SmartFABER: Recognizing Fine-grained Abnormal Behaviors for Early Detection of Mild Cognitive Impairment

Daniele Riboni[*] Claudio Bettini[*] Gabriele Civitarese[*]
Zaffar Haider Janjua[*] Rim Helaoui[◇]

[*]Università degli Studi di Milano, Dept. of Computer Science, EveryWare Lab, Italy
[◇]Philips Research Personal Health, Eindhoven, The Netherlands
Email: {daniele.riboni, claudio.bettini, gabriele.civitarese, zaffar.janjua}@unimi.it,
rim.helaoui@philips.com

Abstract

In an ageing world population more citizens are at risk of cognitive impairment, with negative consequences on their ability of independent living, quality of life and sustainability of healthcare systems. Ubiquitous computing and intelligent data analysis can provide innovative methods and tools for detecting early symptoms of the onset of cognitive impairment and for monitoring its evolution. In this paper we present SmartFABER, an activity recognition system integrated with a novel method to detect abnormal activity routines of elderly people living at home. The method relies on medical models, provided by cognitive neuroscience researchers, describing some of the anomalies in carrying out daily activities that may indicate the onset of early symptoms of mild cognitive impairment. A non-intrusive sensor-based infrastructure acquires low-level data about the interaction of the individual with the home environment including objects, appliances and furniture. Based on those data, a novel hybrid statistical and knowledge-based technique is used to detect the abnormal behaviors of the senior, which are stored by the system and presented through a dedicated dashboard to the clinicians. Differently from related works, our method can detect abnormal behaviors at a fine-grained level, thus providing an important tool to support the medical diagnosis. We have fully implemented the system and we evaluated our method with significant datasets, partly generated by performing activities in a smart home laboratory, and partly acquired during several months of monitoring of the instrumented real home of a senior diagnosed with MCI. We report the output of extensive experiments confirming the superiority of SmartFABER with respect to previously published methods; Qualitative feedback from clinicians also confirms the validity of the approach.

1 Introduction

Independent living and pro-active healthcare are becoming strategic application areas for major research programmes all over the world, considering that the senior population (aged over 65) is projected to double as a percentage over the whole population in the next decades [1]. Among the most frequent threats to independent living is cognitive decline, whose early symptoms often lead to a Mild Cognitive Impairment (MCI) diagnosis. According to the International Working Group on MCI, there are evidences of subtle differences in performing instrumental activities of daily living (IADLs) among MCI patients compared to both healthy older adults and individuals with dementia [2]. Hence, monitoring of daily living activities and recognition of abnormal behaviors may help practitioners to early detect the onset of cognitive impairment.

Ubiquitous computing technologies coupled with intelligent data analysis have a recognized potential in the automatic recognition of IADLs. Indeed, several research projects, and numerous research papers have tried to detect behavioral markers of MCI onset through ubiquitous computing technologies, obtaining a correlation between the predicted and actual cognitive status of the patient. Some of these approaches require the execution of ability tests about the performance of IADLs in an instrumented smart home of a medical institution; hence, they incur in high costs and cannot be applied on a continuous basis. Some of them deploy cameras and sensor networks in controlled environments and use video and audio for activity recognition: these systems are often perceived as too invasive for the elderly's privacy. Other works rely on continuous monitoring of low-level behavioral markers (steps taken, walking speed, ...) and trigger alarms whenever they detect situations sufficiently distant from the expected (modeled) behavior: they provide little support to the diagnosis, since they do not report fine-grained descriptions of the anomalies occurred during the execution of IADLs. Moreover, very few systems have been deployed and tested effectively in the elderly's home. In the following section we provide a short overview of the relevant literature.

In this paper we describe SmartFABER, a novel technique for Fine-grained Abnormal BEhavior Recognition aimed to overcome the limitations of existing techniques in supporting early detection of mild cognitive impairment (MCI) for elderly people living independently at home. SmartFABER exhibits the following features: a) it heavily relies on indicator models built by cognitive neuroscience experts, b) it continuously acquires data from non-intrusive sensors deployed in the senior's home, c) it features an effective abnormal behavior recognition technique coupling state-of-the-art machine learning with knowledge-based inferencing, d) it provides clinicians with a dashboard identifying fine-grained short-term abnormal behaviors (e.g., inappropriate timing in assuming food or medicine intake, improper use of equipment, unnecessary repetitions of actions).

Our main contributions can be summarized as follows:

- SmartFABER is an innovative hybrid activity and anomaly recognition framework. It overcomes the shortcomings of purely statistical methods and it has major improvements over the state-of-the-art hybrid systems [3] in terms of accuracy of activity recognition, and enhanced reporting features;
- SmartFABER has been experimented on real datasets, including a three months deployment in the house of a senior diagnosed with MCI. The results show that SmartFABER is able to detect most anomalies that we have targeted while producing a small number of false positives;
- We show the superiority of SmartFABER with respect to the only other hybrid system we are aware of through a direct comparison over the same datasets.

The rest of the paper is structured as follows. Section 2 discusses related work. Section 3 reports our model of activities and abnormal behaviors. Section 4 presents the SmartFABER method. Section 5 reports experimental results. Finally, Section 6 concludes the paper.

2 Related work

Activity recognition systems proved to be effective for supporting the diagnosis and improving healthy ageing [4, 5]. In the literature, various strategies have been proposed to devise effective and unobtrusive activity monitoring systems by exploiting pervasive computing technologies [6]. A popular research direction for activity recognition consists in exploiting audio-visual information recorded by cameras and microphones with the help of sound, image and scene recognition software. However, those methods are considered too invasive in a home environment, due to the privacy issues that they determine. Hence, in the following we restrict our attention to non-invasive sensor-based techniques.

2.1 Recognition of simple activities

Several techniques were proposed to recognize *simple* activities, which rely on data acquired from body-worn sensors and on the application of supervised learning methods [7, 8]. Early attempts in this sense were mainly based on the use of data acquired from multiple body-worn accelerometers [9], possibly coupled with biometrical sensors and integrated in clothes [10], to recognize locomotion types and simple physical activities. A major limitation of these early systems is that they did not consider contextual information, such as current location, environmental conditions, and surrounding objects, that could be usefully exploited to improve the accuracy of recognition. Hence, other activity recognition approaches take into account the user's context by acquiring environmental data from several sensors [11]. For instance, in [12] the authors proposed the use of machine learning and data acquired from body-worn sensors (an ear microphone, sensor collar integrating electromyogram and microphone, and four upper body accelerometers) to accurately monitor food intake activities (movement, chewing and swallowing). However, being mainly based on body-worn sensors, those methods are not well suited to recognize more complex activities, like IADLs executed at home, which are characterized by the interaction of the individual with several objects and furniture.

2.2 Recognition of complex activities

The recognition of complex activities, like ADLs that we consider in our work, relies on the usage of sensors to detect the interaction of the individual with objects and furniture. However, even when those sensors are available, the recognition of complex activities turns out to be challenging using solely supervised learning methods. Indeed, complex activities are characterized by large inter- and intra-personal variability of execution, and it is very hard to acquire a sufficiently comprehensive training set to include most of the possible ways of executing activities. Hence, different frameworks for knowledge representation and reasoning have been investigated to appropriately model complex human activities by means of ontologies. In particular, description logics [13] have emerged among other symbolic formalisms, mostly because they provide complete reasoning supported by optimized automatic tools. However, as illustrated in [14], both expressiveness and efficiency issues strongly limit the feasibility of ontological approaches to activity recognition. Moreover, the recognition of complex activities through ontological reasoning has to start from some basic observations (e.g., "the user is in a given room", "he is sitting on a chair", etc.); this task requires the use of statistical methods to derive semantic information from raw sensor data.

2.3 Hybrid activity recognition techniques

Given the limitations of both statistical and symbolic approaches, a few hybrid activity recognition systems have been proposed in the literature, which vary on the adopted reasoning techniques and on their interaction mechanisms. An interesting instance of those approaches is Markov Logic Networks (MLN) [15], a probabilistic first-order logic. Given a training set, and a set of probabilistic formulas, with MLN it is possible to learn a weight for each grounded formula by iteratively optimizing a pseudo-likelihood measure. Those weights represent the confidence value of the formula. Deterministic formulas can be added to probabilistic ones to express deterministic knowledge about the domain of interest. Different reasoning tasks can be executed to infer additional information based on formulas and facts. MLN has been used in a previous version of our work [3]. A similar approach was adopted in [16] to model and recognize activities at different levels of complexity using probabilistic description logic. Hybrid ontological and statistical reasoning is proposed in [17] to continuously assess the fall risk of a senior at home, by integrating data acquired from different fall detection systems and environmental sensors. In this work, we propose a hybrid method to recognize the start- and end-time of ADLs based on a combination of supervised learning and knowledge-based conditions to refine the statistical predictions.

2.4 Applications of activity recognition to MCI diagnosis

Several studies in the neuropsychology research field show that it is possible to distinguish between cognitively healthy adults and cognitively impaired individuals based on subtle differences in their behavioral patterns [2]. There is a growing interest in exploiting pervasive computing technologies to automatically capture and measure those differences [5]. For instance, motion sensors and contact sensors have been used in [18] to measure low-level activity patterns, such as walking speed and activity level in the home; results have shown that the coefficient of variation in the median walking speed is a statistically significant measure to distinguish MCI subjects from healthy seniors. More recently, a sensor-based infrastructure has been used to unobtrusively monitor the execution of IADLs by older adults in a smart-home [19]; the results have shown a significant correlation between the cognitive health status of the subject and the level of assistance that he needs to complete the activities.

Based on this line of research, different works have proposed to apply machine learning techniques on data acquired in sensor-rich environments, for assessing the cognitive health status of an individual performing a set of ADLs. In the work of Dawadi et al. [20], patients were invited to execute a list of routines (e.g., write a letter, prepare lunch) inside a hospital smart-home. Different kinds of sensors were used to detect movements inside the home and to track the use of furniture and appliances. Based on data coming from the home sensors, supervised learning methods were used to assign a score to each performed activity; the score measures the ability of the subject to perform the activity correctly. The achieved scores were used to predict the cognitive status of the patient (cognitive health or dementia). The supervised learning approach has been applied in other works, including [21, 22, 23], using several other learning methods. However, while a significant correlation exists between the inferred activity scores and the cognitive health status of the individual, those methods do not provide a description of the observed behavioral anomalies. On the contrary, the medical assessment would benefit from detailed knowledge of the abnormal behavior of the patient. For this reason, in our approach we do not rely on statistical deviations from the "normal" behavior; instead, we aim at recognizing fine-grained anomalies, modeled according to neuroscience experts as possible indicators of MCI, using a hybrid statistical and knowledge-based approach. To the best of our knowledge, this work is the first one (except our preliminary results presented in [3]) that applies this approach to cognitive health assessment.

2.5 Long-term analysis of activity data

In the aforementioned works, the detection of abnormal behaviors is mostly done on a short-term basis and does not take into account the patient's personal habits. Other works have proposed methods to model the patient's usual behavior from the activities performed in the past and use this model to detect anomalies as changes from his/her usual behavior. In [24], a method has been proposed to monitor the circadian (24-hours) variability of the patient's activities using location sensors and statistical calculations were performed regularly on sensors data to recognize possible deviations in the patient's behavior. In [25] in-home activities and sleep restlessness were captured using different sensors and a simple alert system was implemented to detect changes in the activity patterns and generate health alerts that were sent to clinicians to be rated for their clinical relevance. These ratings were then used as ground truth in developing classifiers to recognize relevant alerts. In [26], the authors propose a technique to detect recurrent ADLs patterns, as well as their variations, by mining heterogeneous multivariate time-series from sensor data acquired in a smart home. Another approach based on temporal data mining was presented in [27]. Frequently-occurring temporal relationships between activities were extracted from the observed history of sensor events and used to model the probability that a particular event should or should not occur on a given day. A technique based on unsupervised learning is proposed in [28] to automatically discover ADL patterns and their variations. That technique is coupled with an activity recognition module and with visualization tools to allow practitioners inspecting the trend of activity patterns. Visualization of spatio-temporal data extracted from the long-term observation of elderly's activities at home is used in [29] to identify potential risk situations. In our work, we also aim at monitoring the elderly's behavior on the long term. However, as explained before, we do not rely on statistical measures about activity patterns, but we consider specific anomalies modeled by neuroscience experts.

3 Modeling activities and abnormal behaviors

In the following, we explain how we model human activities and fine-grained abnormal behaviors.

3.1 Human activities

In order to model activities, we adopt the multilevel framework proposed in [16]. In particular, each IADL consists of a sequence of simple actions. For instance, a patient could perform the IADL "taking medicines" by executing this sequence of actions: open the medicine repository, retrieve the medicine box, return the medicine box, and close the medicine repository. Of course, the same activity can be performed by executing different sequences of actions. For instance, "taking medicines" can be performed by these actions: open the medicine repository, retrieve the medicine box, and return the medicine box (leaving the medicine repository open). Since we concentrate on IADLs, we assume that each action corresponds to a manipulative gesture or other body movement involving an object (e.g., "open the silverware drawer", "sit on the kitchen chair').

3.2 Fine-grained abnormal behaviors

By *fine-grained abnormal behaviors* (also called *anomalies* for short) we define those behaviors, observed within a relatively short time period (from a few seconds to a day), which diverge from the expected ones, according to a given model provided by clinicians. In general, a fine-grained abnormal behavior can be detected based on the observation of: the sequence of actions composing a performed activity; the sequence of performed activities; the duration of individual activities and actions; the execution of an improper activity, or the non-execution of some expected activities.

In particular, in this work we consider models of abnormal behaviors that may indicate the onset of MCI, and more generally of a cognitive decline. In order to acquire those models, we collaborated with cognitive neuroscience experts from the Institute Fatebenefratelli[1], Lombardy –a leading center in the field of mental

[1] IRCCS (Research and Care Institute) St John of God Clinical Research Centre, Brescia – http://www.irccs-fatebenefratelli.it

health research and research on neurodegenerative disorders– within the SECURE[2] research project funded by Lombardy region and MIUR Italian ministry. A list of fine-grained abnormal behaviors has been identified during different project meetings among the technical and medical partners of the project, based on the medical practice and relevant literature [30]. For the sake of this work, we have considered anomalies related to food preparation, food consumption, and compliance to medical prescriptions. The anomalies are defined in natural language by the clinicians; e.g., "an anomaly occurs when the patient prepares a meal but forgets to consume it".

In order to facilitate their analysis, we classified anomalies in the following categories:

- Non-critical anomaly. An anomaly is considered as non-critical when the patient skips a relevant action while performing a IADL, or spends too much time to perform the activity, but still he is able to complete the activity correctly. For instance, we consider a non-critical anomaly to occur when the patient forgets to close a repository after taking something from it.

- Critical anomaly. A critical anomaly occurs when the patient skips some necessary action while performing an activity, forgets to execute a required activity, or performs the activity more times than expected. Critical anomalies are stronger indicators of possible MCI onset than non-critical ones. These anomalies are further divided into four categories:

  - Omission: there are some actions in each IADL which are necessary and it is mandatory for the patient to perform them in order to complete the activity correctly: a critical omission occurs when the patient skips any of such actions. For instance, a critical omission related to the activity "taking medicines" is: "the patient forgets to retrieve the prescribed medicine during the prescribed time'.
  - Replacement: this anomaly occurs when, during a IADL, a patient replaces a correct action with a wrong one; for instance, "the patient has placed the butter inside a non-refrigerated cabinet".
  - Improper activity: it occurs when the patient performs an activity that is not consistent with the model. For instance, this anomaly occurs when the patient takes a medicine that was not prescribed.
  - Repetition: this anomaly occurs when the patient repeats the same activity more times than expected; for instance, when the patient consumes the morning breakfast twice in a day.

[2] SECURE: Intelligent System for Early Diagnosis and Follow-up at Home, `http://secure.ewlab.di.unimi.it/`

Of course, human behaviors are characterized by wide variability; factors such as contextual conditions, individual habits and personality traits may determine the execution of various anomalies that are not necessarily due to cognitive impairment. This is especially true for non-critical anomalies, as leaving repositories open, which may be normally done by cognitively healthy people for negligence or hastiness. Hence, while the considered anomalies are indicators of possible abnormal behaviors, they are not intended to provide an automatic diagnosis of the patient's cognitive status. For instance, consider the aforementioned example of improper activity: the fact that the patient takes a medicine that was not prescribed is critical if he does it unintentionally (e.g., for a memory disorder). In other cases it may be a normal behavior; e.g., if the patient intentionally takes an over-the-counter drug that does not interfere with his medical prescriptions. Therefore, the frequency of detected anomalies and their temporal trend are used as a mean to trigger alarms to the practitioners for further inspecting the history of detected anomalies and their fine-grained descriptions.

4 The SmartFABER hybrid technique

In this section, we illustrate the SmartFABER method to recognize fine-grained abnormal behaviors. The overall framework is shown in Figure 1. The hybrid reasoning framework of SmartFABER, shown in Figure 2, exploits both statistical and knowledge-based methods. In particular, knowledge-based methods are used by the SEMANTIC INTEGRATION LAYER to recognize simple events from raw sensor data; by the SMART AGGREGATION module to identify activity instances; and by the KNOWLEDGE-BASED INFERENCE ENGINE to detect the anomalies according to the clinicians' models. Statistical reasoning, taking into account temporal features, is used to classify events into activities and to recognize activity instances. In the following, we describe in detail the main components of SmartFABER.

4.1 Sensor data acquisition

A non-invasive sensor network system including environmental, presence, and contact sensors is deployed at the elderly's home. The smart-home monitoring system, running on a mobile device (e.g., a tablet) within the home, collects in real-time the raw sensed events data from the sensor nodes in order to execute the SmartFABER recognition algorithms. The SEMANTIC INTEGRATION LAYER module is in charge of applying simple inference methods to derive pre-processed events from raw sensor data. For instance, a rule states that "if the presence sensor detects a presence near the kitchen table, and the sensor on the kitchen's chair de-

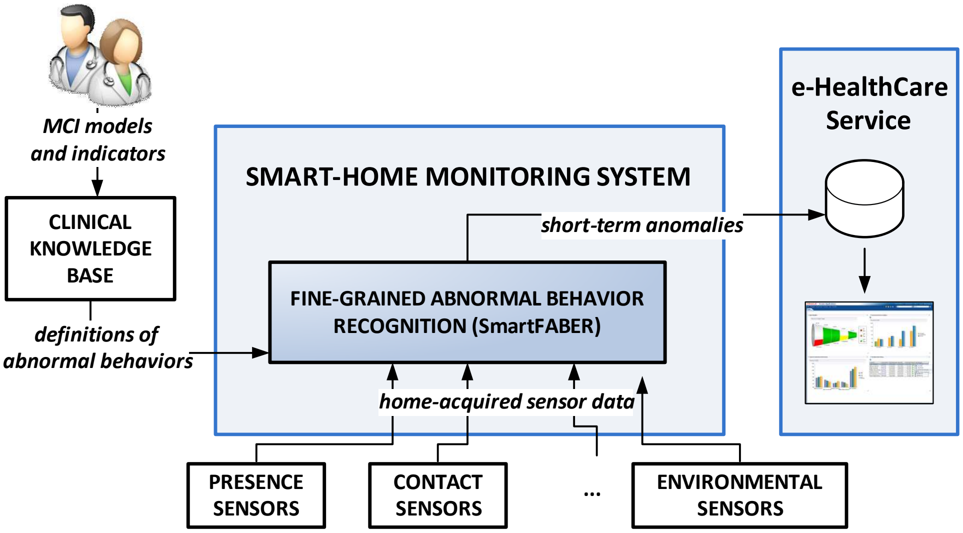

MCI models and indicators
CLINICAL KNOWLEDGE BASE
definitions of abnormal behaviors
SMART-HOME MONITORING SYSTEM
short-term anomalies
FINE-GRAINED ABNORMAL BEHAVIOR RECOGNITION (SmartFABER)
home-acquired sensor data
e-HealthCare Service
PRESENCE SENSORS
CONTACT SENSORS
...
ENVIRONMENTAL SENSORS

Figure 1: The framework to recognize short-term abnormal behaviors

tects a weight of more than 40Kg, then the current event is *sitting on a chair at the kitchen table*". We define $\mathbf{E}$ as the set of all the considered event types (e.g. $\mathbf{E} = \{$Door is opened, Door is closed$\}$). We adopt a simple temporal model for the events. We denote as $\mathbf{T}$ the set of all the possible timestamps. A sequence of events is represented as follows:

$$\langle\, ev(E_1, t_1), ev(E_2, t_2), \ldots, ev(E_m, t_m)\,\rangle,$$

where $ev(E_i, t_i)$ indicates that the event $E_i \in \mathbf{E}$ occurred at timestamp $t_i \in \mathbf{T}$. A unique timestamp is assigned to each event, based on the time at which the related raw sensor events are received by the central mobile device. In this way we impose a total order on event timestamps $\langle\, t_1, t_2, \ldots, t_m\,\rangle$. The sequence of high-level events is then provided to the SmartFABER recognition algorithms.

4.2 Modeling activity classes and instances

We define $\mathbf{A} = \{A_1, A_2, \ldots, A_k\}$ as the set of k considered high-level activity classes (e.g.: $\mathbf{A} = \{$*Preparing Meal, Eating Meal, Taking Medicines*$\}$). An instance a of an activity $A \in \mathbf{A}$ is an occurrence of A during a timespan. Intuitively, a timespan is a particular time interval represented by a start time and an end time, where not every timestamp between the boundaries necessarily belongs to the timespan. This representation allows us to consider activities performed in an interleaved fashion. More formally, we define a timespan ts as a non-convex

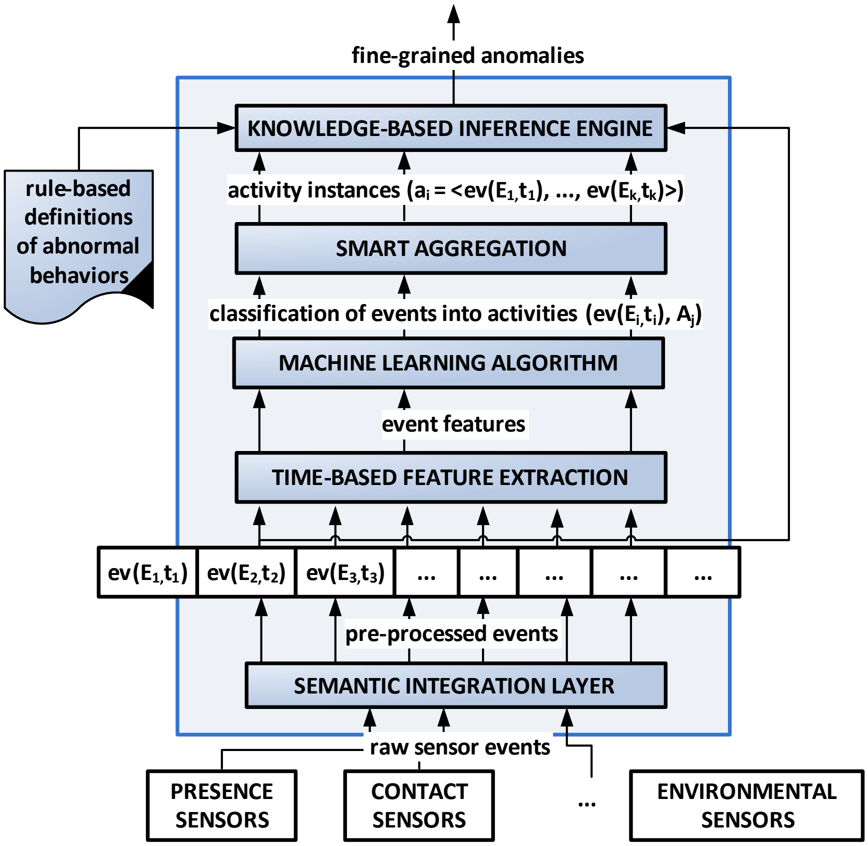

fine-grained anomalies
KNOWLEDGE-BASED INFERENCE ENGINE
rule-based definitions of abnormal behaviors
activity instances (a_i = <ev(E_1,t_1), ..., ev(E_k,t_k)>)
SMART AGGREGATION
classification of events into activities (ev(E_i,t_i), A_j)
MACHINE LEARNING ALGORITHM
event features
TIME-BASED FEATURE EXTRACTION
ev(E_1,t_1)
ev(E_2,t_2)
ev(E_3,t_3)
...
...
...
...
...
pre-processed events
SEMANTIC INTEGRATION LAYER
raw sensor events
PRESENCE SENSORS
CONTACT SENSORS
...
ENVIRONMENTAL SENSORS

Figure 2: The SmartFABER technique to recognize short-term abnormal behaviors

time interval characterized by a finite set of non overlapping temporal intervals $\{[x_1, y_1], [x_2, y_2] \ldots, [x_n, y_n]\}$ where $\forall i \; x_i, y_i \in \mathbf{T}$. Given a timestamp $t \in \mathbf{T}$ and a timespan ts, we say that t belongs to the timespan ts if it belongs to one of its intervals. We denote with a^A_{ts} an instance of an activity $A \in \mathbf{A}$ occurred during a timespan ts. Given a sensor-equipped environment, an activity instance a^A_{ts} generates a sequence of events that we call "a^A_{ts} observations", formally denoted by $Obs(a^A_{ts}) = \langle ev(E_1, t_1), ev(E_2, t_2), \ldots, ev(E_k, t_k)\rangle$, where $\forall i \; E_i \in \mathbf{E}$ and $t_i \in ts$.

4.3 Activity recognition

At each pre-processed event, SmartFABER applies a time-based supervised learning technique to assign the most probable activity class. The classified events are then post-processed in order to identify the most probable activity instances.

4.3.1 Classification of events

The events produced by the SEMANTIC INTEGRATION LAYER are communicated to the TIME-BASED FEATURE EXTRACTION module. For each event $ev(E_i, t_i)$, this module is in charge of building a feature vector v_i representing the n most recent events $\langle\, ev(E_{i-n+1}, t_{i-n+1}), \ldots, ev(E_{i-1}, t_{i-1}), ev(E_i, t_i)\,\rangle$. In particular, we adopt the feature extraction technique proposed in [31], since it takes into account temporal aspects, and proved to be effective in recognizing activities based on streams of sensor events. The produced feature vectors are given as input to a supervised MACHINE LEARNING ALGORITHM to infer for each v_i the most probable class of the activity instance carried out at t_i. The algorithm is trained using a dataset of activities and generated events.

4.3.2 Naive aggregation

The next step is to infer the actual activity instances from the output of the MACHINE LEARNING ALGORITHM by grouping together those events which can be considered observations generated by the same activity instance. Intuitively, temporally close events classified with the same activity class are most likely generated by the same activity instance. We first discuss the baseline approach, named *naive aggregation*. The basic idea of this algorithm is the following: if two consecutive events occurred respectively at t_i and t_{i+1} are classified with the same activity class $A_i = A_{i+1}$, they are considered as observations generated by the same instance of an activity of class A_i. Otherwise, they are considered observations generated by different activity instances. Consider, for instance, the case illustrated in Table 1. This table illustrates in the first three columns a sequence of events associated with activity classes predicted by the MACHINE LEARNING ALGORITHM; in the fourth column, the ground truth about activity instances; and in the last column the output of the *naive aggregation* method. The algorithm produces 5 different instances of activities. However, it is easy to see that this aggregation is not correct. Indeed, the events E_2, E_4 and E_6 share the same activity class and are temporally close. With high probability, the inferred activity classes for the events E_3 and E_5 are mispredictions, since the "Eating meal" and the "Preparing meal" activity instances would have a too short duration. Moreover, consider the case where events E_1 and E_2 correspond respectively with *Presence in the kitchen* and *Open the medicine repository*. These two events alone can not be considered as the only observations generated by an instance of a *Taking Medicines* activity: the medicine repository can possibly contain items not related with medicines and it is also possible that it may be opened just to check the content. Another issue of this technique is that two consecutive events labeled with the same activity class but temporally distant (like E_6 and E_7) would be grouped together, while they most likely belong to separate activity instances. Hence, the particular sequence of events illustrated in Table 1 should identify a single instance of "Taking medicines" that generated the events from E_1 to E_6.

Table 1: An example of *naive aggregation* based on a sequence of classified events

Event type	Timestamp	Predicted activity class	Actual instances	Predicted instances
E_1	08:32:31	Taking medicines	$takingMedicines_1$	$takingMedicines_1$
E_2	08:32:48	Taking medicines		
E_3	08:32:55	Eating meal		$eatingMeal_1$
E_4	08:33:02	Taking medicines		$takingMedicines_2$
E_5	08:34:11	Preparing meal		$preparingMeal_1$
E_6	08:34:13	Taking medicines		$takingMedicines_3$
E_7	11:34:27	Taking medicines	$takingMedicines_2$	
$\dots$	$\dots$	$\dots$		
$\dots$	$\dots$	$\dots$		

4.3.3 Smart aggregation

To overcome the problems of the approach illustrated in 4.3.2, we refined our recognition method. We introduce for each activity class $A \in \mathbf{A}$ a set of conditions that are necessary for a sequence of events to be considered observations generated by an instance of that class. For example, assuming that the infrastructure includes sensors to detect the stove usage, any instance of the activity *Preparing a hot meal* should generate some observations related to the usage of the stove. Other examples of conditions may be constraints on the duration of the activity instance or on the number of generated events. Among those conditions, we consider the upper bound about the duration of activity interruptions: the time distance between every pair of consecutive events within the observations generated by an activity instance a^A_{ts} must be lower than $maxGap_A$. The value of the upper bound $maxGap_A$ depends on the activity class A. Those values are determined statistically; as a result, for example, "Preparing meal" will have a higher $maxGap$ than "Taking medicines". Formally, let $C^{(A)} = \{c_1, c_2, \dots, c_k\}$ be a set of necessary conditions expressed in logic over a sequence of events that are observations of any instance of a class $A \in \mathbf{A}$ (e.g. {"The sequence of events must last more than 3 minutes", "The sequence of events must contain an event regarding the usage of the stove", …}). A sequence of events $s = \langle ev(E_1, t_1), ev(E_2, t_2), \dots, ev(E_k, t_k) \rangle$ can be considered as observations generated by an activity instance a^A_{ts} if it satisfies every condition in $C^{(A)}$. The set of conditions for each class are determined after a detailed analysis of the semantics of the activity class and on statistics about the available observations acquired from the sensor infrastructure. For instance, the previously discussed condition about the duration of the interruption of an activity $A \in \mathbf{A}$ over a sequence of events s can be expressed as:

$$\forall i:\ ev(E_i, t_i), ev(E_{i+1}, t_{i+1}) \in s \rightarrow (t_{i+1} - t_i < maxGap_A)$$

We now introduce the SMART AGGREGATION algorithm, a refined activity instance recognition method. The pseudo-code is shown in Algorithm 1. The first step of the algorithm is a segmentation over the output of the MACHINE LEARNING ALGORITHM: all the events associated with the same activity class A and temporally close (according to $maxGap_A$) are grouped together. For each group g of events classified with an activity A, it is checked if it satisfies all the conditions in $C^{(A)}$. If all the conditions are satisfied, an activity instance a^A_{ts} that generated the observations contained in g is recognized. All the events contained in those groups which did not satisfy the conditions of their class are considered as mispredictions. Hence, the algorithm tries to include them in one of the activity instances recognized at the previous step. For each misprediction $ev(E, t)$, the algorithm builds a set I of activity instances a^A_{ts} (that have been recognized in the previous step) such that t lies between the boundaries of the timespan ts and $\{ev(E, t)\} \bigcup Obs(a^A_{ts})$ satisfies all the conditions in $C^{(A)}$. When $|I| > 1$ we choose the most probable instance based on a function $freq$, which computes the frequency of an event being an observation of the instances of a particular activity class. The values of $freq$ for each possible combination of event type and activity class are computed offline based on the annotated dataset. The event $ev(E, t)$ is added to the observations of the instance $a^A_{ts} \in I$, where A is the most frequent activity class. If I is empty, $ev(E, t)$ is considered an observation of an “other activity” instance.

4.4 Anomaly detection

We remind that we denote as “anomalies” the deviations from the “normal” way of carrying out activities. Anomalies are described in natural language by cognitive neuroscience experts. In order to automatically reason with anomalies, we represent them in propositional logic. Anomalies are represented by the predicate *anomaly(Categ, Obj, Time)*. *Categ* defines the category of the anomaly. *Obj* defines the objects or activities involved in the anomaly; for example, in case of a critical omission, the missed medicine may be the object related to that anomaly. *Time* defines the time instant at which the anomaly is detected.

Table 2 shows the representation of a few anomalies. The semantics of *not* is the one of *negation as failure* [32]. Predicate *prescribed(m,*t_1*,*t_2*)* states that the patient must take medicine m from time t_1 to time t_2 of the current day.

ALGORITHM 1: Smart aggregation

```
Input: A set $S = \{(ev(E_1, t_1), A_1), (ev(E_2, t_2), A_2), \ldots, (ev(E_n, t_n), A_n)\}$ of
       events associated with the predicted current activity classes, where $\forall i$
       $E_i \in \mathbf{E}$, $t_i \in \mathbf{T}$ and $A_i \in \mathbf{A}$.
Output: A set $\mathcal{A}$ of activity instances.

$\mathcal{A} \leftarrow \emptyset$;
$mispredictions \leftarrow \emptyset$;
foreach $A \in \mathbf{A}$ do
    $X \leftarrow$ the events in $S$ predicted with $A$;
    $G \leftarrow segmentation(X, maxGap_A)$;
    foreach $g \in G$ do
        if $g$ satisfies all the conditions in $C^{(A)}$ then
            $a \leftarrow$ an activity instance of class $A$ that generated the observations $g$;
            $\mathcal{A} \leftarrow \mathcal{A} \bigcup \{a\}$;
        else
            $mispredictions \leftarrow mispredictions \bigcup g$;
        end
    end
end
foreach $ev(E, t) \in mispredictions$ do
    $I \leftarrow \emptyset$;
    foreach $a_{ts}^{A} \in \mathcal{A}$ do
        if $t$ lies between the boundaries of $ts$ then
            $x = Obs(a_{ts}^{A}) \bigcup \{ev(E, t)\}$;
            if $x$ satisfies all the conditions in $C^{(A)}$ then
                $I \leftarrow I \bigcup \{a_{ts}^{A}\}$;
            end
        end
    end
    if $I \neq \emptyset$ then
        $a' = \underset{a_{ts}^{A} \in I}{\text{argmax}}\ freq(E, A)$;
        add $ev(E, t)$ to the observations of $a'$;
    else
        consider $ev(E, t)$ as observation of an activity instance of class "other
        activity";
    end
end
return $\mathcal{A}$;
```

Medicine(o) (resp. *Food(o)*) states that object o is a medicine box (resp. food item). *Action(a, o, o', t)* states that the patient executed action a on objects o and

ALGORITHM 2: Segmentation of activity instances

```
Input: A set of events $X = \{ev(E_1, t_1), ev(E_2, t_2), \ldots, ev(E_n, t_n)\}$ and the
       threshold $maxGap$.
Output: A partition of $X$ according to $maxGap$.

$groups \leftarrow \emptyset$;
$currGroup \leftarrow \{ev(E_1, t_1)\}$;
for $i \leftarrow 2$ to $n$ do
    if $t_i - t_{i-1} < maxGap$ then
        $currGroup \leftarrow currGroup \bigcup \{ev(E_i, t_i)\}$;
    else
        $groups \leftarrow groups \bigcup \{currGroup\}$;
        $currGroup \leftarrow \{ev(E_i, t_i)\}$;
    end
end
return $groups \bigcup \{currGroup\}$;
```

Table 2: Examples of rules modeling abnormal behaviors

No.	Rule	Anomaly type
1	$anomaly(cr, fridge, T_2) \leftarrow action(return, RF, S, T_1) \wedge action(close, door, S, T_2) \wedge RefFood(RF) \wedge NonRefStorage(S) \wedge (T_1 < T_2)$.	Critical replacement: the patient has placed a food item that needs refrigeration inside a non-refrigerated cabinet.
2	$anomaly(nca, prepBF, T_1{+}45$ minutes$) \leftarrow startActivity(prepBreakfast, T_1) \wedge endActivity(prepBreakfast, T_2) \wedge ((T_2 - T_1) > 45$ minutes$)$.	Non-critical anomaly: the patient spent too much time to prepare breakfast.
3	$anomaly(co, medicine, T_2) \leftarrow prescribed(M, T_1, T_2) \wedge not((action(retrieve, M, C, T) \wedge MedCabinet(C) \wedge (T_1 \leq T \leq T_2)$.	Critical omission: the patient has not retrieved a prescribed medicine in due time.
4	$anomaly(wa, medicine, T) \leftarrow not(prescribed(M, T_1, T_2)) \wedge action(retrieve, M, C, T) \wedge MedCabinet(C) \wedge Medicine(M)$.	Wrong activity: the patient has taken a medicine that was not prescribed.

o' at time t. *Holds(s,o,t_1,t_2)* states that the status of object o has been "s" from t_1 to t_2 (for instance, "the microwave oven has been on from 11:30 to 11:55"). The Holds predicate allows us to express temporal conditions that are useful in the definition of different anomalies. Temporal expressions that we use in our rule-based definitions include the interval of time during which an action is performed, the temporal distance between two actions, the temporal duration of an activity, the temporal order among activities.

Abnormal behaviors are recognized by the KNOWLEDGE-BASED INFERENCE ENGINE, which periodically (e.g., at the end of each day) evaluates the rule-based anomaly definitions considering the data acquired and inferred during the considered time period: inferred activity instances and preprocessed events, as well as external knowledge including the medical prescriptions of the patient and the classification of objects in categories. Those data are represented by the aforementioned predicates, and added to the propositional logic knowledge base.

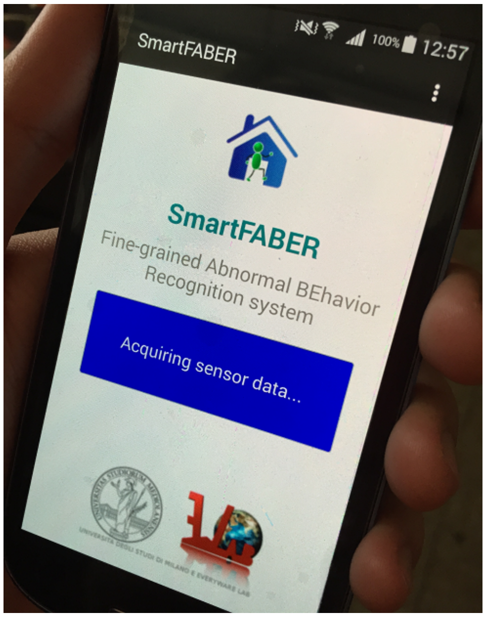
Example 1 *Consider an elderly person living independently at home. Suppose that furniture and devices, including food cabinets and the fridge, are equipped with a magnetic sensor to detect the* open *and* close *actions. An RFID tag is attached to some food boxes to identify their content (e.g., rice, milk, coffee, sugar). RFID readers in the proximity of the cabinets and fridge are in charge of detecting which item has been retrieved or returned. Suppose that at 08:05 AM the patient opens the fridge f and retrieves the milk box m to prepare breakfast. After a few minutes, he mistakenly puts the milk box in the non-refrigerated food cabinet c and closes its door. Hence, based on the sensed events, the following axioms are automatically added to the knowledge base:*

$$\begin{aligned}\text{action(open, door, } f\text{, 8:05:00 AM).}\\ \text{action(retrieve, } m\text{, } f\text{, 8:05:07 AM).}\\ \text{action(return, } m\text{, } c\text{, 8:12:30 AM).}\\ \text{action(close, door, } c\text{, 8:12:35 AM).}\end{aligned}$$

Since the knowledge base contains the axioms RefFood(m) *(stating that the milk box contains a food item that must be kept refrigerated) and* NonRefStorage(c) *(stating that c is a non-refrigerated storage), Rule* 1 *in Table 2 fires, recognizing an abnormal behavior.*

5 Experimental evaluation

We have developed a prototype of SmartFABER, and we have extensively evaluated our techniques with two large datasets of both normal and abnormal behaviors: one acquired in a smart home lab with actors simulating the daily routines of 21 patients, and one acquired during three months of experimentation in the home of an elderly diagnosed with MCI.



SmartFABER

Fine-grained Abnormal BEhavior Recognition system

Acquiring sensor data...

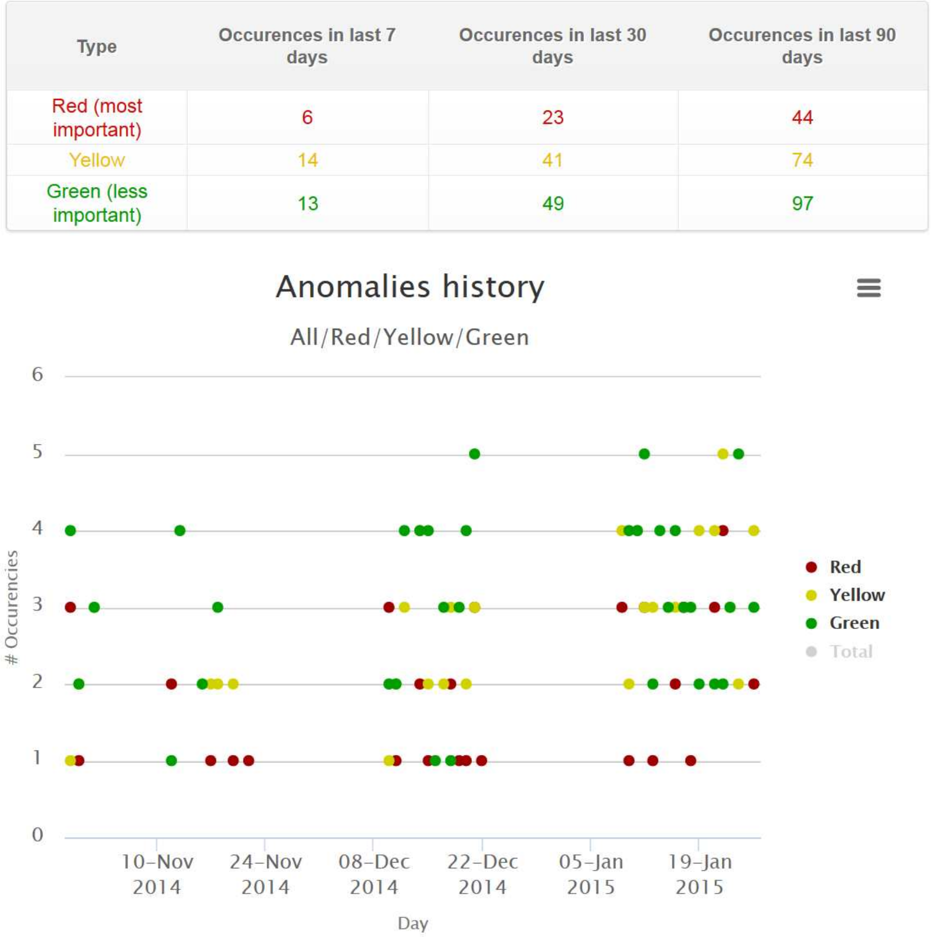
Figure 3: The SmartFABER app.

5.1 Hardware and software implementation

A prototype implementation of the whole system has been developed within the activities of the SECURE project. Since the SmartFABER system is intended to run on a mobile device at the patient's home, the core software modules have been implemented in Java for the Android platform. Figure 3 shows the application running the SmartFABER software. In particular, in order to implement the technique for activity recognition we have used the machine learning libraries of Weka[3]. In order to evaluate the rule-based definitions of anomalies we used the APIs of TuProlog [33], a lightweight Java implementation of an inference engine for the well-known Prolog logic programming language.

For this prototype, we use sensor motes available on the market, which communicate using the ZigBee protocol. Since currently there is no standard interface for that protocol on most Android devices, we use a gateway installed in the home to receive ZigBee messages from sensors and forward them via Bluetooth to the Android device. Sensor motes have been programmed in the C++ language to communicate new data to the gateway at the occurrence of each sensor event. For instance, the pressure sensor attached to the kitchen chair seat communicates the

[3] `http://www.cs.waikato.ac.nz/ml/weka/`



Type	Occurences in last 7 days	Occurences in last 30 days	Occurences in last 90 days
Red (most important)	6	23	44
Yellow	14	41	74
Green (less important)	13	49	97

Figure 4: SmartFABER Clinicians' dashboard.

measured pressure when it exceeds or falls behind given thresholds, to detect when the patient stands up or sits down on the chair. Such thresholds have been determined empirically. The sensor event message includes the timestamp of the sensor reading, the sensor ID and the detected value. A C++ application running on the gateway is in charge of: receiving data from sensors, assigning the unique timestamps, locally storing the data in a PostgreSQL database, and periodically communicating the data to the Android application. At the end of each day, the Android app runs the SmartFABER algorithms for activity boundary detection and anomaly recognition, and communicates the results through the Internet to the backend of the hospital center, where the data are stored.

We have also developed a Web-based dashboard, shown in Figure 4, to allow practitioners analyzing the history and trends of IADLs as well as the anomalies' history.

5.2 Datasets

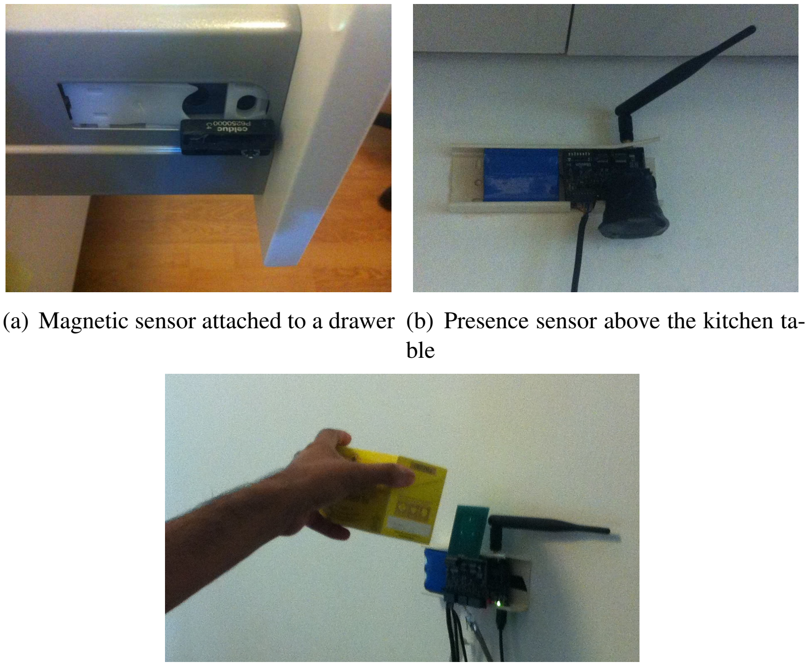
We experimented our method using two datasets; one acquired in a smart home laboratory, and one acquired in the instrumented home of a senior with an MCI diagnosis.

5.2.1 Smart home lab dataset

We have acquired a dataset of IADLs and anomalies, asking to voluntary actors to reproduce the daily routine of 21 elderly persons in our smart home lab. Executed IADLs and anomalies have been carefully designed in collaboration with neuroscience experts to realistically mimic the behavior of two groups: 7 healthy seniors (group 1), and 14 elderly persons with early symptoms of MCI (group 2). We assume that individuals of both groups live alone and independently in their respective homes. During their one-day routine, individuals in group 1 do not execute any critical anomaly, but may execute a few non-critical ones. Individuals in group 1 are mainly used to evaluate the number of false positives produced by our anomaly recognition method. Group 2 individuals may perform several non-critical and critical anomalies during the day.

During the execution of the daily routines, we have acquired the timestamped data coming from the sensors deployed in the smart home and manually annotated the dataset with the start- and end-time of specific activities and anomalies. The following IADLs have been selected to validate our method:

- Preparing food: the patient has to prepare the daily meals (breakfast, lunch, dinner) at appropriate times.



(a) Magnetic sensor attached to a drawer (b) Presence sensor above the kitchen table

(c) RFID reader for medicine boxes and food items

Figure 5: Some sensors used in the smart home lab.

- Consuming meal: when the patient prepares a meal, he has to consume it within a reasonable time period.
- Taking medicines: the patient has to take the prescribed medicines in the due time. We assume that no smart dispenser is used; instead, we assume that the patient keeps all the medicines in a dedicated cabinet.

We have considered the following anomalies:

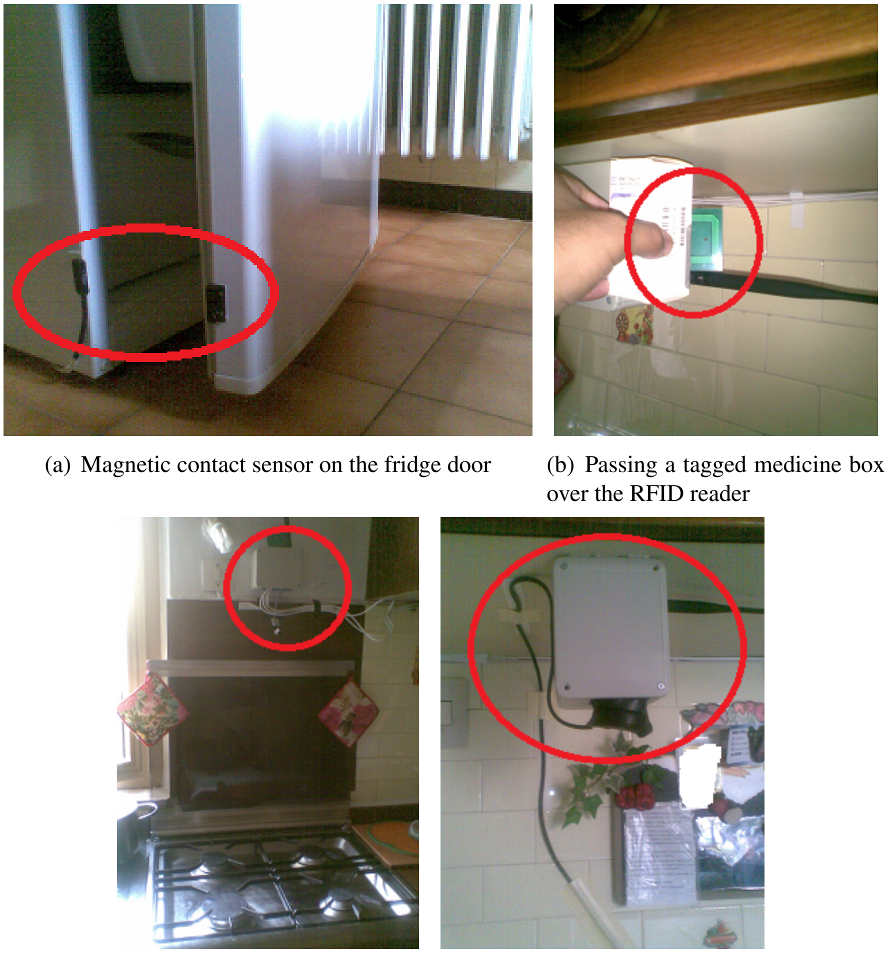
- Non-critical anomalies. They happen when the individual: (NC1) forgot a repository open; (NC2) did not return a medicine to its cabinet; (NC3) retrieved a food item which must be cooked, but did not use the stove burner; (NC4) does not prepare a meal.
- Critical anomalies. They happen when the individual: (C1) did not retrieve a prescribed medicine in the due time; (C2) took a medicine that was not prescribed; (C3) took a prescribed medicine in the due time but multiple times, resulting in inappropriate dosage; (C4) did not turn off the stove burner after finishing to prepare a hot meal; (C5) did not take the silverware before consuming meal; (C6) did not consume the meal after having prepared it; (C7) turned the stove burner on but did not take any cooking pan.

Overall, our dataset contains 21 days of IADLs and anomalies. Group 1 individuals did 7 non-critical and 0 critical anomalies; group 2 individuals did 24 non-critical anomalies and 36 critical ones.

5.2.2 Real home dataset

As a first step towards the evaluation of our methods in the actual home of elderly persons, we took advantage of our cooperation with a medical institution and a telemedicine company as partners of the SECURE project, and deployed our prototype inside the home of an elderly woman aged 74, with a diagnosis of MCI and medical co-morbidities, who lives alone. We will call her Mary in the following. Details about the technical implementation of the system in Mary's home are reported in [34].

We acquired a dataset consisting of 55 days of IADLs performed by Mary. In that period of time, we collected data for about 200 instances of activities. We considered the same type of IADLs as for the smart home lab dataset. For this experimentation, the clinicians provided us with a set of fine-grained abnormal behaviors to be detected, together with Mary's time prescriptions for meals (i.e., breakfast, lunch, and dinner) and medicines intakes. These anomalies are divided in three levels of seriousness:



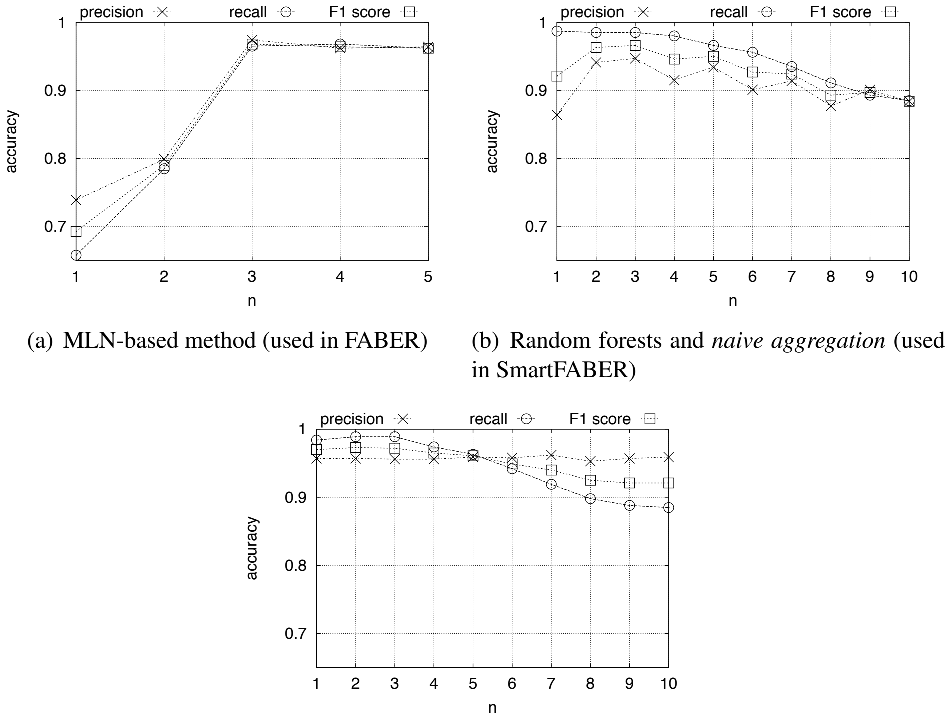
(a) Magnetic contact sensor on the fridge door

(b) Passing a tagged medicine box over the RFID reader

(c) A board with temperature sensor over the stove

(d) Passive infrared presence sensor over the kitchen table

Figure 6: Part of the sensors deployed at the elderly's home



(a) MLN-based method (used in FABER)

(b) Random forests and *naive aggregation* (used in SmartFABER)

(c) Random forests and *smart aggregation* (used in SmartFABER)

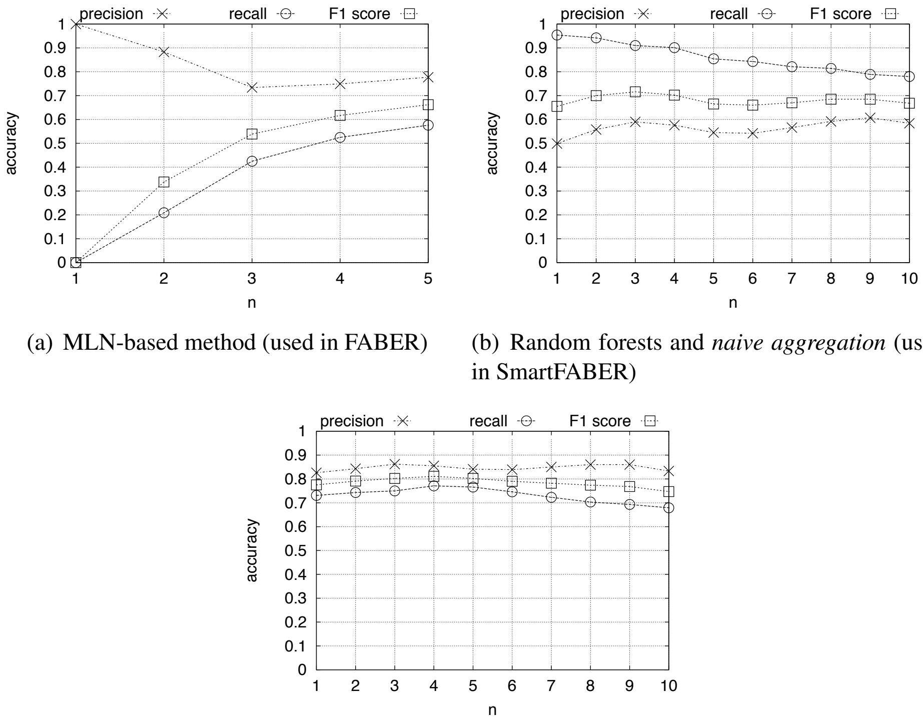
Figure 7: Smart home lab dataset. Accuracy of activity boundary detection; n is the length of the considered temporal sequence of sensor events

- Green anomalies (low level). This type of anomalies occur when the individual: prepares (G1) or consumes (G2) a meal at a different time than prescribed.

- Yellow anomalies (medium level). This type of anomalies occur when the individual: misses to consume (Y1) or prepare (Y2) a meal; takes a prescribed medicine outside the prescribed time (Y3); consumes (Y4) or prepares (Y5) the same meal multiple times during the same day.

- Red anomalies (high level). This type of anomalies occur when the individual: does not take a prescribed medicine (R1); takes a medicine that was not prescribed (R2).

Totally, 605 anomalies were detected, most of them being green and yellow ones.



(a) MLN-based method (used in FABER)

(b) Random forests and *naive aggregation* (used in SmartFABER)

(c) Random forests and *smart aggregation* (used in SmartFABER)

Figure 8: Real home dataset. Accuracy of activity boundary detection; n is the length of the considered temporal sequence of sensor events

For the sake of this project, it was not feasible to directly observe the execution of the activities, except for limited periods of time during the setup of the system, due to obvious privacy reasons. Hence, we manually labeled most of the activities offline, based on the observation of raw sensor data; this was possible since the considered activities are relatively easy to distinguish by a human observer based on the collected sensor readings. We labeled the anomalies by executing their respective rule-based definitions on the dataset of sensor events and labeled activities.

5.3 Recognition of activity boundaries

Recognizing (possibly approximate) activity boundaries is a prerequisite to detect anomalies. We performed an extensive experimental evaluation to compare Smart-

FABER with the preliminary method –named FABER– that was proposed in [3]. For both datasets, we applied those activity recognition techniques:

- The method used in FABER, which was based on Markov Logic Network (MLN) [35];
- The machine learning algorithm used in SmartFABER with the *naive aggregation* algorithm;
- The machine learning algorithm used in SmartFABER with the *smart aggregation* algorithm.

For each technique, we performed a leave-one-day-out cross-validation, evaluating the prediction's quality in terms of the standard measures of precision, recall and F_1 score (the latter is the harmonic mean of precision and recall). This type of cross-validation has also been preliminarily performed to select the most appropriate classifier for the machine learning module of SmartFABER. We selected a Random Forests classifier [36].

Since the anomaly recognition technique relies on detected activity boundaries, we first needed to experimentally choose the value of parameter n, corresponding to the length of the temporal sequence of sensor events to be used by our algorithms. The results of activity boundary recognition on the smart home lab and real home datasets are shown in Figures 7 and 8, respectively. With the smart home lab dataset, very positive results have been achieved (with F_1 score that exceeds 0.96) with all the three considered methods. With the MLN-based technique used in FABER, the highest recognition rate is achieved with $n = 3$. This means that, with this dataset, the temporal sequence of the 3 most recent sensor events is sufficient to reliably detect the start or end of an activity. This is due to the quite repetitive way in which activities have been executed in the lab; longer sequences of sensor events may be needed when activities are executed in more variable ways. Values of n lower than 3 produce worse results, while larger values strongly increase the execution times of the learning phase, without increasing recognition rates. With the new boundary detection method used in SmartFABER, the highest recognition rates are achieved using with $n = 2$ or $n = 3$. The *naive aggregation* and the *smart aggregation* methods achieve similar recognition rates; however, the former produces a larger number of false positives (i.e., detection of activity instances that did not actually occur), which may result in the recognition of several anomalies (especially repetitions of activities) that did not actually happen. On the contrary, the *smart aggregation* method provides more balanced and slightly higher values of precision and recall.

In general, with the real home dataset we achieve lower recognition rates. This is due to the intrinsic variability of activity execution in a real-world situation, with respect to the relatively stable activity execution patterns reproduced in the smart home lab. Moreover, the level of sensor noise due to missing or incorrect sensor readings is inevitably larger in a real home environment than in the lab. With the real home dataset, the MLN-based method used in FABER is the least effective among the ones that we evaluated. The highest recognition rates with MLN are achieved with $n = 5$. We were not able to test the performance with larger values of n, since the execution of the learning algorithm did not terminate in a reasonable amount of time. Independently from the value of n, the FABER method achieved particularly low values of recall; this could result in the detection of several false anomalies related to missed execution of expected activities. The technique used in SmartFABER achieves better results. In particular, the *smart aggregation* method leads to the highest values of F_1 score (slightly above 0.8 with $n = 4$), and very well balanced values of precision and recall. The *naive aggregation* method achieves lower recognition rates, producing a large number of false positives.

Summarizing, with both datasets, the *smart aggregation* algorithm of SmartFABER reduces the number of false positives with respect to the FABER method, and improves the overall accuracy.

5.4 Recognition of fine-grained anomalies

We performed experiments about anomaly recognition using the activity boundaries detected by the different techniques with the most effective value of n.

Table 3: Smart home lab dataset. Accuracy of abnormal behavior recognition with FABER

ANOMALY	GROUP 1			GROUP 2		
	TP	FP	FN	TP	FP	FN
NC1: Repository left open	5	0	2	18	0	0
NC2: Medicine not returned	0	0	0	4	0	0
NC3: Food item not cooked	0	0	0	2	0	0
NC4: Meal not prepared	0	2	0	0	1	0
C1: Missed a prescr. medicine	0	2	0	10	0	0
C2: Took a wrong medicine	0	0	0	7	0	0
C3: Repeated medicine intake	0	0	0	3	0	0
C4: Stove burner left on	0	0	0	0	0	0
C5: Had meal with no silverware	0	0	0	7	0	0
C6: Prepared meal not consumed	0	0	0	1	1	0
C7: Burner turned on by mistake	0	0	0	8	0	0
TOTAL	5	4	2	60	2	0

Table 4: Smart home lab dataset. Accuracy of abnormal behavior recognition with SmartFABER and *smart aggregation*

ANOMALY	GROUP 1			GROUP 2		
	TP	FP	FN	TP	FP	FN
NC1: Repository left open	7	0	0	18	1	0
NC2: Medicine not returned	0	0	0	4	0	0
NC3: Food item not cooked	0	0	0	2	0	0
NC4: Meal not prepared	0	0	0	0	0	0
C1: Missed a prescr. medicine	0	0	0	10	0	0
C2: Took a wrong medicine	0	0	0	7	0	0
C3: Repeated medicine intake	0	0	0	3	0	0
C4: Stove burner left on	0	0	0	0	0	0
C5: Had meal with no silverware	0	0	0	7	0	0
C6: Prepared meal not consumed	0	0	0	1	0	0
C7: Burner turned on by mistake	0	0	0	8	1	0
TOTAL	7	0	0	60	2	0

Table 5: Smart home lab dataset. Results of fine-grained abnormal behavior recognition based on different boundary detection methods

TECHNIQUE	PRECISION	RECALL	F1 SCORE
FABER (MLN)	0.915	0.97	0.942
SmartFABER-SmartAggregation	**0.971**	**1**	**0.985**

Detailed results with the smart home lab dataset are shown in Tables 3 and 4. The results of SmartFABER with the *naive aggregation* method are identical to the ones obtained with the *smart aggregation* method; hence, we omit them. Each row of the table corresponds to a specific anomaly considered in our experiments. The TP column reports the number of true positives for that anomaly; i.e., the number of actual occurrences of that anomaly that were recognized by the technique. FP reports the number of false positives; i.e., the number of anomalies reported by the technique that did not actually occur. FN reports the number of actual occurrences of that anomaly that were not recognized by the technique.

As anticipated, group 1 individuals performed a few non-critical anomalies (NC) and no critical anomaly (C). FABER correctly recognized 5 NCs out of 7. During the 7 days activities of group 1 individuals, the system did 4 false positives. Two of them regarded NC4 (meal not prepared), while the other two regarded C1 (missed a prescribed medicine). Those errors were due to mispredictions of the MLN-based activity boundary detection technique, which in two cases did not recognize the occurrence of activity "preparing meal" and in two cases did not recognize the occurrence of "taking medicine". With SmartFABER we were able to avoid the occurrence of both false positives and false negatives. This result was due to the high accuracy of activity recognition in the smart home lab environment.

Group 2 individuals performed a larger number of NCs and several Cs. For this group, FABER correctly recognized all the occurrence of both critical and non-critical anomalies; i.e., no false negatives happened. During the 14-days activities of that group, the system reported only 2 false positives: one was related to NC4 (meal not prepared) and the other one to C6 (prepared meal not consumed). Even in these cases, false positives were due to mispredictions of the activity boundary detection technique. The performance of SmartFABER was comparable: all the anomalies were recognized, and only two false positives occurred.

Overall, FABER produced 6 false positives during the 21-days activities, while SmartFABER produced only 2 false positives. We claim that the number of false positives is compliant with the requirements of clinicians, especially considering that the individuals totally performed more than 150 instances of activities during the experimentation. Table 5 summarizes the results in terms of precision, recall and F1 score.

Table 6: Real home dataset. Accuracy of abnormal behavior recognition with FABER

ANOMALY	TP	FP	FN	PREC.	REC.	F1 SCORE
G1	37	10	68	0.787	0.352	0.487
G2	36	21	51	0.632	0.414	0.5
Y1	60	46	7	0.566	0.896	0.694
Y2	41	69	0	0.373	1	0.543
Y3	73	4	15	0.948	0.83	0.885
Y4	1	3	5	0.25	0.167	0.2
Y5	0	2	0	0	/0	/0
R1	29	4	14	0.879	0.674	0.763
R2	164	16	4	0.911	0.976	0.943
TOTAL	441	175	164	0.716	0.729	0.722

Results regarding the real home dataset are shown in Tables 6, 7, and 8, and summarized in Table 9. The SmartFABER method with *smart aggregation* achieves the best results in terms of F_1; moreover, the measures of precision and recall are well balanced. Those measures are less balanced using the *naive aggregation* method, which achieves high recall (only a few actual anomalies were not recognized) at the price of low precision (several anomalies have been predicted by mistake). This result depends on the fact that the *naive aggregation* method produces a relatively large number of false positives, which fire the recognition of several

Table 7: Real home dataset. Accuracy of abnormal behavior recognition with SmartFABER and *naive aggregation*

ANOMALY	TP	FP	FN	PREC.	REC.	F1 SCORE
G1	81	22	24	0.786	0.771	0.779
G2	64	54	23	0.542	0.736	0.624
Y1	57	9	10	0.864	0.851	0.857
Y2	29	8	12	0.784	0.707	0.744
Y3	84	3	4	0.966	0.955	0.96
Y4	5	182	1	0.027	0.833	0.052
Y5	0	40	0	0	/0	/0
R1	40	1	3	0.976	0.93	0.952
R2	167	3	1	0.982	0.994	0.988
TOTAL	527	322	78	0.621	0.871	0.725

Table 8: Real home dataset. Accuracy of abnormal behavior recognition with SmartFABER and *smart aggregation*

ANOMALY	TP	FP	FN	PREC.	REC.	F1 SCORE
G1	70	59	35	0.543	0.667	0.598
G2	47	38	40	0.553	0.54	0.547
Y1	60	17	7	0.779	0.896	0.833
Y2	32	17	9	0.653	0.78	0.711
Y3	75	6	13	0.926	0.852	0.888
Y4	0	0	6	/0	0	/0
Y5	0	0	0	/0	/0	/0
R1	39	0	4	1	0.907	0.951
R2	167	17	1	0.908	0.994	0.949
TOTAL	490	154	115	0.761	0.81	0.785

Table 9: Real home dataset. Results of fine-grained abnormal behavior recognition based on different boundary detection methods

TECHNIQUE	PRECISION	RECALL	F1 SCORE
FABER (MLN)	0.716	0.729	0.722
SmartFABER-SimpleAggregation	0.62	**0.871**	0.725
SmartFABER-SmartAggregation	**0.76**	0.81	**0.785**

anomalies about repetition of activities (Y4 and Y5). The F_1 results with FABER and the MLN-based method are comparable with those of SmartFABER with *naive aggregation*, but precision and recall are better balanced with the former.

Overall, we can conclude that SmartFABER with *smart aggregation* achieves the best results in terms of F_1, obtaining a good balance between precision and recall. A preliminary clinicians' assessment of our system can be found in [34].

6 Conclusions and future work

In this paper, we addressed the challenging issue of unobtrusively recognizing behaviors exhibited by elderly persons at home that have been identified by clinicians as relevant for the early diagnosis of MCI. Our SmartFABER hybrid technique to recognize abnormal behaviors differs from previous approaches for combining supervised learning with knowledge-based reasoning to more precisely recognize specific anomalies in carrying out daily living activities. We designed the models of anomalies collaborating with cognitive neuroscience experts. Hence, instead of identifying only generic deviations from normal behavior as most related works do, we provide clinicians with a fine-grained description of the recognized abnormal behaviors identified as indicators of MCI. We implemented a prototype of the system in both a smart home lab and in the real home of an elderly person. Experiments with datasets of activities and anomalies show that SmartFABER achieves high recall while generating a small number of false positives.

The achieved results are promising, but we plan to improve this work in several directions. Our current anomaly recognition method is based on logic rules that strictly determine the detection of an abnormal behavior based on a user-defined set of observations. We consider extending this rigid system with probabilistic reasoning. We are working on integrating data analysis tools in the dashboard offered by the system to clinicians to automatise some of the reasoning that they currently do by looking at long term activity data. Other future work also includes addressing the case of multi-inhabitants. Finally, we plan to work closely with clinicians to extend the set of activities and associated significant anomalies to be monitored, and to extend the experiments to multiple real homes.

Acknowledgments

This work has been partially supported by the project "SECURE: Intelligent System for Early Diagnosis and Follow-up at Home", funded by a grant of Lombardy Region and Italian Ministry of Education, University and Research.

References

[1] B. Rechel, E. Grundy, J.-M. Robine, J. Cylus, J. P. Mackenbach, C. Knai, and M. McKee, "Ageing in the European Union," *The Lancet*, vol. 381, no. 9874, pp. 1312–1322, 2013.

[2] B. Winblad et Al., "Mild cognitive impairment - beyond controversies, towards a consensus," *J Intern Med*, vol. 256, no. 3, pp. 240–246, 2004.

[3] D. Riboni, C. Bettini, G. Civitarese, Z. H. Janjua, and R. Helaoui, "Fine-grained recognition of abnormal behaviors for early detection of mild cognitive impairment," in *Proc. of PerCom'15*. IEEE Comp. Soc., 2015, pp. 149–154.

[4] M. C. A. Klein, N. M. Mogles, and A. van Wissen, "Intelligent mobile support for therapy adherence and behavior change," *J Biomed Inform*, vol. 51, pp. 137–151, 2014.

[5] C. Parsey and M. Schmitter-Edgecombe, "Applications of technology in neuropsychological assessment," *Clin Neuropsychol*, vol. 27, no. 8, pp. 1328–1361, 2013.

[6] J. Ye, S. Dobson, and S. McKeever, "Situation identification techniques in pervasive computing: A review," *Pervasive Mob Comput*, vol. 8, no. 1, pp. 36–66, 2012.

[7] O. Aziz, L. Atallah, B. P. L. Lo, E. Gray, T. Athanasiou, A. Darzi, and G. Yang, "Ear-worn body sensor network device: an objective tool for functional postoperative home recovery monitoring," *J Am Med Inform Assoc*, vol. 18, no. 2, pp. 156–159, 2011.

[8] H. Kalantarian, N. Alshurafa, T. Le, and M. Sarrafzadeh, "Monitoring eating habits using a piezoelectric sensor-based necklace," *Comput Biol Med*, vol. 58, pp. 46–55, 2015.

[9] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Proc. of Pervasive Computing*, ser. Lecture Notes in Computer Science, vol. 3001. Springer, 2004, pp. 1–17.

[10] R. Paradiso, G. Loriga, and N. Taccini, "A wearable health care system based on knitted integrated sensors," *IEEE Trans Inf Technol Biomed*, vol. 9, no. 3, pp. 337–344, 2005.

[11] J. Lester, T. Choudhury, N. Kern, G. Borriello, and B. Hannaford, "A hybrid discriminative/generative approach for modeling human activities," in *Proc. of IJCAI*, 2005, pp. 766–772.

[12] O. Amft and G. Tröster, "Recognition of dietary activity events using on-body sensors," *Artif Intell Med*, vol. 42, no. 2, pp. 121–136, 2008.

[13] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, Eds., *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003.

[14] A. Agostini, C. Bettini, and D. Riboni, "Experience report: Ontological reasoning for context-aware internet services," in *Proc. of IEEE PerCom Workshops*. IEEE Comp. Soc., 2006, pp. 8–12.

[15] M. Richardson and P. Domingos, "Markov logic networks," *Mach Learn*, vol. 62, pp. 107–136, 2006.

[16] R. Helaoui, D. Riboni, and H. Stuckenschmidt, "A probabilistic ontological framework for the recognition of multilevel human activities," in *Proc. of UbiComp*. ACM, 2013, pp. 345–354.

[17] F. D. Backere, F. Ongenae, F. V. D. Abeele, J. Nelis, P. Bonte, E. Clement, M. Philpott, J. Hoebeke, S. Verstichel, A. Ackaert, and F. D. Turck, "Towards a social and context-aware multi-sensor fall detection and risk assessment platform," *Comput Biol Med*, vol. to appear.

[18] T. L. Hayesa, F. Abendroth, A. Adami, M. Pavel, T. A. Zitzelberger, and J. A. Kaye, "Unobtrusive assessment of activity patterns associated with mild cognitive impairment," *Alzheimers Dement*, vol. 4, no. 6, pp. 395–405, 2008.

[19] A. M. Seelye, M. Schmitter-Edgecombe, D. J. Cook, and A. Crandall, "Naturalistic assessment of everyday activities and prompting technologies in mild cognitive impairment," *J Int Neuropsychol Soc*, vol. 19, no. 4, pp. 442–452, 2013.

[20] P. N. Dawadi, D. J. Cook, M. Schmitter-Edgecombe, and C. Parsey, "Automated assessment of cognitive health using smart home technologies," *Technol Health Care*, vol. 21, no. 4, pp. 323–343, 2013.

[21] P. Dawadi, D. J. Cook, and M. Schmitter-Edgecombe, "Automated cognitive health assessment using smart home monitoring of complex tasks," *IEEE Trans Syst Man Cybern*, vol. 43, no. 6, pp. 1302–1313, 2013.

[22] P. Dawadi, D. J. Cook, C. Parsey, M. Schmitter-Edgecombe, and M. Schneider, “An approach to cognitive assessment in smart home,” in *Proc. of the 2011 DMMH Workshop*. ACM, 2011.

[23] D. J. Cook and M. Schmitter-Edgecombe, “Assessing the quality of activities in a smart environment,” *Methods Inf Med*, vol. 48, no. 5, pp. 480–485, 2009.

[24] G. Virone, N. Noury, and J. Demongeot, “A system for automatic measurement of circadian activity deviations in telemedicine,” *IEEE Trans Biomed Engineering*, vol. 49, no. 12, pp. 1463–1469, 2002.

[25] M. Skubic, R. D. Guevara, and M. Rantz, “Testing classifiers for embedded health assessment,” in *Proceedings of ICOST*, ser. Lecture Notes in Computer Science, vol. 7251. Springer, 2012, pp. 198–205.

[26] F. Duchêne, C. Garbay, and V. Rialle, “Learning recurrent behaviors from heterogeneous multivariate time-series,” *Artif Intell Med*, vol. 39, no. 1, pp. 25–47, 2007.

[27] V. Jakkula and D. J. Cook, “Anomaly detection using temporal data mining in a smart home environment,” *Methods Inf Med*, vol. 47, no. 1, pp. 70–75, 2008.

[28] P. Rashidi and D. J. Cook, “COM: A method for mining and monitoring human activity patterns in home-based health monitoring systems,” *ACM Trans Intell Syst Technol*, vol. 4, no. 4, p. 64, 2013.

[29] J. M. Juarez, J. M. Ochotorena, M. Campos, and C. Combi, “Spatiotemporal data visualisation for homecare monitoringof elderly people,” *Artif Intell Med*, vol. to appear.

[30] A. Burns, B. Lawlor, and S. Craig, “Rating scales in old age psychiatry,” *Br J Psychiatry*, vol. 180, pp. 161–7, 2002.

[31] N. C. Krishnan and D. J. Cook, “Activity recognition on streaming sensor data,” *Pervasive Mob Comput*, vol. 10, pp. 138–154, 2014.

[32] P. Norvig and S. Russell, *Artificial Intelligence. A Modern Approach*. Prentice Hall Series in Artificial Intelligence, 2003.

[33] E. Denti, A. Omicini, and A. Ricci, “Multi-paradigm java-prolog integration in tuprolog,” *Sci Comput Program*, vol. 57, no. 2, pp. 217–250, 2005.

[34] D. Riboni, C. Bettini, G. Civitarese, Z. H. Janjua, and V. Bulgari, "From lab to life: Fine-grained behavior monitoring in the elderly's home," in *Proc. of PerCom Workshops*. IEEE Comp. Soc., 2015.

[35] M. Richardson and P. Domingos, "Markov logic networks," *Mach Learn*, vol. 62, pp. 107–136, 2006.

[36] L. Breiman, "Random forests," *Mach Learn*, vol. 45, no. 1, pp. 5–32, 2001.