Predictive Behavioural Monitoring and Deviation Detection in Activities of Daily Living of Older Adults

Citation for published version:
Zheng, J & Papapanagiotou, P 2022, Predictive Behavioural Monitoring and Deviation Detection in Activities of Daily Living of Older Adults. in N Bier, A Fred & H Gamboa (eds), *15th International Joint Conference on Biomedical Engineering Systems and Technologies, Volume 5 - HEALTHINF.* vol. 5, SCITEPRESS, pp. 899-910, 15th International Conference on Health Informatics 2022, 9/02/22. https://doi.org/10.5220/0010974500003123

Digital Object Identifier (DOI):
10.5220/0010974500003123

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
15th International Joint Conference on Biomedical Engineering Systems and Technologies, Volume 5 - HEALTHINF

Predictive Behavioural Monitoring and Deviation Detection in Activities of Daily Living of Older Adults

Jiawei Zheng[a] and Petros Papapanagiotou[b]
School of Informatics, University of Edinburgh, 10 Crichton Street, Edinburgh, EH8 9AB, U.K.

Keywords: Activities of Daily Living, Behaviour Monitoring, Deviation Detection, Predictive Modelling, Sensor Data.

Abstract: Predictive behaviour monitoring of Activities of Daily Living (ADLs) can provide unique, personalised insights about an older person's physical and cognitive health and lead to unique opportunities to support self-management, proactive intervention and promote independent living. In this paper, we analyse ADL data from ambient sensors to model behaviour markers on a daily basis. Using a number of machine learning and statistical methods we model a predicted daily routine for each marker, detect deviations based on a set of relative thresholds and calculate long-term drifts. We further analyse the causal factors of deviations by investigating relationships between different activities. We demonstrate our results using data from a sample of 11 participants from the CASAS dataset. Finally, we develop a dashboard to visualize our computed daily routines and quantified deviations in an attempt to offer useful feedback to the monitored person and their caregivers.

1 INTRODUCTION

Loss of independence in Activities of Daily Living (ADL) is associated with adverse health outcomes, both physical and mental, and mortality in older adults (Albanese et al., 2020; Cohen-Mansfield and Perach, 2012). Adverse health events, including heart failures, falls, strokes, etc. and the onset of cognitive and physical frailty are not random occurrences, but a consequence of long-term health deterioration or unhealthy lifestyle. Gradual cognitive and physical decline can significantly affect the capacity of people in advanced age to perform ADLs independently (Akram et al., 2020). Proactive monitoring and analysis of short and long term deviations from a regular routine in ADLs can provide vital insights on an older person's health and a continuous evaluation of their physical and cognitive ability (Sepesy Maučec and Donaj, 2021). These can not only inform the decision making of care givers towards timely, proactive interventions, but also support person-centred self-management and motivate healthier living.

Modern and emerging smart home and wearable sensor technologies allow us to collect continuous data on daily living in an unobtrusive and affordable manner. This involves time series data with status information and timestamped event data when a status change occurs (Cook et al., 2013a). Activity Recognition and Machine Learning (ML) techniques can then produce fine-grained daily activity data with temporal and spatial information. Further AI modelling can allow us to develop rich temporal and spatial profiles of daily routines, including sleep duration, number of meals, and levels of active movement. We can detect temporal and spatial deviations on individual days or in the long term, such as staying in the toilet too long, too frequent toilet visits during the night, activity delays due to reduced mobility, sleep disruption or decline over time etc. Ultimately, our goal is to link these detected deviations to health outcomes, towards health monitoring and timely, proactive interventions. More importantly, figuring out the causes of the deviations and the relationships between daily activities can provide key insights towards preventative instructions and effective care provisions, both by care givers and the people themselves, to avoid potential negative effects and prevent adverse events.

Until recently, studies that investigate the relationship between ADLs and health outcomes were based on questionnaires and self-reporting (Kanti Majumdar, 2014; Cook and Schmitter-Edgecombe, 2021). The result of this may be affected by experimenter observations or retrospective memory limitations (Palmer, 2018). Instead, sensor-based, passive and continuous monitoring of ADLs, associated with

[a] https://orcid.org/0000-0002-6515-6423
[b] https://orcid.org/0000-0003-0928-6108

Zheng, J. and Papapanagiotou, P.
Predictive Behavioural Monitoring and Deviation Detection in Activities of Daily Living of Older Adults.
DOI: 10.5220/0010974500003123
In *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2022) - Volume 5: HEALTHINF*, pages 899-910
ISBN: 978-989-758-552-4; ISSN: 2184-4305

behavioural analysis of patterns and routines, can provide concrete, objective insights about the relationship between daily activities and health.

Our work is part of the Advanced Care Research Centre, a large multi-disciplinary project focused on the support of people in later life living in their own homes and in supported care environments. (University of Edinburgh, 2021). The main contributions of this paper are the following: *Firstly*, we extract behaviour markers from activity-labelled time series sensor data to model daily behaviours with rich temporal information. *Secondly*, we propose a deviation detection approach for daily behaviours, including a deviation score for a given day and a long-term drift from the normal routine. *Thirdly*, we investigate the relationship between different activities towards a causal explanation of deviations. *Finally*, we develop an interactive dashboard, which can visualize personal temporal profiles of daily behaviours including daily routines, trends of each behaviour marker and potential deviations. The dashboard provides intuitive behavioural statistics of the monitored older adults that may be useful both for self-monitoring and management and as useful information to care providers. We demonstrate the results of our approach using ADL data from 11 participants over 2 months from the CASAS project (Cook et al., 2013a).

2 RELATED WORK

Research in the field of ADLs has recently been receiving increasing attention, especially in the context of supporting independent living for older people and providing effective care. Aminikhanghahi et al. propose a change point detection algorithm for identifying activity transition points, which is used to improve the performance of activity recognition (Aminikhanghahi et al., 2019). In this paper, we focus on predictive daily behaviour monitoring and deviation detection based on the activity-labelled time series data. Yahaya et al. provide a comprehensive list of recent efforts on deviation detection (Yahaya et al., 2019).

Classification methods treat deviation detection as a binary classification problem. They require ADL data labelled as *normal* or *abnormal*, where the latter reflects either a pre-specified pattern of behaviour, such as the behavioural difficulties of people with dementia (Arifoglu and Bouchachia, 2019). Due to the scarcity of abnormal data in real datasets, it is common to train and generate synthetic data for the abnormal class. More recent approaches consider as abnormal any deviation from a normal routine that is learned from historical data (Yahaya et al., 2019; Pazhoumand-Dar et al., 2020; Yahaya et al., 2021b).

Some classification approaches rely on detecting whether the sensor information exceeds a fixed **threshold** of values. Collected data is used as training data representing the normal behaviours and subsequent activity data are used as testing data for the learned model. Data that have significant variations past certain thresholds are defined as outliers. For example, Pierleoni et al. propose a fall detection method based on the fusion data collected from a triaxial accelerometer, gyroscope, and magnetometer on wearable devices (Pierleoni et al., 2015). If the body orientation falls below a pre-defined threshold for a certain period of time, the system will issue an alarm.

However, approaches that rely on wearable sensors are not always applicable in practice. For instance, some people may not feel comfortable constantly wearing a device or may forget to put it on or charge the battery. Moreover, missing data and false positives can make these approaches less reliable (Pazhoumand-Dar et al., 2020). For example, lying down on a bed suddenly may be mistaken for the movement pattern of an accidental fall. Some approaches have overcome these limitations by using ambient sensors in a smart home (Cook et al., 2013a). The data are collected from environmental sensors when the subjects have interactions with their environment. For example, Pazhoumand-Dar et al. use Kinect sensors composed with power consumption data to monitor daily behaviours (Pazhoumand-Dar et al., 2020). Their training data is aimed to model the regularity and frequency of important activities and does not need to be labelled in advance. Howedi et al. use ADL data collected from ambient sensors to detect deviations based on entropy measures (Howedi et al., 2020). Activities with entropy values exceeding a certain range are detected as abnormalities. Similarly, Yahaya et al. propose an ensemble of abnormal detection approach by detecting if the test data differs significantly from the training data based on a threshold for a defined Normality Score (Yahaya et al., 2019).

Finally, **clinical score-based** approaches involve an assessment of older people by clinical experts through various factors of their daily activities, such as cognitive health, functional mobility, etc (Dawadi et al., 2016; Alberdi Aramendi et al., 2018; Cook and Schmitter-Edgecombe, 2021). The assessments are usually conducted at regular time intervals, for example every 6 months, and a total score representing the health status of the participant is calculated. After data collection, a computation model is trained to map the clinical score to the data collected by ambient sensors and predict future scores based on that.

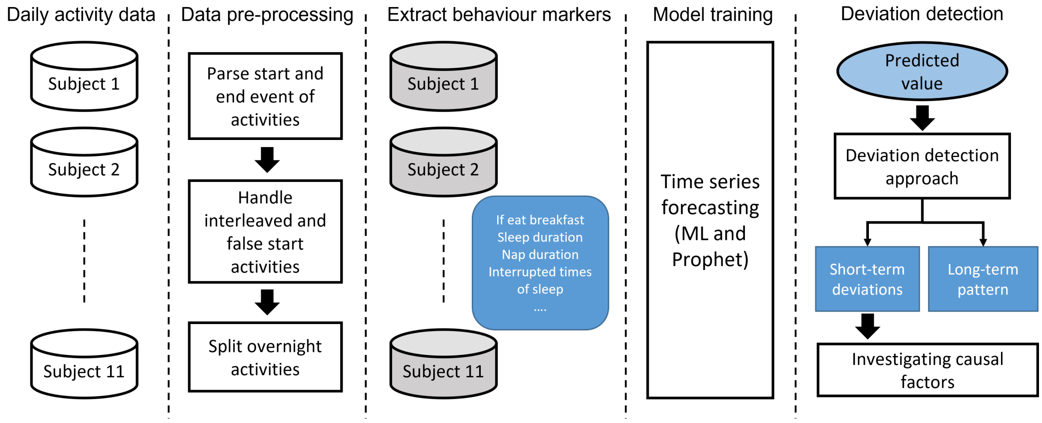


Figure 1: An overview of the proposed approach.

The assessments are carried out by self-reporting, using an appropriate questionnaire. However, the result from this type of assessment may be affected by the respondent's subjective view and state of mind when filling in the questionnaire (Yahaya et al., 2021a). As a consequence, the score may not accurately reflect the actual health condition of the respondent.

In our work, we also adopt a threshold-based classification approach using unlabelled data.

3 PROPOSED APPROACH

Our work is aimed towards proactive health monitoring and predictive deterioration and deviation detection. Our proposed approach consists of four stages, as shown in Figure 1: (1) data preprocessing, (2) extraction of behaviour markers, (3) predictive modelling, and (4) deviation detection, including investigating causal factors of deviations. We describe each stage in more detail next.

3.1 Data Description

In this paper, we analyse available ADL data collected from ambient sensors in smart homes, published by Cook et al. (Cook et al., 2013a). The data sets contain continuous data from unobtrusive sensors including motion sensors, door sensors, light switches and light sensors, deployed in single-family residences.

Each sensor event is labelled with a corresponding activity, including sleeping, cooking, eating, napping, going to the toilet, and working, in a total of 33 different activities. A sample of data is shown in Table 1a.

Our paper is based on data from 11 individuals over 2 months. The raw sensor data is recorded in a time-series format with the following fields:

- **Timestamp:** The date and time of the event.
- **Sensor:** The name of the sensor, as found in the sensor floor plan.
- **Room:** The room-level sensor location.
- **Location:** The fine-grained location of the sensor, such as bed, chair, etc.
- **Message:** The value generated by the sensor, such as `on`, `off`, etc.
- **Sensor Type:** The type of sensor generating the event (e.g. `Control4-Motion`, `Control4-LightSensor`, etc.), such that provides context to the generated message.
- **Activity:** A manual label of the corresponding activity of this event, such as sleeping, eating, etc.

Due to inherent uncertainty in the environment and human behaviours, the dataset is noisy. We specifically identified 4 categories of potential noise below:

- **Accuracy of Labelling:** Since the particular dataset we are examining was manually labelled post-hoc, there is no measure of the accuracy of the labelling. In fact, newer versions of the dataset seem to have updated some of the labels. More generally, even with an automated activity recognition algorithm, such as the one by Cook et al. (Cook et al., 2013b), there is still some level of uncertainty that may affect the results.
- **Lack of end Event:** The observed activity events indicate when an activity is detected. However, there is no clear indication of the exact end time of the activity or the transition time to the next one. This causes some inaccuracy in our calculated activity intervals (see Section 3.2).
- **Distinguishing Activities:** In some cases, the same small set of 1-2 sensors may be used to detect multiple types of activities with similar action

Table 1: Sample of ADL sensor data.

(a) Raw ADL sensor data labelled with activities.

	Timestamp	Sensor	Room	Location	Message	Sensor type	Activity
1	2011-06-15 09:58:27	M007	Kitchen	Kitchen	ON	Control4-Motion	Cook_Breakfast
2	2011-06-15 09:58:42	M007	Kitchen	Kitchen	OFF	Control4-Motion	Cook_Breakfast
3	2011-06-15 09:58:45	D005	Kitchen	Refrigerator	Close	Control4-Door	Cook_Breakfast
4	2011-06-15 09:58:46	M005	DiningRoom	DiningRoom	ON	Control4-Motion	Eat_Breakfast
5	2011-06-15 09:58:47	M008	Kitchen	Kitchen	ON	Control4-Motion	Cook_Breakfast
6	2011-06-15 09:59:05	LS001	Kitchen	Kitchen	4	Control4-LightSensor	Cook_Breakfast
7	2011-06-15 09:59:06	LS007	Kitchen	Kitchen	6	Control4-LightSensor	Cook_Breakfast
8	2011-06-15 09:59:10	M005	DiningRoom	DiningRoom	ON	Control4-Motion	Eat_Breakfast
9	2011-06-15 10:00:29	LS005	DiningRoom	DiningRoom	13	Control4-LightSensor	Eat_Breakfast
10	2011-06-15 10:00:46	LS015	DiningRoom	DiningRoom	14	Control4-LightSensor	Eat_Breakfast
11	2011-06-15 10:00:47	M008	Kitchen	Kitchen	OFF	Control4-Motion	Cook_breakfast

(b) ADL data resulting after pre-processing.

Activity	Start time	End time	Duration	Daycase	Interval
Cook_Breakfast	2011-06-15 09:58:27	2011-06-15 09:59:06	00:00:39	2011-06-15	00:00:04
Eat_Breakfast	2011-06-15 09:59:10	2011-06-15 10:00:46	00:01:36	2011-06-15	00:00:01

patterns or taking place in the same location, such as washing the dishes and preparing a meal. This can lead to some misclassification.

- **Sensor Modalities:** Given the selection of sensors, there are only a few modalities available, mainly movement and lighting. This means that certain activities, such as the exact time an individual falls asleep, cannot be detected accurately. Additional modalities, such as the ones offered by wearable devices, can help provide a finer grained activity detection. However, there will always be some uncertainty in activity recognition due to the granularity of the involved sensors.

Whilst we do not address this noise explicitly in our modelling, it needs to be taken into consideration when making potential decisions or interventions based on our produced insights.

3.2 Data Preprocessing

The ADL dataset consists of raw sensor data, which represents all sensor events that occur during a period of time, along with their specific location, timestamps, sensor ID, message, etc, as shown in Table 1a.

The first processing step involves the removal of noise, as described in the previous section, to the extent possible, as well as the removal of unknown activities labelled as *"other activity"*.

We then focus on particular patterns of sequences of events representing two different, interleaved activities during the same period of time, as shown in Table 1a. A person is *cooking breakfast* when a sequence of *eating breakfast* events occurs. Subsequent events show that the person continues to cook breakfast. This means that the person is cooking while eating and the two activities are interleaved during this time.

In row 4 of Table 1a, the *eat breakfast* event is a single orphan event, which prevents us from measuring the (assumed to be short) duration of the corresponding activity. We treat such events as *false occurrences* and remove them from the dataset.

In rows 8-10 of Table 1a, we observe a continuous sequence of *eat breakfast* events. We consider such cases as an *interleaved activity* and we treat them as though the initial activity ended in row 7 and a new *cook breakfast* activity started in row 11.

Based on the above, we detect consecutive events of the same activity and choose the timestamp of the first event as the start of the activity interval and the timestamp of the last event as the end of the activity interval. This results in a processed dataset (shown in Table 1b) that includes the following temporal features of each activity in order:

- **Activity:** the name of the ADL.
- **Start Time:** the start time of the activity.
- **End time:** the end time of the activity.
- **Duration:** the duration of the activity in seconds.
- **Daycase:** the date of the activity.
- **Interval:** the interval in seconds until the next activity starts.

3.3 Modelling Behaviour Markers

Our aim is to process activities performed on a daily basis in order to detect daily routines and deviations. We model behaviour markers for each day.

This daily segmentation requires us to split events occurring at midnight into two events, each belonging to the previous and new day respectively. For example, a *Watch TV* event from 22:10:00 to 01:12:23 is split into two events, one from 22:10:00 to 23:59:59 for the previous day and one from 00:00:00 to 01:12:23 belonging to the new day. The only exception is the sleep activity, which we count from noon to noon of the next day, as we consider overnight sleep to be a significant indicator of healthy living.

The modelled behaviour markers are extracted according to the features related to each daily activity such that provide key insights of the individual's lifestyle and potential links to health indicators.

The total number of behaviour markers we observed is 26, with some examples shown in Table 2. We subsequently focus our efforts to detect deviations on these particular behaviour markers.

Table 2: Examples of daily behaviour markers.

Behaviour markers	Description
`nap_duration`	total *nap* duration
`sleep_time`	time of *sleep*
`wake_up_time`	morning *wake up* time
`last_personal_hygiene`	time of the last *personal hygiene*
`last_nap_endtime`	end time of the last *nap*
`if_eat_breakfast`	whether *eat breakfast* occurred or not
`bed_toilet_times`	times of *toilet* visits during sleep

3.4 Detecting Deviations

The main goal of our approach is to detect and measure short and long term deviations of ADLs from a routine, such that may indicate health deterioration. A key input here is the *ground truth* compared to which a certain behaviour is considered a *deviation*. Each individual may have a significantly different routine and lifestyle, particularly at a later age. This variance can be exacerbated by the consideration of the large variety of different health conditions that may apply to each person. For instance, the connections between frailty and sleep disturbances have been well studied (Piovezan et al., 2013). We therefore adopted a personalised approach and observe daily living deviations from an individual routine, irrespective of its relation to a healthy standard. This means we are detecting and measuring when and individual is not behaving according to their regular routine, even if that routine is unhealthy.

More specifically, given a set of daily behaviour markers for each individual, we create a computational model of their expected measures each day. Taking `nap_duration` as an example, we develop a model that can predict the total duration of a person's naps during the day given their historical data. We also account for seasonality by including the day of the week (e.g. some people may nap more during the weekend) and the month (e.g. people tend to sleep longer during the winter, particularly if affected by a seasonal affective disorder) as features (Anderson et al., 1994). This model is then used as the ground truth of expected values, against which any new observed behaviour is compared.

The technical aspects of this approach are described in more detail next.

3.4.1 Building the Predictive Model

Our aim is to build a predictive model of the normal daily behaviour markers of an individual. Such a model needs to be trained with historical data of routine daily living. We therefore begin by removing outliers in the data, consisting of values that deviate from the mean by 2 standard deviations, which is a common cut-off for outliers in practice in a small dataset (Ilyas and Chu, 2019).

We also consider special cases of daily behaviours that we may want to filter out. For example, we use specific patterns for detecting if the person is away from home overnight and sleeps elsewhere. If the last event of that day is *leave home* and sleep duration is zero, we consider that day an outlier and remove it.

We use the remaining data to train a regression model to predict the future value of behaviour markers. We only consider seasonality features in the model, i.e. the full date, the day of the week, the month, and the order of events.

Our next goal is to maximize the accuracy of the model. For this, we compare the predictive accuracy of different regression models in a model selection process. In this process, we use one of the participants as the validation set for choosing best performing model and other participant as the testing set. The best performing model on the validation set is evaluated on the testing set. Finally, the most accurate model is chosen as the predictive model.

For the data of each behaviour marker of a individual, we use K-fold time series split technique in Scikit-learn (Pedregosa et al., 2011) to split 3-fold training and testing sets. It returns 3 split of data. In the kth split, it returns the first k folds as training set and the $(k+1)$th fold as testing set. We train and evaluate different ML models and statistical time series models based on the training and testing sets. Note that we build separate model for each behaviour marker of a individual. The evaluation result is obtained by computing the average performance of 3 pairs of training/testing sets.

All experiments are evaluated using the mean absolute error (*MAE*) and the root mean squared error (*RMSE*) measures. *MAE*, shown in (1), and *RMSE*, shown in (2) both measure the average magnitude of the errors in a set of predictions, with 0 corresponding to perfect accuracy, while *RMSE* magnifies the impact of large errors. In these equations, for each predicted value i, $\hat{y}_i$ represents a size-n vector of the predicted values, y_i is the vector of actual values, and n is the number of test instances. All performance evaluations are conducted using 3-fold time series cross validation (Bergmeir and Benítez, 2012).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{2}$$

3.4.2 Deviation Detection

After we build the predictive model, we can get the predicted value of behaviour markers by fitting them into the training set. To detect potential deviations based on the predicted value, we calculate the distance between the predicted routine value and the actual value as shown in (3).

$$z = \frac{|y_{predict} - y_{true}|}{MAE} \tag{3}$$

MAE is the mean absolute error of the predictive model (see also Section 3.4.1). For example, if the *MAE* of a model of sleep duration is 1800 seconds, then the model predicts on average 30 minutes more or less sleep than the actual value. A lower *MAE* indicates a better performing algorithm with 0 corresponding to perfect accuracy. When comparing models trained with datasets from 2 distinct individuals with the same algorithm, a lower *MAE* score is an indication of less variability in the data and, therefore, a more predictable and steady daily routine.

The z value in (3) represents the distance of the actual value from the predicted value as a fraction of *MAE*. If the calculated z value of a predicted behaviour marker exceeds a chosen threshold, a deviation is detected. The chosen threshold corresponds to a time window proportional to *MAE* within which we consider the behaviour as normal or routine. A distinct threshold can be selected for each behaviour marker in the daily routine, to account for the flexibility we want to allow for each activity. For instance, if a participant decides to read a book at a considerably different time than usual, we might not want to consider this as a significant deviation. However, the time they go to sleep is much more important as a health indicator, and therefore we may want to flag smaller deviations. We could therefore choose a higher z threshold for the former behaviour marker and a lower value for the latter.

For example, given the sleep duration model with *MAE* of 0.5 hours discussed above, assume a predicted sleep duration $y_{predict}$ of 8 hours for a particular day. Setting the threshold of z to 1 means that an actual sleep y_true of less than 7.5 or more than 8.5 hours will be considered a deviation.

Given that the *MAE* reflects the variability in an individual's routine, choosing the same z value across all individuals allows us to account for that variability. In the example above, an *MAE* of 1 hour would lead us to only consider sleep of less than 7 hours or more than 9 hours to be a deviation.

The deviations of an individual across all behaviour markers in a particular day can be quantified in terms of a deviation *cost*. In this, deviations in each behaviour marker may have a different cost, for instance in terms of its potential impact to the person's health. For example, a deviation from the expected sleep duration is likely considered more impactful to health compared to a deviation in the time one chooses to read a book.

For this purpose, we set a customized weight for each behaviour marker to adjust the impact of specific behaviours. For the detected deviations above, we then calculate both the absolute total deviation cost and the weighted deviation cost. The total deviation cost is the sum of the cost of each deviation behaviour marker. The weighted deviation cost is shown in (4), where i represents each behaviour marker, z_i is the z value from (3) for each behaviour marker and w_i is the selected custom weight of the marker. The combination of total deviation cost and weighted deviation cost provides a quantification of deviations for a single person on an individual day. This has the potential to be used as a behaviour performance score for care givers when monitoring the daily living routine of an older adult at a glance.

$$C = \sum_{i}^{n} z_i * w_i \tag{4}$$

3.4.3 Long-term Deterioration

In addition to short-term deviations over a single day, we also analyse the long-term trend of behaviour markers. This type of analysis can help us detect long-term changes such as reduced mobility (some activities taking longer), reduced or disrupted sleep patterns, etc. that may be linked to health outcomes, such as deterioration, physical and cognitive frailty, and reduced independence.

We calculate long-term trends of behavioural markers based on the individual's normal daily routine. In this context, we consider days with detected deviations through the previous analysis as *abnormal* and filter them out of the dataset.

3.5 Investigating Relationship between Activities

The last step in our current work is aimed towards insights on why the detected deviations occur, focusing on individual behaviour markers. Taking sleep duration as an example, we can use the approach described so far to classify the sleep duration of each day as *abnormal* (deviating) or *normal* based on the individual's historical data. Taking this labelling as ground truth, we train a new classifier to use the other behaviour markers as features for normal/abnormal classification. We then calculate the *feature importance* of each marker using the feature selection technique in Scikit-learn (Pedregosa et al., 2011). This technique can assign a score to each feature of the classifier based on their impact in predicting the target label. The higher the score, the more important the corresponding feature is. Features with high importance are then considered to have a higher correlation with sleep duration.

4 EXPERIMENTAL RESULTS

In this section, we present the results we obtained in each stage of our approach using the CASAS dataset. We develop and evaluate our predictive model of daily routines by comparing the performance of different learning models. Additionally, we show the results of detected deviations and relationships between sleep duration and other daily behaviours. Finally, we develop an interactive dashboard for visualizing personal temporal profiles of daily living.

4.1 Predictive Model

In our effort to develop a predictive model for individual ADL routines, we performed an array of experiments to select the best performing algorithm.

First, we evaluated 19 well known ML regression models, including Random Forest Regressor, Linear Regressor, Logistic Regressor, Extreme Gradient Boosting, etc. As mentioned earlier, the features used for regression only contain time-related characteristics, i.e. month, day, weekday and chronological order. Then the top 3 best performing among them are integrated as an ensemble using Stacking and Blending techniques (Maclin and Opitz, 1999) to further improve the performance. We also evaluate one statistical time series approach named *Prophet* (Taylor and Letham, 2017), which is specifically tailored to deal with seasonal effects in time-series data, and compare its performance to the ML models.

We randomly chose *hh102* as an example to present our results. The other participants show similarly interesting results. The performance of different models of sleep duration in terms of the two metrics (*MAE* and *RMSE*) is shown in Table 3. The units of these two metrics are both seconds.

Table 3: *MAE* and *RMSE* of different predictive models of sleep duration for individual *hh102*.

Model	MAE	RMSE
Prophet	3683	4532
RandomForestRegressor	4200	5065
KNeighborsRegressor	4306	5394
ExtraTreesRegressor	4552	5436
StackingRegressor (top 3)	5943	6551
BlendingRegressor (top 3)	4214	5127

As the result shows, Prophet is the best model for forecasting sleep duration. In fact, Prophet outperformed the ML algorithms in all behaviour markers across our dataset, so we selected that algorithm for all our predictive models. This came with added benefits of Prophet, such as the calculation of long-term trends (see Section 4.3).

The model performance on the behaviour markers of the individual *hh102* is shown in Table 4. As an example, the mean absolute error in the prediction of the time when breakfast was cooked is 3745 seconds, so approximately one hour, which is a reasonable level of variability for that activity. Similarly, all of the obtained results show a reasonable level of accuracy, given the high variability in people's daily lives.

4.2 Detected Deviations

We present the results of our deviation detection approach using *hh102* as an example. In this, we set the threshold of the z value of each behaviour marker to 1. We set the weight cost of sleep duration, bathe duration and leave home duration, which we considered more important in our particular use case, to 2, and the weights of the rest of the behaviour markers to 1.

Table 4: Performance measures of the models on the behaviour markers of *hh102*.

Behaviour markers	MAE	RMSE
cook_breakfast_time	3745.318	3938.721
eat_breakfast_time	4185.226	4611.391
cook_lunch_time	4232.662	5127.29
eat_lunch_time	6120.533	6651.227
cook_dinner_time	1263.975	1580.311
eat_dinner_time	1687.354	2115.673
sleep_time	2726.95	3160.01
sleep_duration	3875.959	4543.112
take_medicine_time	4615.162	4933.677
morning_medicine_time	2250.793	2825.039
bathe_duration	187.1277	274.4289
leave_home_duration	3415.511	4944.483

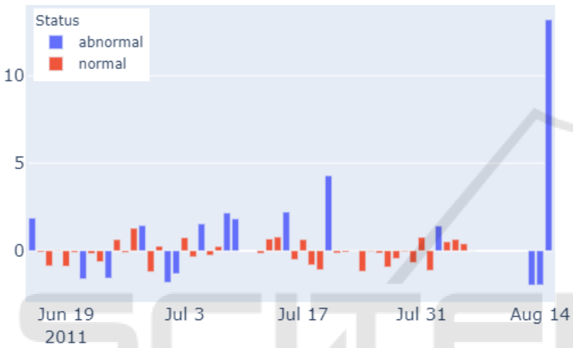


Figure 2: Detected deviations of sleep duration for *hh102*.

Figure 2 shows the calculated z values for sleep duration on different dates. Negative values mean that the actual sleep duration was longer than the predicted value, while positive values mean the expected sleep duration was longer than the actual value. We detect deviations on 15 days, which we label *abnormal*.

We also visualize the actual values and predicted values of behaviour markers related to time of the day and their MAE range in a timetable in Figure 3. Red marks mean the actual value and the yellow marks mean the predicted value with a line showing the corresponding fault tolerance range based on the threshold of the z value. We can readily figure out which behaviour markers are deviations and the magnitude of each one. For example, we see that participant *hh102* took their morning medicine earlier than the predicted time on this particular day.

We summarize the computed personal temporal profile of the individual person, including their daily routines, long-term trends and detected deviations in an interactive dashboard. Figure 3 shows one of the views of our dashboard that visualizes the absolute deviation cost, the weighted deviation cost, the detected deviations, and the predicted ranges (in yellow) and actual values (in red) of behavioural markers related to the time of the day for a particular, selected date. Other views of our dashboard show a timeline of the detected day routine and the long-term trends of the behavioural markers, but due to space limitations we do not present these here.

4.3 Long-term Deviations

Based on the 2 month data of the individual *hh102*, we analyse the sleep duration trend to detect whether it is deteriorating. The general trend of sleep duration for *hh102* is shown in Figure 4, where the horizontal axis and the vertical axis represent the observed date period and the sleep duration in seconds respectively. The black dots represent the actual sleep duration of that day, while the blue shaded area shows the prediction range (lower bound and upper bound) provided by Prophet. The predicted trend shows the sleep duration follows a relatively steady pattern over time.

We further explore the long-term trend by filtering out deviations that may be skewing the trend of the normal routine. Firstly, we classify the sleep duration as abnormal or normal based on the detected deviations. Secondly, we filter out the dates with abnormal sleep duration and investigate the normal pattern over an extended period of time. More specifically, we calculated the overall linear trend and weekly seasonality (the relative effect of each day of the week to the predicted value) of our predictive model, the results are shown in Figure 5. The normal linear trend (Figure 5a) shows the sleep duration of *hh102* is increasing during this period. The weekly seasonality shows the person gets the least sleep time on Wednesday.

We have also included the trend calculated using all of the data, without filtering deviations (Figure 5a). The 2 trends are visibly different, which demonstrates how deviating behaviour can significantly affect the observed trend. In the context of analysing long term sleep routines with the aim of exploring implications to health, abnormal days become outliers that skew the results. Instead, we choose to filter those out and focus on the "normal" or routine behaviour and how it evolves through time.

4.4 Relationships between Activities

We take sleep duration as an example to investigate the relationship with other activities. Through temporal deviation detection, the sleep duration of each day can be labelled as abnormal and normal. We combine all the labelled sleep duration data of 11 individuals to train a classification model. The goal is to classify sleep duration as normal or abnormal using the other behaviour markers as features.

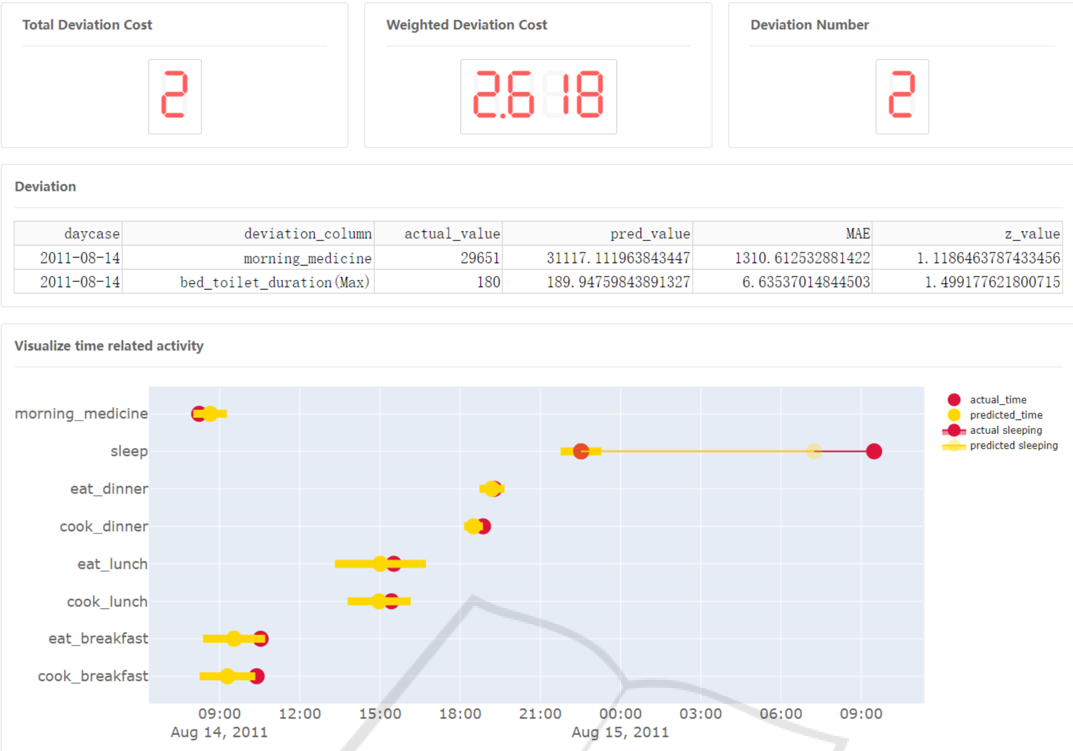


Figure 3: The dashboard for visualizing the temporal profile of each individual.

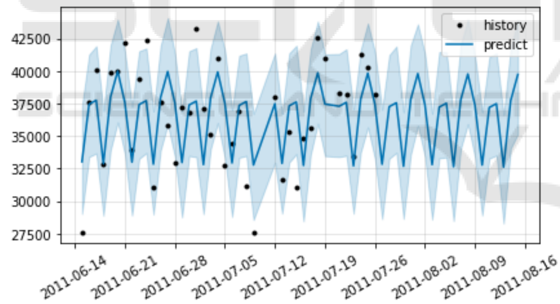
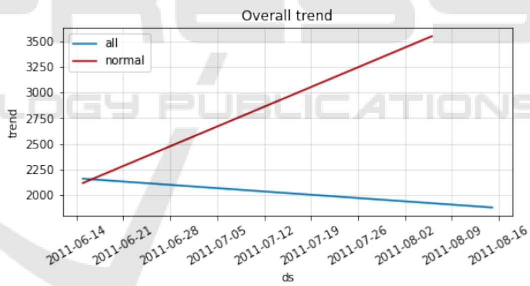


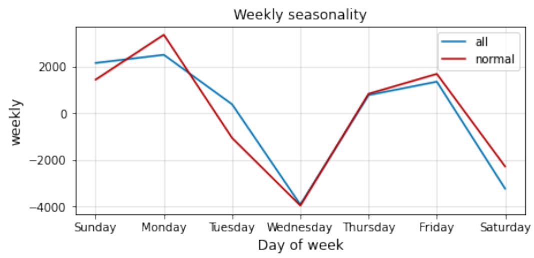
Figure 4: The predicted trend of sleep duration for *hh102* using data from 15 June to 26 July 2011.

To this end, we first train separate predictive models of sleep duration for each person in the dataset. The performance of each model is shown in Table 5. The results show that the error of each predictive model is less than 1 hour, which we consider a reasonable error range for a sleep duration prediction.

Next, we experiment with different z values between 1 and 2 to choose the best classification method and investigate how different z values impact the accuracy. We train separate classification models based on the results of deviations calculated by different z values. A 10-fold cross validation is used to calculate the performance. We choose accuracy, Auc, recall, precision and F1 score of the cross-validated results as



(a) The overall trend of sleep duration for *hh102*.



(b) The weekly seasonality of sleep duration for *hh102*.

Figure 5: Long term pattern of sleep duration for *hh102*.

our metrics to evaluate the classification model. The Auc metric can make a reasonable evaluation of a binary classification problem on imbalanced datasets, i.e. datasets where one class (normal) has a much

Table 5: Individual performance of sleep duration models.

Participant	MAE	RMSE
hh101	3959.30	4863.67
hh102	2802.01	3152.73
hh103	1177.24	1443.22
hh104	2974.59	3491.83
hh105	2013.01	2181.18
hh106	1218.23	1442.67
hh107	1939.70	1956.81
hh108	2671.29	2962.98
hh109	1537.11	2286.03
hh110	559.72	804.56
hh111	1533.05	1784.17

larger population than the other (abnormal).

The results show that the Random Forest Classifier performs best, so we use it as the main model. The comparison of performance for different z values is shown in Table 6. It shows that the higher z value we choose, the better classification performance we get. The higher the performance of the model means the features are more predictable and the results on feature importance are more reliable. We get the highest Auc when the z value is set to 1.8, which means the classification performs best on this imbalanced case. Therefore, we investigate the relationships between sleep duration and other activities for a z value of 1.8.

The calculated (impurity-based) feature importance of the Random Forest Classification model is shown in Figure 6. These particular results show that the total duration and the last time of personal hygiene activities have a strong influence on sleep duration. Also, the morning wake up time has a high correlation with sleep duration, meaning that the time one wakes up in the morning can significantly affect their sleep duration compared to their usual routine at night. Surprisingly, nap duration during the day has a weaker correlation with the nightly sleep duration.

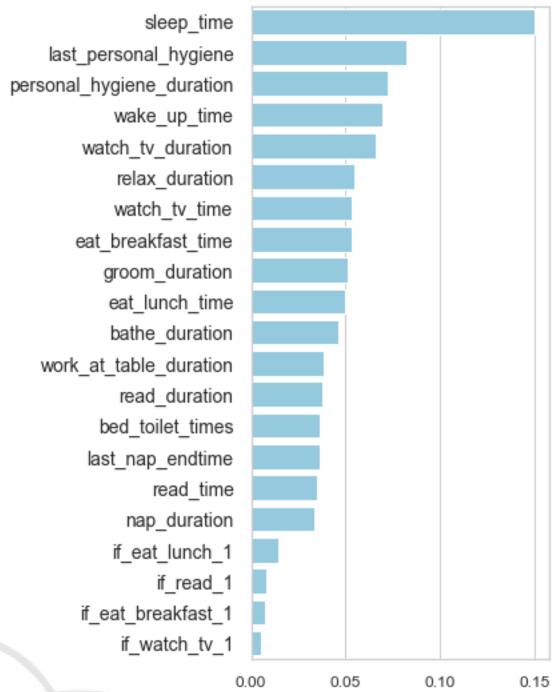


Figure 6: Feature importance of the Random Forest Classifier.

5 CONCLUSIONS

In this paper, a predictive behaviour monitoring approach is proposed for ADLs coupled with a methodology to detect short-term deviations and long-term trends. Moreover, we investigate the causal factors of deviations by exploring the relationship between different daily activities in terms of predictive power. Experiments are conducted on a sample of 11 individuals' ADLs data that is publicly available.

Due to the dynamic nature of human behaviour and the variability between different people, we set an adjustable deviation threshold for each behaviour marker such that the model is tailored to individual human activities. We can also modify the individual weights of each behaviour marker in order to obtain an aggregate score that reflects the importance of each deviation, for instance in terms of health outcomes. Finally, we develop a dashboard which visualizes the various information of the human behaviour, such as daily routine, potential deviations, trend of sleep duration, etc. The dashboard provides both the person involved and their caregivers with key behavioural insights.

The work presented in this paper is indicative of the useful insights that sensor data on ADLs can provide towards health monitoring of older adults. We believe that such insights have the potential to enable an unprecedented capacity for self monitoring and management as well as inform proactive interventions far in advance of any adverse health events.

However, further challenges still exist in dealing with the inherent noise in the data and the variability of people's routines. The models used are naturally sensitive to noise. We therefore believe that future improvements in ADL data collection, especially in the 4 types of noise we identified (see Section 3.1), will bring forth significantly better and more accurate insights. Moreover, results need to be further contextualised to the needs, lifestyles, and medical conditions of the individual participants. Further comparison experiments with literature could get more convincing insights for providing proactive interventions

Table 6: The performance of classification model with different z value.

z value	Normal cases	Abnormal cases	Accuracy	Auc	Recall	Precision	F1
1	268	226	0.5478	0.5892	0.5914	0.5815	0.5858
1.3	330	164	0.6812	0.6145	0.8707	0.7067	0.7799
1.5	359	135	0.7304	0.6705	0.9145	0.7565	0.8279
1.8	390	104	0.8058	0.7001	0.9814	0.8094	0.8871
2.0	405	89	0.8348	0.6703	0.9857	0.8395	0.9067
2.3	422	72	0.8580	0.6709	0.9966	0.8581	0.9222

and making decisions. Data collection and analysis beyond the limited dataset we have explored so far is likely to improve the quality of our algorithms and lead to new types of insights particularly for long-term predictions.

REFERENCES

Akram, U., Gardani, M., Riemann, D., Akram, A., Allen, S. F., Lazuras, L., and Johann, A. F. (2020). Dysfunctional sleep-related cognition and anxiety mediate the relationship between multidimensional perfectionism and insomnia symptoms. *Cognitive Processing*, 21(1):141–148.

Albanese, A. M., Bartz-Overman, C., Parikh, MD, T., and Thielke, S. M. (2020). Associations Between Activities of Daily Living Independence and Mental Health Status Among Medicare Managed Care Patients. *Journal of the American Geriatrics Society*, 68(6):1301–1306.

Alberdi Aramendi, A., Weakley, A., Aztiria Goenaga, A., Schmitter-Edgecombe, M., and Cook, D. J. (2018). Automatic assessment of functional health decline in older adults based on smart home data. *Journal of Biomedical Informatics*, 81:119–130.

Aminikhanghahi, S., Wang, T., and Cook, D. J. (2019). Real-Time Change Point Detection with Application to Smart Home Time Series Data. *IEEE Transactions on Knowledge and Data Engineering*, 31(5):1010–1023.

Anderson, J. L., Rosen, L. N., Mendelson, W. B., Jacobsen, F. M., Skwerer, R. G., Joseph-Vanderpool, J. R., Duncan, C. C., Wehr, T. A., and Rosenthal, N. E. (1994). Sleep in fall/winter seasonal affective disorder: Effects of light and changing seasons. *Journal of Psychosomatic Research*, 38(4):323–337.

Arifoglu, D. and Bouchachia, A. (2019). Detection of abnormal behaviour for dementia sufferers using Convolutional Neural Networks. *Artificial Intelligence in Medicine*, 94:88–95.

Bergmeir, C. and Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191:192–213.

Cohen-Mansfield, J. and Perach, R. (2012). Sleep Duration, Nap Habits, and Mortality in Older Persons. *Sleep*, 35(7):1003–1009.

Cook, D. J., Crandall, A. S., Thomas, B. L., and Krishnan, N. C. (2013a). CASAS: A Smart Home in a Box. *Computer*, 46(7):62–69.

Cook, D. J., Krishnan, N. C., and Rashidi, P. (2013b). Activity Discovery and Activity Recognition: A New Partnership. *IEEE Transactions on Cybernetics*, 43(3):820–828.

Cook, D. J. and Schmitter-Edgecombe, M. (2021). Fusing Ambient and Mobile Sensor Features Into a Behaviorome for Predicting Clinical Health Scores. *IEEE Access*, 9:65033–65043.

Dawadi, P. N., Cook, D. J., and Schmitter-Edgecombe, M. (2016). Automated Cognitive Health Assessment From Smart Home-Based Behavior Data. *IEEE Journal of Biomedical and Health Informatics*, 20(4):1188–1194.

Howedi, A., Lotfi, A., and Pourabdollah, A. (2020). An Entropy-Based Approach for Anomaly Detection in Activities of Daily Living in the Presence of a Visitor. *Entropy*, 22(8):845.

Ilyas, I. F. and Chu, X. (2019). *Data Cleaning*. Association for Computing Machinery, New York, NY, USA.

Kanti Majumdar, K. (2014). Relationship of Activity of Daily Living with Quality of Life. *British Medical Bulletin*, BBB[2][4][2014]:757–764.

Maclin, R. and Opitz, D. (1999). Popular Ensemble Methods: An Empirical Study. *Journal of Artificial Intelligence Research*, 11:169–198.

Palmer, M. G. (2018). Experimenter presence in human behavior analytic laboratory studies: Confound it? *Behavior Analysis: Research and Practice*, 19(4):303.

Pazhoumand-Dar, H., Armstrong, L. J., and Tripathy, A. K. (2020). Detecting deviations from activities of daily living routines using kinect depth maps and power consumption data. *Journal of Ambient Intelligence and Humanized Computing*, 11(4):1727–1747.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830.

Pierleoni, P., Belli, A., Palma, L., Pellegrini, M., Pernini, L., and Valenti, S. (2015). A High Reliability Wearable Device for Elderly Fall Detection. *IEEE Sensors Journal*, 15(8):4544–4553.

Piovezan, R. D., Poyares, D., and Tufik, S. (2013). Frailty and sleep disturbances in the elderly: possible connections and clinical implications. *Sleep Science*, 6(4):175–179.

Sepesy Maučec, M. and Donaj, G. (2021). Discovering Daily Activity Patterns from Sensor Data Sequences and Activity Sequences. *Sensors*, 21(20):6920.

Taylor, S. J. and Letham, B. (2017). Forecasting at scale. Technical Report e3190v2, PeerJ Inc.

University of Edinburgh (2021). Advanced Care Research Centre. https://edin.care. Accessed on Dec 12, 2021.

Yahaya, S. W., Lotfi, A., and Mahmud, M. (2019). A Consensus Novelty Detection Ensemble Approach for Anomaly Detection in Activities of Daily Living. *Applied Soft Computing*, 83:105613.

Yahaya, S. W., Lotfi, A., and Mahmud, M. (2021a). Detecting Anomaly and Its Sources in Activities of Daily Living. *SN Computer Science*, 2(1):14.

Yahaya, S. W., Lotfi, A., and Mahmud, M. (2021b). Towards a data-driven adaptive anomaly detection system for human activity. *Pattern Recognition Letters*, 145:200–207.