustness and effectiveness of the results, we randomly split the datasets into 5 subsets, and report the average performance of 5-fold cross-validation. For the prediction of each disease category, we make sure that it distributes uniformly in each subset.

5.2 Result Analysis

5.2.1 Performance Comparison with Baselines. The overall experimental results are summarized in Table 4. For the proposed *HealthWalks* model, we conduct statistical analysis to examine the significance of performance gain over the best baseline. Specifically, we adopt Student's t-test with Bonferroni correction to figure out the p-value p and effect size ES respectively, where $p < 0.05$ and $ES > 0.8$ indicate the performance gain is statistically significant [63]. According to the results of the performance, we have the observations and conclusions as follows.

- First of all, the demographic and manual feature based methods show notable performance gain over the random guess method and deep neural network methods, including MLP and LSTM which simply utilize the frequency distribution vectors and the points of interest sequence. A possible reason is that the complex correlation between mobility patterns and disease category cannot be preserved by the simple frequency distribution vectors or the points of interest (POI) sequence, which indicates the importance of generating numerous features to boost the performance of the disease category prediction.

- Secondly, our proposed *HealthWalks* significantly outperforms all baselines. *HealthWalks* reaches a 22.84% relative gain over the best baseline (i.e., demo+manual) in terms of Micro-F1 and 31.79% relative gain in terms of Macro-F1 and 11.85% relative gain in terms of Top-2 Accuracy. It should be noted that all the relative performance gains mentioned above have passed the Student's t-test with Bonferroni correction ($p < 0.05, ES > 0.8$).
- Finally, we compare our complete *HealthWalks* model with the degraded versions, and the results indicate that the performance grows consistently as more parts are added to the framework, proving that each part of our framework plays its own role in enhancing the prediction of the disease category. It should be noted that the ego version of our *HealthWalks* has already gained higher performance than the traditional method using raw demographic features, which showcases the effectiveness of our model to capture the mobility patterns linked with health conditions.

In conclusion, our model achieves great performance in disease category prediction compared with simple deep neural network methods and models only based on demographics and intuitive mobility metrics, indicating that our model can better capture the health signals reflected by the mobility trace. Furthermore, the results suggest that our model can successfully aggregate the semantic information and demographics, which demonstrates the effectiveness and scalability of our framework.

Table 4. Performance comparison with baseline models, where (**) indicates $p < 0.05$ (with Bonferroni correction) and $ES > 0.8$ over the best baseline.

Method	Micro-F1	relative gain	Macro-F1	relative gain	Top-2 Acc	relative gain
Random guess	0.1112	-66.10%	0.1007	-58.04%	0.2692	-43.35%
MLP	0.2115	-35.52%	0.1771	-26.21%	0.3681	-22.54%
LSTM	0.2463	-24.91%	0.2108	-12.17%	0.3842	-19.15%
demo	0.3213	-2.04%	0.2245	-6.46%	0.4605	-3.09%
manual	0.2624	-20.00%	0.2091	-12.88%	0.3936	-17.17%
demo+manual	0.3280	0.00%	0.2400	0.00%	0.4752	0.00%
HealthWalks (ego)	0.3226	-1.65%	0.2266	-5.58%	0.4203	-11.55%
HealthWalks (ego+demo)	0.3561	8.57%	0.2554	6.42%	0.4685	-1.41%
HealthWalks (ego+POI)	0.3574	8.96%	0.2652	10.50%	0.4605	-3.09%
HealthWalks (ego+demo+POI)	0.3722	13.48%	0.2944	22.67%	0.4926	3.66%
Demo+HW (ego+demo+POI)	**0.4029****	22.84%	**0.3163****	31.79%	**0.5315****	11.85%

5.2.2 Performance across Different User Groups. Now, we evaluate how the model performance varies across user groups with different disease category and demographics (i.e., gender, age, education and income). The observations and conclusions can be summarized as follows.

- First of all, the performance across different disease categories is displayed in Figure 5 in the form of confusion matrix, where the elements of the main diagonal represent the precision for each disease category. As can be seen clearly in Figure 5, the healthy people yield the highest precision score (56%), followed by cardiology (41%) and endocrinology (36%), while the orthopedics group falls last (13%). One plausible reason is that the healthy people show a far more active mobility pattern from all other people with various disease categories, while the orthopedics seems to be confusingly similar to the birth, endocrinology and dental groups, since they also show less active mobility pattern. This confusion matrix suggests that we should explore more about why some categories are easy to pick out, while some disease categories are confusingly similar to each other, which will be discussed in Section 6.

- Secondly, we showcase the performance across different demographics in Figure 6. Figure 6(a) indicates that female group has a slightly higher F1 score than the male group. Figure 6(b) shows that the F1-score of disease category prediction decreases as the age increases, which is probably because older people tend to get into more diseases as well as become less physically active, which results in less information to distinguish different health conditions. The performance across different income groups is displayed in Figure 6(c), indicating that the performance consistently increases as the income level increases. One possible reason is that people with a higher income level might lead a life with higher standard and pay more attention to their well-being. A similar finding is displayed in Figure 6(d), where the postgraduate group shows higher performance than those with lower education level. The possible reason is that people with a higher education level might lead to a higher income level, which might improve their life quality.

To conclude, our proposed model generally performs better in the user group who is healthy, younger and with higher income level and higher education level. The variation in the performance across different demographics, especially age, income and education level showcases the necessity to combine demographics with mobility features to enhance the model performance.

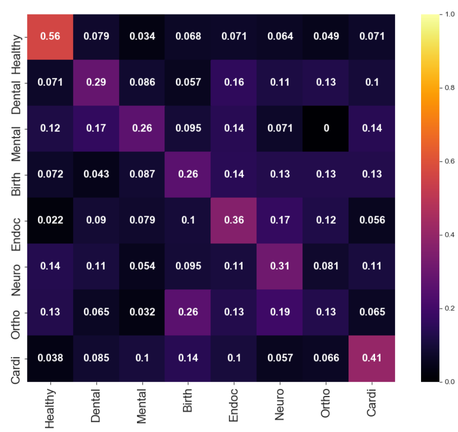
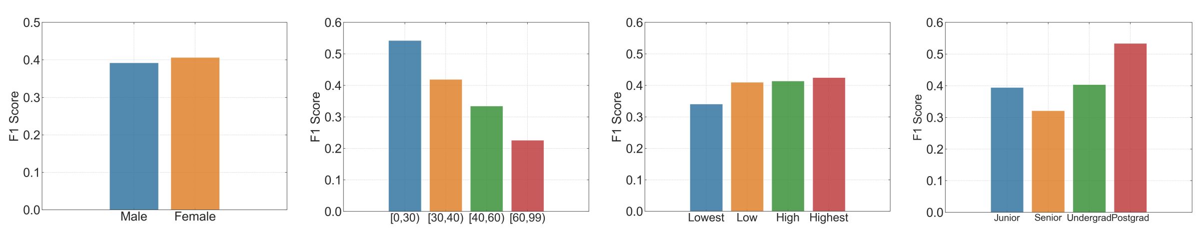
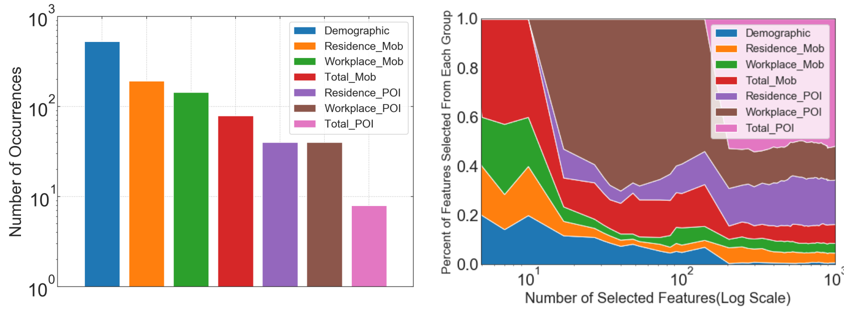


Fig. 5. Confusion matrix of model performance across different disease category.



(a) gender (b) age (c) income (d) education

Fig. 6. Performance of disease category prediction across users groups with different characteristics.



(a) (b)

Fig. 7. The distribution of the feature category 7(a) and the influence of regularization on feature selection and model performance 7(b)

6 UNDERSTANDING THE MOBILITY FEATURES FOR HEALTH SENSING

In this section, we aim at understanding the mobility features generated by *HealthWalks* for health risk prediction. First of all, we concentrate on those features generated by *HealthWalks*. Since these features can be used for many downstream tasks, i.e., representing an individual's mobility behavior and predicting the disease categories, while preserving decent interpretability, it should be of vital importance to give full analysis of these features and cast insight on the health signals reflected by the mobility trace.

6.1 Feature Categorization

Before diving into the analysis of the training process, we should first classify the numerous mobility metrics generated by *HealthWalks*, since the quantity and fine granularity of these features might complicate the interpretation, which is a weak point compared with models based on several intuitive features. In order to overcome this possible disadvantage, we classify each feature with a certain category from a more macroscopic perspective. Previous work [36] has indicated that residence and workplace are two vital locations since people spend most of their time around their residence and workplace during daytime and nighttime, respectively. Therefore, we manually classify the feature according to the two legal transitions of the filtering operation 'filter mobility records around residence' and 'filter mobility records around workplace' in Section 3. Also, we should differentiate the features between physical mobility features and semantic mobility features using points of interest (POI) information. Finally, we label the 1,032 features into 7 groups: demographic, mobility metric around the residence, mobility metric around the workplace, mobility metric over the total mobility trace, POI-related metric around the residence, POI-related metric around the workplace and POI-related metric over the total semantic mobility trace. The distribution of the features across the categories is displayed in 7(a). Furthermore, we plot the influence of the regularization on the feature selection from each group. As is shown in Figure 7, the set of features selected becomes more diverse as the penalty decreases. When the penalty is extremely high, only demographics and physical mobility features are selected. Semantic mobility features are gradually selected when the penalty decreases. Finally, the optimal regularized model includes 750 features, and these features are distributed almost uniformly in each feature category.

6.2 Feature Analysis

6.2.1 Global Interpretability. First of all, we apply multi-class XGBoost with explainable technique (SHAP) to offer global interpretability and show the distribution of the SHAP value across different feature categories. The global feature importance is computed by the average of the SHAP values aggregated across all the training samples for each disease category, which represents the average impact of each feature on the model output magnitude. Feature names in the graphs to follow are interpreted according to descriptions in Section 4.3.

Overview: We present the overall global feature importance in Figure 8(a). POI-related metrics over a total mobility trace have the largest influence on the outcome due to their relatively large count and rich semantic information. Each category has reasonably balanced impacts on each of the classification outcomes. **Demographic:** Impacts of demographic information are shown in Figure 8(b). Among them, age has the largest impact on prediction results, followed by gender due its strong correlation with the birth department. **Residence Mobility:** The top 5 physical mobility features around residence are displayed in Figure 8(c). The number of locations, uncorrelated entropy, and real entropy averaged across all days turn out to be impactful. **Workplace Mobility:** As is displayed in Figure 8(d), top contributing features are still entropy and number of locations, but radius of gyration around workplace has shown increased impact on the prediction. **Total Mobility:** The top 5 physical mobility metrics on the entire mobility trace are shown in Figure 8(e), where distance-related metrics seem to be more vital. **Residence POI:** As is shown in Figure 8(f), the number of restaurants, shops and life services visited around residence, which implies the frequency of visits to the places with specialized functions, showcases great impact on the prediction outcomes. **Workplace POI:** The top 5 semantic mobility features around the workplace are displayed in Figure 8(g). Also, daily averages, sum and maximum of restaurants, shopping, recreation and education POIs visited are of great importance. **Total POI:** The top 5 semantic mobility metrics across the whole mobility trace are presented in Figure 8(h), where visits to educational, shopping and residential areas have shown prominent importance.

Summary: On one hand, physical mobility metrics (distance, entropy etc.) are impactful in the classification of potential patients. A possible reason is that such physical metrics effectively reflect physical activeness, jointly forming a health signal that largely contributes to the prediction of patients. One the other hand, POI-related metrics contribute more to disease categorization, as they are closely linked with the users' life habits that indirectly affect various aspects of health. The impacts and local explanations of these features will be inspected more carefully in the next section.

6.2.2 Local Explanations. We adopt additional two strategies for more fine-grained explainable classification, one versus rest (OVR) and one versus one (OVO) to offer local explanations for the influence of each feature on each prediction result. For OVR classification models, we present the Top-5 contributing features that lead to consistent results with reasonable explanations. The observations and conclusions are summarized as follows.

Healthy: The top 5 contributing features for healthy-vs-rest classification are given in Figure 9(a). Compared to patients, healthy users show notably different mobility patterns. The most prominent factors in recognizing healthy people are the total entropy and the daily average of entropy [44] of visited locations around the workplaces during working hours. Larger entropy contributes to a higher probability of being healthy, which leads to the understanding that those who are able to visit various locations around workplaces tend to be physically more active and consequently more healthy. Another influential feature is the daily average of company POIs visited on weekdays away from the workplace, a higher value of which indicates a higher tendency of being unhealthy. The plausible reason is that visits to companies outside the workplace are usually due to business, a large number of such visits indicate a higher probability of overworking and even overfatigue, leading to health issues.

Dental: The Top 5 contributing features are shown in Figure 9(b). Frequent visits to sports-related locations are closely associated with relatively higher chances of dental problems. Although the correlations are unclear

(a) Overview (b) demographic

(c) residence mobility (d) workplace mobility

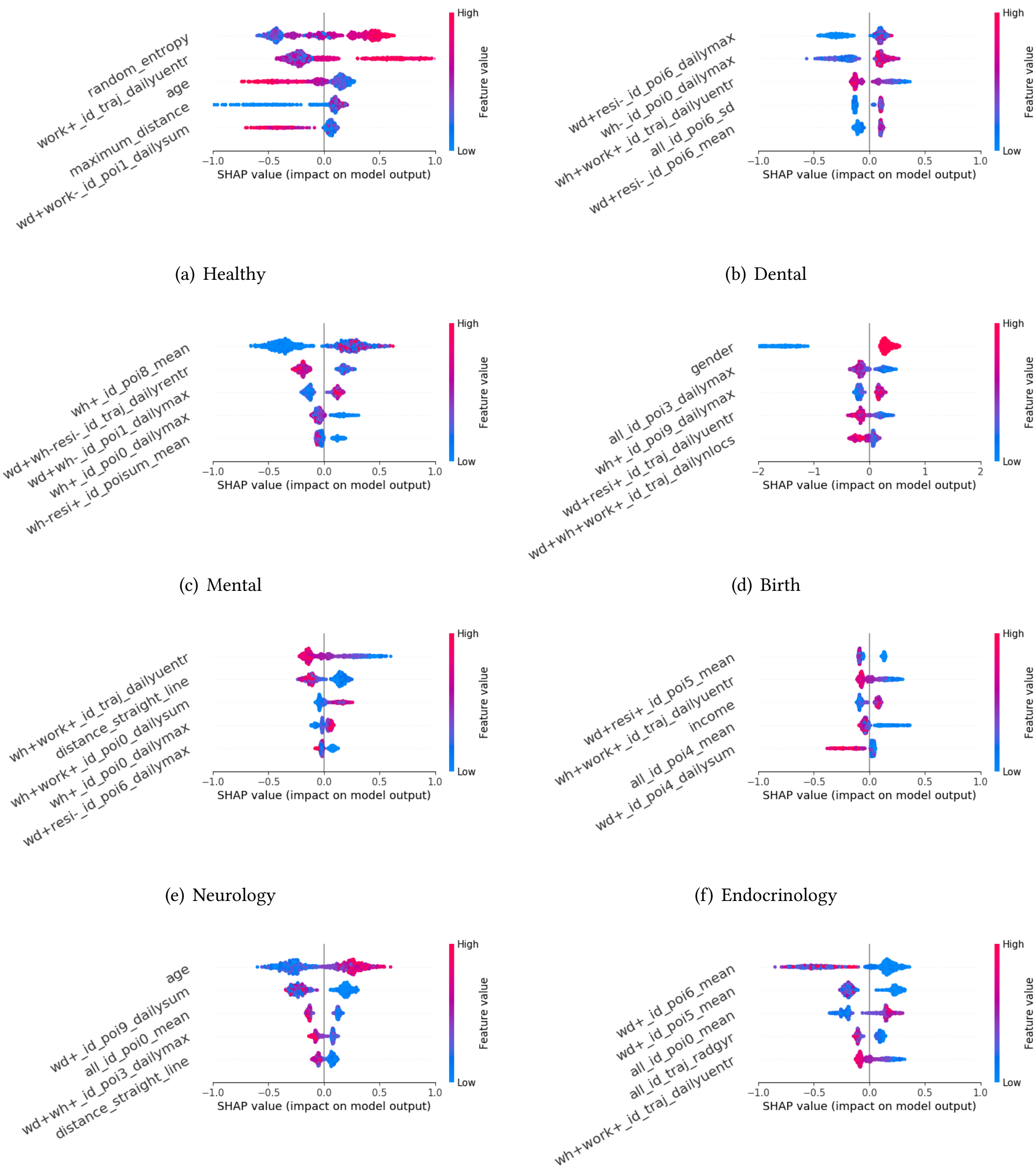
(e) total mobility (f) residence POI

(g) workplace POI (h) total POI

Fig. 8. The global feature importance across different feature categories.

without further case studies, it is worth noting that existing research has reported increased dental injuries [64] and oral dehydration [65] during sports activities.

Mental: The Top 5 contributing features are displayed in Figure 9(c). Longer stays around education and company POIs in the daytime imply a higher likelihood for mental issues. The pressure from heavy work or study are the most plausible reasons. Further indicators include decreased real entropy and less visits to POI hot



(a) Healthy

(b) Dental

(c) Mental

(d) Birth

(e) Neurology

(f) Endocrinology

(g) Orthopedics

(h) Cardiology

Fig. 9. Top-5 features for OVR classification models. Each dot in the graph corresponds to a user, and its color denotes the value of the feature and its horizontal position denotes the SHAP value of the feature linked with the outcomes. For example, in Figure 9(b), the rightmost dots are considered more likely to be healthy samples by the model, and the converse holds.

areas, which showcase the inactive mobility pattern of people with mental disorders. Besides, visits to restaurants often correlate with social activeness and are negatively correlated to the risk of mental issues.

Birth: The Top 5 features are given in Figure 9(d). Gender turns out to be the most effective predictor to eliminate all male users from birth cases. Fewer visits to shopping areas also indicate increased likelihood of hospital visits due to birth, as women in ordinary social roles tend to visit shops frequently, and a notable decrease may indicate noteworthy changes in physical conditions.

Endocrinology: The Top 5 features are displayed in Figure 9(e). Indicators of low physical activity increase the risk of endocrine diseases. Such indicators include low entropy, low total distance, and increased stays around companies during daytime, while more visits to sports-related locations reduce the risk [1].

Neurology: The Top 5 features are shown in Figure 9(f). Increased visits to recreational locations and life services lower the likelihood of neurological disorders, as well as increased entropy indicating variety in mobility.

Orthopedics: The Top 5 contributing features are showcased in Figure 9(g). Age has a large impact on hospital visits due to orthopedic issues, as the elderly are less resilient to external damage. Decreased visits to restaurants, shops, and residence, implying limited mobility, reduced movement, and possibly less external injuries, contribute to decreased likelihood of orthopedic issues.

Cardiology: The Top 5 features are displayed in Figure 9(h). Frequent visits to sport facilities imply sufficient physical exercise leading to less predicted likelihood of heart diseases [1]. Frequent visits to restaurants are persumably linked to irregular and unhealthy eating habits and consequently more risk of chronic heart problems [11].

Summary: To conclude, the local explanations based on the OVR classification can reveal more meaningful and fine-grained health signals reflected by the mobility trace, which can be good knowledge for human experts to gain better understanding of the disease mechanisms and the role of physical activities in health. For example, mobility entropy has shown great association with health conditions, with lower values suggesting lower physical activeness and consequently higher likelihood to have various categories of diseases. There may also be cases where physical conditions lead to inactive mobility, i.e. pregnancy will definitely lower the activeness of mobility behavior. Higher risks of endocrine and cardiovascular diseases tend to be linked with fewer visits to sports facilities and more visits to restaurants. Features related to different kinds of POIs exhibit different associations with all disease categories, presumably due to their strong connection to users' life habits and social activities.

6.2.3 A Detailed Case Example. Finally, in order to extract more detailed local interpretability, we closely inspect one of the users, plot the SHAP values combining the OVR and OVO outcomes via a force plot, and display the reasonings for the prediction of our model. As is shown in Figure 10, the patient is predicted to have the Endocrinology-related disease. The OVR classification result displays the reason for choosing endocrinology as having frequent visits to restaurants during working hours and having a low daily average of mobility entropy during working hours. Next, the results of the OVO classification explain against predicting the birth department by highlighting the notable frequent visits to restaurants and an age unlikely to give birth, and against predicting as healthy by frequent visits to restaurants and scarce visits to sports-related locations. To conclude, the results combined consistent results with reasonable and comprehensive explanations, showcasing a successful attempt at department suggestion with tailored reasoning as a reference.

7 DISCUSSION

7.1 Health vs. Mobility

The most important finding from our analysis is that individual health conditions are indeed closely related to user mobility patterns. While there have been numerous studies in the medical/health domain and HCI communities that discuss such potential associations [11, 57], we are the first to empirically show that there are surprising levels of signal overlap between individual health conditions and user mobility data. Specifically, we

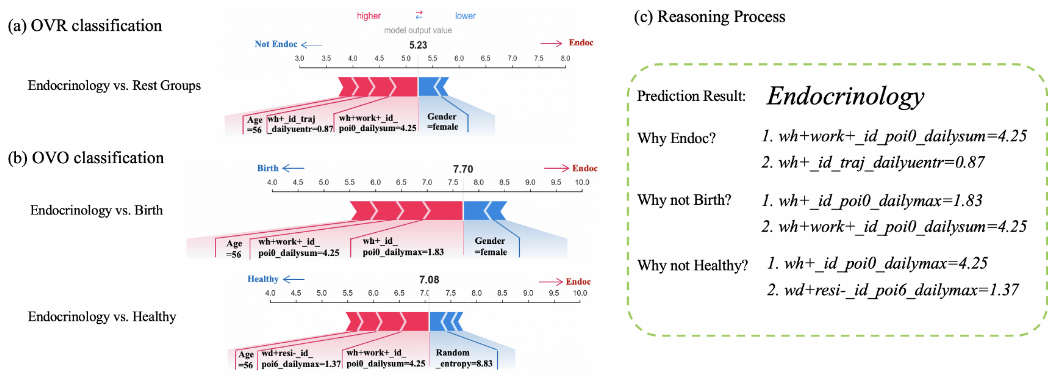


Fig. 10. An Endocrinology case example which displays a reasoning procedure with local explanations. Red arrows indicate the feature effects that drive the XGBoost outcomes higher while blue arrows show the feature effects that drive results lower.

demonstrate that with mobility features and social demographic features, we can achieve 40.29% in Micro-F1 and 31.63% in Macro-F1 when predicting 8 disease categories of individual users. Moreover, our analyses reveal insightful associations between mobility features and certain disease categories. In particular, mobility entropy has shown great association with health conditions, with lower values suggesting lower physical activeness and consequently a higher likelihood to be unhealthy. Higher risks of endocrine and cardiovascular disorders tend to be associated with fewer visits to sports facilities and more visits to restaurants. Features related to various POIs exhibit different correlations with different disease categories including mental, endocrine and cardiology, etc., probably due to their connection to life habits and social activities. To conclude, while the neural based models focus on achieving better performance [66, 67] on health sensing, our approach is much more interpretable and can help guide domain experts to better understand disease mechanisms and the role of physical activities in health.

On the other hand, it would be interesting to discuss our findings from the perspective of user privacy. Our study essentially demonstrates that there is a strong link between two extremely sensitive data sources, e.g. individual mobility data and health condition. Thus improperly revealing user trajectories may lead to even more severe privacy leakage as the user's health condition may be potentially inferred. Therefore, our analysis indicates that more careful steps should be taken with regard to mobility data to ensure that user privacy be protected in mobile applications.

7.2 Implications and Application

Our research has important design implications for ubiquitous computing as well. As demonstrated by our study, individual mobility features can help infer fine-grained individual health risks. As an increasing number of devices can collect user mobility data, e.g. cellphone, and wearable devices like smart watches, we envision future mobile health sensing applications which leverage such collected user mobility data, along with other useful user social demographic information, to reveal the potential health risks to the user. This would provide much more accessible and timely health alerts to users, as a complement of traditional medical examination, which is more accurate yet more expensive and time-consuming. Moreover, we envision that our sensing techniques can be combined with gamification systems to promote individual user positive behavior changes [68].

Our findings can also be useful on a more aggregated level, e.g. at the neighborhood level or city level. For the application of the scheduling urban resources, a feasible way is to randomly collect large-scale mobility data from people (i.e., including healthy people and people who have certain kinds of disease) and check their recent outpatient information from the hospitals. Then we could give a rough estimation of health condition for each region based on our model. This could be achieved by aggregating the individuals' health conditions predicted by our model for each region. For example, assuming that we could estimate from our model that 80% of the randomly selected samples from District A are predicted to have high risk in endocrinology. After being aware of the interpretability extracted from the model that frequent endocrinology-related disease cases are highly correlated with restaurants and sports facilities, the urban planners in District A will be suggested to build more sports facilities and less restaurants to help improve the situation. In the similar way, urban planners can derive useful insights, e.g. better planning of green space and transportation route, medical resources, etc.

Furthermore, interpretability extracted from our proposed model could also be helpful for individuals, healthcare office experts and machine learning model developers. For machine learning model developers, it is obvious that they could be benefited from the extracted interpretability to select vital features and fine-tune their algorithms more efficiently. For individuals, they could be told from the monitoring system what specific mobility features lead to the prediction of having a certain kind of disease. For example, if a user is predicted to have cardiology-related disease, and our model indicates the features that contribute most to the prediction are 'the high frequency of going to restaurants' and 'low frequency of going to the sports facilities', then the user will try to change the situation by paying less frequent visits to restaurants and doing more physical exercises. We believe through conveying this interpretability of the model to individuals, their living habits could be improved! In the similar way, this interpretability could bring insight to healthcare experts or even urban planners for their policy making and further research.

7.3 Limitation and Future Work

Our work has some limitations as well. First, the experiment results are not sufficient enough for an accurate patient classification system. Therefore, the intended applications of our work would be the prediction of potential health risks for a more personalized physical examination, and region-level health conditions from a macroscopic view. One main reason for the relatively low performance is the limited number of training samples. In the future, we plan to collect real-world dataset at larger scale to achieve better performance and see how mobility behavior is linked with broad health condition at region level. Secondly, we carried out all the analysis in an observational way thus we are not able to derive any causal relationship thus we focus on correlation in our paper, e.g. we are unable to distinguish whether mobility pattern lead to certain disease or vice versa, yet we argue the correlation can be helpful enough for sensing of individual health condition.

To conclude, the current approach may not be refined enough to assist with accurate medical diagnosis, however it does suggest feasibility of the approach and the value of continued research in leveraging ubiquitous mobility data for better health monitoring systems. In the future, we plan to incorporate more controlled experiments with our data analysis so as to better understand the causal relationship. Besides, our analysis is limited to one single city, in the future we plan to run analysis on multiple cities to ensure the generalizability of our findings. our work has proved this possibility and has achieved relatively good performance with limited information. Furthermore, if we could combine the mobility data with more disease-related information like symptom data or historical health records, the performance is expected to be further improved for more accurate and more fine-grained health sensing system.

8 CONCLUSION

In this paper, we explore the problem of leveraging large-scale urban mobility data to predict fine-grained health condition. To make this study possible, we collect health conditions of 747 users who visit hospital and tracked their mobility for 2 months. To automatically capture the correlation between mobility patterns and health conditions, we propose *HealthWalks* to combine points of interest data with raw trajectory data to generate a large number of physical and semantic mobility metrics. Features are then selected by regularization technique for the final classification. Empirical experiments on a real-world dataset show that our model achieves 40.29% in Micro-F1 and 31.63% in Macro-F1 for the 8-class disease category prediction, which outperforms the best baseline by 22.84% in Micro-F1 and 31.79% in Macro-F1. In addition, the mobility metrics generated by *HealthWalks* also captures the health signal reflected by the mobility trace, which is of vital importance for numerous applications, especially for a more customized physical examination and the scheduling of the medical resources in the urban area.

REFERENCES

[1] David M Buchner, Shirley AA Beresford, Eric B Larson, Andrea Z LaCroix, and Edward H Wagner. Effects of physical activity on health status in older adults ii: Intervention studies. *Annual review of public health*, 13(1):469–488, 1992.

[2] Tim Morris, David Manley, and Clive E Sabel. Residential mobility: Towards progress in mobility health research. *Progress in human geography*, 42(1):112–133, 2018.

[3] Nick W Ruktanonchai, Patrick DeLeenheer, Andrew J Tatem, Victor A Alegana, T Trevor Caughlin, Elisabeth zu Erbach-Schoenberg, Christopher Lourenço, Corrine W Ruktanonchai, and David L Smith. Identifying malaria transmission foci for elimination using human mobility data. *PLoS computational biology*, 12(4), 2016.

[4] Zhiwen Yu, En Xu, He Du, Bin Guo, and Lina Yao. Inferring user profile attributes from multi-dimensional mobile phone sensory data. *IEEE Internet of Things Journal*, 2019.

[5] Bjarke Mønsted, Anders Mollgaard, and Joachim Mathiesen. Phone-based metric as a predictor for basic personality traits. *Journal of Research in Personality*, 74:16–22, 2018.

[6] HA Groot De. Patient classification system evaluation. part 1: Essential system elements. *The Journal of nursing administration*, 19(6):30–35, 1989.

[7] Benjamin Somberg, Kenneth Hoyme, Howard Simms, Muralidharan Srivathsa, and David Johnson. System and method for managing patient triage in an automated patient management system, November 9 2006. US Patent App. 11/121,594.

[8] Sungkyu Park, Cheng-Te Li, Sungwon Han, Cheng Hsu, Sang Won Lee, and Meeyoung Cha. Learning sleep quality from daily logs. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2421–2429, 2019.

[9] Michael Leo Birnbaum, Sindhu Kiranmai Ernala, AF Rizvi, Elizabeth Arenare, AR Van Meter, M De Choudhury, and John M Kane. Detecting relapse in youth with psychotic disorders utilizing patient-generated and patient-contributed digital data from facebook. *NPJ schizophrenia*, 5(1):1–9, 2019.

[10] Luca Canzian and Mirco Musolesi. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pages 1293–1304, 2015.

[11] Yingzi Wang, Xiao Zhou, Anastasios Noulas, Cecilia Mascolo, Xing Xie, and Enhong Chen. Predicting the spatio-temporal evolution of chronic diseases in population with human mobility data. In *IJCAI*, pages 3578–3584, 2018.

[12] Qiang Gao, Goce Trajcevski, Fan Zhou, Kunpeng Zhang, Ting Zhong, and Fengli Zhang. Trajectory-based social circle inference. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 369–378, 2018.

[13] Jie Feng, Yong Li, Chao Zhang, Funing Sun, Fanchao Meng, Ang Guo, and Depeng Jin. Deepmove: Predicting human mobility with attentional recurrent networks. In *Proceedings of the 2018 world wide web conference*, pages 1459–1468, 2018.

[14] Fengli Xu, Tong Xia, Hancheng Cao, Yong Li, Funing Sun, and Fanchao Meng. Detecting popular temporal modes in population-scale unlabelled trajectory data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(1):46, 2018.

[15] Nikos Mamoulis, Huiping Cao, George Kollios, Marios Hadjieleftheriou, Yufei Tao, and David W Cheung. Mining, indexing, and querying historical spatiotemporal data. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 236–245, 2004.

[16] Yanwei Yu, Hongjian Wang, and Zhenhui Li. Inferring mobility relationship via graph embedding. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):1–21, 2018.

[17] Cheng Yang, Maosong Sun, Wayne Xin Zhao, Zhiyuan Liu, and Edward Y Chang. A neural network approach to jointly modeling social networks and mobile trajectories. *ACM Transactions on Information Systems (TOIS)*, 35(4):1–28, 2017.

[18] Scott M Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. Explainable ai for trees: From local explanations to global understanding. *arXiv preprint arXiv:1905.04610*, 2019.
[19] Zhidan Liu, Zhenjiang Li, Kaishun Wu, and Mo Li. Urban traffic prediction from mobility data using deep learning. *IEEE Network*, 32(4):40–46, 2018.
[20] Fengli Xu, Zongyu Lin, Tong Xia, Diansheng Guo, and Yong Li. Sume: Semantic-enhanced urban mobility network embedding for user demographic inference. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(3):1–25, 2020.
[21] Jürgen Hackl and Thibaut Dubernet. Epidemic spreading in urban areas using agent-based transportation models. *Future Internet*, 11(4):92, 2019.
[22] Pedro C Hallal, Cesar G Victora, Mario R Azevedo, and Jonathan CK Wells. Adolescent physical activity and health. *Sports medicine*, 36(12):1019–1030, 2006.
[23] Steven N Blair and Jeremy N Morris. Healthy hearts and the universal benefits of being physically active: physical activity and health. *Annals of epidemiology*, 19(4):253–256, 2009.
[24] Adrianne E Hardman and David J Stensel. *Physical activity and health: the evidence explained*. Routledge, 2009.
[25] Steven N Blair, Harold W Kohl, Neil F Gordon, and Robert S Paffenbarger Jr. How much physical activity is good for health? *Annual review of public health*, 13(1):99–126, 1992.
[26] Kenneth E Powell, Amanda E Paluch, and Steven N Blair. Physical activity for health: What kind? how much? how intense? on top of what? *Annual review of public health*, 32:349–365, 2011.
[27] Tim Althoff, Jennifer L Hicks, Abby C King, Scott L Delp, Jure Leskovec, et al. Large-scale physical activity data reveal worldwide activity inequality. *Nature*, 547(7663):336–339, 2017.
[28] Ya-Li Zheng, Xiao-Rong Ding, Carmen Chung Yan Poon, Benny Ping Lai Lo, Heye Zhang, Xiao-Lin Zhou, Guang-Zhong Yang, Ni Zhao, and Yuan-Ting Zhang. Unobtrusive sensing and wearable devices for health informatics. *IEEE Transactions on Biomedical Engineering*, 61(5):1538–1554, 2014.
[29] Javier Andreu-Perez, Daniel R Leff, Henry MD Ip, and Guang-Zhong Yang. From wearable sensors to smart implants—toward pervasive and personalized healthcare. *IEEE Transactions on Biomedical Engineering*, 62(12):2750–2762, 2015.
[30] Agnes Grünerbl, Patricia Oleksy, Gernot Bahle, Christian Haring, Jens Weppner, and Paul Lukowicz. Towards smart phone based monitoring of bipolar disorder. In *Proceedings of the Second ACM Workshop on Mobile Systems, Applications, and Services for HealthCare*, pages 1–6, 2012.
[31] Gerald Bauer and Paul Lukowicz. Can smartphones detect stress-related changes in the behaviour of individuals? In *2012 IEEE International Conference on Pervasive Computing and Communications Workshops*, pages 423–426. IEEE, 2012.
[32] Hancheng Cao, Fengli Xu, Jagan Sankaranarayanan, Yong Li, and Hanan Samet. Habit2vec: Trajectory semantic embedding for living pattern recognition in population. *IEEE Transactions on Mobile Computing*, 2019.
[33] Hancheng Cao, Zhilong Chen, Fengli Xu, Yong Li, and Vassilis Kostakos. Revisitation in urban space vs. online: A comparison across pois, websites, and smartphone apps. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(4):1–24, 2018.
[34] Zhilong Chen, Hancheng Cao, Huangdong Wang, Fengli Xu, Vassilis Kostakos, and Yong Li. Will you come back/check-in again? understanding characteristics leading to urban revisitation and re-check-in. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(3):1–27, 2020.
[35] Ilaria Liccardi, Alfie Abdul-Rahman, and Min Chen. I know where you live: Inferring details of people's lives by visualizing publicly shared location data. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1–12. ACM, 2016.
[36] Yuchen Qiu, Yuanyuan Qiao, Shengmin Wang, and Jie Yang. Residence and workplace recovery: User privacy risk in mobility data.
[37] Hongzhi Shi, Yong Li, Hancheng Cao, Xiangxin Zhou, Chao Zhang, and Vassilis Kostakos. Semantics-aware hidden markov model for human mobility. *IEEE Transactions on Knowledge and Data Engineering*, 2019.
[38] Yu Zheng, Lizhu Zhang, Xing Xie, and Wei-Ying Ma. Mining interesting locations and travel sequences from gps trajectories. In *Proceedings of the 18th international conference on World wide web*, pages 791–800, 2009.
[39] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *nature*, 453(7196):779–782, 2008.
[40] Kai Zhao, Sasu Tarkoma, Siyuan Liu, and Huy Vo. Urban human mobility data mining: An overview. In *2016 IEEE International Conference on Big Data (Big Data)*, pages 1911–1920. IEEE, 2016.
[41] Hancheng Cao, Jagan Sankaranarayanan, Jie Feng, Yong Li, and Hanan Samet. Understanding metropolitan crowd mobility via mobile cellular accessing data. *ACM Transactions on Spatial Algorithms and Systems (TSAS)*, 5(2):1–18, 2019.
[42] Luca Pappalardo. Scikit-mobility: A python library for the analysis, generation and risk assessment of mobility data. *arXiv preprint arXiv:1907.07062*, 2019.
[43] Sahar Hoteit, Stefano Secci, Stanislav Sobolevsky, Guy Pujolle, and Carlo Ratti. Estimating real human trajectories through mobile phone data. In *2013 IEEE 14th International Conference on Mobile Data Management*, volume 2, pages 148–153. IEEE, 2013.

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

[45] Hancheng Cao, Jie Feng, Yong Li, and Vassilis Kostakos. Uniqueness in the city: Urban morphology and location privacy. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(2):1–20, 2018.

[46] Christian Albert Hammerschmidt, Sicco Verwer, Qin Lin, and Radu State. Interpreting finite automata for sequential data. *arXiv preprint arXiv:1611.07100*, 2016.

[47] Joshua Blumenstock, Gabriel Cadamuro, and Robert On. Predicting poverty and wealth from mobile phone metadata. *Science*, 350(6264):1073–1076, 2015.

[48] Fan Li, Yiming Yang, and Eric P Xing. From lasso regression to feature vector machine. In *Advances in Neural Information Processing Systems*, pages 779–786, 2006.

[49] Andrew Y Ng. Feature selection, l 1 vs. l 2 regularization, and rotational invariance. In *Proceedings of the twenty-first international conference on Machine learning*, page 78, 2004.

[50] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[51] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[52] Mikel Galar, Alberto Fernández, Edurne Barrenechea, Humberto Bustince, and Francisco Herrera. An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recognition*, 44(8):1761–1776, 2011.

[53] Tae Keun Yoo, Ik Hee Ryu, Hannuy Choi, Jin Kuk Kim, In Sik Lee, Jung Sub Kim, Geunyoung Lee, and Tyler Hyungtaek Rim. Explainable machine learning approach as a tool to understand factors used to select the refractive surgery technique on the expert level. *Translational Vision Science & Technology*, 9(2):8–8, 2020.

[54] Simon Meyer Lauritsen, Mads Kristensen, Mathias Vassard Olsen, Morten Skaarup Larsen, Katrine Meyer Lauritsen, Marianne Johansson Jørgensen, Jeppe Lange, and Bo Thiesson. Explainable artificial intelligence model to predict acute critical illness from electronic health records. *arXiv preprint arXiv:1912.01266*, 2019.

[55] Paul Covington, Jay Adams, and Emre Sargin. Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM conference on recommender systems*, pages 191–198, 2016.

[56] Felix A Gers, Jürgen Schmidhuber, and Fred Cummins. Learning to forget: Continual prediction with lstm. 1999.

[57] Chunting Zhou, Chonglin Sun, Zhiyuan Liu, and Francis Lau. A c-lstm neural network for text classification. *arXiv preprint arXiv:1511.08630*, 2015.

[58] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864. ACM, 2016.

[59] Jinpeng Wang, Gao Cong, Wayne Xin Zhao, and Xiaoming Li. Mining user intents in twitter: A semi-supervised approach to inferring intent categories for tweets. In *AAAI*, 2015.

[60] Maksim Lapin, Matthias Hein, and Bernt Schiele. Top-k multiclass svm. In *Advances in Neural Information Processing Systems*, pages 325–333, 2015.

[61] Andrew I Schein, Alexandrin Popescul, Lyle H Ungar, and David M Pennock. Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 253–260. ACM, 2002.

[62] Monica Rogati and Yiming Yang. High-performing feature selection for text classification. In *Proceedings of the eleventh international conference on Information and knowledge management*, pages 659–661, 2002.

[63] Paul D Ellis. The essential guide to effect sizes: An introduction to statistical power, meta-analysis and the interpretation of research results. 2010.

[64] Jill Inouye and Chris McGrew. Dental Problems in Athletes. *Current Sports Medicine Reports*, 14(1):27–33, 2015.

[65] Ian Needleman, Paul Ashley, Peter Fine, Fares Haddad, Mike Loosemore, Akbar de Medici, Nikos Donos, Tim Newton, Ken van Someren, Rebecca Moazzez, Rod Jaques, Glenn Hunter, Karim Khan, Mark Shimmin, John Brewer, Lyndon Meehan, Steve Mills, and Stephen Porter. Oral health and elite sport performance. *British Journal of Sports Medicine*, 49(1):3–6, 2015.

[66] Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F Stewart, and Jimeng Sun. Gram: graph-based attention model for healthcare representation learning. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 787–795, 2017.

[67] Junyi Gao, Cao Xiao, Yasha Wang, Wen Tang, Lucas M Glass, and Jimeng Sun. Stagenet: Stage-aware neural networks for health risk prediction. In *Proceedings of The Web Conference 2020*, pages 530–540, 2020.

[68] Sunny Consolvo, David W McDonald, Tammy Toscos, Mike Y Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, et al. Activity sensing in the wild: a field trial of ubifit garden. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1797–1806, 2008.