# FIID: Feature-Based Implicit Irregularity Detection Using Unsupervised Learning From IoT Data for Homecare of Elderly

Cuijuan Shang, Chih-Yung Chang, *Member, IEEE*, Jinjun Liu, Shenghui Zhao, and Diptendu Sinha Roy

***Abstract*—Advances in wireless sensor networks and increasing Internet-of-Things devices give great opportunities for smart homecare of the elderly. Smart homecare has been a promising issue and received much attention recently. Irregularity detection is one of the most important issues in smart homecare for assessing the health condition of the elderly. However, most of the researches focused on the explicit irregularity detections which are usually based on the drastic changes of sensor data, such as falling. Existing mechanisms for detecting implicit irregularity rely on the subjective assessment of behaviors' importance by the elder and simply outputs the binary detection results. This article proposes a feature-based implicit irregularity detection mechanism (FIID), which extracts the regularity features using unsupervised learning and outputs the probability of implicit irregularity. The proposed FIID identifies the regular behaviors which satisfy the *time-regular* and *happen-frequently properties* as the regularity features of daily behaviors. These features then construct a multidimensional feature space to calculate the implicit irregularity probability of the daily health condition. Performance results show that the proposed FIID outperforms the existing implicit irregularity mechanism in terms of precision, recall as well as F-measure.**

***Index Terms*—Daily behavior, homecare, irregularity detection, regularity, unsupervised learning.**

## I. Introduction

THE AGING of the world population is becoming more and more serious. Many challenges, such as increasing the cost of healthcare and insufficient caregiver resources, push people to find ways to cope with the aging problem. Advances in wireless sensor networks and Internet of Things promote the smart healthcare of elderly move from hospitals and clinics to their homes. Meanwhile, the elderly prefer to live in their own home. Consequently, homecare has become an active research domain which mainly involves the many kinds of technologies, including monitoring system [1]–[3], behavior recognition [4]–[6], and irregularity detection [7]–[9]. The monitoring systems aim to record the activities of the resident using video, wearable devices, or wireless sensor networks. The wireless sensor networks have been paid much attention considering the privacy and user friendliness in the sense without requiring the elderly to wear any devices. Furthermore, plenty of behavior recognition algorithms are proposed to identify behaviors by analyzing the relationship between behaviors and sensor data. Irregular behavior can indicate some unsafe situation, health condition deterioration, or unstable state of mind of the elderly [10]. Therefore, irregularity detection can efficiently improve the life quality of the elderly living alone at home.

Irregularity detection is mainly classified into two types: 1) explicit and 2) implicit. The explicit irregularity detection aims to identify the irregular behaviors in real time and requires the corresponding caregivers to immediately process the irregularity, such as racing heart and falling detection [11], [12]. Usually, some features of explicit irregularity are known and consequently some specific sensors can be deployed to capture those known features to identify the irregularity. However, the features of implicit irregularity are uncertain leading to the difficulty in detection. For instance, uncomfortableness is a kind of implicit irregularity, which might result in the changes of occurrence time or order of some behaviors, such as sleep and eat, but might not cause drastic changes in the values of sensors. In this situation, a deep analysis of behaviors or behavior patterns is required to identify the irregularity to alert caregivers.

Since it is difficult to describe the characteristics of implicit irregularity, the idea of implicit irregularity detection is changed to extract the features of regular behavior patterns. Furthermore, if the behaviors do not conform to the extracted features of regular behavior patterns, the implicit irregularity might exist. Some existing algorithms of this type only considered the irregularity of single behavior without discussing the irregularity of a combination of behaviors [13]–[15]. Shang *et al.* [9] treated the behaviors of

Manuscript received February 8, 2020; revised April 3, 2020; accepted April 22, 2020. Date of publication April 27, 2020; date of current version November 12, 2020. This work was supported in part by the Natural Science Foundation of Anhui Higher Education under Grant KJ2019ZD44, Grant KJ2019A0648, and Grant KJ2018B01; and in part by the Anhui Provincial Natural Science Foundation under Grant 1908085MF211. *(Corresponding author: Chih-Yung Chang.)*

Cuijuan Shang, Jinjun Liu, and Shenghui Zhao are with the School of Computer and Information Engineering, Chuzhou University, Chuzhou 239000, China (e-mail: shangcuijuan@chzu.edu.cn; jinjunl@chzu.edu.cn; zsh@chzu.edu.cn).

Chih-Yung Chang is with the Department of Computer Science and Information Engineering, Tamkang University, New Taipei 25137, Taiwan (e-mail: cychang@ mail.tku.edu.tw).

Diptendu Sinha Roy is with the Department of Computer Science and Engineering, National Institute of Technology Meghalaya, Shillong 793003, India (e-mail: diptendu.sr@nitm.ac.in).

Digital Object Identifier 10.1109/JIOT.2020.2990556

one day as a basic analysis element and considered the different impacts of different behaviors on the implicit irregularity detection. However, the extraction of regular behavior patterns in [9] is rough and not delicate. Meanwhile, the weights of different behaviors in [9] are determined based on the subjective judgment of residents, which is not reliable. Therefore, implicit irregularity detection still needs more effort for further research.

This article proposes a feature-based implicit irregularity detection (FIID) mechanism which can be used to issue a warning alert of the elderly to the caregivers or their children. The proposed FIID extracts the regularity features of the behavior patterns for the elderly by using unsupervised learning. Based on the features, the probability of irregularity of one day can be calculated and the implicit irregularity can be identified. The contributions of the proposed FIID are itemized in the following.

1) *Extraction of Regularity Features:* A regular behavior is defined with two properties, including the *time-regular* and *happen-frequently properties*. The proposed FIID adopts unsupervised learning clustering to identify the regular behaviors which can be treated as the regularity features of the behavior patterns for an elder.
2) *Quantification of Regularity:* As different behaviors have different regularities and further have different impacts on irregularity, this article quantifies the regularities of regular behaviors by using histograms. Compared with *IIRD*, the proposed FIID is more reliable since it adopts the degree of regularity to determine the weights of different behaviors rather than the subjective judgment.
3) *Construction of New Model (R-Space) for Identification of Regular Behavior:* This article constructs a multidimensional *R*-space based on the extracted regularity features. Each dimension of the constructed new model is applied to identify the probability of irregularity of important behavior. Then, the probability of implicit irregularity is further calculated by integrating the irregularity probabilities on all the dimensions. Compared with the binary result of *IIRD*, the probability result obtained by applying the proposed FIID is more flexible.
4) *Avoid Overfitting Problem:* The existing *IIRD* adopts the string comparing mechanism to identify the irregularity, which might raise the overfitting problem when some behaviors occurred disorder or missing. This article proposed a FIID mechanism that can avoid the overfitting problem as found in the existing *IIRD*. The experiments verify the improved performance of FIID in terms of precision, recall, and F-measure.

The remainder of this article is organized as follows. Section II summarizes related work. Section III formalizes the assumptions and problem statement of irregularity detection. In Section IV, the FIID mechanism is introduced. Section V presents the performance study and Section VI gives conclusions and future work of this article.

## II. Related Work

Many behavior irregularity detection algorithms have been proposed to assist caregivers or children to understand the health condition of the elderly. In the following, two main types of algorithms, namely, explicit irregularity detection [16]–[20] and implicit irregularity detection [9], [13], [21], are reviewed.

Most of the existing researches focused on the detection of explicit irregularity which can be distinguished by identifying the drastic changes of sensor data. Two major issues of explicit irregularity, including abnormal values of physiological parameters and fall action, have gained lots of attention. The following reviews these researches and compare them with this article.

Some researchers have been proposed for the detection of abnormality of physiological parameters. Dey *et al.* [16] proposed a home-based ECG monitoring system to monitor the heart electrical activity. Compared with [16], Fanucci *et al.* [17] proposed a comprehensive system to collect vital signs of the chronic heart failure patients at home. The values of vital parameters, including electrocardiogram, SpO2, blood pressure, and weight, are processed to asset if they lay in a safety zone established by the thresholds. Furthermore, Nia *et al.* [3] explored the continuous personal health monitoring system that uses eight biomedical sensors. If the value of the health parameter is out of the safety zone formed according to the prior knowledge, alerts will be triggered and assigned to the corresponding caregiver to immediately handle the irregularity. However, these algorithms can only cope with the explicit problems of the physiological parameter.

Fall detection is another crucial issue for the elder living alone at home. Plenty of fall detection algorithms have been proposed aiming to ensure that the caregiver can respond immediately to the fall accidents. To detect fall action at home, different technologies, including visual, wearable, and motion sensors, were adopted for collecting motion data of the elderly. Based on the visual sensor, Rhuma *et al.* [18] designed a fall detection system by recognizing the postures. In [18], the features of the human body foreground were extracted and put into multiclass SVM to classify different behaviors. Furthermore, the floor information was considered to determine falls or other activities. Gutiérrez-Madroñal *et al.* [12] and Wang *et al.* [19] adopted the wearable acceleration sensors to detect falls. A threshold of sensor values is used to identify the occurrence of falls. Guan *et al.* [20] proposed a fall detection system based on the pyroelectric infrared (PIR) sensors by extracting the salient spatial–temporal features of falls. Moreover, Wang *et al.* [7] proposed a truly unobtrusive fall detection system, named WiFall, employing physical layer channel state information (CSI) as the indicator of activities. The local outlier factor is used to detect the falling movements in [7].

The above-mentioned researches focused on the explicit irregularity detection which mainly depends on the high correlations existed between the irregularity and the sensory data. However, the detection of implicit irregularity encounters is

TABLE I
COMPARISONS OF THE MAIN CHARACTERISTICS OF THE PROPOSED FIID WITH THE EXISTING RELATED WORK

| Studies | Implicit Irregularity | Analyzed Data Type | Behavior Combination | Behavior Weight | Multi-Dimensional Feature Space | Algorithms |
|---|---|---|---|---|---|---|
| [7] | No | Channel State Information | No | No | No | one-class Support Vector Machine, Random Forest |
| [16], [17] | No | Sensor data | No | No | No | Threshold-based |
| [18] | No | Sensor data | No | No | No | Multiple SVMs and Rule-based |
| [20] | No | Sensor data | No | No | No | Mask Strategy |
| [13] | Yes | Sensor data | No | No | No | Combination of Regression Analysis, Histogram and Probabilistic Model |
| [21] | Yes | Behavior data | No | No | No | Gaussian Distribution and Fuzzy Fule-based |
| [9] | Yes | A Day's Behavior Data | Yes | Subjective | No | Unsupervised Learning |
| Our Proposed *FIID* | Yes | A Day's Behavior Data | Yes | Statistic | Yes | Unsupervised Learning, Histogram |

more challenging because implicit irregularity usually does not directly lead to the drastic changes of sensor data and needs to analyze a sequence of changes from sensor data or behaviors.

Some researches [13], [21] explored the implicit irregularity detection by investigating long-term sensor data collected from various sensors based on statistic methods or machine learning algorithms. Elbert *et al.* [13] utilized the sensor data collected for a long time period to extract behavior patterns by combining the cosinor analysis, histogram-based approach, and probabilistic model. A model represented by the probabilistic density function was trained for each interested behavior and employed to determine whether a behavior is irregular. However, the impact of training data on performance is significant, which indicates that a large amount of sensor data are needed to ensure performance. Meanwhile, this algorithm cannot be applied to the complex irregularity which involves with multiple behaviors.

Considering context-aware information, Forkan *et al.* [21] adopted a hidden Markov model-based approach to detect irregularity in routine behaviors from statistical histories and further employed an exponential smoothing technique to predict future changes in various vital signs. Then, a fuzzy rule-based model was used to fuse the outcomes of these different models for making the final decision and sending an accurate context-aware warning to the caregivers. However, it is difficult to transform the huge amount of context-aware information into a prior knowledge base and to set fuzzy rules. Furthermore, similar to [13], the proposed mechanism in [21] did not take into account the behavior combination on irregularity and thus cannot cope with the complex irregularity, which only can be identified by investigating a sequence of behavior changes.

Different from [13] and [21], Shang *et al.* [9] proposed an implicit irregularity detection, called *IIRD*, which treated a day's behaviors as a basic element of analysis. Two important features, including the distance and similarity between daily behaviors, were used to classify the regular and irregular daily behaviors and detect the implicit irregularity of elderly health conditions. *IIRD* also considered the combination of behaviors and introduced different weights for different behaviors for detecting the irregularity. Based on the defined features and behaviors' weights, an unsupervised learning algorithm was used to identify the implicit irregularity. However, the features defined in [9] are rough since the occurrence times of daily behaviors were simplified. Besides, the weights of different behaviors were determined according to the subjective judgment of residents, which was not reliable.

In the context of [9], this article proposes an FIID mechanism which can extract the features of regular daily behaviors using unsupervised learning and quantify the regularity of each regular behavior based on the statistic tool. The comparison of existing works and the proposed FIID referring to the irregularity detection is given in Table I. The term "behavior combination" refers to whether the study considered the irregularity issue which is caused by the combination of multiple behaviors. The term "behavior weight" aims to identify whether the study assigned different weights to different behaviors, aiming to emphasize that different behaviors have different importance when identifying the irregularity. The term "algorithm" refers to the key techniques applied in the study. As shown in Table I, the proposed FIID exhibits several good features, including being able to detect implicit irregularity, analyzing a set of behaviors or their combinations, and using unsupervised learning. Compared to *IIRD* [9], the proposed FIID employs the statistic tool to extract delicate features and further defines the weights of behaviors based on the regularities of the extracted features. Besides, to the best of our knowledge, the proposed FIID first constructs a multidimensional feature space as the new model to detect the daily irregularity.

## III. ASSUMPTIONS AND PROBLEM STATEMENT

This section initially introduces the system model and assumptions. Then, the problem statements are described.

### A. System Model and Assumptions

This article considers the scenario that the daily behavior monitoring system as described in [4] has been installed in a home aiming to collect activity data of the elder living alone. Furthermore, the activity recognition algorithm proposed in [4] was applied to identify the behaviors of the elder based on the collected sensor data.

Let $B = \{b_1, b_2, \ldots, b_{|B|}\}$ denote the set of interested behaviors. Let $B_i^c = (b_1^c, b_2^c, \ldots, b_{|B_i^c|}^c)$ denote the sequence of behavior occurrences which actually occur in the $i$th day. For simplification, the notation $B_i^c$ is called *daily behavior sequence* of the $i$th day. The detailed information of a behavior occurrence $b_j^c \in B_i^c$ can be represented by the form $b_j^c(\text{name}, \text{start\_time}, \text{end\_time})$, where name $\in B$ denotes an interested behavior name, start_time denotes the time when the behavior $b_j^c$ occurs, and end_time denotes the time when the behavior $b_j^c$ ends. Furthernore, the notation $\mathbb{B} = \{B_1^c, B_2^c, \ldots, B_m^c\}$ denotes the set of daily behavior sequences in the past $m$ days.

In the next section, the goal of the investigated problem will be presented.

### B. Problem Statements

Given a set $\mathbb{B} = \{B_1^c, B_2^c, \ldots, B_m^c\}$ of daily behavior sequences collected in the consecutive past $m$ days, this article aims to develop a mechanism $\mathcal{M}$ which can identify *irregular days*.

Let $\delta_i^A$ denote a Boolean variable which represents the fact that whether daily behavior sequence $B_i^c = \{b_1^c, b_2^c, \ldots, b_{|B_i^c|}^c\}$ is irregular. This is

$$\delta_i^A = \begin{cases} 1, & B_i^c \text{ is actually irregular} \\ 0, & \text{otherwise.} \end{cases}$$

Assume that an irregularity detection mechanism $\mathcal{M}$ can identify the irregular daily behavior sequence. Let $\delta_i^{\mathcal{M}}$ denote a Boolean variable which represents the detection result of mechanism $\mathcal{M}$ for a daily behavior sequence $B_i^c$. That is

$$\delta_i^{\mathcal{M}} = \begin{cases} 1, & B_i^c \text{ is identified to be irregular by } \mathcal{M} \\ 0, & \text{otherwise.} \end{cases}$$

Given a set $\mathbb{B}$, the true positive (TP) of mechanism $\mathcal{M}$, denoted by $\text{TP}^{\mathcal{M}}$, is defined by the number of identifications that the daily behavior sequence is actually irregular and is also identified to be irregular by $\mathcal{M}$. $\text{TP}_I^{\mathcal{M}}$ can be calculated by

$$\text{TP}^{\mathcal{M}} = \sum_{i=1}^{m} \delta_i^{\mathcal{M}} * \delta_i^A. \quad (1)$$

For the mechanism $\mathcal{M}$, there are two types of error identification, namely, false positive (FP) and false negative (FN). The FP identification, denoted by $\text{FP}^{\mathcal{M}}$, is the number of identifications that the daily behavior sequence is actually regular but is identified to be irregular by applying $\mathcal{M}$. $\text{FP}^{\mathcal{M}}$ is measured by

$$\text{FP}^{\mathcal{M}} = \sum_{i=1}^{m} \max\left(0, \delta_i^{\mathcal{M}} - \delta_i^A\right). \quad (2)$$

Another type of error identification, called FN identification of $\mathcal{M}$, denoted by $\text{FN}^{\mathcal{M}}$, is defined by the number of identifications that the daily behavior sequence actually is irregular but is identified to be regular by applying $\mathcal{M}$. $\text{FN}^{\mathcal{M}}$ can be calculated by

$$\text{FN}^{\mathcal{M}} = \sum_{i=1}^{m} \max\left(0, \delta_i^A - \delta_i^{\mathcal{M}}\right). \quad (3)$$

Then, the recall and precision of irregularity identification for mechanism $\mathcal{M}$, denoted by $\text{RC}^{\mathcal{M}}$ and $\text{PC}^{\mathcal{M}}$, respectively, can be calculated as follows:

$$\text{RC}^{\mathcal{M}} = \frac{\text{TP}^{\mathcal{M}}}{\text{TP}^{\mathcal{M}} + \text{FN}^{\mathcal{M}}} \quad (4)$$

$$\text{PC}^{\mathcal{M}} = \frac{\text{TP}^{\mathcal{M}}}{\text{TP}^{\mathcal{M}} + \text{FP}^{\mathcal{M}}}. \quad (5)$$

Furthermore, the F-measure, denoted by $F_\beta^{\mathcal{M}}$, is used to adjust the weights of FPs and FNs. The following equation gives the calculation of $F_\beta^{\mathcal{M}}$:

$$F_\beta^{\mathcal{M}} = \frac{(\beta^2 + 1) * \text{PC}^{\mathcal{M}} * \text{RC}^{\mathcal{M}}}{\beta^2 * \text{PC}^{\mathcal{M}} + \text{RC}^{\mathcal{M}}}. \quad (6)$$

If the mechanism $\mathcal{M}$ aims to penalize the FNs more strongly than FPs and give more weight to recall, the parameter $\beta$ can be set at a value more than 1. On the contrary, if parameter $\beta$ is set at a value smaller than 1, the FPs will be penalized more strongly and the precision will be given more weight.

Let $\mathbb{M}$ denote the set of all possible mechanisms each of which can identify the implicit irregularity from the set of daily behavior sequences collected in the consecutive past $m$ days for the elderly. This article aims to propose an unsupervised learning mechanism $\mathcal{M}$ which can minimize the number of errors and maximize the F-measure metric when identifying the implicit irregularity. Then, the objective function of this article can be expressed by the following equation:

$$\max_{\mathcal{M} \in \mathbb{M}} F_\beta^{\mathcal{M}}. \quad (7)$$

## IV. Proposed Mechanism

This article aims to identify the irregular days using unsupervised learning from a given set $\mathbb{B}$ of daily behavior sequences collected in the consecutive past $m$ days. Herein, we assume that each irregular daily behavior sequence has a big difference with any regular daily behavior sequence. In other words, if we have the features of regular daily behavior sequences, irregular daily behavior sequences can be identified. Based on this concept, this article first extracts the regularity of daily behaviors. Then, the irregularity will be identified by checking behaviors of each day according to the extracted regularity features.

The proposed FIID mechanism is composed of three phases, namely, selection of regular behaviors (SRB), mapping of daily behavior sequence (MDBS), and identification of irregular days (IID). The SRB phase aims to choose some regular behaviors as the features from daily behavior sequences and further construct a multidimensional feature space. In the MDBS phase, each daily behavior sequence will be mapped to a point in the constructed feature space. Finally, the IID phase identifies the irregular days by checking each point in the feature space according to the regularities of the regular behaviors. The following presents the details of each phase.

### A. Selection of Regular Behaviors Phase

The goal of this phase is to select some regular behaviors which can represent the regularity features of regular days. In the following, a regular behavior is defined. Then, the algorithm that identifies the regular behavior is presented. A behavior is said to be *regular* if it has two regular properties. The following presents properties *time-regular* and *happen-frequently*.

*Time-Regular Property:* A behavior has the *time-regular property* if it occurs regularly at a certain time period on most days.

*Happen-Frequently Property:* A behavior has the *happen-frequently property* if it happens on most days.

For instance, behavior "GetUp" is a regular behavior since the elder gets up every day and the "GetUp" behavior almost happens around 6 o'clock.

This phase mainly consists of two steps. The first step aims to examine whether the behavior has the *time-regular property.* Then, the *happen-frequently property* is further checked.

*1) Examining the Time-Regular Property:* To examine the *time-regular property*, some clustering operations will be applied. Let $\mathbb{B}^c = \{b_1^c, b_2^c, \ldots, b_{|\mathbb{B}^c|}^c\}$ denote the set of all behavior occurrences collected in the past $m$ days. Initially, the set $\mathbb{B}^c$ will be partitioned into several subsets according to the behavior name. Let subset $\mathbb{B}_i^c = \{b_j^c.\text{name} = b_i | b_j^c \in \mathbb{B}, b_i \in B\}$ denotes the set of behavior occurrences of behavior $b_i \in B$. It is a challenge that one behavior might regularly happen many times in different time periods in one day. A typical example is behavior "BrushTeeth" which always happens once in the morning and once in the evening. This implies the invalid identification that behavior "BrushTeeth" is irregular. To conclude that "BrushTeeth" is regular, the two occurrences of "BrushTeeth" should be treated as different behaviors. That is, each $\mathbb{B}_i^c$ should be further partitioned according to the occurred time period.

To achieve this, the unsupervised clustering algorithm, named DBSCAN, is applied to each subset $\mathbb{B}_i^c$ to obtain several behavior occurrence groups. Each behavior occurrence group, denoted by $G^c$, might be associated with a regular behavior. In order to check whether two behavior occurrences $b_j^c \in \mathbb{B}_i^c$ and $b_l^c \in \mathbb{B}_i^c$ belong to the same behavior occurrence group, the distance between them, denoted by $d(b_j^c, b_l^c)$, will be defined according to the occurrence time. That is

$$d\left(b_j^c, b_l^c\right) = \left|b_j^c.\text{start\_time} - b_l^c.\text{start\_time}\right|. \tag{8}$$

DBSCAN has two main parameters: *min_pts* and *Eps*. The parameter *min_ps* is the minimum number of occurrences required in a cluster while *Eps* is the minimum distance between two behavior occurrences for them to be considered in the same cluster. After applying DBSCAN, each behavior occurrence can be either included in a behavior occurrence group or treated as noise. If a behavior occurrence $b_j^c$ belongs to a behavior occurrence group $G^{TR}$, it means that $b_j^c$ has a time-regular property. On the contrary, if $b_j^c$ is treated as noise, it means $b_j^c$ does not have the time-regular property. The parameters *min_pts* and *Eps* will be discussed in Section V.

Let $\mathbb{G}^{TR}$ denote the set of all clusters formed after applying DBSCAN on each $\mathbb{B}_i^c$. We have

$$\mathbb{G}^{TR} = \left\{G_1^{TR}, G_2^{TR}, \ldots, G_{|\mathbb{G}^{TR}|}^{TR}\right\}.$$

Let $b_i^{TR}$ be the behavior whose occurrence belongs to $G_i^{TR}$. That is, $b_i^{TR}$ has been examined to have the *time-regular property* and will be the candidate to be further examined the *happen-frequently property.* Let $B^{TR}$ be the set of all $b_i^{TR}$, for $1 \le i \le |B^{TR}|$. That is

$$B^{TR} = \left\{b_1^{TR}, b_2^{TR}, \ldots, b_{|\mathrm{B}|^{TR}|}^{TR}\right\}.$$

It is noted that if a behavior usually occurs around 24 o'clock, the corresponding occurrences might be clustered into two groups. Let $G_i^{TR,\text{before}}$ and $G_i^{TR,\text{after}}$ denote the set of behaviors that occurred before 24 o'clock and the set of behaviors occurred after 24 o'clock, respectively. The two groups $G_i^{TR,\text{before}}$ and $G_i^{TR,\text{after}}$ will be merged by

$$G_i^{TR} = G_i^{TR,\text{before}} \cup G_i^{TR,\text{after}}.$$

Next, we will further present the examination of the *happen-frequently property* for each $b_i^{TR} \in B^{TR}$.

*2) Examining the Happen-Frequently Property:* To examine the *happen-frequently property*, the histogram is applied to analyze the statistic characteristic of each $G^{TR} \in \mathbb{G}^{TR}$. A histogram of behavior occurrence group $G^{TR}$, denoted by $H^{TR}$, is a graphical display of data using bars with different heights. The width of each bar is also called the *class interval*, denoted by $\tau$, which might be an hour or 10 min. Let $T^D$ denote the time length of one day and is divided into $q$ periods by $\tau$. That is, $T^D = (\tau_1, \tau_2, \ldots, \tau_q)$. For instance, if $\tau = 10$ min, we have $q = 1440$. The height of each bar, denoted by $h_x$, is the *frequency* of behavior occurrence $b_j^c \in G^{TR}$ occurred in time period $\tau_x$.

Let $\textbf{Floor}(b_j^c.\text{start\_time}/\tau)$ denote a function which rounds value $b_j^c.\text{start\_time}/\tau$ down to the nearest integer number. Let $\eta_{x,j}$ denote a Boolean variable representing whether or not the event $b_j^c \in G^c$ occurs in the time period $\tau_x \in T^D$. That is

$$\eta_{x,j} = \begin{cases} 1, & \text{Floor}\left(\frac{b_j^c.\text{start\_time}}{\tau}\right) + 1 = x \\ 0, & \text{otherwise.} \end{cases}$$

Then, the histogram $H^{TR}$ of $G^{TR}$ can be calculated by

$$H^{TR} = \left\{ h_x | \ h_x = \sum_{b_j^c \in G^{TR}} \eta_{x,j} \quad \forall \tau_x \in T^D \right\}. \tag{9}$$

We assume that the occurrence time of a regular behavior follows a normal distribution. Let $\bar{t}_s$ and $\sigma$ denote the mean and standard deviation values of the start times of behavior occurrences in group $G^{TR}$, respectively. That is

$$\bar{t}_s = \frac{\sum_{b_j^c \in G^{TR}} b_j^c.\text{start\_time}}{|G^{TR}|}$$

$$\sigma = \sqrt{\frac{\sum_{b_j^c \in G^{TR}} \left(b_j^c.\text{start\_time} - \bar{t}_s\right)^2}{|G^{TR}|}}.$$

For the ease of presentation, the distribution of group $G^{TR}$ is denoted by $G^{TR}(\bar{t}_s, \sigma)$. Here, we will take $[\bar{t}_s \pm 2\ \sigma]$ as the confidence interval. Corresponding to the histogram, the confidence interval of $H^{TR}$, denoted by $H^{TR}.CI$, can be calculated by $[(\text{Floor}(\bar{t}_s/\tau) + 1) \pm (\text{Floor}(2\sigma/\tau) + 1)]$. Let $\Omega_{trh}$ denote the threshold to determine whether group $G^{TR}$ has the *happen-frequently property.* That is, a group $G^{TR}$ is said to have the *happen-frequently property* if the number of $b_j^c \in G^{TR}$ that fall in confidence interval of $H^{TR}$ should be not smaller than threshold $\Omega_{trh}$. Recall that each group $G_i^{TR}$ associated with a behavior $b_i^{TR}$ which already has the *time-regular property.* Let $H_i^{TR}$ be the histogram of $G_i^{TR}$. The following criterion further examine the *happen-frequently property* of behavior $b_i^{TR}$.

*Criterion for Having Happen-Frequently Property:* Behavior $b_i^{TR}$ is said to have the *frequent-happen property* if it satisfies the following criterion:

$$\sum_{x \in H_i^{TR}.CI} h_x \geq \Omega_{trh}. \tag{10}$$

Each behavior $b_i^{TR}$ will be checked if the above criterion is satisfied. If it is the case, the behavior $b_i^{TR}$ has both *time-regular* and *happen-frequently properties* and hence the behavior $b_i^{TR}$ is called a *regular* behavior $b_i^{TR}(\text{name}, G_i^{TR})$, where name $\in B$ is the behavior name and $G_i^{TR}$ is the corresponding behavior occurrence group. The value of threshold $\Omega_{trh}$ can impact the identification of a regular behavior and will be discussed in Section V.

After the execution of this task, we have obtained a set of $n$ *regular* behaviors, denoted by $B^R = \{b_1^R, b_2^R, \ldots, b_n^R\}$, where $b_i^R(\text{name}, G^{TR})$ denote some regular behavior $b_j^{TR}$. Each regular behavior $b_i^R \in B^R$ can be treated as a feature of *regular day*. Given $n$ regular behaviors, we can construct an $n$-dimensional feature space. For the ease of presentation, the multidimensional space based on the regular behavior set $B^R$ is simply called $R$-space and each regular behavior $b_i^R$ is called a *feature dimension* of $R$-space. The value of each dimension ranges from $-24$ to 24 where [0, 24] represents the occurrence time of each behavior and $[-24, 0)$ represents the absence of the behavior in that dimension.

Fig. 1 gives an example of the $R$-space construction. Assume that two regular behaviors, including "HaveBreakfast" and "Sleep," have been identified in the SRB phase. As shown in Fig. 1, the constructed $R$-space consists of two dimensions which are marked with "HaveBreakfast" and "Sleep." Assume that the confidence intervals of $H^{TR}$ corresponding to the regular behaviors "HaveBreakfast" and "Sleep" are [7, 9] and [21, 23], respectively. As shown in Fig. 1, the center of the confidence interval of regular behavior "HaveBreakfast" is 8 o'clock, which indicates that the regular behavior "HaveBreakfast" usually occurs around 8 o'clock. Similarly, the center of the confidence interval of regular behavior "Sleep" is 22 o'clock. The next task will map the daily behavior sequences onto the $R$-space.

Fig. 2 depicts the procedure of the SRB phase. As shown in Fig. 2, the behavior data collected in the past $m$ days are treated as the inputs of this phase. In step 1, the input data are partitioned into several subsets based on the behavior name. Each subset $\mathbb{B}_i^c$ is further partitioned by using DBSCAN based on the start time of behavior occurrence. The output of step 1 is the set of all clusters $\mathbb{G}^{TR}$ which acts as the input of step 2. In step 2, the histogram of each cluster $G_i^{TR} \in \mathbb{G}^{TR}$ is constructed. Finally, the regular behaviors can be obtained by checking the criterion for having the *happen-frequently property*, as shown in Fig. 2.

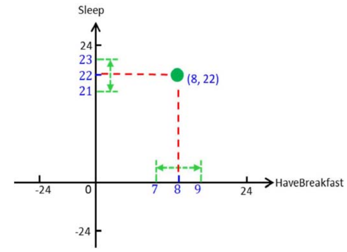



Fig. 1. Example of the $R$-space construction.

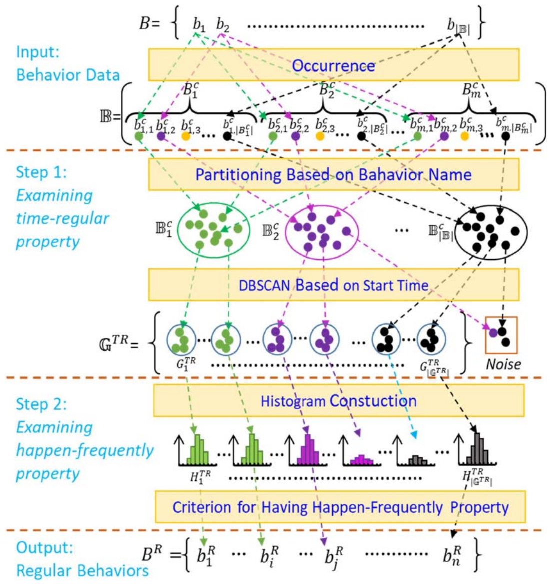



Fig. 2. Procedure of the SRB phase.

### B. Mapping of Daily Behavior Sequence Phase

Given a set $\mathbb{B} = \{B_1^c, B_2^c, \ldots, B_m^c\}$ of daily behavior sequences of the past $m$ days, this phase aims to map the set of daily behaviors of the $i$th day, say $B_i^c \in \mathbb{B}$, to a point in the $R$-space aiming to distinguish that the $i$th day is a regular or irregular day. The mapping operation is detailed in the following.

Consider a daily behavior sequence $B_i^c = (b_1^c, b_2^c, \ldots, b_{|B_i^c|}^c)$. Let $V_i = (v_1, v_2, \ldots, v_n)$ denote a point with the $n$-dimensional coordinates in the $R$-space, where each $v_j \in [-24,\ 24]$. Let $f^R$ be a mapping function which maps a behavior sequence $B_i^c$ to a point in the $R$-space. That is

$$f^R(B_i^c) = V_i.$$

Each coordinate $v_j \in V_i$ is the mapping result of $B_i^c$ on the feature dimension $b_j^R \in B^R$. Recall that each behavior occurrence $b_j^c \in B_i^c$ is represented by the form $b_j^c$(name, start_time, end_time), where name, start_time, and end_time denote the behavior name, occurrence starting time, and occurrence end time, respectively. The value of $v_j \in V_i$ can be a mapping of occurrence starting time or end time of $b_k^c \in B_i^c$. Here, we will take the occurrence starting time to explain the mapping operation of $f^R$.

The mapping operation of $f^R$ consists of two steps. The first step aims to map each $b_j^c \in B_i^c$ on a feature dimension in the $R$-space. Then, the value of each coordinate $v_j \in V_i$ is calculated in the second step.

*1) Mapping of Behavior Occurrence:* In this step, each behavior occurrence $b_j^c \in B_i^c$ will be checked whether or not it can be mapped to a feature dimension $b_k^R \in B^R$. If the behavior name of $b_j^c$ is identical to the behavior name of $b_k^R$, it is allowed to be mapped to the feature dimension $b_j^R$. Let $\rho_{j,k}$ denote a Boolean variable representing whether or not the behavior occurrence $b_j^c \in B_i^c$ can be mapped to the feature dimension $b_k^R$. That is

$$\eta_{j,k} = \begin{cases} 1, & b_j^c.\text{name} = b_k^R.\text{name} \\ 0, & \text{otherwise.} \end{cases}$$

It is possible that one $b_j^c$ is mapped to multiple feature dimensions. That is, the condition $\sum_{k=1}^{n} \eta_{j,k} > 1$ holds. Therefore, the best feature dimension should be selected for $b_j^c$. To achieve this, the distance between behavior occurrence $b_j^c$ and regular behavior $b_k^R$ will be defined by the following equation:

$$d\left(b_j^c, b_k^R\right) = \boldsymbol{Abs}\left(b_j^c.\text{start\_time} - b_k^R.G^{TR}.\bar{t}_s\right) \tag{11}$$

where $\boldsymbol{Abs}(x)$ denotes a function which returns the absolute value of input $x$.

The regular behavior $b^{R,\text{best}}$ which has the minimal distance with $b_j^c$ will be selected to be the best feature dimension for $b_j^c$. That is

$$b^{R,\text{best}} = \arg\min_{\eta_{j,k}=1} d\left(b_j^c, b_k^R\right). \tag{12}$$

Till now, each behavior occurrence $b_j^c$ has been checked whether or not it can be mapped to a feature dimension. If it is the case, the value of each coordinate $v_k \in V_i$ should be further calculated which will be discussed in what follows.

*2) Calculation of Mapping Value:* Let $B_k^R$ denote the set of behavior occurrences in $B_i^c$ which have been already mapped to the feature dimension $b_k^R$. Herein, it is noticed that two or more behavior occurrences might be mapped to a feature dimension. For instance, two occurrences of behavior "Sleep" are identified because the elder took a nap around 2 P.M. and slept around 10 P.M. in a day. Meanwhile, only one regular behavior named "Sleep" is selected in the SRB phase. According to (11), two occurrences of behavior "Sleep" will be mapped on to the same feature dimension. In this situation, the occurrence which has the nearest distance to the dimension $b_k^R$, denoted by $b_{\text{nearest}}^c$, will be selected from $B_k^R$. That is

$$b_{\text{nearest}}^c = \arg\min_{b_j^c \in B_k^R} d\left(b_j^c, b_k^R\right).$$

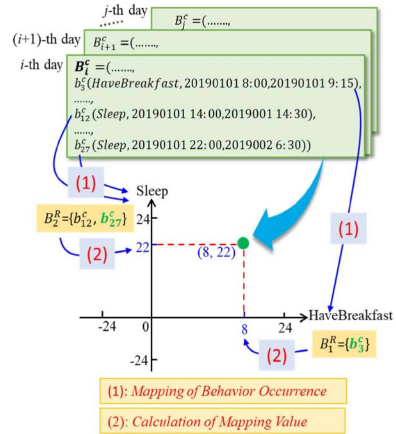



Fig. 3. Example of MDBS.

Then, the coordinate $v_k$ of $B_i^c$ on the feature dimension $b_k^R$ will be decided by the occurrence time of $b_{\text{nearest}}^c \in B_k^R$. Besides, if $B_k^R$ is empty, the value of coordinate $v_k$ will be set at $-24$. That is

$$v_k = \begin{cases} b_{\text{nearest}}^c.\text{start\_time}, & B_k^R \neq \emptyset \\ -24, & \text{otherwise.} \end{cases} \tag{13}$$

Fig. 3 gives an example to explain the mapping from a daily behavior sequence onto the constructed $R$-space. Assume that a 2-D $R$-space has been constructed as shown in Fig. 1. Assume that the behavior occurrences of the daily behavior sequence $B_i^c$ in the $i$th day are given in Fig. 3, where $B_i^c$ contains two "Sleep" occurrences $b_{12}^c$ and $b_{27}^c$. $b_{12}^c$ and $b_{27}^c$ are mapped to the "Sleep" dimension according to (12) and the set $B_2^R$ is constructed after executing step *Mapping of Behavior Occurrence*, denoted by (1) in Fig. 3. In step *Calculation of Mapping Value*, denoted by (2), the behavior occurrence $b_{27}^c \in B_2^R$ is selected as $b_{\text{nearest}}^c$ and $v_2$=22 is calculated according to (13). Similarly, the "HaveBreakfast" occurrence $b_3^c$ is mapped to the "HaveBreakfast" dimension and $v_1$=8 can be calculated. Finally, a point $V_i = (v_1, v_2)$ =(8,22), as marked by a green vertex, can be obtained in the $R$-space, which is the mapping result of $B_i^c$.

Finally, the mapping operation of $f^R$ has already been done. A set of $m$ points $\mathbb{V} = \{V_1, V_2, \ldots, V_m\}$ on the $R$-space will be obtained. In the next task, the irregular days will be identified based on the set $\mathbb{V}$ in the $R$-space.

### C. Identification of Irregular Days Phase

This phase aims to identify the irregular days by examining the mapping result $\mathbb{V}$ of daily behavior sequence set $\mathbb{B}$ onto the $R$-space. A day is said to be *regular* if each regular behavior $b_k^R \in B^R$ happens in this day and the starting time of the occurrence satisfies the *time-regular property* of $b_k^R$. It indicates that a regular day has two properties. The first one is the *behavior-regular property* which means that each regular behavior $b_k^R \in B^R$ happens on this day. The second one is the *occurrence-regular property* which means that the occurrence of each regular behavior $b_k^R$ satisfies the *time-regular property* of $b_k^R$. Different from the regular days, an irregular day might not satisfy the *behavior-regular* and *occurrence-regular properties*. It indicates that the mapping coordinates of an irregular sequence on the feature dimension $b_k^R$ is under zero or has a big difference with that of regular sequence. Based on the above conception, the regularity of the $i$th day can be identified by examining the coordinates of the point $V_i$ obtained by mapping $B_i^c$ to the $R$-space. Two factors will be considered when identifying the irregularity. The first one is the irregularity probability on each feature dimension and the other is the weight of each feature dimension. The following presents the details of the two factors.

*1) Irregularity Probability on Each Feature Dimension:* Consider a feature dimension corresponding to a regular behavior $b_k^R(\text{name}, G^{TR})$. Let $[t_k^{\min}, t_k^{\max}]$ denote the time-regular interval of $b_k^R$ dimension. Let $\boldsymbol{TC}(Datetime)$ denote the *Time Calculation* function which returns the following value:

$$Datetime.\text{hour} * 3600 + Dateime.\text{minute} * 60 + Datetime.\text{second},$$

where *Datetime* is a variable of timestamp type.

Then, the values of $t_k^{\min}$ and $t_k^{\max}$ can be calculated as follows:

$$t_k^{\min} = \min_{b_j^c \in b_k^R.G^{TR}} \boldsymbol{TC}\left(b_j^c.\text{start\_time}\right)$$

$$t_k^{\max} = \max_{b_j^c \in b_k^R.G^{TR}} \boldsymbol{TC}\left(b_j^c.\text{start\_time}\right).$$

Let $V_i = (v_{i,1}, v_{i,2}, \ldots, v_{i,n})$ denote the point obtained by mapping $B_i^c$ to the $R$-space. Let $p_{i,k}$ denote the irregularity probability of $B_i^c$ on feature dimension $b_k^R$. Let $P_i$ denote the set of irregularity probabilities of $B_i^c$ on all the feature dimensions in $R$-space. Obviously, we have $P_i = (p_{i,1}, p_{i,2}, \ldots, p_{i,n})$. The value of $p_{i,k}$ is determined by the value of $v_{i,k}$. If the coordinate $v_{i,k} \in V_i$ falls in the interval $[t_k^{\min}, t_k^{\max}]$, the behavior is treated as regular and hence, the value of $p_{i,k}$ is set at 0. If the value of $v_{i,k}$ is smaller than zero, it indicates that the occurrence of the important behavior is missing. Hence, the value of $p_{i,k}$ is set at 1. Otherwise, the value of $p_{i,k}$ is computed based on the minimal distance, denoted by $d(v_{i,k}, b_k^R)$, between $v_{i,k}$ and the regular interval $[t_k^{\min}, t_k^{\max}]$ of $b_k^R$ dimension. Since the coordinate $v_{i,k}$ might fall in the interval $[0, t_k^{\min}]$ or interval $[t_k^{\max}, 24]$, the value of $d(v_{i,k}, b_k^R)$ can be calculated by

$$d\left(v_{i,k},\ b_k^R\right) = \min\left(Abs\left(v_{i,k} - t_k^{\min}\right), Abs\left(v_{i,k} - t_k^{\max}\right)\right). \quad (14)$$

Let $L_k^{IR}$ denote the max length of intervals $[0, t_k^{\min}]$ and $[t_k^{\max}, 24]$. That is

$$L_k^{IR} = max\left(t_k^{\min}, 24 - t_k^{\max}\right). \quad (15)$$

Then, the value of $p_{i,k}$ can be calculated according to the following equation:

$$p_{i,k} = \begin{cases} 0, & v_{i,k} \in \left[t_k^{\min},\ t_k^{\max}\right] \\ 1, & v_{i,k} < 0 \\ \frac{d(v_{i,k},\ b_k^R)}{L_k^{IR}}, & \text{otherwise.} \end{cases} \quad (16)$$

Fig. 4 gives an example of irregularity probability calculation. Assume that three regular behaviors, including $b_1^R$, $b_2^R$, and $b_3^R$, are obtained in the SRB phase. Consequently, a 3-D feature space is constructed as shown in Fig. 4. The time-regular interval of each dimension is given and marked in Fig. 4. According to (16), the irregularity probability on each dimension for the points which are located in the green rectangle has the same value zero. For instance, the point $V_1$, marked as a green point in Fig. 4, has its irregularity probability set $P_1 = (0, 0, 0)$. The irregularity probability set of the points which are located outside the rectangle, has at least one nonzero irregularity probability element, such as $P_2$ of $V_2$ marked as a purple point and $P_3$ of $V_3$ marked as a red point in Fig. 4.

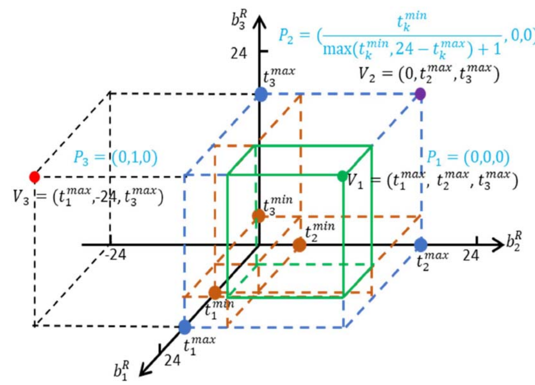



Fig. 4. Example of $R$-space and irregularity probability calculation.

*2) Weight of Each Feature Dimension:* The weight of each feature dimension can be represented by the regularity of the corresponding regular behavior. Recall that each regular behavior $b_k^R$ has both the *time-regular* and *happen-frequently properties*. Two properties of each regular behavior $b_k^R$ can be quantified according to the characteristic of the histogram $H^{TR}$ of $b_k^R.G^{TR}$. Recall that $H^{TR}.CI$ denotes the confidential interval of $H^{TR}$. Actually, the size of $H^{TR}.CI$ indicates the degree of the *time-regular property* of $b_k^R$. If the size of $H^{TR}.CI$ is small, $b_k^R$ has a strong *time-regular property*. On the contrary, if the size of $H^{TR}.CI$ is large, $b_k^R$ has a weak *time-regular property*. Recall that the *happen-frequently property* can be examined by (10) which means that the number of $b_k^R$ occurrences occurred in $H^{TR}.CI$, denoted by $\sum_{x\in\ H^{TR}.CI} h_x$, should not be smaller than a threshold. Similarly, the value of $\sum_{x\in\ H^{TR}.CI} h_x$ indicates the degree of the *happen-frequently property* of $b_k^R$. If the value of $\sum_{x\in\ H^{TR}.CI} h_x$ is large, the *happen-frequently property* of $b_k^R$ is strong. On the contrary, if the value of $\sum_{x\in\ H^{TR}.CI} h_x$ is small, the *happen-frequently property* of $b_k^R$ is weak. Let $r_k$ denote the quantified value of two properties of regular behavior $b_k^R$. The value of $r_k$ can be calculated by

$$r_k = \frac{\sum_{x\in H^{TR}.CI} h_x}{H^{TR}.CI}. \quad (17)$$

Then, the weight of each feature dimension $b_k^R$, denoted by $w_k$, can be calculated by the following equation:

$$w_k = \frac{r_k}{\sum_{k=1}^{n} r_k}. \quad (18)$$

Obviously, we have $\sum_{k=1}^{n} w_k = 1$.

Furthermore, let $\mathbb{P}_i$ denote irregularity probability of $B_i^c$ in $R$-space. Considering the irregularity probability on each

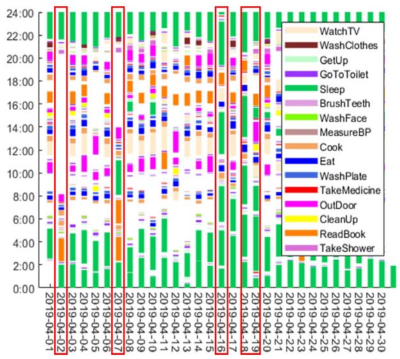



Fig. 5. Information of actual behaviors of a month.

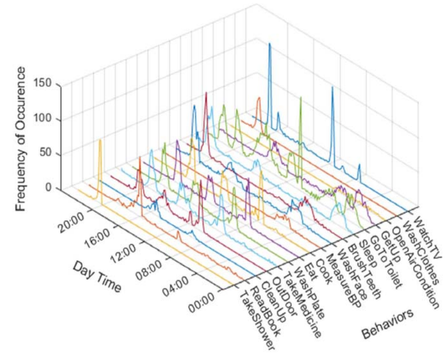



Fig. 6. Occurrence frequency of sensor events.

feature dimension and weight of each feature dimension, the value of $\mathbb{P}_i$ can be calculated by the following equation:

$$\mathbb{P}_i = \sum_{k=1}^{n} w_k * p_k. \tag{19}$$

Herein, the irregular days can be identified according to the irregularity probability of each day. Let $\Delta_{trh}$ denote a probability threshold. If the value of $\mathbb{P}_i$ is larger than $\Delta_{trh}$, then the $i$th day will be identified as an irregular day. The value of $\Delta_{trh}$ will be discussed in Section V.

## V. SIMULATION

This section discusses the parameters of the proposed FIID mechanism and investigates the performance improvement of the proposed mechanism against *IIRD*, DBSCAN, and $k$-means in terms of recall, precision, and F-measure. The following presents the data description and performance results.

### A. Data Description

This section describes the simulation data [9] which were collected based on the behavior patterns of interviewed elders. Then, the daily behaviors of 12 months are selected from one behavior pattern. To observe the identification capabilities of different approaches on irregular behaviors, the irregular days are labeled by a human. Fig. 5 depicts the derived behaviors of one month. As shown in Fig. 5, some behaviors of irregular days, enclosed by the red rectangle, occurred irregularly, as compared with those of regular days. For example, behavior "ReadBook" occurs around 2 o'clock on April 2 while it occurs around 16 o'clock on most days; behavior "Eat" only occurs once on April 2 while it happens three times on most days.

Fig. 6 shows the occurrence frequency of derived behaviors for 12 months. The occurrences of each behavior form a curve which likely indicates the habit of this behavior during 12 months. For instance, behavior "TakeShower" usually occurs around 20 o'clock and behavior "ReadBook" mostly happens around 16 o'clock.

### B. Experimental Results

Recall that the proposed FIID mainly consists of three phases. The SRB and IID phases remain some parameters which should be carefully adjusted will be discussed in the experiments.

Recall that the SRB phase aims to construct an $R$-space which is a multidimension for identifying the irregular day. The key idea is to employ DBSCAN to cluster behavior occurrences into many groups each of which satisfies the time-regular property. The two parameters *Min_pts* and *Eps* of DBSCAN are important which can impact the number of features constructing the $R$-space. A large number of features can result in a rigid $R$-space which might mistake a regular day for an irregular day while a small number of features can result in a weak $R$-space which might identify an irregular day as a regular day. The good values of parameters *Min_pts* and *Eps* can impact the number of groups and the number of points in each group. The result containing a small number of groups will yield a small number of features and each feature has low regularity, which leads to a weak $R$-space. Therefore, the best values for setting parameters *Min_pts* and *Eps* of DBSCAN should result in a suitable number of groups and high regularity of each group. To achieve this, the harmonic mean [22] of the average regularity and the number of obtained groups is applied. Recall that the result of executing DBSCAN is represented by

$$\mathbb{G}^{TR} = \left\{G_1^{TR}, G_2^{TR}, \ldots, G_{|\mathbb{G}^{TR}|}^{TR}\right\}.$$

Let $r_i$ denote the regularity of group result $G_i^{TR}$. Let $n = |\mathbb{G}^{TR}|$ and $\bar{r}$ denote the number and the average regularity of obtained groups, respectively. That is

$$\bar{r} = \frac{\sum_{i=1}^{n} r_i}{n}.$$

Then, the value of the harmonic mean, denoted by $C_H$, can be calculated by

$$C_H = \frac{2 * \bar{r} * n}{\bar{r} + n}. \tag{20}$$

Fig. 7(a)–(c) presents the impact of parameters *Min_pts* and *Eps* of DBDACN on the clustering performance in terms of the average regularity, the number, and the harmonic mean of obtained groups, respectively. In general, when the parameter *Min_pts* increases and the parameter *Eps* decreases, the

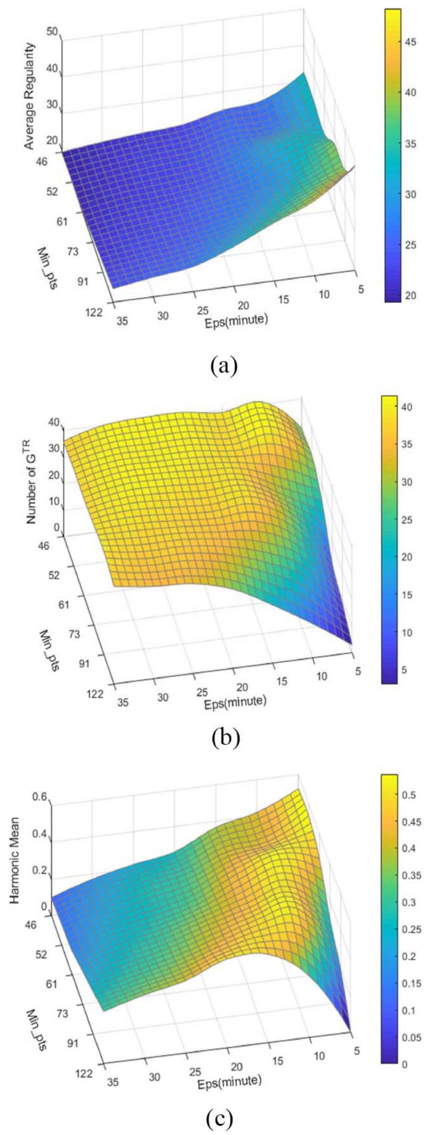



Fig. 7. Performance of DBSCAN by varying *Eps* and *min_pts*. (a) Average regularity. (b) Number of groups $G^{TR}$. (c) Harmonic mean.

average regularity increases while the number of groups and the harmonic mean of the obtained groups increase first and then decrease. This occurs because that the large values of *Min_pts* and small values of *Eps* tend to be rigid while small values of *Min_pts* and large values of *Eps* tend to be loose. Consequently, the parameters *Min_pts* and *Eps* should be set at the appropriate values for obtaining a high value of $C_H$. As shown in Fig. 7(c), the values of *Min_pts* and *Eps* correspond to the value of $C_H$ larger than 0.4 can be adopted. In the following experiments, *Min_pts* and *Eps* will be set at 61 and 15, respectively, because they achieve a high value $C_H = 0.43$.

Fig. 8 investigates the impacts of the threshold $\Omega_{trh}$ on the average regularity, the number, and the harmonic mean of regular behaviors. Recall that threshold $\Omega_{trh}$ is used to check the criterion for having the *happen-frequently property* in the SRB phase. The threshold $\Omega_{trh}$ is set ranging from $(1/12)$ to $(12/12)$ of the number of days. As shown in Fig. 8, the number of regular behaviors, marked in blue, decreases with the threshold $\Omega_{trh}$. It is obvious that some behaviors do not occur every

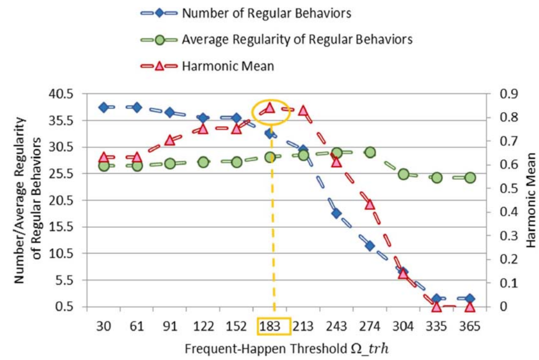



Fig. 8. Impacts of the threshold $\Omega_{trh}$.

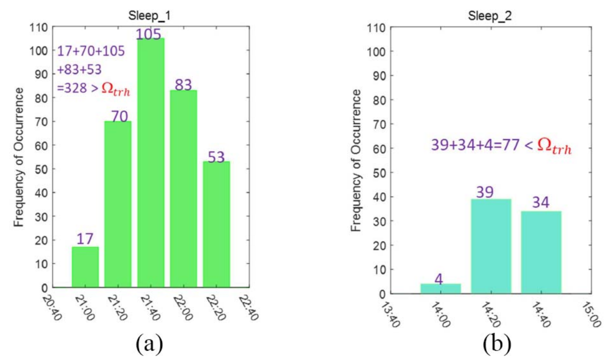



Fig. 9. Snapshot of checking the criterion for having the happen-frequently property. (a) "Sleep" behavior at night. (b) Snap behavior.

day, such as "TakeShower" and "WashClothes," especially in winter. When the frequency threshold $\Omega_{trh}$ increases, these behaviors cannot satisfy the threshold. However, the average regularity of regular behaviors, marked in green, first increases with the threshold $\Omega_{trh}$ and then decreases from the point $\Omega_{trh} = 274$. The increment of regularity occurs because some behaviors with low regularity have been excluded. However, when $\Omega_{trh}$ is larger than 274, the decrement of regularity occurs because some behaviors which occur regularly but not every day are excluded. The curve of harmonic mean, marked in red, reflects the integration of blue and green lines. As shown in Fig. 8, when the threshold $\Omega_{trh}$ is set at 183 which is equal to $(6/12)$ of the number of days, the value of harmonic mean is largest. In the following experiments, the threshold $\Omega_{trh}$ is set to $(6/12)$ of the number of days.

Fig. 9 further examines the execution of *happen-frequently property* applied to the behavior "Sleep." In the SRB phase, there are two groups of behavior "Sleep" obtained by executing the DBSCAN clustering algorithm. Fig. 9(a) and (b) shows their histograms. As shown in Fig. 9(a), the group called "Sleep_1" actually denoting the sleep behavior at night for an elder, satisfies the threshold value of $\Omega_{trh}$ and thus can be considered as one feature of the regular behavior. On the contrary, as shown in Fig. 9(b), the group $G^{TR}$, named "Sleep_2," which actually denotes the snap behavior for an elder, does not satisfy the threshold value of $\Omega_{trh}$. Therefore, this behavior does not to be considered as the feature of the regular behavior.

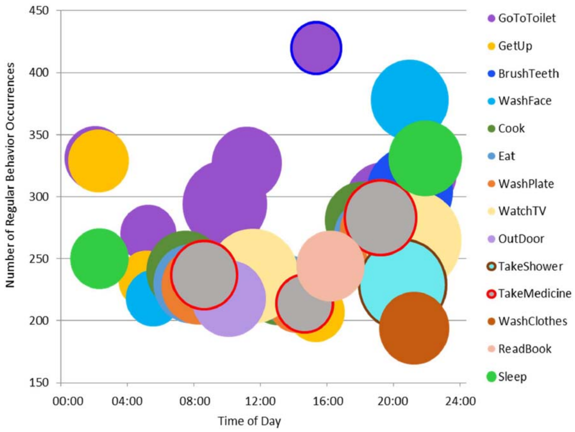



Fig. 10. Snapshot of regular behaviors.

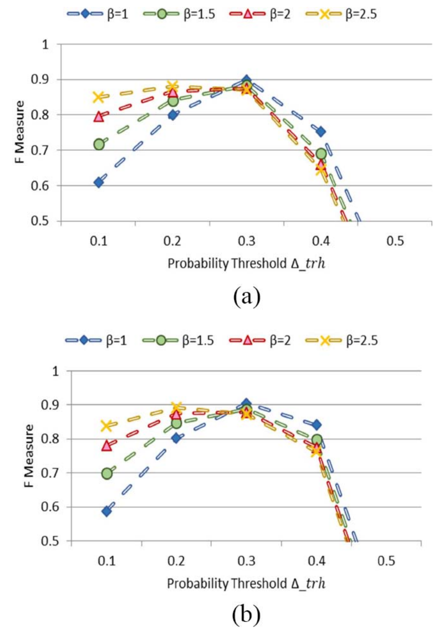



Fig. 11. Performance of the proposed FIID by varying the probability threshold $\Delta_{trh}$. (a) Proposed FIID without the weight of feature dimension. (b) Proposed FIID without the weight of feature dimension.

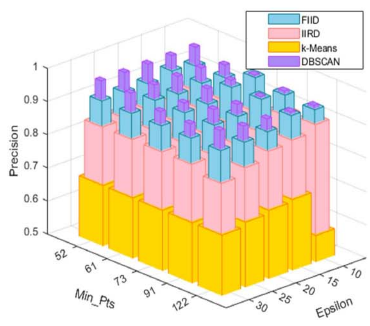
To closely observe the result of the SRB phase, Fig. 10 further presents the number of occurrences of each group using the bubble graph. The color of the bubble indicates the type of behavior. For instance, the three gray bubbles with red outline indicate the type of "TakeMedicine," which indicates that the elder usually takes medicine three times a day. The size and height of the bubble denote the regularity of the behavior and the number of occurrences belonging to the group $G^{TR}$ corresponding to the behavior, respectively. Herein, it is noticed that the behavior with a large number of occurrences does not imply that it has high regularity. This is because that the occurrence time of that behavior might be distributed over a wide time range. As shown in Fig. 10, the regularity of behavior "TakeShower", marked as the cyan bubble in brown outline, is large even though the number of the behavior occurrences in the corresponding group is relatively small. On the contrary, the regularity of behavior "GoToToilet," marked as the purple bubble with the blue outline, has small regularity but a large number of occurrences in the corresponding group. This further verifies the trend of the average regularity of regular behaviors in Fig. 10.

Fig. 11 observes the F-measure of the proposed FIID with different $\beta$ by varying the probability threshold $\Delta_{trh}$ which is used to determine irregularity in the IID phase. Recall that the calculation of irregularity probability involves the weight of feature dimension according to (19). Fig. 9(a) and (b) presents the performances of FIID without and with considering the weight of feature dimension, respectively. In general, the F-measure increases first and then decreases with the increase of probability threshold $\Delta_{trh}$ in both situations. It occurs because that the small value of $\Delta_{trh}$ leads to the increase of FP and the decrease of precision while the large value of $\Delta_{trh}$ leads to the increase of FN and the decrease of recall. As shown in Fig. (9), the F-measure reaches the maximum value when $\Delta_{trh}$ is equal to 0.3.

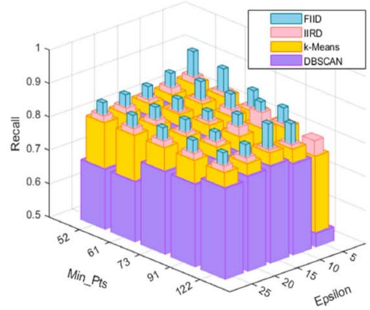
Untill now, the appropriate values of parameters in the proposed FIID have been determined from the previous experiments. The following further compares the performances of the proposed FIID against *IIRD*, DBSCAN, and *k*-means.

Fig. 12 compares the four mechanisms in terms of precision, recall, and F-measure by varying the values of parameters *Min_pts* and *Eps*. The related parameter $\beta$ is set at 2, which indicates the recall is paid more attention than the precision. In general, the appropriate values of parameters *Min_pts* and *Eps* can improve the precision and recall of irregularity detection for FIID, as shown in Fig. 10(a) and (b), respectively. This occurs because the appropriate regularity features can be extracted with the appropriate values of parameters *Min_pts* and *Eps*, which significantly reduce the numbers of FP and FN. In comparison, the proposed FIID outperforms *IIRD*, DBSCAN, and *k*-means in terms of recall and F-measure in most cases. This occurs because the proposed FIID extracts regularity features according to the regularity of behaviors and removes the impact of behaviors with low regularity. Besides, *IIRD* takes all behaviors of one day into account by adopting string alignment, which leads to a large number of FN. The DBSACN has the worst recall since the irregular data are clustered into the regular data group while *k*-means has the worst precision because the differences between irregular days might be large, leading to a large number of FP.

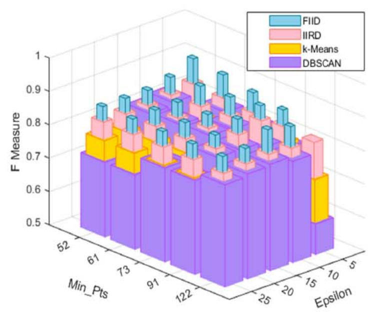
Fig. 13 further compares the four algorithms in terms of precision, recall, and F-measure by varying the data size. In general, the performances of the proposed FIID and *k*-means increase with the size of data size while the performance of the *IIRD* increases first and then decreases with the data size. This occurs because the regular behaviors extracted by applying the FIID mechanism and *k*-means from large size data are likely to be more reliable than that from small size data. Besides, the model constructed by the *IIRD* mechanism using a small set of training data inclines to be sensitive while the overfitting problem tends to appear when the number of training data is too large. DBSCAN tends to include the irregular days into the regular group with the large size data. In comparison, the proposed FIID outperforms the other compared algorithms in




(a)

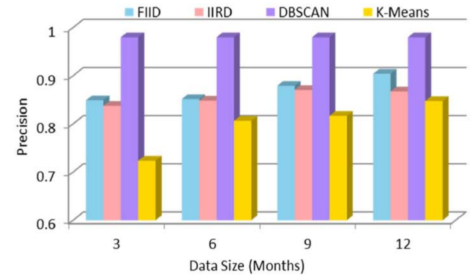



(b)

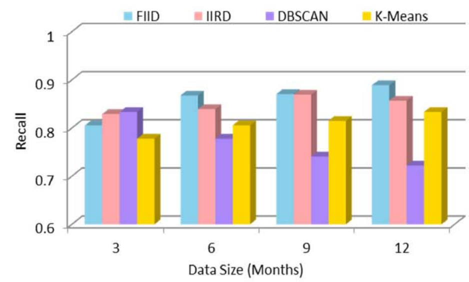



(c)

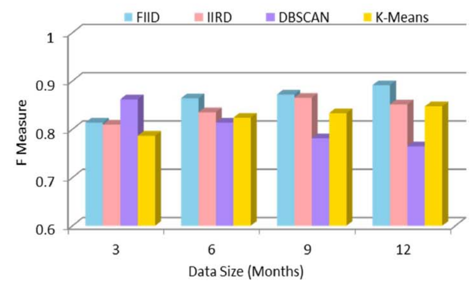
Fig. 12. Performance of four compared algorithms by varying *Min_pts* and *Eps*. (a) Precision. (b) Recall. (c) F-measure.




(a)




(b)




(c)

Fig. 13. Performance of four compared algorithms by varying data size. (a) Precision. (b) Recall. (c) F-measure.

terms of F-measure. This occurs because that the proposed FIID algorithm extracts the regular behaviors and further takes the regularity of each regular behavior into account, leading to the effective decrease of FP.

## VI. Conclusion

This article proposed a feature-based implicit irregularity identification mechanism, called FIID, which can be used to support the homecare of the elderly living alone. The proposed FIID consists of three phases, namely *SRB, MDBS,* and *IID*. The SRB phase aims to select some regular behaviors with *time-regular* and *happen-frequently properties* by employing the unsupervised learning algorithm. The obtained regular behaviors are treated as kernel features which are used to construct a multidimensional feature space. Then, the MDBS phase maps each daily behavior sequence to a point in the constructed feature space. Finally, the IID phase detects the implicit irregularity by considering the irregularity probability of each feature dimension for each day. Extensive experiments show that the proposed FIID outperforms the compared algorithms in terms of precision, recall, and F-measure. Our future work will further analyze and extract refined regular behavior patterns to provide better homecare for the elderly.

## References

[1] Y. Zhang, L. Sun, H. Song, and X. Cao, "Ubiquitous WSN for healthcare: Recent advances and future prospects," *IEEE Internet Things J.*, vol. 1, no. 4, pp. 311–318, Aug. 2014.
[2] H. Habibzadeh, K. Dinesh, O. R. Shishvan, A. Boggio-Dandry, G. Sharma, and T. Soyata, "A survey of healthcare Internet of Things (HIoT): A clinical perspective," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 53–71, Jan. 2020.
[3] A. M. Nia, M. Mozaffari-Kermani, S. Sur-Kolay, A. Raghunathan, and N. K. Jha, "Energy-efficient long-term continuous personal health monitoring," *IEEE Trans. Multi-Scale Comput. Syst.*, vol. 1, no. 2, pp. 85–98, Apr. 2015.
[4] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, "Sensor-based activity recognition," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 6, pp. 790–808, Nov. 2012.
[5] M. Hasan and A. K. Roy-Chowdhury, "A continuous learning framework for activity recognition using deep hybrid feature models," *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 1909–1922, Nov. 2015.

[6] L. Yao *et al.*, "Compressive representation for device-free activity recognition with passive RFID signal strength," *IEEE Trans. Mobile Comput.*, vol. 17, no. 2, pp. 293–306, Feb. 2018.

[7] Y. Wang, K. Wu, and L. M. Ni, "WiFall: Device-free fall detection by wireless networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 581–594, Feb. 2017.

[8] S. Aminikhanghahi, T. Wang, and D. J. Cook, "Real-time change point detection with application to smart home time series data," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 5, pp. 1010–1023, May 2018.

[9] C. J. Shang, C.-Y. Chang, G. Chen, S. Zhao, and J. Lin, "Implicit irregularity detection using unsupervised learning on daily behaviors," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 1, pp. 131–143, Jan. 2020.

[10] F. Zhou, J. R. Jiao, S. Chen, and D. Zhang, "A case-driven ambient intelligence system for elderly in-home assistance applications," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.* , vol. 41, no. 2, pp. 179–189, Mar. 2011.

[11] G. Valenza *et al.*, "Wearable monitoring for mood recognition in bipolar disorder based on history-dependent long-term heart rate variability analysis," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 5, pp. 1625–1635, Sep. 2014.

[12] L.Gutiérrez-Madroñal, L. L. Blunda, M. F. Wagner, and I. Medina-Bulo, "Test event generation for a fall-detection IoT system," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6642–6651, Aug. 2019.

[13] D. Elbert, H. Storf, M. Eisenbarth, Ö. Ünalan, and M. Schmitt, "An approach for detecting deviations in daily routine for long-term behavior analysis," in *Proc. 5th Int. Conf. Pervasive Comput. Technol. Healthcare*, 2011, pp. 426–433.

[14] H.-H. Hsu and C.-C. Chen, "Rfid-based human behavior modeling and anomaly detection for elderly care," *Mobile Inf. Syst.*, vol. 6, no. 4, pp. 341–354, Dec. 2010.

[15] M. Novák, F. Jakab, and L. Lain, "Anomaly detection in user daily patterns in smart-home environment," *J. Sel. Areas Health Inform.*, vol. 3, no. 6, pp. 1–11, Jun. 2013.

[16] N. Dey, A. S. Ashour, F. Shi, S. J. Fong, and R. S. Sherratt, "Developing residential wireless sensor networks for ECG healthcare monitoring," *IEEE Trans. Consum. Electron.*, vol. 63, no. 4, pp. 442–449, Nov. 2017.

[17] L. Fanucci *et al.*, "Sensing devices and sensor signal processing for remote monitoring of vital signs in CHF patients," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 3, pp. 553–569, Mar. 2013.

[18] M. Yu, A. Rhuma, S. Naqvi, L. Wang, and J. Chambers, "A posture recognition-based fall detection system for monitoring an elderly person in a smart home environment," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 6, pp. 1274–1286, Nov. 2012.

[19] J. Wang, Z. Zhang, B. Li, S. Lee, and R. S. Sherratt, "An enhanced fall detection system for elderly person monitoring using consumer home networks," *IEEE Trans. Consum. Electron.*, vol. 60, no. 1, pp. 23–29, Feb. 2014.

[20] Q. Guan, C. Li, X. Guo, and B. Shen, "Infrared signal based Elderly fall detection for in-home monitoring," in *Proc. 9th Int. Conf. Intell. Human–Mach. Syst. Cybern.*, 2017, pp. 373–376.

[21] A. R. M. Forkan, I. Khalil, Z. Tari, S. Foufou, and A. Bouras, "A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient assisted living," *Pattern Recognit.*, vol. 48, no. 3, pp. 628–641, Mar. 2015.

[22] M. O. Hasna and M.-S. Alouini, "Harmonic mean and end-to-end performance of transmission systems with relays," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 130–135, Jan. 2004.


**Cuijuan Shang** received the B.S. degree in biomedical engineering from Zhengzhou University, Zhengzhou, China, in 2008, and the M.S. degree in biomedical engineering from Shandong University, Jinan, China, in 2011. She is currently pursuing the Ph.D. degree with the Department of Computer Science and Information Engineering, Tamkang University, New Taipei, Taiwan.

She is currently a Lecturer with the School of Computer and Information Engineering, Chuzhou University, Chuzhou, China. Her current research interests include wireless sensor network and healthcare.

**Chih-Yung Chang** (Member, IEEE) received the Ph.D. degree in computer science and information engineering from the National Central University, Taoyuan, Taiwan, in 1995.

He is currently a Full Professor with the Department of Computer Science and Information Engineering, Tamkang University, New Taipei, Taiwan. His current research interests include Internet of Things, wireless sensor networks, artificial intelligence, and deep learning.

Prof. Chang has served as an Associate Guest Editor for several SCI-indexed journals, including the *International Journal of Ad Hoc and Ubiquitous Computing* from 2011 to 2019, the *International Journal of Distributed Sensor Networks* from 2012 to 2014, *IET Communications* in 2011, *Telecommunication Systems* in 2010, the *Journal of Information Science and Engineering* in 2008, and the *Journal of Internet Technology* from 2004 to 2008.

**Jinjun Liu** received the Ph.D. degree in computer application technology from the School of Computer and Information, Hefei University of Technology, Hefei, China, in 2019.

He is currently a Professor with the School of Computer and Information Engineering, Chuzhou University, Chuzhou, China. His current research interests include gerontechnology, Internet of Things, and eHealth.

**Shenghui Zhao** received the M.S. degree from Hefei University of Technology, Hefei, China, in 2003, and the Ph.D. degree from Southeast University, Nanjing, China, in 2013.

She is currently a Professor with the School of Computer and Information Engineering, Chuzhou University, Chuzhou, China. Her current research interests include trusted computing, wireless networks, smart healthcare, and Internet of Things.

**Diptendu Sinha Roy** received the Ph.D. degree in engineering from Birla Institute of Technology Mesra, Ranchi, India, in 2010.

He is currently an Associate Professor with the Department of Computer Science and Engineering, National Institute of Technology Meghalaya, Shillong, India, as an Associate Professor. His current research interests include Internet of Things, big data, machine learning, and soft and evolutionary computing.