Personal Health Assistance for Elderly People via Smartwatch Based Motion Analysis

Rainer Lutze
Dr.-Ing. Rainer Lutze Consulting
Langenzenn, Germany
Rainer.Lutze@lustcon.eu

Klemens Waldhör
FOM University of Applied Sciences
Essen / Nuremberg, Germany
Klemens.Waldhoer@fom.de

***Abstract* — A new approach is presented for a personal health assistant for elderly people utilizing smartwatches. On the smartwatch, an app featuring an artificial neuronal net (ANN) analyzes the motion patterns of the smartwatch wearer. The ANN recognizes health relevant events and activities of daily living (EDLs, ADL). The system architecture of the app, the data acquisition process, the selection and design of suitable data models and the advantages of ANNs versus other recognition engines are elaborated. The characteristics of the recognized ADLs will be utilized for continuously calculating the wellbeing of the smartwatch wearer, safeguarding a self-determined living in the familiar home up to the very old age.**

Keywords* — *wearables, smartwatches, ambient assisted living (AAL), assistance systems, automatic recognition of activities, events of daily living (ADLs, EDLs), artificial neuronal nets (ANN), prediction model, universal model, individual model,

I. Introduction

A self-determined and safe living of elderly people in their familiar home as long as possible is a desirable objective for many of us. Ambient intelligent assistance technologies safeguard such a life by regularly monitoring the wellbeing and potential health hazards. Programmable smartwatches are one of the most promising devices for such health assistance technologies, because i) they carry many of the necessary sensors for monitoring wellbeing and health parameters on board, ii) do not require expensive demolition / construction work at home and iii) can be used at home as well as outdoors. Moreover, they are available at reasonable costs. In our work, we focus on mainstream smartwatches with an integrated mobile cellular radio (like the Samsung Gear™ 3G, LG Urbane LTE™ 2 or Sport™, Huawei Watch 2™). These smartwatches allow to establish autonomously a speech connection in order to clarify the situation on the spot in case of a concluded emergency [1]. Moreover, relevant data (e.g. current geographic position of the smartwatch wearer, the heart rate) can be transferred directly and without the (necessary) additional utilization of a smartphone (as it is the case for the Apple Watch 2™).

Current smartwatches directly can only measure the performed steps of the smartwatch wearer and/or the heart rate, pulse. All other aspects of the wellbeing and potential health hazards for the smartwatch wearer must be concluded form condensed sensor data and suitable comparisons with data acquired, learned from the past. A common approach is to recognize - by the smartwatch sensors - those *activities of daily living* (ADLs) which are present in a healthy life of everyone and structure the days and nights. The conclusions about the wellbeing will then be based on the *presence, duration* and *intervals* between those recognized activities of daily living. In addition, direct health hazards – like tumbles/falls, heart palpitations – have to be additionally considered and recognized by the smartwatch sensors (as events of daily living, EDLs).

II. Previous Work

Activities of daily living (ADLs) (see [2]) have been a central issue in organizing professional nursing practice and for determining the independency status of elderly people, introduced by Sidney Katz more than 60 years ago.

Suryadevara and Mukhopadhyvay ([3]) have proposed a wellbeing function w based on the components *absence* and *excess duration of ADLs.* Their approach stems on a rich instrumentation of the household by a net of wireless sensors. The wellbeing function w maps the recognized ADLs and their characteristics into $[0,1] \subset \Re$. For ideal wellbeing, the function value of w is 1; if the function value falls below a defined threshold (e.g. 0.5), a health alert is issued. In [4], we have extended the w definition for accounting a third independent wellbeing component *agility*, which measures the typically step distribution walked over the course of the day by the smartwatch wearer. The nominal values for a typical interval between the ADLs, the typical execution time of a specific ADL and the typical step sum achieved by a specific hour of the day all are individual values and specific for a certain day of the week. These nominal values have to be acquired by an initial training phase of the system with at least a one-week duration and will be further adapted by time series analysis (see [3], [4]), taking into account also seasonal factors.

When no recognized ADL in the household is taking place, $ß_1$, the wellbeing sub-function for *inactivity* measurement, is applied based on the definition in [3] as $ß_1(t, T) = e^{-t/2*T}$, where t is the current (time) duration of inactivity since completion of

978-1-5090-4881-6/17 $31.00 © 2017 IEEE
DOI 10.1109/ICHI.2017.79

the last recognized ADL, and T is the specific average inactivity between ADLs learned from the past for the current day of the week. On the opposite, as long as a recognized ADL is ongoing, $ß_2$, the wellbeing sub-function for the measurement of *excess duration* of this specific ADLs will be applied, which has been defined in [3] to $ß_2(TN, ta) = e^{(TN - ta) / TN}$, for ta > TN; 1, otherwise where ta is the actual duration of the (ongoing) ADL and TN is the specific maximum duration of the corresponding recognized ADL in a normal situation learned from the past for the current day of the week. The agility subfunction $ß_3$ measuring the movement profile of a person at current time t is defined in [4] to $ß_3$(t, stp, STP) =

$$\begin{array}{ll} e^{(stp(t) - STP(t)) / STP(t)}, & \text{for } stp(t) < STP(t) \text{ and NOT } (E_1) \\ 1, & \text{for } stp(t) \geq STP(t) \text{ and NOT } (E_1) \\ 0, & \text{if event } E_1 \text{ (= fall) as been detected} \end{array}$$

where stp(t) is the sum of steps performed during the current day until actual time t, F(t) is the cumulative distribution function of steps over the day, SN is the specific total number of steps learned from the past for the current day of the week and with STP(t) = SN*F(t) estimated from the nominal step sum for the current day at time t. $ß_3$ will be calculated all over the day.

The wellbeing function w: SensoricEvents -> [0,1], $[0,1] \subset \Re$, will be formally defined as:

$$w = \min \{ß_1, ß_2, ß_3\}$$

This means that whenever the *inactivity* (missing any recognized ADL) or the *excess duration* of an ongoing activity category or the lacking *agility* gets critical and the w value falls below 0.5, a *health hazard alert* will be issued.

In our context, we focus on the subset of basic ADLs, which

- can be recognized by the usual integrated smartwatch sensors (*3D accelerometers, and gyros, magnetometer, barometer, heart rate monitor / pulsometer, GPS for the smartwatch class chosen)* and communication technologies *(Bluetooth, Wi-Fi, 3G/4G cellular)*. In contrast to usual fitness apps, we are interested in more complex activities, which can be concluded from a motions analysis than walking or cycling. In the focal point of our interest are ADLs like »combing*«, »liquid ingestion / drinking«,* »*(midday) nap, rest«,* ... see section III.B for a discussion of those ADLs.
- will be typically carried out each single day and by everyone, independent from culture and/or sex, ideally independent from a dominant hand (on the wrist of which the smartwatch has to be worn),
- will be carried out multiple times within a day and thus allow for a preferably equidistant partitioning and structuring of the day / night.

With respect to their eminent negative health consequence for the targeted user group, additionally the EDL »*tumbling«* needs to be considered. It is common knowledge that one third of all elderly persons of age of 65 or more fall at least one time per year.

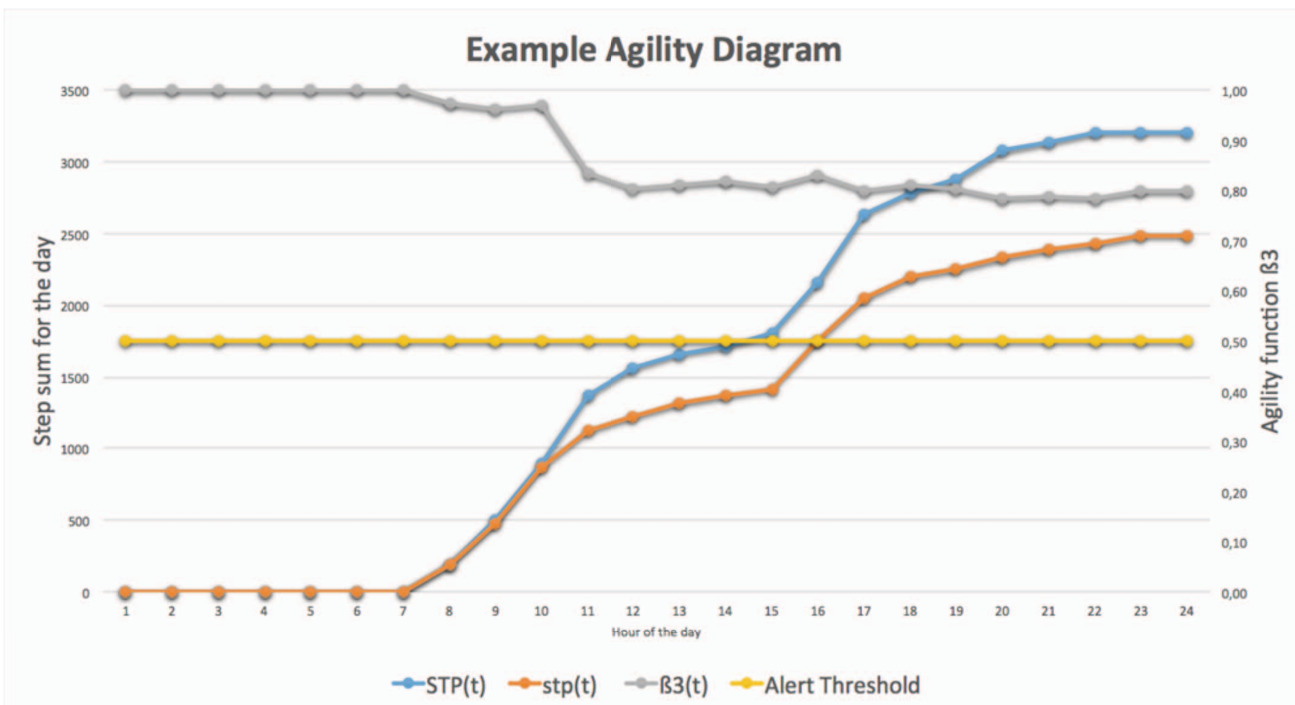


Fig. 1: Typical agility for a 24-hour period based on an $\alpha = 0.1$ (giving heavy weights for historical values). As can been seen the agility value is far above the threshold indicating that no agility problems are present. Left scale denotes the accumulated steps (orange: actual steps of the day, blue: the estimated accumulated steps for the period, grey: the agility value $ß_3$ and yellow: the threshold for $ß_3$).

III. System Design Considerations

A. System Structure

The implemented system, smartwatch app, utilizes the Samsung Gear™ S smartwatch device for providing assistance in the four dimensions: I) *communication* (manually and automatically established speech connections to family members on duty or a home emergency call center), II) *orientation*, III) *localization* and IV) *health hazard detection.* The implemented scope of personal health assistance is described in [1], [4] in closer details, Fig. 2 shows some screenshots. Fig. 14 depicts a block diagram of the app architecture, more architectural details can be found in [5], [18].

Fig. 2: Smartwatch Samsung Gear S with app displaying *communication* and *orientation* information, internal pre-alert, external alert

The application architecture is based on a hierarchical structure. On the lower layer the EDL, ADL recognition via a ANN takes place (see section III). The recognized EDLs, ADLs will evoke actions in a structured description of the *health hazard handling* process executed on the upper layer. Health hazard handling is described in a declarative way via UML (finite) state machines. This declarative description is well suited for maintaining and updating the volatile, best practice *health hazard handling* knowledge [5]. The decisive advantages of medical process modelling via UML is summarized in [20]: *"Process modeling gives the methodological basis for quality improvement in medicine because it is able to break out complexity and provide transparency to medical processes."*

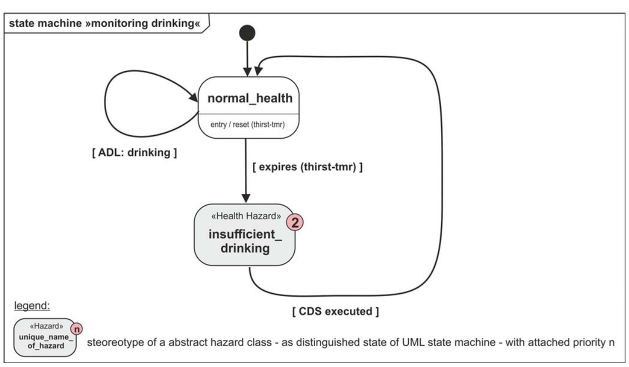


Fig. 3: Simple state machine for concluding about insufficient liquid ingestion

Example: In Fig. 3, the health hazard handler »monitoring drinking« consists of two states. Upon (re)entering the initial state »normal_health«, the timer *thirst-timer* is reset. Thus, the state machine stays in this specific state, as long as the ADL »*liquid ingestion«* occurs in sufficient frequency. If the *thirst-timer* expires because the time period since last recognized drinking ADL is exceeded, the state machine transfers to the new state »insufficient_drinking«. This new state is of special type *critical dialogue section* and will be posted on the central blackboard (see below). Such a posting indicates an execution request with attached priority "2" for the associated internal dialogue activity flow for this state. The intended dialogue sequence with the smartwatch wearer is described in a corresponding UML activity diagram, see Fig. 15 for the principal schema of the dialogue. This schema systematically covers all necessary dialogue steps for: a) *informing* the smartwatch wearer about the specific situation ("pre-alert"), b) *requesting a decision* from the smartwatch wearer, c) *responding* with the dialogue in case of an affirmative or rejecting user reaction as well as a non-reaction of the user. In addition, the potential data transmission of relevant health data, which takes place in the background of the dialogue will be covered by the schema as well as the follow-up clarification call ("external alert").

The implemented app gains its complexity by the fact that a multitude of potential health hazards have to be monitored by the smartwatch app simultaneously. Single health hazards are typically modelled and described by a separate handler, in order to alleviate an independent representation and maintenance of the incorporated pragmatic handling knowledge. But, there are also joint hazard handlers for contextually combined health hazards (see Fig. [16]). Thus, the situation may occur that more than one health hazard handler wants to communicate to the smartwatch wearer rsp. a distant family member / home emergency call center at the same time. This situation will be dealt with a priority – or exactly: severity – based health hazard communication management. The I/O devices of the smartwatch (touchscreen, bezel, mic, buzzer / loudspeaker) may be attached to at most one health hazard handler at time and for a short time interval. To realize this, we have introduced the concept of a *critical dialogue section* (cf. [18]). As soon as the smartwatch I/O resources have been granted to a selected health hazard handler, the handler will use them *exclusively* until termination of the execution of the corresponding critical dialogue section. The selection of the suitable health hazard handler for executing a dialogue sequence with the smartwatch wearer is supported via a central blackboard, on which all current communication requests are posted by the different handlers. From all current requests on the blackboard, a central scheduler selects the most appropriate health hazard handler for execution based on the medical or situational severity of the posted request and all other present requests on the blackboard (see [5], [18] for details of the scheduling algorithm).

B. Determining a Suitable Set of ADLs

A tradeoff has to be made between the plenitude of ADLs, which shall be recognized, and the reliability of the recognition results. The more ADLs the system knows and is looking for, the more fine-grained the course of the day can be partitioned into different ADLs and periods of time in between. The shorter these periods of time are, until the next ADL will typically occur, the earlier a deviant behavior influencing wellbeing and / or indicating potential health hazards can be detected. But, the more ADLs need to be discriminated by the recognition engine, the less reliable the recognition result will typically be.

Based on the aforementioned criteria, we have decided to recognize the following eight ADLs:

1. Nightly sleep
2. (midday) nap, rest
3. absence form home (for social activities/visits, strolling, shopping, …)
4. liquid ingestion, drinking (see [6], [7], [15] for details)
5. hand washing / drying (typically carried out after toilet activity, before eating)
6. teeth brushing
7. shaving
8. combing

ADL no. 1 »*nightly sleep« can* only be observed indirectly, in that the smartwatch is typically not worn during this period due to nightly battery recharging and usual sleeping habits. But, placing down the smartwatch when retiring to bed at night and reattaching the watch in the morning after rising can be reliably detected via movement analysis and the heart rate sensor, pulsometer.

Assuming that the smartwatch will be worn all over the day, ADL no. 2 »*(midday) nap, rest«* can be directly observed and easily detected by the smartwatch app via its characteristic non-movement pattern.

Also, ADL no. 3 »*absence from home*« can be directly followed by the smartwatch app via loss of the known home Wi-Fi signal and GPS. GPS will be further used for tracking and geofencing outdoor activities (see [1], [4] for details).

For the recognition of ADLs no. 4 to 8, these ADLs can only be discriminated from each other by their characteristic movement pattern. This holds also for the recognition of falls. We have decided to do this recognition process via data mining and artificial neuronal nets (ANNs). Input layer of the feed forward ANN are the (condensed) specific signals from the smartwatch sensors. The ANN has one hidden layer and each ADL no. 4 to 8 will be represented by a specific output neuron, with the EDLs *»tumbling«, »heart palpitations«* as additional 6^{th} and 7^{th} output neuron of the ANN and a 8^{th} output neuron for any other unclassified activity. ANNs have been chosen with respect to their favorable recognition quality and renunciation of additional runtime packages in comparison to logistic regression and other tested methods (see [7] for details). Another strength of ANN is their suitability for incremental training with the *backward propagation* algorithm (cf. [8], chap. 5.2).

ANN have successfully been used for detecting falls with quite a high precession, [11] reports an accuracy of 91%. [12] gives an overview about various sensor based implementations, most of them with an accuracy rate about 90% using various data mining techniques like multilayer perceptrons, support vector machines or naïve Bayes approaches. A good discussion about the challenges of fall detection is given in [13], focusing not only a wearable system but also on camera bases approaches. It is important to note at least in Europe any detection techniques based on video is not accepted because of privacy concerns. Thus, only foot mat related sensor technologies which require expensive hardware investments remain as an option or any kind of wearable sensor. [14] shows that Convolutional Neuronal Network (CNN) perform best for supervised learning techniques, while overall the differences to other approaches like SVM are not very high. Our work differs from those as we aim for detecting several different ADLs in one model, and not just the binary decision between fall and not not-fall. [19] compares smartwatch based ADL detection with smartphone based detection showing that smartwatches can detect a wider variety of ADLs. Smartwatches gain their strength in fall detection in that they are reliable worn at the wrist and will be on duty during the whole course of a day. In contrast, smartphones are typically put aside from time to time, especially during accident susceptible activities like showering.

C. Handedness and Relevance of ADLs.

ADL no. 5 »*hand washing / drying«* and the EDL *»tumbling«* are typically independent from the arm resp. wrist, on which the smartwatch will be worn. *ADL* no. 4 »*drinking« and* no. 6 »*teeth brushing«* will be typically carried out only with the dominant hand. It turned out for the test persons that it is not a problem to wear their smartwatches on the wrist of their dominant hand (see chapter IV below). This is alleviated by the fact that smartwatches can rotate their display so that sideward control elements always remain at the familiar location pointing towards the hand of the wearer.

ADL no. 4 »*drinking«* has been selected with respect to the dangerous effects of dehydration for elderly people caused by the decreasing natural thirst sensation at higher age ([6]).

ADL no. 5 »*hand washing / drying«* have been primarily included for technical reasons because they are typically executed several times a day and provide a good partitioning of the day into shorter time spans between those ADLs.

The importance of ADLs no. 5 »*hand washing / drying«* and 6 »*teeth brushing«* is not only given by the fact, that they have a characteristic movement pattern which makes it suited for automatic activity recognition, but also for their social relevance. Regular teeth brushing and hand washing are significant symptom for a well-managed life. Stopping these activities typically indicate a loss of self-esteem / self-control and might be symptoms of progressing mental disorientation or dementia ([9]).

IV. Research Questions

A central question is whether a universal, person independent model of the ADL / EDL recognition process is sufficient or if an individually trained model will be necessary, at least for personal activities like *teeth brushing?* The additional effort for processing and building an individual model will be counter-balanced by the prospect to utilize this individual model for an authentication of the smartwatch wearer.

From this central question, several follow-up research questions have been derived:

1. Which is the best prediction model? Candidates are neuronal networks, regression models or decision trees [10].
2. Is one universal model sufficient to recognize the relevant ADLs/EDLs based on a target recognition rate of at least 90%?
3. Are there differences in the acquired sensor data between the various smartwatch types (operating systems like Android Wear or Tizen)?
4. Are there differences based on the ADLs with regard to universal / individual model, thus while one ADL just requires a universal model, another ADL requires individual training?
5. How many training data have to be collected per person?
6. How many different persons are required to create a stable model?

In the analysis of this paper we concentrate on hypothesis 2 and neuronal net models. Hypothesis 1 has been tested in [16], results show that neuronal nets perform at least as good as logistic regression, while decision trees perform much worse. To verify hypotheses 3 to 6 the number of data currently available are not sufficient for a definite answer. First results show that at least 20 – 30 activity instances have to be collected per person for stable trainings models with a high recall and precision. Hypothesis 6 is partially answered by the results for hypothesis 2.

V. Experiment

1) In a first step, we developed a sensor data gathering app (Fig. 6) where we collected about 2375 different activity instances starting from the above-mentioned core ADLs and EDLs like drinking, brushing, tumbling, combing, shaving, washing etc. The distribution of the ADLs is shown in Fig. 4 and Fig. 5.

   The user was instructed to perform a specific activity (e.g. drinking or tumbling). The activity was selected by the user from the app (Fig. 6: 1) and the user entered an abbreviated user name. Next screen (Fig. 6: 2) was presented to the user. The user started the data collection by pressing "Start Collecting Data". A timer was started which gave both video, tactile and audio feedback when the user should start the activity (Fig. 6: 3). The app reads the sensor data for about 10 seconds (Fig. 6: 4, 5). pending on the activity another measuring cycle was started automatically (Fig. 6: 6). This was done e.g. for tooth brushing while for tumbling only one cycle was performed. Finally, the app stores the gathered data in a file (Fig. 7 and 8). The data were collected using several different smartwatches running Android Wear (Sony Smartwatch 3, Samsung Gear Live) and Tizen (Samsung Gear S).

   Experiments were run with different set of users in various ages. Not all users performed all activities.

   For the purpose of this analysis we have only used the data from the Gear S into account.

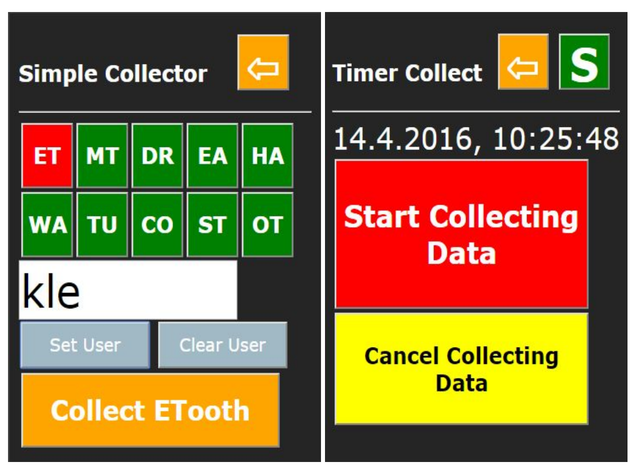
Activity / User	U1	U2	U3	U4	U5	Sum
Drink	25	0	108	10	3	**146**
Other	0	0	533	1	24	**558**
Tooth	59	40	261	0	100	**460**
Tumble	226	0	45	67	15	**353**
Sum	**310**	**40**	**947**	**78**	**142**	**1517**

Fig. 4: Distribution of frequencies of the collected 1517 ADLs used for the training of the neural net model. For this analysis, ADLs which did not fit into the category analyzed (like brushing, combing, washing, shaving etc.) were mapped into Other ADLs. Users are denoted by Un.

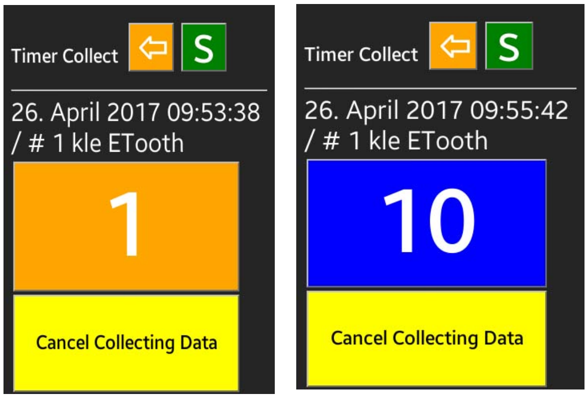
Activity / User	U6	U7	U3	U8	U9	U10	U11	Sum
Drink	2	0	14	0	0	2	3	**21**
Other	1	0	11	0	0	0	0	**12**
Tooth	0	0	5	0	5	0	0	**10**
Tumble	0	2	3	3	0	0	0	**8**
Sum	**3**	**2**	**33**	**3**	**5**	**2**	**3**	**51**

Fig. 5: Distribution of frequencies of the collected 51 test ADLs used for the testing the neural net model. For this analysis, ADLs which did not fit into the category analyzed (like brushing, combing, washing, shaving etc.) were mapped into Other ADLs. Users are denoted by Un. User U3 in Fig. 4 and 5 denote the same user while the others are different from Fig. 4.

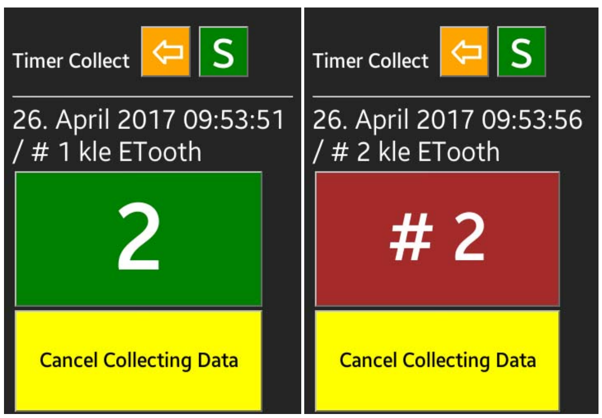
The next steps are based in a standard CRISP (Cross Industry Standard Process for Data Mining) process [17].



(1) (2)



(3) (4)



(5) (6)

Fig. 6: Samsung Gear S GUI for collecting ADLs. The user selects the ADL, enters a user name, pushes Collect (1) and from (2) he can actually start the data collection process. (3) is a pre-countdown start, informing the user that the data collection will start in 1 second. (4) indicates that data collection has started and will last for 10 seconds. (5) indicates data collection will be finished in two seconds. (6) informs the user that now a second data collection will start. This step (6) depends on the type of activity recorded. In case of a short activity like tumbling or drinking just one ten second interval is recorded at once, in case of tooth brushing as an example of a longer activity up to five ten second interval are recorded at once.

2) In the *second step* the gathered sensor data are normalized:

   a) All sensor data are standardized and interpolated into a fixed time interval (20 milliseconds). This was achieved by applying some filters, e.g. a high/low pass filter.

   b) A core set of statistical attributes (39 attributes like means, standard deviations, minimum, maximum, inter quartiles…) are computed for each ADL. Dependent variable is ADL type (Activity), independent variables are the 39 statistical attributes.

   c) For each ADL (experiment) a data record is written into a new csv summary file together with the information which type of ADL is performed and the user name. This resulted in several ADL summary files depending on the hypothesis (Fig. 8).

3) In a *third* the data were checked for missing values (e.g. sometimes the smartwatch did not collect gyrometer or magnetometer data for whatever reason). Those cases were ignored from the analysis.

4) In a *fourth step,* we applied several data mining techniques using R and Rapid Miner: multinomial logistic regression, clustering, decision trees and for the results presented in this paper neuronal networks using the normalized sensor data from Fig. 9. For the data mining process, we grouped the ADLs in two categories: a) drinking, teeth brushing, and tumbling and b) all other activities recorded like walking, running, washing, sitting etc. into a common ADL category called other.

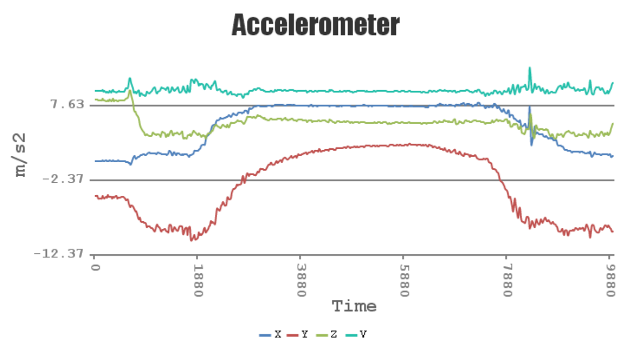


Fig. 7: Typical sensor measurements of a 10 second training period for accelerometer values

```
0;0;1460610085887;GST;true;3;-151;Wash;Clean;kle;true
1;1;1460610089138;GA;0.5407779216766357;-
5.319244384765625;-4.048655986785889;0.39050865173339844;-
3.066258192062378;-2.5658717155456543
2;2;1460610089138;GR;22.58479881286621;62.87120056152344;2
8.322559356689453 …
0;22;1420537162398;WG;0.058594197;0.07191106;-0.56729835
```

Fig. 8: Example raw data retrieved from the smartwatch. Each sensor in a line together with time stamp in milliseconds, sensor type and x, y, z axis sensor values, WA= accelerometer, WG = gyrometer, WM = magnetometer;

ACC.ZC_X	ACC.ZC_Y	ACC.ZC_Z	...	ACC.MEAN_Z	ACC.SD_Y	GYR.ZC_X	ACTIVITY
0	82	39	...	4.34083573	2.55596739	87	Comb
0	81	40	...	5.3805763	2.04349355	107	Comb
82	0	0	...	4.50360789	1.83203915	94	Drink
24	0	0	...	3.74075199	5.65520525	120	Drink
0	0	6	...	4.20948926	5.46076168	71	Eat
0	6	0	...	4.13021336	6.34172876	84	EShave
0	0	53	...	3.43561234	5.98092861	90	EShave
30	2	39	...	4.00581057	5.60997578	128	ETooth
0	6	6	...	3.59402789	6.50202609	91	EShave
24	28	77	...	3.80754762	6.10485199	71	Wash
41	26	79	...	4.03990255	4.79287513	58	Wash

Fig. 9: Example summary sensor data for various activities. One line represents one experiment. MEAN = mean of 10 second period, IQ= interquartile, SD = standard deviation etc.

VI. Results

A. *Is one universal model sufficient to recognize the relevant ADLs/EDLs based on a target recognition rate of at least 90%?*

For this hypothesis, we created four ADL classes: Drink, Tooth brushing, Tumble and Other.

The results of training the neural net model is shown in Fig. 10. It shows that all relevant recognition rates are above 90%. This is also supported by the cross validation runs (Fig. 11). This supports the argument that a general model with data trained from several persons is sufficient. Further analysis done (not shown here) show that models trained for a specific person perform better, esp. for those ADLs where the underlying movements are similar (e.g. distinguishing between tooth brushing and combing).

Training General Model	true Other	true Drink	true Tooth	true Tumble	class precision
pred. Other	541	2	12	7	96.26%
pred. Drink	2	134	0	7	93.71%
pred. Tooth	9	5	446	1	96.75%
pred. Tumble	6	5	2	338	96.30%
class recall	96.95%	91.78%	96.96%	95.75%	**1517**

Fig. 10: Recall and precision of the trained general model.

Cross Validation General Model	true Other	true Drink	true Tooth	true Tumble	class precision
pred. Other	184	0	3	3	96.84%
pred. Drink	1	48	0	0	97.96%
pred. Tooth	4	4	158	0	95.18%
pred. Tumble	2	1	1	107	96.40%
class recall	96.34%	90.57%	97.53%	97.27%	516

Fig. 11: Recall and precision of the trained general model using cross validation based on 66/33% repetitive split of the data.

Test Model	**true Drink**	**true Tooth**	**true Other**	**true Tumble**	**class precision**
pred. Drink	17	0	0	0	**100.00%**
pred. Tooth	3	10	1	0	**71.43%**
pred. Other	1	0	11	0	**91.67%**
pred. Tumble	0	0	0	8	**100.00%**
class recall	**80.95%**	**100.00%**	**91.67%**	**100.00%**	**51**

Fig. 12: Recall and precision of test data. Note that all tumbles are recognized correctly although most users were different from the training set.

Applying the general model to the test data give a different picture. Tumble performs perfect although five of the eight users performing tumbles are not part of the training set. Tooth brushing and drinking show some false positives. One reason could be that those ADLs need more additional training data as both ADLs operate near the head and therefore some gestures associated have high similarity.

Test Model U3	**true Drink**	**true Tooth**	**true Other**	**true Tumble**	**class precision**
pred. Drink	12	0	0	0	**100.00%**
pred. Tooth	0	5	0	0	**100.00%**
pred. Other	2	0	11	0	**84.62%**
pred. Tumble	0	0	0	3	**100.00%**
class recall	**85.71%**	**100.00%**	**100.00%**	**100.00%**	**33**

Fig. 13: Recall and precision of test data just for user U3.

Running the same analysis as above just with the User U3 shows even better results. User U3 was part of the training set. The only exception is drinking recognition, the reason here was that two drinking ADLs where performed totally different from the normal drinking behavior. As one can conclude from that the recognition gets much better if a user is part of the training set. For real world application, this could induce that before really using the smartwatch as an ADL recognizing device users should be encouraged to train typical activities and use an improved NN model.

VII. Discussion

A. *ADL Recognition and Smartwatch System Support*

Continuous monitoring of EDL, ADL recognition in the smartwatch app requires an ongoing execution and adaption of the ANN, as soon as there will be new sensor signals. This requires a reliable background operation of the smartwatch app, even when the user is not looking at the smartwatch screen and the display therefore will be shut off. Unfortunately, and for energy saving purposes, smartwatch OSs tend to hibernate the app execution in situations, were the display is shut off. Smartwatch OS like Tizen™ or Android Wear 2.0™ are featuring such (background) service operations in their most advanced versions. Reliable background operations are mandatory and of crucial importance for a wide acceptance and trust in assistance apps for the elderly.

B. *EDL* »Falling«

This EDL entails a lot of difficulties. First of all, the detection of the EDL requires a barometric sensor in the smartwatch. This sensor typically is only present in "high-end" smartwatches. Second, the training of the EDL is inherently *dangerous* for the test persons with respect to potential injuries. Trained stuntmen or young people would be no alternative because their tumbling behavior will deviate too much from falls of elderly people. For the same reasons, crash dummies from the automotive field also would be no alternative, in that they would remain passive and would not show the characteristic last fraction of a second active (panic) reaction against the ongoing fall which is typical for humans. Therefore, we used "young elderly" of about sixty years of age for our tumble tests. But, it is still an open issue whether our trained falls – as planned, conscious event – really are representative for the majority of everyday accidents, sudden falls in the household. Unfortunately, practically no video sequences are available for such real falls as objective illustrative evidence.

VIII. Conclusion and Future work

EDL, ADL recognition based on an ANN works on today commercial smartwatches and delivers the necessary input for calculating the wellbeing of the smartwatch wearer. Continuous reliable detection of the EDL »*tumbling«, the ADLs* described requires durable background operations of the smartwatches, which only now will be supported by the most advanced smartwatch operating systems (OSs).

Universal models collected from different smartwatches, OSs and test persons are sufficient for achieving the targeted minimal 90% precision and recall rate for EDL, ADL detection. Best rates could be achieved by an individual model trained for a specific smartwatch. Such an individual model could be used even for user identification (e.g. in the scope of a benefit plan for good teeth brushing practice). But, the sensitivity of the individual model will require a substantial retraining even in cases of a smartwatch model change or even a major OS update.

A future application of the ANN based motion analysis will be dedicated to the combined analysis of the motion patterns, heart rate and blood glucose data from continuous glucose monitoring (CGM) systems. This will allow to conclude whether a change of glucose measurement data can be explained by the agility patterns of the smartwatch wearer. CGM systems like DEXCOM G5™ and Abbott's FreeStyle Libre™ already are rsp. will be capable to deliver these data via Bluetooth or NFC to smartwatches. The unobtrusive presence of those data on the wrist will support better self-management of the widespread diabetes mellitus type 2.

IX. References

[1] R.Lutze and K.Waldhör, "Smartwatches as Next Generation Home Emergency Call Systems, *8th German AAL Conference* 2015, 29.-30.4.2015, Frankfurt/Germany, Paper A1, VDE Press *(in German)*

[2] https://en.wikipedia.org/wiki/Activities_of_daily_living

[3] N.K.Suryadevara and S.C.Mukhopadhyvay, "Determining Wellness through an Ambient Assisted Living Environment", *IEEE Intelligent Systems,* Vol. 29(2014), No3, pp.30-37

[4] R.Lutze and K.Waldhör: "Integration of Stationary and Wearable Support Services for an Actively Assisted Liiving of Elderly People: Capabilities, Achievements, Limitations, Prospects – A Case Study", in: Wichert, R., Mand.E. (eds) "Ambient Assisted Living", Springer Int. Publishing 2017, pp. 3-26

[5] R.Lutze and K.Waldhör: "A Smartwatch Software Architecture for Health Hazard Handling for Elderly People*", 3rd IEEE International Conference on HealthCare Informatics (ICHI),* 21.-23.10.2015, Dallas, USA

[6] R.Lutze, R. Baldauf, K.Waldhör*:* "Dehydration Prevention and Effective Support for the Elderly by the Use of Smartwatches"*, 17th IEEE International Conference on E-Health Networking, Application & Services (HealthCom),* 14.-17.10.2015, Boston, USA

[7] K.Waldhör and R.Baldauf: "Recognizing Trinking ADLs in Real Time using Smartwatches and Data Mining", *Rapid Miner Wisdom / Europe Conference,* Ljubljana, Slovenia, 2015

[8] C.M.Bishop "Pattern Recognition and Machine Learning", Springer Science+Business Media, 2nd ed., 2006

[9] Dental care and oral health. Report of the British Alzheimer's Society, 2015

[10] D. T. Larose and C. D. Larose, "Data mining and predictive analytics" Hoboken, New Jersey: John Wiley & Sons, 2015.

[11] [H. Kerdegari, K. Samsudin, A. Rahman Ramli, and S. Mokaram, "Development of wearable human fall detection system using multilayer perceptron neural network", *International Journal of Computational Intelligence Systems*, vol. 6, no. 1, pp. 127–136, 2013.

[12] T. Prajapati, N. Bhatt, and D. Mistry, "A Survey Paper on Wearable Sensors based Fall Detection," *IJCA*, vol. 115, no. 13, pp. 15–18, 2015.

[13] R. Igual, C. Medrano, and I. Plaza, "Challenges, issues and trends in fall detection systems," *BioMed Eng OnLine*, vol. 12, no. 1, p. 66, 2013.

[14] A. Lisowska, G. Wheeler, V. C. Inza, and I. Poole, "An Evaluation of Supervised, Novelty-Based and Hybrid Approaches to Fall Detection Using Silmee Accelerometer Data", 2015 *IEEE Int. Conf. on Computer Vision (ICCV), Workshops, pp.10-16, Chile*

[15] R. Baldauf, "Mobile sensorbasierte Erkennung von Trinkbewegungen", Thesis, FOM, Nürnberg, 2015 (in German)

[16] K. Waldhör and R. Lutze, "Smartwatch Based Tumble Recognition – A Data Mining Model Comparison Study", *18th IEEE Int. Conference on E-Health, Networking, Application & Services (HealthCom*), 14.-16.9.2016, Munich, Germany

[17] D. S. Putler and R. E. Krider, "Customer and Business Analytics: Applied Data Mining for Business Decision Making Using R", CRC Press, Taylor & Francis Group, 2012.

[18] R. Lutze and Klemens Waldhör, "Model Based Diaglogue Control for Smartwatches", 17th HCI International Conference, 9-14.7.2017, Vancouver, BC, Canada, Springer Int. Publishing 2017

[19] Weiss, G.M., Timko J., Gallagher, C. Yoneda, K. and Schreiber A.. "Smartwatch-based Activity Recognition: A Machine Learning Approach", *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI) 2016,* 24.-27.2.2016, p. 426-429, Las Vegas, USA,

[20] Ferrante, S., Bonacina, S., Pozzi, G., Pinciroli, F., Marceglia, S., "A Design Methodology for Medical Processes", Applied Clinical Informatics, vol 7, no. 1, p.191–210, Schattauer Publishing, 2016

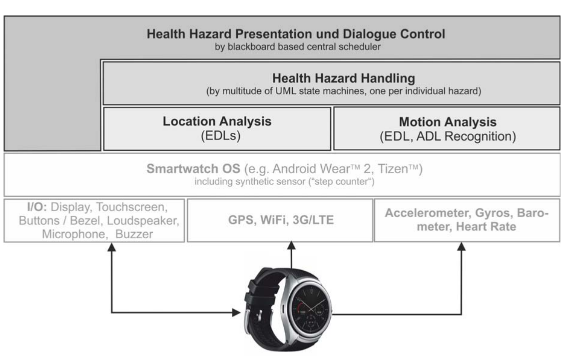


Fig. 14 (left): Block diagram of the smartwatch personal health assistance app with its layered architecture: *layer 0:* smartwatch HW with sensors, I/O; *layer 1*: smartwatch OS; *layer 2:* motion analysis via ANN and location analysis via GPS monitoring – geofencing; l*ayer 3:* simultaneous health hazard recognition handling via a multitude of simultaneously running finite state machines; *layer 4:* health hazard presentation and dialogue control vis a blackboard based scheduler

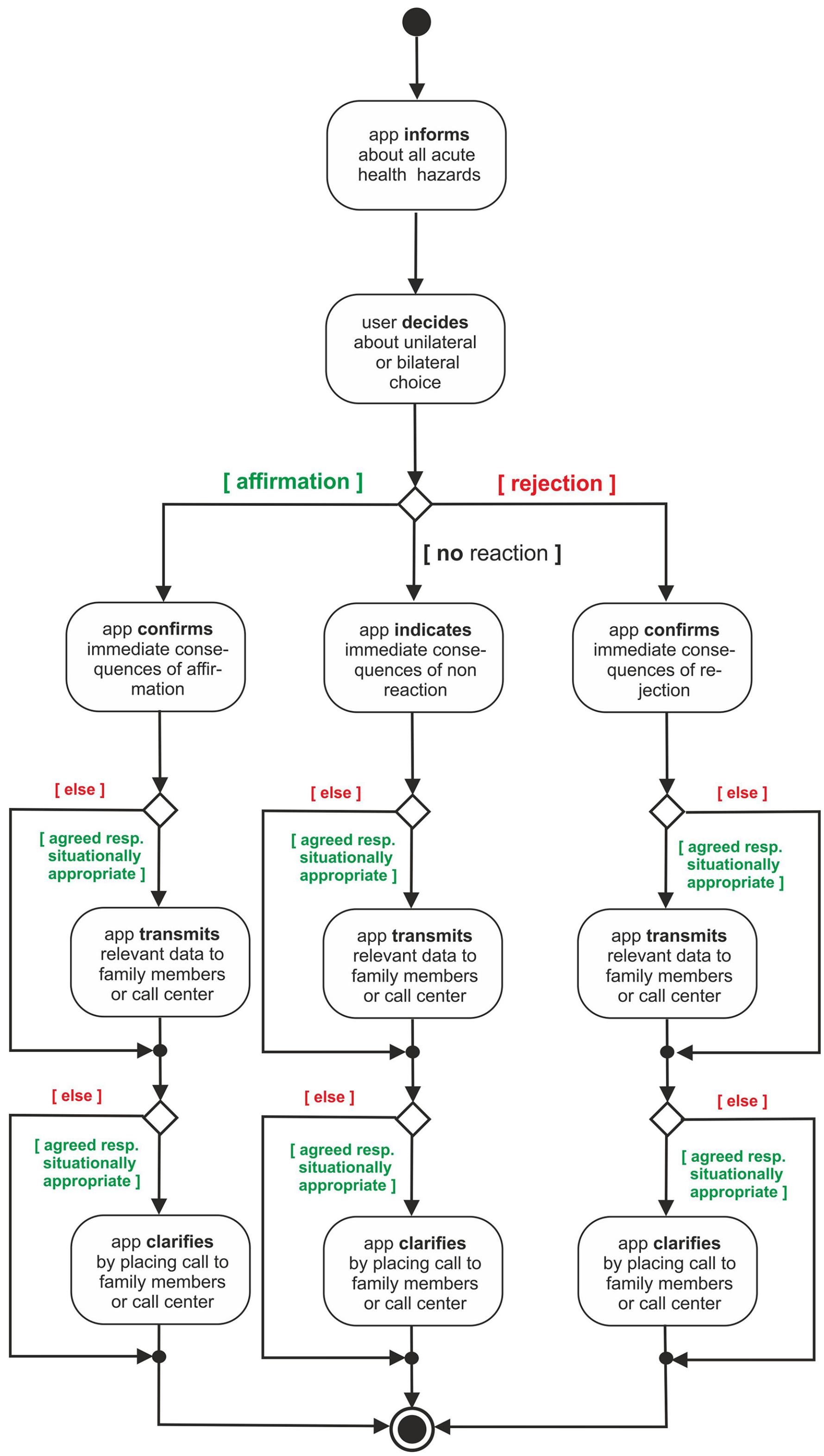


Fig. 15: Schematic dialogue within a critical dialogue section

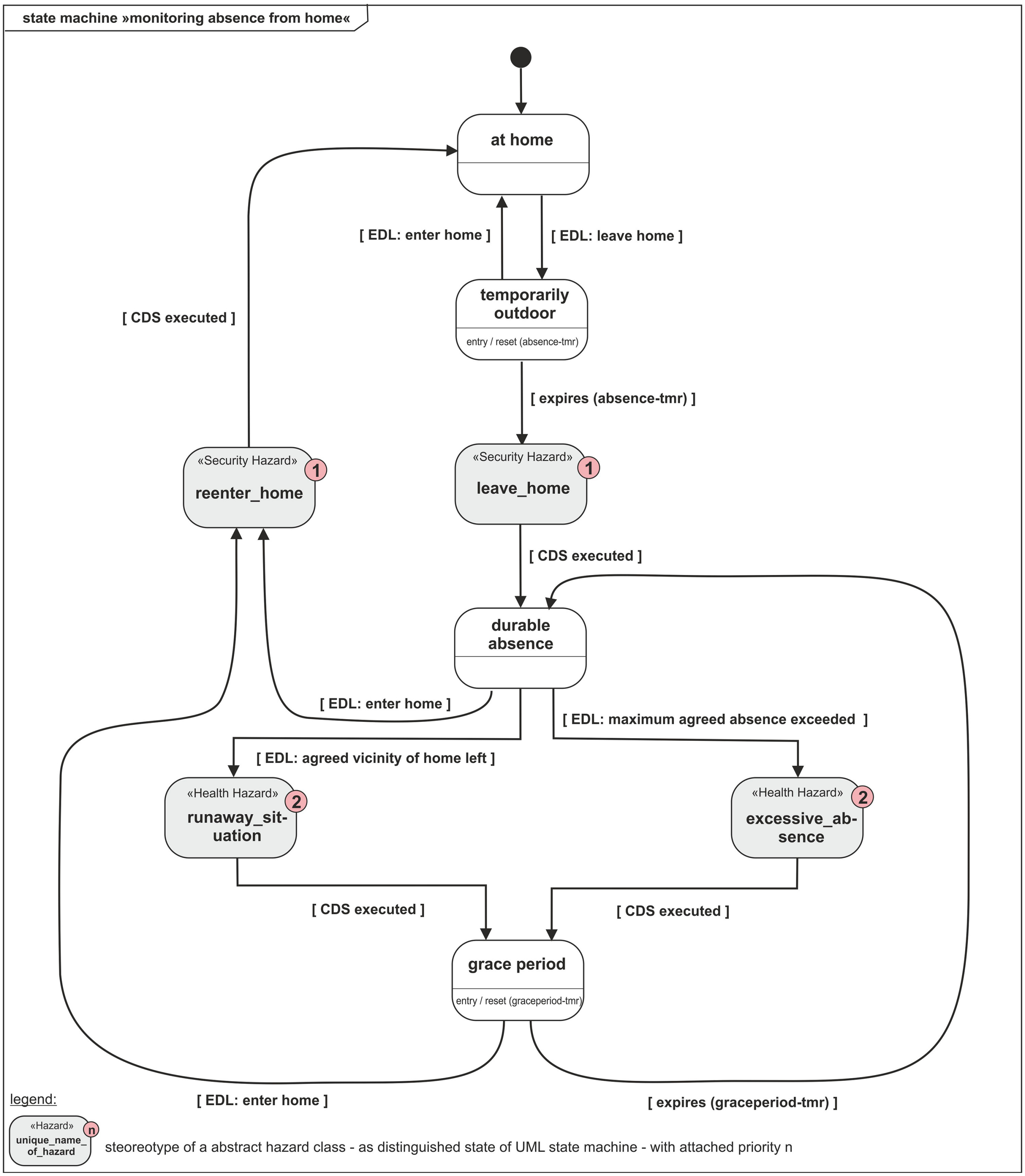


Fig. 16: Complex joint health hazard handler for runaway situations and excessive absence from home