Behavioral Anomaly Detection of Older People Living Independently

Carlos A. S. Cunha(✉), Rui Pedro Duarte, and David Mota

Polytechnic Institute of Viseu, Viseu, Portugal
{cacunha,pduarte,dmota}@estgv.ipv.pt

Abstract. Older people living independently represent one significant part of the population nowadays. Most of them have family or friends interested in being informed about changes in their routine. Considering these changes signal some physical or mental problem, they can trigger a contact or action from the interested persons to provide support. This paper presents an approach for non-intrusive monitoring of older people to send alerts after detecting anomalous behaviors. An analysis of seven months of data gathered using PIR sensors in a couple's living house has shown regularities in their presence in compartments along the day. We validated the adequacy of an outlier detection algorithm to build a model of the persons' behavior, exhibiting just 3.6% of outliers interpreted as false positives.

Keywords: Autonomous Living · Anomaly Detection · Behavioral Analysis

1 Introduction

According to the report of the United Nations [14], the number of older persons over the next three decades is predicted to more than double compared to 2020, reaching over 1.5 billion in 2050. The same report shows that older persons often live independently when their health is good enough. Nevertheless, they rely on the support of children, siblings, and other kin (friends, neighbors) not residing with them.

Older persons need to be able to perform activities of daily living (ADLs) [9] unassisted to live independently (e.g., eating, bathing, and mobility). ADLs represent the elements forming each person's routine. The spatiotemporal distribution of ADLs models a regular behavior pattern, which identifies behavioral

This work is funded by National Funds through the FCT—Foundation for Science and Technology, I.P., within the scope of the project Ref. UIDB/05583/2020. Furthermore, we would like to thank the Research Centre in Digital Services (CISeD) and the Polytechnic of Viseu for their support. Moreover, the authors greatly thank IPV/CGD-Polytechnic Institute of Viseu/Caixa Geral de Depositos Bank, within the scope of the projects PROJ/IPV/ID&I/002 and PROJ/IPV/ID&I/007.

S. Longfei and P. Bodhi (Eds.): MobiQuitous 2022, LNICST 492, pp. 571–582, 2023.
https://doi.org/10.1007/978-3-031-34776-4_30

anomalies. People's families and friends are interested in these anomalies because they express alerts to conditions that may require their intervention.

Previous work on monitoring ADLs of the elderly focuses mainly on long-term behavioral analysis—e.g., for detection of dementia [11,13]. Some techniques are intrusive (e.g., video-vigilance) [10] or change the person's routine (e.g., wearables [3,8]). This paper presents the infrastructure based on low-cost PIR sensors installed in older adults' homes living alone to gather data and a machine learning approach to detect behavioral anomalies in their everyday behaviors.

This paper addresses the following research questions:

- How to gather data required for detecting anomalous behaviors in older adults living independently?
- Are there regularities in the daily living behaviors captured by the permanence of individuals in house compartments of a couple living together?
- How to model anomalous events from data representing temporal permanence of individuals in house compartments?

This remaining of this paper is structured as follows. Section 2 presents the related work. Section 3 describes the data gathering and preprocessing methodology. Section 4 analyses behavioral patterns from the data and presents the anomaly detection model. Section 5 discusses experimental results. Finally, Sect. 6 presents the conclusion.

2 Related Work

Most work on behavioral analysis explores data gathered in the context of ADLs. In [11] ADLs are studied for diagnosis of dementia. Statistical methods and machine learning techniques have shown accuracy higher than 90% from sensor data – door, motion, temperature, humidity, vibration and lidar sensors, and smart plugs. The use of household appliances during late-night hours has revealed an essential feature in predicting early-stage dementia patients. Also, in [13], the authors explore anomalies as a predictor of dementia. They perform pattern extraction over activity vectors of whole days. The Hamming distance and Levenshtein distance determine the distance between two sequences of activities, while the entropy measures the activity uncertainty at each time slot. The approach showed the presence of regularities in the dataset used in the experimental work. In [2], the authors present a simulation tool to create normal and wandering trajectories and related sensor activities to detect overnight wandering in patients with dementia in home environments. The resultant simulation datasets created a decision tree classifier to perform binary classification of normal and wandering pathways. An unobtrusive activity recognition classifier employing deep convolutional neural networks (DCNNs) is studied in [6]. The approach uses anonymous binary sensors to recognize different ADLs. The approach was evaluated with high accuracy using a single older woman living inside for eight months. In [15] the authors address detection of anomalous days for ADLs and evolution trends. They propose a similarity index to measure deviations from maps of pheromone distributions, entailing the activity intensity, the places where the activities occur, and relative weights of places related to the spatial distribution of activity. Another work [5] studies unsupervised machine learning to cluster users' behavioral patterns. Evaluation using inexpensive and unlabelled sensors in a community-based housing with 17 residents exhibited accuracy higher than 85%.

Root cause analysis pursues the problem behind the anomaly. Understanding the cause of anomalies unveils the problem behind ADL patterns. The authors of [17] investigated a similarity measure approach to identify single anomaly sources.

A review of techniques for abnormal human activity recognition based on an analysis of video surveillance is presented in [4]. Deep learning dominates this area, focusing mainly on fall detection. The review pointed to vision-based single-person behavior recognition as an area not yet explored due to the non-availability of large datasets to train deep networks.

The work presented in this paper distinguishes itself from previous work as it addresses the detection of behavioral anomalies based just on the position of persons in the house to ensure their privacy. Moreover, it analyzes the regularities of a couple instead of individuals.

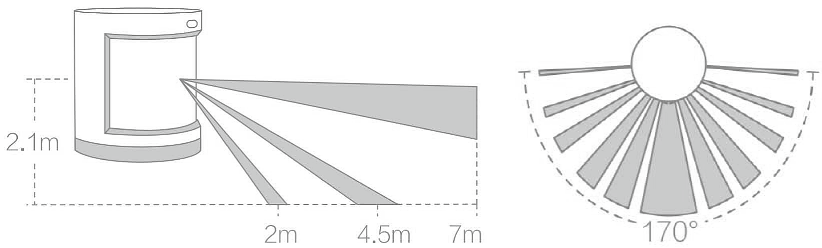
3 Data Gathering and Preprocessing

The testbed used to gather data was deployed at a husband and wife's house for approximately seven months – from February until September of 2021. They live alone in a village in the center of Portugal. Their ages are between 68 and 70 years old, and they have a college education. They rarely receive regular family visits and are often at the house. Participants consented and accepted their participation in this study, as long as the data are used for scientific research.

3.1 Data Gathering

The placement of sensors followed the criteria of covering as much monitoring area as possible and reducing intrusiveness in people's daily activities to a minimum. The PIR sensors used in the testbed are small enough (30 mm $\times$ 30 mm $\times$ 33 mm) to be ingrained in the environment without being intrusive to the aesthetics or functionality of house compartments. They provide a maximum coverage distance of seven meters (Fig. 1(a)) and angle of 170° (Fig. 1(b)).

Sensors were installed in house compartments where people perform their normal daily activities. Exceptions are transition zones between compartments. Following that criteria, we considered the room (Fig. 2(b)), WC (Fig. 2(a)), living room (Fig. 2(d)), and kitchen (Fig. 2(c)). Also, each sensor location was chosen to avoid undesired displacement (e.g., during cleaning activities), leading to erroneous data.



(a) Coverage distance (b) Coverage angle

Fig. 1. PIR sensors specification.

3.2 Data Integration

The sensor data gathered by the hub is only accessible through third-party applications. We adopted OpenHab [1] for that purpose. This platform provides connectors specific to several IoT sensor types, including those used in our testbed.

The Linux OpenHabian image installed in the Raspberry PI (version 3) with 512 Mb, contains the openHAB platform. This platform configures each sensor connected to the hub as a *thing*, through a specific *channel* responsible for transmitting the respective sensor data. These data are incrementally written in logs in text format, as shown in Listing 1.1.

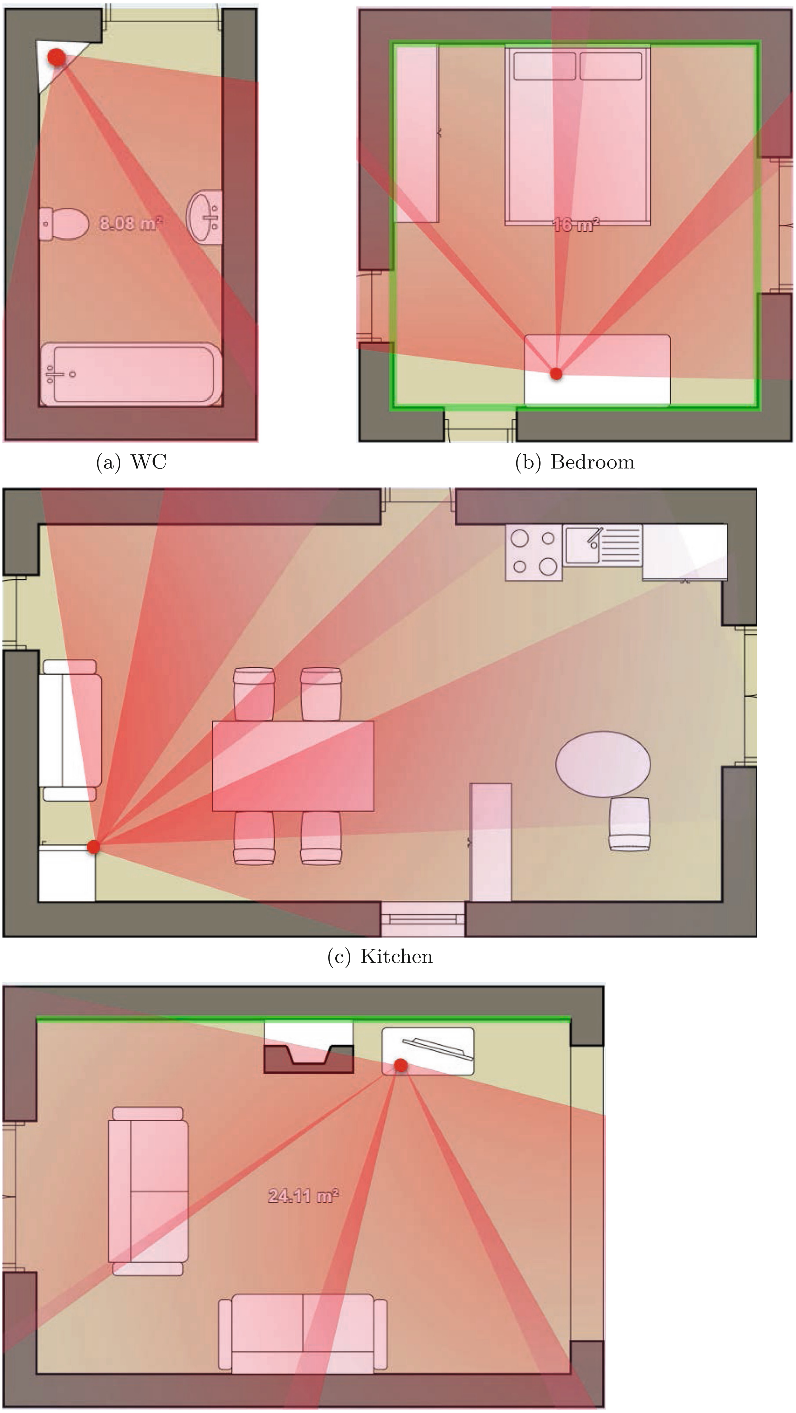
Listing 1.1. OpenHab event

```
"$oid":"60284c135ed15f380a0c35ab",
"date":"2021-01-30",
"time":"17:07:55.544",
"sensorID":"XiaomiMiMotionSensor4",
"event":"ON"
```

A Python script performs data preprocessing. It starts by querying the MongoDB database and transforming the openHAB log into an accessible format for data processing. Then it labels each record with the compartment name, removes incomplete data blocks (days with records missing), and aggregates data temporally (Listing 1.2).

3.3 Incomplete Data Removal

Each record in the dataset represents an event triggered by a sensor at a specific time. Thus, residents' absence of movement or failure periods explain periods of lacking data. To shield the model performance, the preprocessing activity must remove the records associated with failure periods from the training dataset. However, discriminative separation of failure periods from the absence of movement is complex just from data analysis.



(a) WC (b) Bedroom

(c) Kitchen

(d) Living room

Fig. 2. Placement of Sensors

Empirical data analysis shows that only faulty days in the dataset have less than 200, 70, and 20 movement triggers for the living room, restroom, and bedroom, respectively. Accordingly, we excluded days from the dataset not exhibiting these minima.

3.4 Data Aggregation

The temporal granularity of each record impacts the performance of behavioral models. For too fine-grain temporal aggregation, the data variance may compromise the identification of patterns. Conversely, too coarse-grain time aggregation may hide important data for pattern extraction.

We aggregated the number of triggers in one-hour intervals after analyzing the 5th, 50th, and 95th quantiles, respecting the number of triggers for intervals of 1, 15, 30, and 60 min.

3.5 Data Formatting

Data gathered from sensors are commonly represented as time series. Two preprocessing transformations are necessary before processing the records provided by sensors: (1) generate timestamp representations adequate for processing; (2) create records for periods of missing data.

Each dataset record indicates the time of the day a person triggers sensors inside a compartment. Representing the timestamp feature as a numerical variable is essential to specify the time distance between triggers. Thus, to perform behavioral analysis using the position of persons during the day, we decomposed the timestamp into several features, namely, the month, weekday, and time. Also, to process the time of the day as a numerical variable, we converted it to the format of (1).

$$feature_{hour} = hour + minutes/60 \tag{1}$$

Listing 1.2. OpenHab event

```
Date,Time,Weekday,Month,Time,C1,C2,C3,C4
2021-02-21,02:00:00,Sun,Feb,2.0,1,0,0,0
2021-02-21,02:20:00,Sun,Feb,2.33,1,0,0,8
2021-02-21,02:40:00,Sun,Feb,2.67,1,1,1,1
2021-02-21,03:00:00,Sun,Feb,3.0,8,0,1,1
2021-02-21,03:20:00,Sun,Feb,3.33,0,0,0,8
2021-02-21,03:40:00,Sun,Feb,3.67,0,0,0,8
2021-02-21,04:00:00,Sun,Feb,4.0,8,0,0,0
2021-02-21,04:20:00,Sun,Feb,4.33,0,0,0,8
2021-02-21,04:40:00,Sun,Feb,4.67,0,0,0,8
2021-02-21,05:00:00,Sun,Feb,5.0,8,0,0,0
2021-02-21,05:20:00,Sun,Feb,5.33,0,0,0,8
2021-02-21,05:40:00,Sun,Feb,5.67,0,0,0,8
```

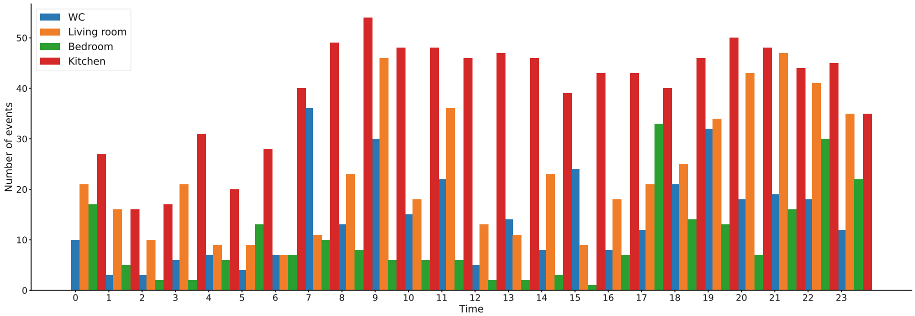


Fig. 3. Number of triggered events per hour.

```
2021-02-21,06:00:00,Sun,Feb,6.0,8,0,0,0
2021-02-21,06:20:00,Sun,Feb,6.33,0,0,0,8
2021-02-21,06:40:00,Sun,Feb,6.67,0,0,0,8
2021-02-21,07:00:00,Sun,Feb,7.0,0,0,0,0
```

4 Analysis of Results

The research objectives described in this paper depend on behavioral regularities in house inhabitants. We start by analyzing the number of triggers for each period and compartment to evaluate these regularities. Then, we perform the same analysis for the lower and higher limits, exposed by the percentile 5th and 95th. These percentiles filter the 5% bottom and top values, which are susceptible to representing outliers that compromise the identification of patterns. Finally, we evaluate the occupation probability of each compartment, indicating the likelihood of an inhabitant being present there at a specific time.

4.1 Analysis of Events Per Hour

Figure 3 presents the number of events registered in the house compartments per hour. The periods with a higher presence in all compartments are observable from the graphics analysis. These values expose the regularities of the presence of persons in each compartment.

4.2 Occupation Probability

The probability of compartments occupation by persons represents another valuable analysis perspective since it exhibits the occupation likelihood of a compartment during a specified period. Hence, a behavioral anomaly results from an unlikely presence in a compartment.

Some events in the data represent transient activity filtered from the analysis since they represent temporary presence in the compartment (e.g., when

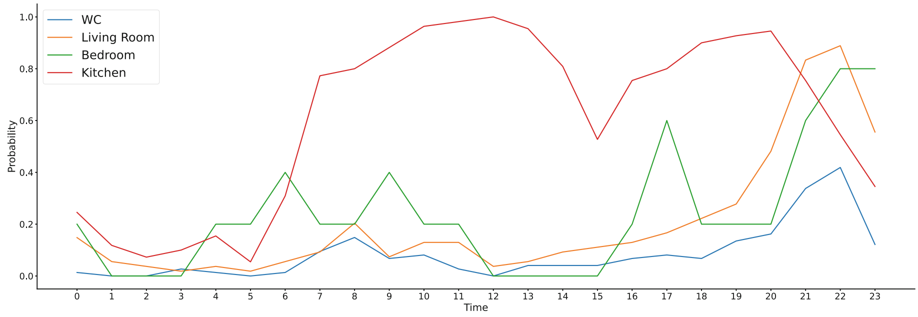


Fig. 4. Probability of occupation of each compartment during the day.

somebody passes by a compartment) instead of effective permanence. Since each sensor triggers a new event at each minute, we consider five events (i.e., five minutes) as the threshold for events to account for the presence in the compartment at the corresponding hour, as specified in (2).

$$Prob_t = \frac{\#Days\{\#triggers_t >= 5\}}{\#Days} \tag{2}$$

Figure 4 presents the presence probability in each house compartment. It reveals the following patterns:

– from midnight to 6 am, the likelihood of presence in any compartment is low;
– the presence in the kitchen is high between 7 am and 2 pm and between 4 pm and 9 pm;
– occupation in the living room is higher between 9 pm and 11 pm;
– the room is inhabited at 5 pm, and after 10 pm – the location of the sensor position (close to the room's entrance) explains the low activity registered after midnight;
– regarding the occupation of WC, there is more occupation at the end of the day, reaching its peak around 10 pm.

4.3 Outlier Detection

A visual analysis of patterns shows regularities in the use of house compartments. However, building an outlier detection model able to predict unusual behaviors exposed by the occupation of spaces in the house requires combining the presence of persons in all compartments at a specific time. Additionally, the error created by hourly data aggregating can compromise the model prediction performance. These constraints lead the solution to an anomaly detection model combining data from several compartments that exploits time as a numerical variable calculated as in (1), to avoid the error originating in temporal data aggregations.

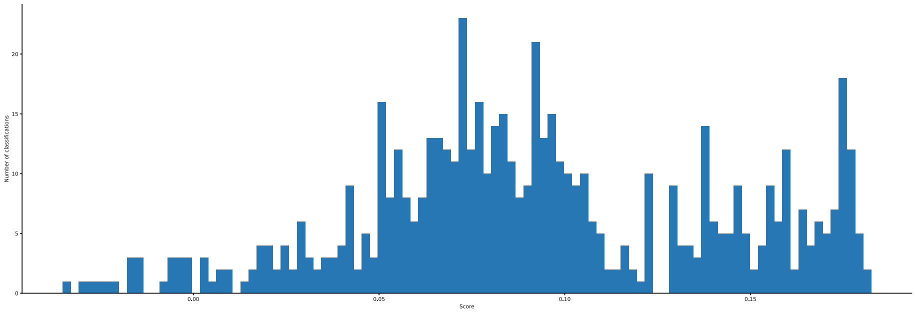


Fig. 5. Outlier classification using Isolation Forests.

Machine learning algorithms classification is two-fold: supervised and unsupervised. Supervised learning algorithms build models from samples of all classes (e.g., normal and anomalous behaviors). In contrast, unsupervised algorithms model one class (normal samples) and classify anomalous samples when they unfit the model. Unsupervised algorithms suit the problem addressed in this paper since the expected behavior is observable, but the anomalous behavior space is practically unlimited.

We decided on the Isolation Forest algorithm [12] to create an anomaly detection model due to its exemplary performance in several domains [7,16]. This algorithm isolates anomalies based on the characteristics that make them unique. Accordingly, it fits a decision tree on all the observations and regards outliers as those instances close to the root of trees.

The Isolation Forest requires parameterization of the contamination factor. This parameter sets the sensitivity of the algorithm to outlier classification. We evaluated several contamination factors and decided on 0.1 (10%) due to its lowest number of outliers with the training dataset, showing the best model adjustment to the data. The resultant model has 3.6% of outliers – shown in Fig. 5 as the number of samples classified with a score inferior to zero. Table 1 lists these outliers, representing false positives if used in an alert system.

5 Discussion of Results

Section 4 presents ADL regularities defined in the context of people's temporal occupation of house compartments. Inference of fine-grain ADLs (e.g., reading a book, drinking water) demands intrusive appliances with low impact on the answer to our research questions. Informal caretakers or familiars of older adults may want to be notified of changes in their daily routine without interfering and disrespecting their privacy. A marker signaling an abnormal condition that requires attention is a change in an individual's home routine, described in terms of the temporal distribution of permanence in house compartments.

Detection of abnormal conditions explores patterns of two older adults' presence in house compartments. However, ordinary people's routine outliers may

Table 1. Outliers identified during the training process.

weekday	hour	kitchen	living-room	room	wc	score	prediction
4	23.0	1	0	1	0	−0.004654	−1
6	2.0	0	1	1	1	−0.016904	−1
6	14.0	0	1	1	0	−0.015443	−1
6	23.0	0	0	1	0	−0.021297	−1
7	5.0	0	1	1	0	−0.027471	−1
7	17.0	1	0	1	1	−0.025279	−1
7	23.0	1	1	1	0	−0.014824	−1
5	5.0	0	1	0	1	−0.006145	−1
6	4.0	0	1	0	1	−0.009018	−1
7	0.0	1	0	1	0	−0.002024	−1
7	5.0	0	1	1	1	−0.035581	−1
6	23.0	1	0	1	0	−0.001213	−1
5	5.0	0	1	1	1	−0.003214	−1

lead to false notifications. That may happen because they change their routine for some reason not explained by a problem that requires attention from their family.

False positives leading to false notifications may impact the trust of the system. Experimental results showed 3.6% of outliers respecting the period of analysis. Since the main system goal is to trigger notifications to familiars interested in being informed about abnormal conditions, the consequence of an anomaly would be an unnecessary contact. However, the value of an accurate notification may compensate for the cost of unnecessary notifications respecting false positives when they are rare.

The observed regularities in the data may depend on the couple's activities rather than the individuals. Patterns in data produced by PIR sensors mean that a couple's routine can be exploited to trigger alerts for anomalous conditions.

6 Conclusion

The increasing percentage of older persons living independently nowadays makes monitoring their ADLs valuable, allowing alerts to family or friends not living with them when something goes wrong. However, sensors attached to the person and other physical objects can simplify ADLs identification and processing but create intrusiveness issues.

This paper presents a solution based on simple PIR sensors placed in house compartments to gather information about the presence of persons in those compartments without compromising their daily routine. The events generated by these sensors are used to build a model of the location of persons in the house.

Deviations from the expected behavior (e.g., going to the kitchen several times during the night) provide the basis to trigger notifications.

We gathered and analyzed the activity of an older couple during seven months of living together in the same house. Experimental results unveiled behavioral regularities in the routine of persons that can be explored to detect abnormal conditions.

We plan to design the solution to trigger notifications based on the model created in future work. Despite being rare, the number of false positives may lead to distrust by the most demanding users. Combining several detected irregularities would potentially reduce the number of unnecessary alerts. A significant problem is determining how to explore the combination of some outliers detected without extending the notification latency.

References

1. openHAB. https://www.openhab.org
2. Casaccia, S., Rosati, R., Scalise, L., Revel, G.M.: Measurement of Activities of Daily Living: a simulation tool for the optimisation of a Passive Infrared sensor network in a Smart Home environment. In: 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), pp. 1–6 (2020). https://doi.org/10.1109/i2mtc43012.2020.9128409
3. Cunha, C.A., Duarte, R.P.: Multi-device nutrition control. Sensors **22**(7), 2617 (2022)
4. Dhiman, C., Vishwakarma, D.K.: A review of state-of-the-art techniques for abnormal human activity recognition. Eng. Appl. Artif. Intell. **77**, 21–45 (2019). https://doi.org/10.1016/j.engappai.2018.08.014
5. Fiorini, L., Cavallo, F., Dario, P., Eavis, A., Caleb-Solly, P.: Unsupervised machine learning for developing personalised behaviour models using activity data. Sensors **17**(5), 1034 (2017). https://doi.org/10.3390/s17051034
6. Gochoo, M., Tan, T.H., Liu, S.H., Jean, F.R., Alnajjar, F.S., Huang, S.C.: Unobtrusive activity recognition of elderly people living alone using anonymous binary sensors and DCNN. IEEE J. Biomed. Health Inform. **23**(2), 693–702 (2018). https://doi.org/10.1109/jbhi.2018.2833618
7. He, Y., Zhu, X., Wang, G., Sun, H., Wang, Y.: Predicting bugs in software code changes using isolation forest. In: 2017 IEEE International Conference on Software Quality, Reliability and Security (QRS), pp. 296–305 (2017). https://doi.org/10.1109/QRS.2017.40
8. Kantoch, E., Augustyniak, P., Markiewicz, M., Prusak, D.: Monitoring activities of daily living based on wearable wireless body sensor network. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 586–589. IEEE (2014)
9. Katz, S.: Assessing self-maintenance: activities of daily living, mobility, and instrumental activities of daily living. J. Am. Geriatr. Soc. **31**(12), 721–727 (1983). https://doi.org/10.1111/j.1532-5415.1983.tb03391.x
10. König, A., et al.: Validation of an automatic video monitoring system for the detection of instrumental activities of daily living in dementia patients. J. Alzheimers Dis. **44**(2), 675–685 (2015)

11. Kwon, L.N., et al.: Automated classification of normal control and early-stage dementia based on activities of daily living (ADL) data acquired from smart home environment. Int. J. Environ. Res. Public Health **18**(24), 13235 (2021). https://doi.org/10.3390/ijerph182413235
12. Liu, F.T., Ting, K.M., Zhou, Z.H.: Isolation forest. In: 2008 Eighth IEEE International Conference on Data Mining, pp. 413–422 (2008). https://doi.org/10.1109/ICDM.2008.17
13. Maučec, M.S., Donaj, G.: Discovering daily activity patterns from sensor data sequences and activity sequences. Sensors **21**(20), 6920 (2021). https://doi.org/10.3390/s21206920
14. United Nations: World population ageing 2020. Technical report (2020). https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/undesa_pd-2020_world_population_ageing_highlights.pdf
15. Susnea, I., Dumitriu, L., Talmaciu, M., Pecheanu, E., Munteanu, D.: Unobtrusive monitoring the daily activity routine of elderly people living alone, with low-cost binary sensors. Sensors **19**(10), 2264 (2019)
16. Xu, S., Zhu, J., Jiang, J., Shui, P.: Sea-surface floating small target detection by multifeature detector based on isolation forest. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **14**, 704–715 (2021). https://doi.org/10.1109/JSTARS.2020.3033063
17. Yahaya, S.W., Lotfi, A., Mahmud, M.: Detecting anomaly and its sources in activities of daily living. SN Comput. Sci. **2**(1), 1–18 (2020). https://doi.org/10.1007/s42979-020-00418-2