Non-Intrusive human activity recognition and abnormal behavior detection on elderly people: a review

Athanasios Lentzas[1] and Dimitris Vrakas[1]

[1] School of Informatics, Aristotle University of Thessaloniki, Greece
`[alentzas, dvrakas]@csd.auth.gr`

Abstract. With the world population aging at a fast rate, ambient assisted living systems focused on elderly people gather more attention. Human activity recognition (HAR) is a component connected to those systems, as it allows identification of the actions performed and their utilization on behavioral analysis. This paper aims to provide a review on recent studies focusing on HAR and abnormal behavior detection specifically for seniors. The frameworks proposed in the literature are presented. The results are also discussed and summarized, along with the datasets and metrics used. The absence of a universal evaluation framework makes direct comparison not feasible, thus an analysis is made trying to divide the literature using a taxonomy. Solutions on the challenges identified are proposed, while discussing future work.

Keywords: Human activity recognition (HAR), abnormal behavior detection, elderly ambient assisted living, non-intrusive human activity recognition

1   Introduction

Modern societies, especially in developed countries, are aging at a high rate. In 2017, 13 percent of world population was over 60 years old - approximately 962 million people - expecting an annual growth at a rate of about 3 percent (United Nations, Department of Economic and Social Affairs, 2017). High income countries will experience an increase of 5.6 percent of people aged 60 or more, by 2030, while upper-middle income countries will face a higher increase rate of 7.8 percent (United Nations, Department of Economic and Social Affairs, 2015). United Nations projections estimate that the global population will increase to 9.8 billion by 2050, with the population aged more than 60 years old, being roughly the same as children under 15, approximately 2.1 billion each. Additionally, the increasing life expectancy will raise the number of people aged more than 80 from 137 million in 2017 to 909 million by 2100.

Older adults are more susceptible to body function disorders or age-related diseases, both physical and mental. More than 20 percent of adults aged 60 or more, suffer from a mental disorder, with the most common being dementia and depression (World Health Organization, 2017). Specifically, mental conditions are insufficiently identified by health professionals and patients, with the latter usually being afraid to ask for assistance since those diseases are often stigmatized. Dementia patients are expected to

triple by 2050, while the health care costs for that particular disease will be approximately 10% of the total expected health care costs increase in the next 20 years (World Health Organization, 2015). The need for smart patient observation systems is therefore needed, since monitoring by humans is not only costly but also inefficient with the aging population constantly increasing. A plethora of Ambient Assisted Living systems for older adults have been proposed to the literature (Rashidi & Mihailidis, 2013), including robot assistants (Yusuf, Woo, Botzheim, Kubota, & Tudjarov, 2017), patient monitoring systems etc. Human activity recognition is a key factor for the effectiveness of these solutions. Although most of them are designed having elderly people in mind, they can also be used in mentally/physically impaired persons regardless their age. As a consequence, human activity recognition has been extensively studied in the literature.

Activity recognition aims to identify the actions and goals of agents acting in an environment. The purpose of human activity recognition (HAR), is the identification of common human activities in real life scenarios. Wearables, as Mukhopadhyay (2014) presented in his review, and a variety of sensors, like Radio Frequency Identification (RFID) tags or pressure sensors, have been used in order to recognize activities of daily living, in the field of elderly people monitoring. With the rapid expansion of smartphones and wearables with embedded gyroscope, accelerometer and GPS, as well as cheaper sensing equipment becoming available publicly, HAR systems have been considered more as a smart home or out of hospital patient monitoring solution.

Another aspect of HAR systems is the ability to detect abnormal behaviors. Every individual, especially older adults, develop patterns in their living. That daily routine could be recognized using HAR. Any variations could be detected and the system could alert their relatives or caregivers. A typical example is mild cognitive impairment, a clinical condition causing cognitive changes (such as memory loss), often leading to dementia or Alzheimer disease (Petersen et al., 1999). Dementia, for instance, has specific symptoms that a person could develop such as, difficulties to perform motor activities, identify objects, difficulties for abstract reasoning, valid judgement, reduced speaking ability, or verbal / written language understanding (Thies & Bleiler, 2013). Most of these symptoms could be detected by observing a person's activities of daily living and creating a behavioral model. Although aforementioned conditions are not curable, early detection could lead to proper treatment, hindering them, thus reducing treatment costs, and promote independent living for elders. There are also evidence that smart home installations with the ability to perform behavioral anomaly detection, could prevent, in some cases, hospitalization (Kornowski et al., 1995).

The purpose of this paper is to review the current advancements on the HAR / abnormal behavior detection area. The first section describes a typical HAR system installed in an ambient assisted living environment, presenting the main components and their role inside the system. Anomaly detection in elderly behavior is also described, as well as how it uses HAR data to extract a behavioral model and identify abnormalities. The third section presents current work on the area and compares systems based on their design. Proposed solutions are broken down to sensors used, activities recognized and whether the recognition is offline or online. The fourth section describes the way HAR is realized, diving into the feature selection and extraction techniques used, the classification method and architecture of proposed systems. The fifth section presents

an evaluation, experimental analysis and an empirical analysis of state-of-the-art approaches. Additionally the current literature is divided using a taxonomy based on sensors used, i.e. wearables, ambient sensors (Lara & Labrador, 2013), type of recognition (HAR, abnormal behavior detection and whether the recognition is performed locally or on a remote server. Lastly, conclusion and future work are drawn.

2 Human Activity Recognition / Abnormal Behavior Detection

The purpose of a Human Activity recognition system is to correctly identify human actions in real time and inform the agents interested in those actions. HAR was firstly realized by attaching lights on major joints of a person dressed in black (Johansson, 1973). Johansson used orthographic projection to determine the structure of body rigid parts, where each part was represented by two points. Apart from activities performed, an HAR system can also be used to recognize the interaction between humans and objects as well as humans and humans. This information could provide additional insight about the action's context.

As a core concept of psychology, abnormality refers to deviation from any social norm, typical behavior, cultural or ethical expectations. Anomaly detection is the identification of patterns that do not comfort with the model created (Chandola, Banerjee, & Kumar, 2009). Using the results obtained from activity recognition the aforementioned model can be generated for each individual.

2.1 Human Activity Recognition Framework

HAR systems typically follow an architecture referred as HAR framework. The framework describes the basic components such a system needs in order to achieve its goal. This framework on its simplest form, as shown in Figure 1, consists of a sensory medium, a processing unit and a display to present the results.

a) Sensory medium

A variety of sensors have been proposed in the literature, each providing data to recognize different activities. Wearables and smartphones are already capable of performing activity recognition tasks, i.e. smartwatches tracking physical exercise (jogging, running, stair climbing etc.). That particular feature, as well as the variety of sensors embedded in those devices, made them popular in activity recognition field (Kumari, Mathew, & Syal, 2017). Apart from smart phones and wearables, other sensory devices include: Radio Frequency Identification (RFID) networks (G. Hu, Qiu, & Meng, 2016; Yao et al., 2017), pressure sensors, force sensors etc. The wearable and ambient sensors are summarized on Table 1. A typical HAR system installed in a house can collect readings from various sensors, creating a sensing network.
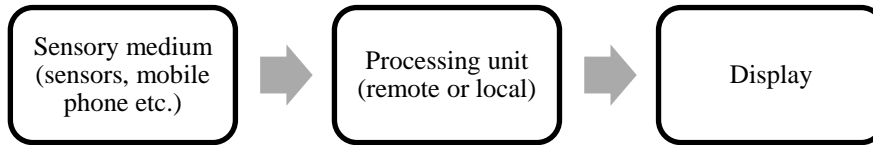
Figure 1. Simple HAR system steps

Sensor type	Sensor
Mobile / wearable sensors	Accelerometer
	Gyroscope
	Heartrate
	Body temperature
	GPS
Ambient sensors	Pressure
	Humidity
	Infrared
	Radio frequency identification
	Magnetic switches (door/cabinet opening/closing
	Ultrasonic
	Force
	Contact
	Electric power usage
	Light
	Temperature

Table 1. Sensors used in HAR

Data collection, using the aforementioned sensors, is the first step of every HAR system. Depending on the purpose of the system, a different sensing network is used. For instance, if only simple activities should be recognized (i.e. walking, standing, sitting) accelerometer and gyroscope data are enough to identify them. On the other hand, when one needs to recognize more complex activities (i.e. drinking water, using appliances etc.), the sensing network should include sensors that can expose information about those actions. Raw data collected, are then send to a processing unit.

b) Processing unit

The processing unit is responsible for data manipulation, feature extraction and selection - if applicable - and activity recognition. These procedures could either happen offline on a local device (i.e. smartphone) or online (i.e. remote server). The data manipulation step includes noise removal and data preprocessing (e.g. raw data representation). Features used in activity recognition can be divided into three categories, time domain, frequency domain and physical features. A variety of features are available for

use on HAR domain. Using more features though, doesn't always lead to better classification results since based on the type of classification, some features may be redundant or irrelevant. Some of the feature selection methods used in activity recognition are ReliefF, an algorithm ranking features based on relevance (Robnik-Šikonja & Kononenko, 2003) and Correlation based Feature Selection (CFS), a technique using a heuristic evaluation function based on correlation to rank feature subsets (Hall, 1999). Feature selection, extraction, such as Principal Component Analysis (PCA), as well as classification are further discussed in section four.

c) Display

The classification results (i.e. the recognized activity), can be communicated to the user using different ways. Depending on the design of the system, an on-screen display can be used, a vocal message etc. There are cases were results are available through an API. This is useful in cases the recognized activity should be used as an input to a separate component, such as a behavior modelling system.

When designing an HAR system, there are specific aspects to take into consideration. Obtrusiveness is an important factor when implementing activity recognition, especially in elderly people. Configurations that require users to carry equipment, or wear sensors, could be invasive, expensive and create discomfort. Lara et al. (2013) also brought system's energy consumption into attention. Processing, communication and visualization require energy with the former being the most power demanding task. That's the reason data are preferably transmitted using short range communication (low energy Bluetooth or Wi-Fi). Another characteristic, is where the recognition is done. Recognition done on a server (online) allows the implementation of more complex models as well as higher storage capabilities. On the other hand, offline recognition, like a mobile phone, may provide less computational power and limited battery, but reduces energy consumption for constantly transferring raw data as well as system's response time. Whether the recognition will be online or offline depends heavily on the classification method used, i.e. bagging (Witten, Frank, & Hall, 2011) requires a lot of resources during evaluation phase, making it unsuitable for offline HAR.

2.2 Abnormal Behavior Detection

Abnormal behavior can be defined as *actions that are unexpected and often evaluated negatively because they differ from typical or usual behavior* (Durand & Barlow, 2003). Five main criteria have been proposed to identify abnormality: statistical criterion, social criterion, personal discomfort, maladaptive behavior and deviation from ideal (Rosenhan & Seligman, 1984). Usually a multi criteria approach is used to define a behavior as abnormal. A statistical infrequent behavior, for example, that prevents the person from "normal functioning" and breaks the social norm will be categorized as abnormal.

An abnormal behavior detection system deployed in ambient assisted living environments, used a behavioral model created by observing the usual activities performed. As soon as the model was generated, each activity was compared against that model to detect the likelihood the person would perform that activity (Aran, Sanchez-Cortes, Do, & Gatica-Perez, 2016).

Apart from a behavioral model, a scenario based system was also proposed in the literature (Amiribesheli & Bouchachia, 2015). Initially, dementia symptoms proposed in the literature were collected. Using those symptoms, a set of scenarios was developed. Out of the initial set only thirteen scenarios were chosen after validation by social caregivers and dementia specialists. Each user's activity was validated using those scenarios for abnormality detection. Different cases had different interventions, such as invite awareness, suggestion, prompts, urges and performs.

Another approach for identification of behavioral divergence, was the use of a rules modeling system (Riboni, Bettini, Civitarese, Janjua, & Bulgari, 2015). First-order logic formulae were used to create a knowledge base, in which axioms were added based on the performed activity. Rules were specified to evaluate the abnormality of an activity or a group of activities. The rules were represented using propositional logic. For example, D. Riboni et al. (2015) in their work, defined the predicate *anomaly (Categ, Obj, Time),* to represent an anomaly, with *Categ* being the category of the anomaly, *Obj* the object involved in the activity and *Time* the instant the anomaly happened.

2.3   Evaluation metrics

Evaluation of an HAR system is a complex task. A well-defined framework that could be used to compare multiple techniques is not available. Depending on the sensors used to collect raw data, different datasets are used to train the classifiers. Even when the sensors employed are the same, there may be differences on the activities recognized. The number of activities identified, as well as similar activities e.g. standing and sitting, could affect the performance, resulting in a big number of misclassified instances.

Metrics used on binary classification problems can be generalized for a problem with n classes. Each metric is calculated for a single class, and the mean for all classes is calculated, for instance when walking is the activity recognized, all instances of walking will be positive with all other instances negative.

Performing cross validation leads to the generation of a confusion matrix. Confusion matrix is a matrix CM_{nxn} in a n class problem, where the $CM_{i,j}$ element is the number of examples that belong to class i and classified as instances of class j. Given that matrix the following numerical values can be extracted:

1. True Positives (TP): correctly classified positive instances
2. True Negatives (TN): correctly classified negative instances
3. False Positives (FP): misclassified negative instances (classified as positives)
4. False Negatives (FN): misclassified positives instances (classified as negatives)

Using the above information several metrics used in HAR can be calculated:
   a)   Accuracy, one of the most standard metrics, represents the percentage of correctly classified examples - positive and negatives (1).

b) Precision and Recall are the ratio of correctly classified instances to the total positive classified examples (2) and the total positive instances respectively (3).

c) F-score is the combination of Precision and Recall in a single metric (4).

It is worth mentioning, that using the mean of a metric, to evaluate the performance of a system may lead to biased measurements (Forman & Scholz, 2009). Forman and Scholz presented an extensive analysis on F-score, Recall and Precision, and a novel way to calculate those metrics. Instead of calculating the mean, the sum of True Positives, True Negatives, False Positives and False Negatives were used to calculate the metrics. The authors concluded that their method was more unbiased and should be used instead of average F-score.

When evaluating abnormal behavior detection two more metrics are used: True Positive Rate (TPR) and False Positive Rate (FPR). The former represents the percentage of correctly detected anomalies out of total number of anomalies and the later the percentage of normal activities detected as anomalies out of total number of normal activities.

$$Accuracy = \frac{\sum TP + \sum TN}{\sum Examples} \quad (1) \qquad Precision = \frac{\sum TP}{\sum TP + \sum FP} \quad (2)$$

$$Recall = \frac{\sum TP}{\sum TP + \sum FN} \quad (3) \qquad F_{score} = 2\frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

2.4   Datasets

Although there are many available datasets used on activity recognition, the majority of them were not collected from elders. As older adults tend to have different motor patterns, it is crucial to have data originating from them. A lot of published research papers, use data gathered specific for their systems. Nonetheless, there are some publicly available datasets.

One of the most commonly used datasets was published by Shinmoto Torres, Ranasinghe, Shi, & Sample (2013). This dataset was obtained by 14 older subjects between 66 and 86 years old. All of them had a wearable sensor attached to their chest (on top of clothing). The attached sensor was an RFID tag equipped with a 3D accelerometer and a microprocessor. Additionally, three or four antennae were placed around the room directed to areas associated with higher risk of falling, such as bed, chair and open area. Apart from the acceleration data the received signal strength indicator (RSSI) was also logged along with the closest antenna. Subjects performed a predefined sequence of actions: lying to sitting, sitting to standing and walking.

Ojetola, Gaura, & Brusey (2015) published a dataset for fall events and daily activities. The subjects used were between 18 and 51 years old. Although data from elderly were not available, authors simulated real life conditions inside laboratory environment. Inertial sensors (i.e. accelerometers) were used to gather the data. More specifically, acceleration and orientation data were obtained from devices strapped to the chest and

thigh of subjects. Subjects were asked to perform two different routines. The first routine simulated falls, walking, sitting and standing while the second one simulated ascending and descending of stairs.

Except from datasets directly obtained from elders or simulating elderly behavior, there are available data that were gathered using both younger and older adults. The RealWorld HAR dataset (Sztyler & Stuckenschmidt, 2016), used such an approach. The older subject was 62 years old while the younger 26. All of the subjects performed predefined drills, with wearables device attached on the head, chest, upper arm, waist, forearm, thigh and shin. All devices were equipped with accelerometer and gyroscope data. From the available data, time and frequency domain features were extracted. It is worth mentioning that the experiments, on the contrary with most datasets, were not conducted on a laboratory, but on realistic conditions (i.e. walking in the city, jogging in a forest etc).

Similar to the RealWorld HAR dataset, the HealthyLife dataset (Do, Loke, & Liu, 2013) obtained data from people range between 6 and 67 years old. All data were gathered from a single 3D accelerometer (smartphone) carried in different ways (pants pockets, jacket pockets, hand-held, shoulder bags). As subjects were left free to perform any activity, they were request to keep a log diary of the action they performed at a time in order to later annotate the data. Given only the accelerometer data, a basic set of activities could be recognized. Authors also included GPS data for each activity in case a more complex set of activities or user preferences were needed.

Dataset	Features		Sensors	Labels
Shinmoto Torres et al.	3-axis acceleration, RFID antenna ID, RSSI, Phase, Frequency		Accelerometer, RFID sensors	Walking to chair/ bed, sitting, getting of the chair, lying on bed, getting of the bed, walking, standing
Ojetola et al.	3-axis acceleration / angular velocity		3D accelerometer/ gyroscope	Standing, fall forward/backward, lying, sitting on bed/chair, near-fall, walking, crouching, fall rightward/leftward, real fall forward/backward
Sztyler & Stuckenschmidt	Time	Correlation coefficient (Pearson), entropy (Shannon), gravity (roll, pitch), mean, mean absolute deviation, interquartile range (type R-5), kurtosis, median,	3D Accelerometer / gyroscope	Climbing stairs down / up, jumping, lying, standing, sitting, running, jogging, walking

		standard deviation, variance			
	Frequency	Energy (Fourier, Parseval), entropy (Fourier, Shannon), DC mean (Fourier)			
Do et al.	3-axis accelerometer, GPS signature		3D accelerometer, GPS	Without GPS	Walk, run, drive, stay still
				With GPS	User at certain place, User preferences

Table 2: Available datasets with their respective features, sensors and labels (activities recognized).

3   System Design

This section presents proposed solutions in literature based on their design choices. Firstly, the sensors used by each system are presented and then the activities recognized are further analyzed. Comparison continues with whether the recognition is offline or online and the different activities recognized by each system.

3.1   Sensors

The first step in HAR is related to sensory input. The most common non-intrusive sensors used in the literature are smartphones, wearables (smartwatches, smart bands etc.), Radio Frequency Identification (RFID) devices, ambient sensors (force sensors, temperature sensors, humidity sensors etc.). It is worth mentioning that cameras and microphones (used for speech recognition) also appear in a plethora of published works. However, they are considered intrusive and beyond the scope of this paper.

A) Smartphones / Wearables
Mobile phones and wearables have been extensively used on HAR domain (Kumari et al., 2017; Labrador & Lara, 2014; Lara & Labrador, 2013), as they are already equipped with sensors such as accelerometer, gyroscope and GPS. Additionally, smartphones are not very expensive and people tend to be familiar with their use since nowadays they are part of people's daily life. Those aspects made them a reliable choice for data gathering and in some cases also data processing (Abdallah, Gaber, Srinivasan, & Krishnaswamy, 2015; Capela et al., 2016; Damaševicius, Vasiljevas, Šalkevicius, & Wozniak, 2016; Ronao & Cho, 2016; Kang et al., 2018). Capela et al. (2016) used smartphones to monitor able-bodied and stroke patients' activities. Participants were wearing a smartphone on their waist. Accelerometer and gyroscope data

were gathered to recognize the activity performed. Their research results shown that although smartphones could be used in HAR for population with a stroke, their unique movement characteristics provided additional challenges.

The problem of elders having different movement characteristics (same as stroke patients), required the use of an elderly simulation kit (Álvarez de la Concepción, Soria Morillo, Álvarez García, & González-Abril, 2017). That kit restricted the movements of the subject, making him move as an elder. Additionally, special glasses were used to obscure the participants' vision. A mobile device was attached on his waist gathering acceleration and gyroscope data.

Wearables have also been used in HAR domain. Santiago et al. (2017) employed a pendant responsible for activity recognition, more specifically fall detection. The pendant was transmitting data to a smartphone responsible for classification. When a fall was detected, the wearable sent an alert to the phone with the latter contacting the person responsible. The advantage of the proposed system was that the elder did not have to carry a smartphone all the time indoors, since the communication between the two devices was based on Bluetooth.

A combination of wearable devices and smartphones have also been used (Sansrimahachai & Toahchoodee, 2017; Sztyler, Stuckenschmidt, & Petrich, 2017). Combined acceleration and gyroscope data from multiple sources, could provide more information and help distinguish activities that were often confused, i.e. sitting and standing. T. Sztyler et al. (2017) used sensors placed on head, chest, upper arm, thigh and shin as well as a smartphone on the waist and a smartwatch on the wrist. The phone and the watch were also used to gather the sensor data and label them using a custom application. Although in their work they used only the data from the aforementioned sensors, their dataset, that is publicly available, contains camera, sound, light and magnetic field data.

B) Radio Frequency Identification

Popularity of device-free activity recognition systems have increased recently. As a consequence, Radio Frequency Identification (RFID) use in HAR domain has been explored. RFID tags and devices are deployed in the house creating a network. The radio signal fluctuations created by a person's movement, are collected and used to extract the performed activity. Radio signal strength indicator (RSSI) and Channel state information (CSI) are used to associate fluctuation with a performed activity. An RFID tag network has been used by Ruan (2016), to identify activities and localize a person in house.

Yao et al. (2017) deployed a similar installation with Ruan. Passive tags were placed in the house walls. The system presented was designed with lower complexity in mind. Tag placement in the environment didn't affect the overall performance. Several experiments were conducted by the authors to evaluate the performance of their solution. Apart from tag density already mentioned, sensitivity to furniture changes, distance between human and tags as well as activity orientation were taken into consideration during evaluation. Results showed that all the previous cases did not have a significant impact on performance.

 An RFID solution, with the assistance of an accelerometer, was presented by G. Hu et al. (2016) Apart from the tags placed in the environment, a passive RFID tag equipped with a 3D-accelerometer was used. The latter had to be worn by the person acting. While the wearable tag was responsible for gathering all the data required to classify the activity, the tags placed on the environment only acted as a localization network, finding the room one is acting by exploiting the signal fluctuations.

C) Ambient Sensors

Ambient sensors include a variety of sensors such as motion, force, pressure etc. Sensors can be binary, returning 0 or 1 based on activation, or digital, transmitting a variable proportional to a physical dimension such as pressure or temperature. Arifoglu & Bouchachia (2017) in their work presented an activity recognition system using data collected by Van Kasteren, Englebienne, & Kröse (2011). The dataset used, contains readings from sensors placed in the environment. Reed switches were placed on doors and cupboards, pressure sensors on couches and beds, mercury contacts on drawers, passive IR to detect motion and float sensors on the toilet. All sensor readings are binary.

An ambient sensor network was set up by R. Hu et al. (2017). Door sensors were used, not only on the entrance, but also on the fridge door to detect activities such as cooking and eating. Additionally, passive infrared motion sensors were incorporated, in order to detect movement in a participant's home. Apart from the ambient network, authors evaluated their work with the enhancement of a Fitbit activity tracker and compared whether it can provide any value or not. The Fitbit tracker is a wearable used to detect activities such as walking with high accuracy (Paul et al., 2015).

Force and motion sensors have already been used in the literature, Aran et al. (2016) added smoke sensors in their ambient sensing network. The target of their system was abnormal behavior detection, so smoke detectors could provide useful data, like cooking on a non-usual time. Sensors firing were used to provide localization information of the person. Using that method to determine the location of someone, could lead to potential problems when two different sensors fired, either because they were overlapping or due to synchronization problems. Aran et al. addressed those issues and proposed a solution, by choosing the location based on the duration of the activity signal.

Zambrana et al. (2016) used eKauri, a commercial smart home kit using ambient sensors. Presence, temperature and luminescence sensors were integrated in each room of the house, as well as door sensors in each entrance. Due to the fact that eKauri was not able to distinguish between visitors and monitored person, the authors removed gathered data from days the subject was not alone.

A different kind of ambient sensor network was incorporated by Nef et al., (2015). Wireless sensing boxes were placed in rooms, dining table, fridge door and toilet flush handle. Each box was equipped with a variety of sensors. Passive infrared for motion sensing, temperature sensors, luminescence, humidity as well as acceleration sensors. For validation purposes each person was equipped with a wearable belt clip fitted with switches for all activities. Subjects were instructed to flip the switch corresponding to the performed activity.

D) Hybrid Sensing

Another method of sensing the environment in order to extract the performed activities or detect behavior abnormalities drawing researchers' attention, is non-intrusive load monitoring (NILM). Monitoring the total power consumption of a household, the power could be disaggregated to the individual appliances (Nalmpantis & Vrakas, 2018). The disaggregated data could then be used to evaluate the behavior and identify activities of daily living (Alcalá, Ureña, Hernández, & Gualda, 2017). Appliance usage could lead to the development of routine patterns of how an elder act, as Alcala et al. (2017) pointed in their paper. Any detection of anomalies in the already learned appliance usage, thus a potential symptom of elderly disorders, would result in a warning raised to a relative or caregiver (phone call, sms etc.).

A smart assistive living environment for elderly living alone, using a combination of the previous mentioned sensors, have been proposed by Meng et al. (2017). In their work, they used pressure, noise, light, temperature and humidity sensors to monitor the environment and elder's interaction with it. Additionally, IR and RFID tags were used as motion sensors. The goal of the authors was to integrate more sensors in their system. A similar approach was presented by Riboni et al. (2015). In their work, environmental sensors were placed indoors as well as presence sensors and RFID tags. Environmental sensors included pressure, temperature, door sensors etc. RFID tags were placed on items that the person was using in order to identify when it used them and if it put them back in place, for example put the milk back in the fridge. These data were used to identify behavioral abnormalities and signs of cognitive impairment.

A solution exploiting the advantages of wearable devices and Bluetooth Low Energy (BLE) technology have been studied by Mighali et al. (2017). Their elderly monitoring system required the person to wear a sensor tag, equipped with a 3-axis accelerometer, a 3-axis gyroscope and a BLE transceiver. The accelerometer and gyroscope were used to identify the performed activity. The transceiver's purpose was communication with smartphones or smart watches and indoor localization. Localization was realized by communicating with BLE beacons placed in each room. Sensor tag was also equipped with a microprocessor and was powered by a single coin cell battery.

Fan et al. (2017) developed an ambient sensing network, combining ambient sensors and wearables. In their work, experiments were conducted by monitoring third age volunteers. Their homes were equipped with a variety of sensors, such as motion, door (including door lock), humidity, temperature, light, water dispenser, air condition monitor, water leak and pressure. Also, elders were wearing a smartwatch with step count, heart rate and sleep monitoring capabilities. Subjects living routines and health data, as well as any alerts were contacted to their doctors and relatives.

Another hybrid method was developed by Sebestyen et al. (2016). Sensors were spread in the house including door and pressure sensors on furniture. All indoor signals generated were transmitted to an Arduino device placed in home. Localization was also performed based on the sensors firing. In addition to that, the subject was equipped with a smartphone device, processing its own data.

Lastly, under the City4Age project, Mainetti et al. (2015) presented an unobtrusive sensing network for elders. Published work was using ambient sensors, smartphones

and wearables as well as IR and BLE technology. Smart plugs were used, a device rarely used. It is worth mentioning City4Age performed activity recognition and abnormal behavior detection not only by sensing elders indoors but also outdoors.

3.2   Data Processing

The medium that houses data processing and classification, is an important consideration when developing an HAR system or performing abnormal behavior detection. Local processing, has the advantage that data transmission cost is either eliminated or greatly reduced as there is no need to contact an external server. Additionally, with the increased computational power and storage capacity modern mobile devices have, it is easier to perform more complex algorithms and store more sensor data for longer time. Another important aspect of offline processing is security. Local storage, for both historical data and sensor data reduces potential security issues as, in most cases, physical access is required to retrieve the data. Robustness is also promoted by offline recognition, since there is no need to transmit data using wireless communication that are prone to failure or interruptions. On the other hand, offline recognition has some limitations. While processing and storage have been significantly improved, they are still not sufficient to support more complex classification models and long-term history storage. Energy consumption is also a consideration when offline recognition is chosen. Data processing, feature extraction and classification are tasks that may require a reasonable amount of energy. Most mobile devices have a limited battery life that deteriorates with time and can affect performance.

With cloud infrastructures becoming cheaper and available to a broader audience as well as Internet of Things expansion, online solutions are getting a lot of attention. Taking advantage of the high storage and processing power, the implementation of more complex models and data processing algorithms is feasible. Providing an activity recognition system as service and using a universal encoding for sensors, as presented by Fan et al. (2017), allows the use of the system regardless of the sensors and the overall architecture. A variation of online processing is the installation of a processing unit at home, usually a local computer or an Arduino/raspberry platform. This method, promotes the advantages of online HAR while getting some benefits of offline recognition. The literature is almost equally divided between the two methods and there are various criteria that authors use to choose one, such as response time, classification method and availability.

3.3   Recognized Activities

An important aspect of any HAR system is the activities it is able to identify. Fall detection is important when creating an ambient assisted environment for elders and it was the activity recognized by Santiago et al. (2017). Falling was also recognized along with other actions in Alvarez's (2017) and Yao et al. (2017) work. The former detected immobility, walking activities as well as riding a bicycle and driving, while the latter detected a plethora of actions including sitting, standing, waving, kicking, bending over and crouching to standing.

Long periods of immobility could be a sign of health issues or lead to potential health problems. Sansrimahachai's et al. (2017) work was focused on informing elders of extensive idle time and urging them to exercise. Inactivity detection was also one of the actions Arifoglou et al. (2017) detected in their paper. Additionally, sleeping, breakfast and dinner eating, drinking, toileting and leaving the home were recognized.

A more basic set of actions, such as standing, sitting, lying down, walking and moving upstairs was recognized by the majority of research work, such as Mighali et al. (2017), Sebestyen et al. (2016), Capela et al. (2016), Hu et al. (2017), Sztyler et al. (2017). The smartphone-based recognition of Capela et al. was capable of detecting small movements such as washing dishes. Those actions were detected both to able bodied as well as stroke participants since their movement motives differ.

Sleep recognition is also implemented in the system presented by Zambrana et al. (2017) and Meng et al. (2017). Sleeping disorders could be an alert for mild cognitive impairment, as well as help with its detection. Moreover, RFID tags attached to objects and power monitoring done by Meng et al., allowed the recognition of object usage, for example taking medicines and activities involving appliances such as washing clothes or watching TV.

A different approach was followed by R. Hu et al. (2017) in their work. Instead of identifying the activities performed by the individual, they focused on recognizing visits as well as the duration spent on each location of the house and the motion on them. Visit detection is important for elderly people leaving alone according to authors. Isolation and decline in a person's social engagement and interactions could lead to health issues and age related disorders (Singh & Misra, 2009). Knowing when one had visitors can also help caregivers plan their regular visits. Since visit detection was achieved using non-intrusive methods (i.e. door sensors), it was crucial to distinguish when the elder was leaving and when a visitor entered. Outing detection, was also performed by the system proposed by Aran et al. (2016) paired with indoor localization.

4   State-of-the-art approaches

A variety of techniques have been proposed in the literature for human activity recognition. While machine learning techniques are getting a lot of attention, there are still solutions based on statistical models and probabilities. Features used to identify activities derive from three domains as already mentioned. The majority of papers exploit time domain features mainly because extracting them requires less computational power, thus allowing real time extraction.

Data gathering techniques vary. Most authors used publicly available datasets. A few papers used data specifically gathered for their work either from a lab with controlled conditions or volunteers from clinics and houses. Ambient and wearable sensors used, return continuous data, thus a segmenting method is required. Sliding window with a specific overlap has been used extensively in the literature.

This section discusses the frameworks present in papers focused on elderly people. A detailed analysis of the classification/recognition method is made. Additionally, features extracted and used are presented. Although many studies have been published in the HAR and abnormal detection domain we limited our study. The selection was made

based on classification method used, activities recognized and sensors used to gather data. Additionally, only recently published works were included (published in the last 3 years) as older works have already been analyzed in previous reviews. The main families reviewed can be seen on figure 2.
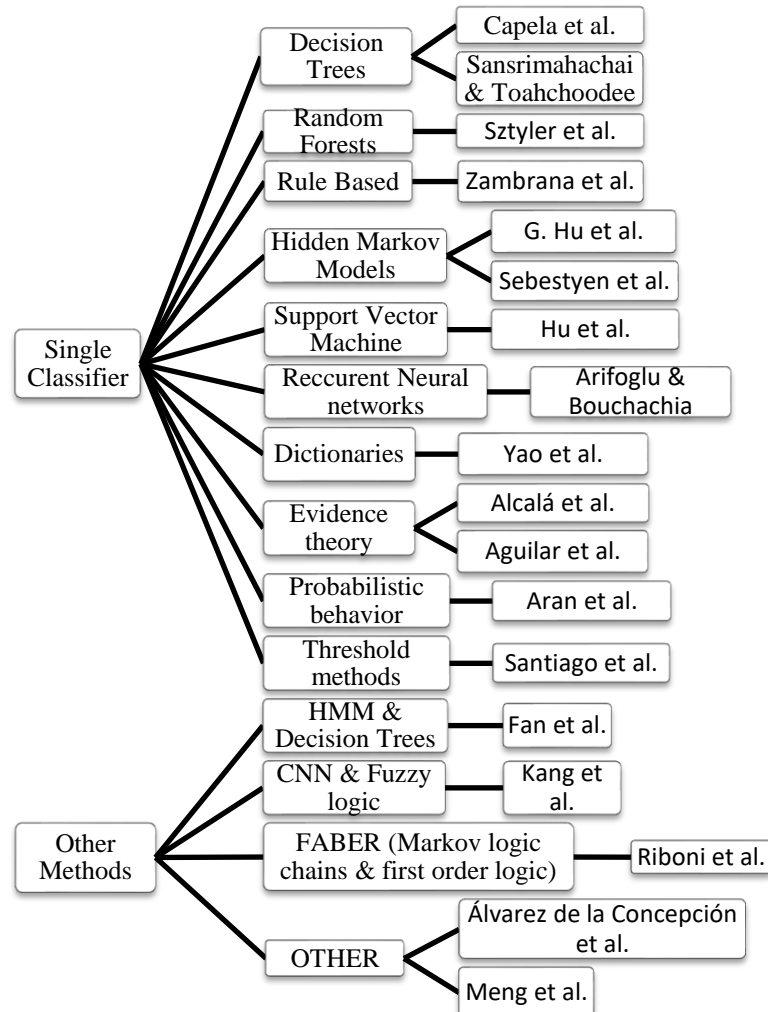

Figure 2: Summary of methods reviewed

4.1   Decision Trees

Decision trees and their variations, have been extensively used for activity recognition. Capela et al. (2016) in their approach employed decision trees to recognize activ-

ities using a smartphone equipped with an accelerometer, gyroscope and magnetometer. Since the orientation may vary among people, a rotation matrix was calculated from one second data while the subject was standing still. After that, all sensory input was corrected using that matrix, thus the phone's orientation was determined. A 1-second sliding window was also used while gathering the raw data from the sensors.

Different sensor data and features were used in each stage of the decision tree. The features used can be seen on the Table 3. Firstly, whether a person was moving or not was identified using a threshold for each corresponding feature. If the person was immobile, the orientation of the trunk was observed. Using specific thresholds, the activity was classified as standing, if the person's body was upright, as sitting, if it was leaning back, or as lying down, if it was horizontal. In case the person was standing, examining how many of the features used on the first step exceeded the threshold, as well as the time the person was standing (more than 3 seconds), the person was considered to perform small movements. If the person was found to be mobile on the first stage, the second stage automatically classified it as walking. On the third stage, walking was the default activity performed. If the maximum slope feature exceeded a threshold for more than 5 seconds, then the activity was classified as stair climbing.

Feature	Formula	Activity used
Simple moving average of sum of range of linear acceleration	$\frac{\sum_{i=1}^{4}\left[\left(range(Xlin_i)\right) \pm \left(range(Ylin_i)\right) \pm \left(range(Zlin_i)\right)\right]}{4}$	Immobile, mobile, stand, small moves
Difference to Y	$Ygrav - Zgrav - Xgrav$	Stand, sit, lie down
Sum of range of linear acceleration	$\left(range(Xlin_i)\right) + \left(range(Ylin_i)\right) + \left(range(Zlin_i)\right)$	Immobile, mobile, stand, small movements
Sum of standard deviation of linear acceleration	$\left(SD(Xlin_i)\right) + \left(SD(Ylin_i)\right) + \left(SD(Zlin_i)\right)$	Immobile, mobile, stand, small movements
Maximum slope of simple moving average of sum of variances of gravity	$SMA_{var} = \frac{\sum_{i=1}^{4}(Var(Xgrav)_i \pm Var(Ygrav)_i \pm Var(Zgrav)_i)}{4}$ $\max(SMA_{var}(2) - SMA_{var}(1), SMA_{var}(3) - SMA_{var}(2), SMA_{var}(4) - SMA_{var}(3))$	Walk, stairs

Table 3. Features used. Gravity acceleration (Xgrav,Ygrav,Zgrav), linear acceleration (Xlin,Ylin,Zlin), Standard Deviation (SD). Capela et al. (2016)

Capela et al. (2016) tested their work on a dataset with data gathered from able bodied and stroke participants. The reason data from both were used, was because it was observed that classifiers trained with younger or able bodied people did not perform so well when used on elderly or people with disabilities (Del Rosario et al., 2014). A smartphone was placed camera forward in the subject's belt (right front) or pant waist.

Sansrimahachai & Toahchoodee (2017) employed C4.5 decision trees to classify activities. Their proposed solution was an online activity/immobility recognition system, using a smartphone and a smartwatch to gather data. Raw acceleration and angular velocity data streams were redirected, through a message broker system, to the online recognition service. A 3-second sliding window was applied to raw data before being sent to the recognition service. The classic C4.5 decision tree algorithm was used for classification with statistical features, such as min, max and standard deviation. Apart from the activities recognized (sleeping, lying, sitting, standing, walking, stairs and running), periods of immobility were also identified. If inactivity or related activities were found for more than two hours, then a notification was sent to caregivers of the elder in order to take action. Also, a web-based application was created for relatives, caretakers etc. The application allowed them to monitor activities and heart rate at real time, access an activity log, track immobility and check the movement efficiency of the elder. Data gathering was done by placing the smartphone on a belt worn at the waist of older adults.

4.2   Random Forests

Decision trees are a solid classifier for activity recognition. Despite their extensive use, they are prone to certain problems such as overfitting. Sztyler et al. (2017) employed random forests, an ensemble method known to solve the overfitting problem (Breiman, 2001). A one second sliding window, overlapping by 50% was used and features from time and frequency domains were extracted from raw data. Discrete Fourier transform, was applied to convert features from time to frequency domain. Additionally, features based on gravity were extracted. Those features could be used to determine the orientation of the device. The acceleration and gravity forces were separated using a low pas filter. Using the gravity vectors obtained from the previous step, the authors computed the angles between them (roll and pitch), revealing device's orientation. It is worth mentioning that whether the device is back or forth could not be determined, as it requires the azimuth angle, which could not be calculated, since the direction of north was not available.

Another issue addressed in their paper, was the ability to use a pre-trained classifier with different people. For example, many HAR frameworks targeting elder adults use data gathered from younger and able-bodied persons, mainly due to lower availability of elderly data and/or difficulties in experiments with elders. Authors presented a cross-subject activity recognition model. Four cross subject approaches were constructed and evaluated. The first approach is a leave-one-out method, where for each subject a classifier is created using all available data except from the one subject left out. The second approach is top pairs, where the five most similar samples for each subject were chosen and used for training. A K-fold (K=10) cross validation was employed to get the average recognition rate. The third approach was top-pairs, where a classifier was trained using data from one subject and was evaluated against all other subjects. The top five matches were paired with the subject used for training. The final classifiers were trained using the data from the pairs already formed. Last approach, and the one authors found

most promising, was the physical approach. Initial dataset was split into groups of people by using various criteria. The criteria were gender and physique of a person, affecting activities like walking and fitness level based on performance on certain activities (i.e. running, jogging). Each person was assigned to a group before classification and the appropriate model was used for recognition. Standing and sitting activities were found not to be related with any physical characteristics.

4.3   Rule Based approach

Zambrana et al. (2016), realized an activity recognition system using a rule-based approach. In their paper a sleep recognition system for elderly people assistance was presented. Sleep activity was defined as the period starting when one went to bed and ends when he woke up. Sleep recognition's target was to find the time an elder went to bed and woke, the duration of the sleep and the duration of actual sleep (total duration minus visits to bathroom or other night activities). The tag used (eKauri), was equipped with a variety of sensors, as already mentioned. The features selected to train the classifier were:

- Time a motion took place
- Number of motions in bedroom/all rooms before the motion detected (four intervals, five minutes each)
- Average luminance in bedroom/all rooms before the motion detected (four intervals, five minutes each)

Rules used in their system were: user is in the bedroom, activity is performed at night, user is inactive and activity duration is more than 30 minutes. The main issue they encountered was that their system was designed under the assumption that night time is between 20:00 and 8:00. This assumption was not true for a majority of people. Their system ended up classifying activities like watching TV or reading at night as sleeping. In order to overcome that issue, a binary classifier was employed to classify periods into the bedroom as awake or sleep.

4.4   Hidden Markov Models

Hidden Markov Models (HMM) have been proven a solid choice in HAR domain. G. Hu et al. (2017) used RFID tags enhanced with a 3-D accelerometer and HMM to perform activity recognition. Data preprocessing was done using Kalman filter, in order to smooth them and focus on one frequency spectrum. Features used were from the time and frequency domain. A 2-second sliding window with 50% overlap was used to group the sequence by frame. Collection rate was set to 101Hz in order to reduce noise and body acceleration was extracted using a Butterworth's low filter. During classification, the Viterbi algorithm was used to calculate the probability of a sequence belonging to an activity class. Baum-welch algorithm was used to obtain the model parameters. In their work, 10299 frames were used with each activity taking 25 frames. Seven HMM were used in total with 6.866 vectors used to train them.

Hidden Markov models with the use of Viterbi algorithm to determine the activity, was also used by Sebestyen et al. (2016). They proposed an online recognition framework, consisting of three different modules communicating through HTTP. Data from home sensors were sent to an Arduino device for processing while accelerometer and gyroscope data were processed directly at the smartphone. Processing included filtering to remove noise. Accelerometer data were denoised with a combination of median and moving average filters. Multiple HMM were used to model the behavior of a person due to its complexity. Authors classified the models needed into two categories, based on the time and the location the activity is happening. Markov's chains were created to model activities and behavior associated with each part of the day (morning, mid-day, afternoon, night), and the location of action. Experiments were performed in a lab with simulated activity scenarios.

4.5   HMM & Decision trees hybrid

Fan et al. (2017), presented a cloud based human activity recognition system for smart homes. They provided a robust activity recognition as a service framework, using a standardized representation of sensor data allowing the integration of heterogeneous sensor networks. Additionally, REST API's were provided, allowing communication in JSON format.

Sensor data were clustered based on the number of sensors and devices in each room. Also, the purpose of each room was labeled during initialization. A three-layer sliding window was used with the first layer determining the working days / holidays and weather, the second layer explicit specifying if a device is on and the third layer detecting the association between sensors.

Classification was achieved by employing an HMM and C4.5 decision trees hybrid. The generated model represented the sequence of actions most likely to happen, given a set of observations. The probability distribution could be seen on (5) with $\{x_1, x_2, ..., x_{n-1}\}$ being the feature sequence and $\{y_1, y_2, ..., y_{n-1}\}$ the behavior model. The maximum probability was found using Viterbi algorithm.

$$P(x_1, x_2, ..., x_{n-1}) = P(y_1)P(x_1|y_1) \prod_{i=2}^{n} P(y_i|y_{i-1})P(x_i|y_i) \quad (5)$$

Decision trees, were used to exploit the fact that they do not require domain knowledge or parameter setting. Results from both classifiers were taken into account, and their intersection was returned as recognized activity.

Anomaly detection was achieved by using the Kullback-Leibler divergence. With q being the trained model and p the true distribution of activity stream, calculating the difference shown on (6) allowed detection of anomalies. The larger the difference the higher the relative entropy. Difference exceeding a predefined threshold, resulted in an anomaly detection and an alert was set of. False alarms may occur, so the authors used intersection set and boosting vote to reduce their rate. It is worth mentioning that based on the system presented on the paper, the elder had the ability to manually cancel an alarm no matter how high the entropy was.

$$D(p||q) = H(p,q) - H(p) = \sum_i p(i) * log\frac{p(i)}{q(i)} \qquad (6)$$

4.6 Support Vector Machines

Hu et al. (2017), on their home visit detection system designed for elderly living alone, relied on support vector machines (SVM) for classification. Data from ambient sensors (IR and door sensors) and Fitbit wearable were gathered from participants. Additionally, a nurse visit log was used to associate caretaker's visits at home. Training data contained both the visit log and the sensory data. Indoor localization was performed by registering the room that the last sensor fired was placed. An assumption was made that until a new sensor fires, the subject remains on the same room. It was also assumed that a visit event happens between two open/close events on the home's entrance. Any of the aforementioned door events happening within one minute or less, were discarded from the observation set. Any visiting event that overlaps a logged visit at least 50% was considered a nurse visit.

A one-class SVM classifier with a Gaussian kernel was employed. Features used included two six-dimension vectors, one including the total duration being on each room and the other the total times a sensor fired during the time segment, the number of room transitions and the step counter from Fitbit. Nurse visit log was used to label training data. Trained SVM was tested on a different home with changes on sensor installation as well as different person and behavior pattern. The reason was to observe if the model could be reused without retraining it, especially in cases were labeled data were hard to find.

4.7 Deep Learning Approaches

Arifoglu & Bouchachia (2017) in their paper presented a recognition system based on recurrent neural networks (RNN). Three different types of RNN were used in their system: Vanilla RNN, Long Short Term Memory (LSTM) RNN (Hochreiter & Urgen Schmidhuber, 1997), Gated Recurrent Unit (GRU) (Cho, Merrienboer, Bahdanau, & Bengio, 2014). The neural networks were trained with labeled activities. Trained model, when a test sequence was given as input, assigned labels to each activity in the sequence. Every labeled activity was also assigned a confidence value. The mean confidence value for each label in the training set was calculated and compared to the value of the instance coming from the test set. If the value was bigger than the mean, it was considered a normal activity, otherwise abnormal.

Data preprocessing was the same for all three classification methods. Using the sliding window technique (60 second window) the raw data in the dataset (Van Kasteren et al., 2010), were split into slices. Three different features were extracted from those slices:

- Binary: representing whether a sensor was fired or not.
- Change-point: returning 1 when a sensor's state was changed or 0 when the state was unchanged.

- Last-fired: information about the last fired sensor (1 if the sensor fired last or else 0).

The goal of their work was to identify dementia related abnormalities in a person's activities. According to authors, no dataset related to the behavior of people with dementia was available. Artificial anomalies were introduced in the dataset. More specifically repeating or forgetting activities, dehydration and sleep disruption abnormalities were used.

A common symptom of dementia suffering elders, is the repetition of already performed activities or forget a performed action thus skipping it. In order to project this in the dataset, Arifoglu & Bouchachia (2017) added manually a set of actions. Activities added was teeth brushing and eating related actions (preparing dinner, eating, getting snack).

Sleeping related disorders and night wandering are considered severe symptoms of dementia. Dehydration could be a result of a person forgetting to drink water. It could potentially result to sleep pattern disruption by reducing the number of times an elder visited the bathroom during night. Those anomalies were created by inserting synthetic activities in the night activity sequence (getting drink, going to toilet). The unmodified dataset was used for training while the modifications were introduced in the test data.

A deep learning method based on Convolution Neural Networks (CNN) and fuzzy logic has been proposed by Kang et al. (2018). In their work, apart from activities, transition activities (i.e. turning left or right) were also identified. Accelerometer and gyroscope data collected from a smartphone were used as input on a CNN. Data were segmented using an overlapping sliding window approach. The degree of sliding and the height of the window were determined using simulated annealing. Two different CNNs were used, one for simple and one for transition activities. The results obtained from the two networks were integrated using fuzzy logic. Bell-shaped membership functions were used to represent the features of each activity and integrate the two different results. The fuzzification was done on the area that simple and transition actions were intersecting.

4.8   Dictionary Learning

Yao et al. (2017), used RFID tags to collect data and recognized the performed activity with a dictionary approach. Dividing the continuous data from RFID tags into segments representing an activity, was achieved using a sliding window. Seven statistical features were extracted from those segments: min, max, mean, variance, root mean square, standard deviation and median. Canonical Correlation Analysis (CCA) was employed for feature selection. Canonical correlation was calculated for every feature pair and a greedy algorithm was used to generate feature subsets. Features that were weakly correlated got a higher ranking, while strongly correlated features a lower one. The greedy algorithm used was forward searching, where starting with an empty feature set, features were added and evaluated using the classification performance.

For every activity one dictionary was learned using training samples. Activity's dictionary is independent from other activities, thus providing flexibility and scalability

on the system, as well as allowing new activity learning without affecting previously learned dictionaries. As authors mentioned, dictionary learning, could be done by using a small amount of training data, compared to other methods, reducing the need for manually labeling and annotating big datasets. Dictionary learning could be represented as an optimization problem formalized as shown on (8). For each class C^k a dictionary matrix $D^k \in \mathbb{R}^{m \times k}$, with m being the feature dimension and k the activity, had to be constructed. The training samples $O^k = \{O_1^k, O_2^k, ..., O_N^k\}$ had to have a sparse representation $X^k = \{X_1^k, X_2^k, ..., X_N^k\}$ over that dictionary. Maximum dictionary vectors the O^k matrix could be represented as, is $T_O^k (T_O^k \ll K)$. K-SVD algorithm was used to solve the optimization problem.

$$\min_{D, X} ||O - DX||_2^2, \ s.t. ||x_i||_0 \leq T_o \qquad (8)$$

In order to assign an incoming signal to a certain activity, authors proposed several ways to use the already learned coefficients. Such ways included, maximal, maximal mean, maximal sum of coefficients, reconstruction error and concatenate coefficients where learned coefficients were stacked with features, forming a new feature vector used on an SVM for classification. Dictionary approach was tested against state-of-the-art classifiers as K-Nearest Neighbors (KNN), LSVM, RF, and Naïve Bayes (NB).

4.9   Dempster-Shafer theory

Dempster-Shafer theory (DST), also known as evidence theory, was employed by Alcalá et al., (2017). Power consumption from each appliance was considered as reading from an independent sensor. Each device's consumption was providing a belief about usage normality. A general belief about how normal was the usage of all appliances, was obtained by merging all the individual beliefs.

While modeling the basic belief function for every device, the day and the time interval of the day were considered, since usage varies during different days and hours. The time interval T_i and the day were used to bin together occurrences of each appliance. Dividing the number of occurrences in each bin for a specific time interval (3 hours), with the total number of occurrences for that day, the probability for that device to be used that time of day was obtained. The same process was done for every appliance, thus modeling the basic belief functions of all devices. Certainty constants were used to multiply the possibilities. Their values were empirically set to 0.9 in case of an event and 0.1 otherwise, meaning that in case of a switch on event there was 10% uncertainty and the opposite.

For evaluation purposes the Household Electricity Survey (HES) database (Zimmermann et al., 2012) and the UK-DALE dataset (Kelly & Knottenbelt, 2015) were used. A non-intrusive load monitoring algorithm was firstly deployed providing the disaggregated data. The authors, considered only appliances that could be manually turned on/off and discarded devices with automatic or continues usage (i.e. fridge).

Evidential networks is the representation of the DST as acyclic oriented graphs (Hong et al., 2009; Simon & Weber, 2009). In their fall detections system, Aguilar,

Boudy, Istrate, Dorizzi, & Mota (2014), used dynamic evidential networks (DEN) based on the temporal belief filter (Ramasso, Rombaut, & Pellerin, 2006). DEN was used to perform data fusion on data originating from two different sources. Data were originating from two separate subsystems: an IR sensing network (F. & S., 2005) and a wearable fall detection device (Baldinger et al., 2004). The approach proposed by the authors, was able to identify soft falls, a functionality that the wearable did not have. Moreover, data fusion improved performance and the usage of dynamic evidential networks allowed the system to work even when one of the two sensing devices was not present.

4.10 Threshold based methods

Santiago et al. (2017), published a fall detection system using a wearable and a cell phone. Their system was based on thresholds to detect falling of elders. A wearable pedant, equipped with accelerometer, gyroscope, Bluetooth, as well as a panic button and a stop alarm button. Proposed architecture was based on offline recognition on a smartphone communicating with the pedant via Bluetooth.

The pedant constantly monitored its accelerometer and gyroscope sensors and transmitted data to the smartphone. When acceleration exceeded a predefined threshold, the gyroscope variation was examined. If the variation was found to exceed the threshold, regardless of the direction of the change, a 3 second timer was started. After that timer, the position variation was again compared to the threshold. If it exceeded it again, a fall detection event was fired and a 30 second timer was started. During that timer's duration, the elder could cancel the alert using the button on the pendant. If the alarm was not canceled, a notification was sent to relatives and/or caregivers.

Another threshold approach was published by Mighali et al. (2017), under the City4Age project. City4Age aimed to promote friendlier cities for elderly or people with mild cognitive impairment. In the published paper, authors presented an architecture designed for indoor positioning and motility detection. The former is executed offline on a wearable using BLE beacons placed indoors to determine the persons location, while the later was achieved using the inertia sensors of the mobile/wearable device. Both results were sent to a cloud infrastructure for further analysis.

As already mentioned, localization was performed with BLE beacons. Each beacon broadcasted a unique identifier (i.e. its MAC address). The receiver device calculated the distance from the beacon using the signal strength. The main issue with that method, that authors addressed, was that the results could be found unreliable especially when transiting from one room to another and the person is at the edge. That problem was solved by adopting a stability period. Stability period introduced a time delay that the user had to be in location, in order to detect a room change.

Motility system, was the main component responsible for activity recognition. A sensor tag was used on the research conducted in the paper, equipped with 3 axis accelerometer and magnetometer as well as a BLE transceiver. Due to tag's limited battery and memory, a simplified threshold design was adopted, able to distinguish between mobility and immobility. Raw accelerometer data were obtained at 25Hz, and a

median filter with n=3 was applied, removing noise spikes. After filtering, the raw data were used to calculate the Signal Magnitude Vector (SMV) using (9). By calculating the SMV, orientation was removed from data. A 3-second sliding window with 30% overlap was employed to segment the data. Standard deviation had been chosen as the feature to compare against the threshold. If standard deviation exceeded the threshold, a moving period was identified, while if it was below, a still period was recognized.

$$SMV = \sqrt{acc_{x^2} + acc_{y^2} + acc_{z^2}} \qquad (9)$$

4.11 Probabilistic behavior model

Aran et al. (2016), developed a system for detecting abnormal behaviors, often leading to potential health issues. Due to the different needs of sensors, depending on the size of house and each room, an abstraction layer was created to overcome different sensor configurations. Abstraction layer was responsible for converting all sensory input into universal events defined by their start/end time and a label. Authors used two event types: locations and outings.

Location of the subject inside the house was found based on which sensor fired. The proposed system, kept track of the current location and when no activity was detected, it assumed that the person remained in the same room. The challenge when using sensor event to detect location was when multiple sensors fired. This could be due to sensor covering the same area or sensors not synchronized. In order to overcome that problem, the location was updated based on the duration of the activity signals.

Knowing when the elder was in the apartment, thus analyzing sensory input and his behavior was also important. Door sensors were used by the authors for that purpose. An assumption was made that the person always closes the entrance after leaving and opened it to enter. Considering that assumption, outing could be easily identified, when two consecutive door events were registered and no actions in the house were detected.

Individual's behavior was modeled using a statistical model (10). Location sequence was represented as $L = \{l_t\}$, hours of the day as $H = \{h_t\}$, $h_t \in \{1,2\,...,24\}$, probability of being at a location during a specific hour as $\theta_{h,l} = P(h|l)$ and the count of locations in a time slot as $n(l,h,L,H) = \sum_i \mathbb{1}(l_i = l \wedge h_i = h)$.

$$p(L;H;\theta) = \prod_h \prod_l \theta_{h,l}^{n(l,h,L,H)} \qquad (10)$$

Above model, could reveal information regarding different activities based on the location at a specific time, such as sleeping, wake up time or sleep disruptions. K-means clustering with 2 clusters was applied to cluster behavior patterns shared across multiple persons. Clustering shown that most people spend time in their living rooms, while sleeping in bedrooms, an observation important according to authors, as it could affect the sensor deployment.

Finally, anomalies were found as deviations from patterns using a cross entropy measure. Lower entropy means that the model created could predict the distribution, while higher entropy means the model could not predict the data with high accuracy. In order to validate their approach, authors split the dataset into weekly intervals. Data

from the weeks before the one used for validation, were used for training. Cross entropy was calculated for each hour of each day.

4.12 Other

Riboni et al. (2015) in their paper, presented the Fine-grained Abnormal Behavior Recognition (FABER) system. FABER is a hybrid technique using Markov logic chains and a knowledge-based inference engine representing knowledge as first order logic formulae. FABER's purpose was the early detection of mild cognitive impairment. First component of their recognition framework was the semantic integration layer. All sensors (ambient, RFID, presence) communicated raw data to that layer. The semantic integration layer then extracted basic activity and/or event information using simple inference methods. The extracted information, represented using a shared vocabulary, were sent to the Markov logic network reasoner. The representation of the sensor events can be seen on (7). In that event sequence representation $event(e_{ji}, t_i)$ is the sensor event e_{ji} at time instance t_i (only one sensor event can fire at one time instance).

$$< event(e_{j1}, t_1), event(e_{j2}, t_2), ..., event(e_{jm}, t_m) > \quad (7)$$

The first order logic knowledge base used could potentially lead to ambiguous results. For example, as the authors state, if the event sequence indicated that the silverware drawer is closing and the glassware cabinet is opening, could be interpreted as setting the table or washing dishes, activities that cannot happen at the same time. That problem was addressed using the Markov logic network (MLN). Logic formulae were given a weight, describing the confidence on their validity. The probability of a formula being true, with respect to axioms representing reality, was used to evaluate the validity. Weights were learned using an observation set with labeled data. MLN's goal was to find, based on the observations and formulae already defined, the most probable set of axioms. First order logic formulae were also employed to detect the activities boundaries, i.e. when an activity starts and ends.

The last component of the system presented, was the inference engine, responsible for abnormal behavior identification. Abnormalities were represented with propositional logic rules. Behavioral analysis includes evaluation of the activity boundaries, anomaly predicates, and external knowledge such as medication prescribed. Anomalies were categorized as non-critical or critical. Non-critical anomalies occured when the elder skipped or forgot a step during an activity sequence execution (e.g. forgot to close a drawer after taking something from inside) or when activities took more time than normal. Those anomalies were signs of mild cognitive impairment, though they were considered minor. Critical anomalies occurred when the patient forgot, skipped or repeated an activity. Critical anomaly was also considered when the person executed the same activity more than once (e.g. eat the same meal twice).

Data were gathered from a laboratory installation but the system had to be retrained when it was tested in a real case scenario. Preliminary testing was performed on a hospital with physicians and care takers assessing the system.

Álvarez de la Concepción et al., (2017) used accelerometer data for activity recognition and fall detection on elderly. Training data were obtained by using a 5 second

sliding window. Features extracted are arithmetic mean, minimum, maximum, median, standard deviation, geometric mean and features from frequency domain by applying a fast Fourier transformation. Activities recognized were not predefined in their system. Instead, users could specify the activities they wanted to identify and provided the data by performing those actions for a specific set of time. Duration of performing an action depended on the activity itself, for example walking needed 20 seconds, driving required 15 minutes and riding a bike 3 minutes mainly because acceleration did not occur at a specific frequency.

Ameva algorithm (Gonzalez-Abril, Cuberos, Velasco, & Ortega, 2009), was employed for variable discretization. Having the class labels C , the continuous attributes L and the statistic features S mentioned before, by applying the algorithm on each statistical feature, a matrix $Dm\{C, L, S\}$ was generated, containing all the set of intervals associated with the activity C. After that the probability of a feature associated with a class was calculated and a matrix was generated, containing instances of the training data that belong to specific interval. Lastly the relative probability matrix was calculated, representing, how likely was, a value associated to a statistic feature, to belong to a certain activity. During the recognition process, a majority voting system was used to find the activity that was performed. Features were considered uncorrelated, providing same value of information.

Due to the usage of discrete variables instead of continuous, as well as the elimination of dependencies between them, battery consumption of the application was improved, according to authors. Considering that elders were not familiar with use of such devices and they were not very keen on following long procedures needed for training, it is emphasized that their approach, allowed quick training on a set of activities the user wanted. In case an unknown activity was found, meaning that the probability that a value belongs to a class was low, an alert was generated. After performing experiments, the threshold that determines whether an action was recognized, was set to 25%, for the joint probability. Fall detection was approached with a different method. Acceleration peaks followed by a 5 second inactivity were monitored. In case one was found it was labeled as fall.

Online Daily Habit Modeling and Anomaly detection (ODHMAD), was a framework designed for elderly activity recognition and behavior analysis proposed by Meng et al. (2017). ODHMAD consisted of three different modules, the sensor gathering module, the activity recognition module and the daily habit modeling and anomaly detection module. Data gathering layer was in charge of raw data processing and information extraction. Second layer was responsible for recognizing the performed activity given the data from the previous component. Lastly, third layer modeled the habits of the elders using probabilistic models and detects whether a performed activity diverged from that pattern, thus recognized as an anomaly.

Online activity recognition layer, as the name implies, was able to identify performed activities. The proposed approach, did not rely on classification to recognize activities thus no training data were needed. Instead the algorithm presented relied on information from sensors to extract the action and metadata (i.e. start/end time, duration, and breaks). Several assumptions were made by the authors, in order to be able to

recognize activities based on sensor activation. While those assumptions limit the overall system, they were true for most ambient and wearable sensors. Those assumptions were:

- Sensors, when not active, returned a stable value
- During an activity, sensors returned higher or lower values
- Activation time could not exceed time in idle state
- Activities had a finite duration
- Short breaks from an activity did not interrupt it
- Same set of sensors should not be the only indicator of more than one activity

Information about sensor activation and duration were what the recognition system used to identify activities. One of the features obtained was the activation period of a sensor, indicating whether a sensor is currently activated and the start/end date/time of activation. Sensor normal status information, allowed the OAR component to identify a sensor activation. By modeling the continuous signal obtained when no activities were performed, the activation signal could be recognized. Another feature used, was the break and pending status of the sensor, allowing the recognition of breaks during activities and recognition of actions with small pauses (i.e. sleeping disruption). Lastly an index was used that maps activities with corresponding sensor status. This index allowed the system to quickly recognize an activity as well as model complex activities requiring more than one set of sensors, in order to identify them.

The next layer of the ODHMAD framework was the dynamic daily habit modeling (DDHM). A tree structure with two layers was generated dynamically by that layer, in order to model daily habits, based on the activities recognized by the previous layer. First layer, contained the activities, while the second layer the probability that an activity would be performed at different time periods. Information regarding start/end time and duration were used to find the similarity of detected activities and modeled activities. The higher the similarity, the more probable was the activity to happen in the modeled period. Similarity, between modeled activities and incoming activities, was calculated by evaluating how close their starting and end times were, and how much they overlap. If the similarity was below a threshold (80%) a new time period for that activity was modeled, i.e. added to the tree. Pruning on rare nodes was also performed, in order to reduce computational cost and prevent node proliferation.

Anomaly detection was based on the tree structure created by the DDHM module and followed a similar approach. Since the system had already modeled the behavior any incoming activity that had a similarity less than 30% with the most similar modeled activity, was detected as potential anomaly and an alarm was raised. As the authors mentioned, anomaly detection worked only after the tree had been stabilized, meaning that the daily behavior of the elder had been completely modeled. Due to anomaly detection and activity modeling relying on the same technique, they could not work simultaneously.

5   Performance analysis

Performance comparison is a difficult task with many challenges. Firstly, the datasets used have major differences not only on the sensors used to gather raw data but, on the subjects and conditions those data were gathered. Also, each framework presented in the literature identified a unique set of activities, or performed behavior modeling based on different elderly actions and interactions with their environment. Most used metrics include Accuracy, F-measure with Precision, Recall following. This section aims to present the results of each method using the metric each paper used.

Decision Trees

Capela et al. (2016) performed experiments using data from both able bodied and stroke participants and used the F-score metric to calculate their system performance. A mobile phone was attached on their waste, gathering data from its sensors while performing activity recognition. Stroke subjects were older adults, while able bodied were younger. Their results showed that the more complex the classification, the lower their score. More specifically, their decision tree approach had better results (>94%) while detecting mobility/immobility. During the second stage of the DT classifier, performance was lower. Classifying immobility as standing, sitting and lying down, was based on a single feature (inclination) and a static threshold. It was observed that stroke participants had different postures due to health conditions and advanced age, resulting in misclassification. Stage 3 had an even lower performance. Authors mentioned that stroke patients walking activity was not classified correctly mainly due to hemiparesis preventing pelvis movement, thus tampering acceleration data. Unique movement patterns between participants also caused activities to be wrongly classified. Authors concluded that in order to identify stair climbing and small movements, additional features were needed. Also, as stroke and older subjects have different mobility levels, a dataset from younger people would over fit to their age group, preventing a broad use of the model. Classification results of each stage can be seen on Table 4.

DT stage	# Activities	Performance (avg)	
1st	2	Able-bodied	98%
		Stroke	97%
2nd	4	Able-bodied	86%
		Stroke	77%
3rd	6	Able-bodied	55%
		Stroke	47%

Table 4. Classification performance per DT stage

Sansrimahachai & Toahchoodee (2017) in order to evaluate the performance of their proposed solution, developed an activity recognition application for android. The mobile phone was attached to the waist of elderly people for data gathering and processing. Seven subjects were used in total with no noise filters applied. Accuracy was used as a metric with 93.5% average and individual accuracies as seen on Table 5. Stair climbing activity had a lower accuracy due to acceleration data, similarity with walking and running activities.

Activity	Accuracy
Sleeping	96.48%
Lying	95.5%
Sitting	96.89%
Standing	97.10%
Walking	90.94
Stair Climbing	85.03%
Running	95.33%

Table 5. Accuracy for each activity recognized.

Random Forests

Sztyler et al. (2017) tested their solution exploiting random forests and smartphones separately for the three different functionalities it provided (on-body position detection, subject specific activity recognition, cross-subject recognition). On-body position classifier was responsible for detecting dynamic and static activities, as well as device's position on the subject, thus improving overall classification performance. The second phase, subject specific recognition, was in charge of classifying the action performed. Two different methods were tested on that phase, a position independent approach and a position aware one. The first one was classifying the activities disregarding the device position. The second approach, used one set of classifiers for each subject and device orientation. Knowing the orientation allowed the authors to extract an activity specific feature set in their experiment, in order to improve accuracy. Cross-subject activity recognition was used to measure how well the proposed system could identify activities performed by subjects, that is hard to gather and label data from (i.e. elders). During the experiments, authors assumed that the orientation was known. A group approach was used, where each group represent certain people whose labeled data could be used to train a classifier for a person with no labelled data. The methods used to create those groups were discussed in 4.2 (leave one out, top pairs, physical, randomly).

The F score was employed by the authors as the evaluation metric. Position recognition had an 81% F-score with the classifier, trained for each person using all activities, performing better when the device was placed on the shin (88%) and worse when placed on the upper arm (78%). Authors, analyzed the results, and concluded that the stronger the acceleration of an activity the better the position recognition. Evaluating their dynamic and static activity separation approach, led to better results. The average F-score when firstly the activity was classified as static or dynamic, was 89% with the highest being 94% on the shin and the lowest 85% on upper arm. The classifier deciding whether an activity was static or dynamic, performed well with a 97% score.

Subject specific recognition was evaluated with and without device position information, showing that knowing the position and orientation of device improved the classification results. Indeed, there was a 4% improvement with the position free classification with F-score being 80% and 84% when position information was used. Several positions were evaluated to find if there was an optimal place to attach the device on

human body. No optimal place was found as the position was highly related to the activity that was performed. A device placed on the chest, for example would provide better results when recognizing stair climbing, while when placed on the thigh, standing was better identified. Proposed random forest classifier was also evaluated against five different classification algorithms outperforming them in the same dataset (NB, kNN, SVM, NN, and DT).

Cross subject recognition was evaluated using all the aforementioned methods for subject grouping. Only dynamic activities (climbing, jumping, running. Walking) were considered as the static activities (standing, sitting, lying), according to authors, are similar for each individual. The smartphone was placed on different position on the body, with the waist having the best score regardless the method that was used to create the test groups. Physical criterion for subject grouping, had the best f-score (78%). When adding the static activities, performance slightly dropped for dynamic activities (79%). Authors addressed that issue by experimenting with adding additional accelerometer and gyroscope enabled devices. Incorporating a smart band, the performance of the classifier improved, scoring 82% on groups created using physical characteristics. An assumption was made that the position of the device was always the waist, as cross subject device's position recognition results were found to be low (best 77%, worst 54%). Authors concluded that cross subject activity recognition is feasible when grouped with physical attributes, but still further investigation was needed.

Rule Based approach

Zambrana et al. (2016) as already mentioned, used a rule-based approach to perform sleeping recognition, while using a binary classifier to classify activities at bedroom during night, as awake or asleep. Authors experimented with three different classifiers, SVM, KNN and RF. After fine tuning the parameters for each classifier, the SVM with a Radial Basis Function (RBF) kernel was found to have the highest accuracy of them (96%). Classifier allowed the authors to predict the time the subject went to bed and woke up, thus allowing them to calculate the duration of sleeping and by subtracting the active hours during night, the duration of actual rest hours. The SVM employed correctly classified all awake instances while misclassifying only 18 instances of sleep examples.

Hidden Markov Models

G. Hu et al. (2017) tested their proposed solution, exploiting HMM, using the Moo RFID tag and a commercial reader to capture the data. A 3-second sliding window with 60% overlap was used for data sampling. After fine tuning the parameters of the HMM, the authors validated their approach on a real-life scenario. Their model was also compared with the work of HMM-based RFID solution (Garcia-Valverde, Garcia-Sola, & Botia, 2010; Tran et al., 2009) already proposed in the literature, with authors stating that they outperformed them by ~2%.

Sebestyen et al. (2016) evaluated their HMM using the morning activities of a person. Although in their work no numerical results were presented, they stated that their results were acceptable and promising, thus expanding to more activities, time and location would be their focus. The recognized activities so far were only associated with

the kitchen usage during morning. Their HMM could recognize the actual activity performed (i.e. making breakfast, eating breakfast, making coffee and drinking coffee) and extract certain observations related to that activity, such as walking, standing, sitting and non-uniform movements.

HMM & Decision trees hybrid

Fan et al. (2017) evaluated their activity as a service solution using a real-life scenario. A volunteer living alone, suffering from hearing and anorectal diseases and living quite far from the hospital, had agreed to have the framework implemented at his home. The elderly participant was monitored for a three-week period. Several ambient sensors as well as wearables were used for monitoring. Data were processed on the cloud with his grandson and doctor being able to receive real time alert and monitor his status. Although overall system's accuracy was 93%, authors insist that further testing is required with more participants, as detecting abnormalities on his behavior was hard to be found in a three-week period.

FABER

FABER's performance, a hybrid framework using machine learning and symbolical reasoning, was assessed using data gathered from actors simulating patients' behavior in a laboratory installation. In total three different activities of daily living had been simulated with anomalies regarding those activities: preparing food, eating and taking medicines. Actors simulating the behavior of elderly people were divided into two groups. First group was simulating the behavior of 7 health seniors while second 14 people with mild cognitive impairment symptoms. As already mention in 4.12, anomalies were divided into critical and non-critical. First group was only performing few non-critical anomalies while the second was prone to both types of them. For non-critical anomalies FABER had achieved a 90% F score, for critical anomalies 96% and the total F score was 93%. Riboni et al. (2015) mentioned that misclassification was due to inaccurate activities' boundaries detection.

Support Vector Machines

Hu et al. (2017) tested their one class SVM, for home visits detection using nurse logs and data from elderly leaving alone. In their work they incorporated a Fitbit tracker in order to evaluate whether a wearable will improve system's accuracy or not. Their model was first tested against labeled data from 1st user, before it was deployed and tested on different users and home installations. Additionally, unlabeled visiting events were used for testing. According to authors, it was crucial to test their pre-trained model and its performance on different scenarios without the need for retraining. Overall results (as shown on Table 6), proved that the Fitbit tracker improved the overall accuracy compared to using only ambient sensors. On the other hand, using the pre-trained model on different installations, had a negative impact on performance.

User	Ambient (labeled)	Ambient (unlabeled)	Ambient/Fitbit (labeled)	Ambient/Fitbit (unlabeled)
1	85.1%	44.8%	87%	49%

2	83.3%	28.4%	87.5%	26.8%
3	58.8%	15%	76.5%	6.1%
4	75%	22,5%	75%	32.5%
5	73,7%	15.3%	84.2%	16.4%

Table 6. Hu et al. results per user, sensors and type of data

Deep Learning Approaches

Three different type of RNN were tested against the dataset from Van Kasteren et al. (2010), vanilla RNN, LSTM and GRU. The original dataset contained data from three different households, and the neural networks were evaluated against all of them. The datasets were split using leave one day out cross validation method to test and training data. Following that method, one day was used for testing and the rest for training. That technique was used for all days and the average score was reported. Arifoglou et al. (2017) compared their performance with the performance of the classifiers used by Van Kasteren et al. (2010) when benchmarking their dataset (HMM, NB, HSMM, CRF, NB). Results for each classifier and dataset can be seen on Table 7, while results for abnormal behavior are presented on Table 8. Authors concluded that RNN approach, especially LSTM, showed promising results especially on abnormal behavior detection but still needed improvements as the method was prone to false negatives.

Classifier	Dataset A	Dataset B	Dataset C
NB	95.3%	86.2%	87%
HMM	92.3%	81%	83.9%
HSMM	91.8%	82.3%	84.5%
CRF	96.4%	92.9%	89.7%
Vanilla RNN	95.5%	87.9%	86.7%
LSTM	96.7%	87.2%	87.4%
GRU	96.1%	87%	86.6%
SVM	96.1%	86.2%	87.5%

Table 7. Classifiers' accuracy on each dataset.

Classifier	TPR	FPR
NB	40.4%	43.5%
HMM	58.36%	96.2%
HSMM	98.85%	32.2%
CRF	66.22%	40.5%
SVM	72.11%	44%
LSTM	91.43%	40.96%

Table 8. TPR and FPR for each classifier

Kang et al. (2018) evaluated their results on data collected from five elders performing predefined activities (walking straight, standing, turning left, turning right), for anytime between 3 and 5 minutes. All data were gathered using an Android smartphone. Results can be seen on the table 9. Authors also compared the performance using the

parameters obtained using simulated annealing with optimal values proposed in the literature (Huynh & Schiele, 2005) and combinations derived from Fourier transformation method. Comparison showed that simulated annealing generated parameters that could significantly increase classification performance.

Activity	F score
Walking (simple	0.9944
Standing (simple)	0.9922
Turn to left (transition)	0.8019
Turn to right (transition)	0.9772

Table 9: F score for each activity

Dictionary Learning

Yao et al. (2017) performed their experiments with data gathered from six subjects, performing 23 different activities. Raw data from RFID tags, were gathered at 0.5 sec intervals. Their dictionary system was compared to Multinomial Logistic Regression with l_1 (MLGL1), KNN, SVM, RF and NB using F score for evaluation. After fine tuning the parameters and features for each method, results showed that the dictionary-based method outperformed all of them. More specifically, although SVM showed similar performance with the method proposed when doing a person depended validation, during person independent validation, dictionary learning had better performance. An issue identified by the authors during evaluation, was the low accuracy on lower body movements (i.e. kicking). This could be an issue with the hardware setting or intra-class variability. Authors, tried to collect a bigger spectrum of RFID data by placing different lines of tags on different height, each corresponding to a distinct body part.

Experiments were performed, regarding the setup of the system and the response time. The average latency of the system was found to be approximately 4.5 seconds, mainly due to data collection and feature selection. As RFID tags were placed indoors, challenges arising from the environment were examined. Firstly, the distance between tags was evaluated. Varying the distance from 0.3 to 1 meter and re-evaluating the system, showed that the proposed method was tag density independent. Since the signal fluctuation was used to gather data, the changes on furniture and the effect on the performance was tested. It was found that furniture could slightly affect the performance. Authors also experimented with the distance between persons and tags, revealing that distance had no particular effect on recognition. Lastly the sensitivity of the system to person orientation was tested, and found that proposed solution could identify most orientation sensitive activities. Orientation had an effect only on actions with similar intra class gap, such as falling left and right. Average accuracy of the system was found to be 96% across 23 different postures and actions.

Dempster-Shafer theory

Using the UKDALE and HES dataset, Alcalá et al. (2015) evaluated their approach on how to exploit non-intrusive load monitoring to perform activity recognition and abnormal behavior detection. Their approach was benchmarked against a Gaussian mixture model (GMM) (Alcalá, Ureña, & Hernández, 2015) and the union probability

was used for scoring. Authors divided the training data on days of week instead of working days and holidays, a change that improved the performance of GMM.

Four different household were used for evaluation, three single pensioner houses from the HES dataset and one family house from the UKDALE. Devices with automatic operation or constant consumption (e.g. fridge) were removed. A threshold was set empirically with any score bellow that would be considered an anomaly. Results showed that Dempster-Shafer theory (DST) could model uncertainty better and detect anomalies more efficiently. Also, a single appliance with a strict routine, found in the family household experiment, could saturate the score, preventing GMM from detecting an anomaly. This issue is not applicable to DST, where the evidence of an abnormal pattern is higher than the evidence of normal one. Another conclusion the authors came to, was that DST approach could detect short term deviations more effectively compared to the GMM approach.

Overall the DST system proposed was evaluated and found to be more reliable as it can detect effectively both short- and long-term deviations. Also, it is more sensitive to anomalies and produces lower number of false alarms due to inactivity. Lastly, strong routines, a case in elderly people households, was examined and the system's performance was better than the already proposed GMM.

Dynamic evidential networks proposed by Aguilar et al. (2014) were compared to the classical evidential networks approach. Comparison was made on the same dataset, consisting of 33 different fall scenarios: 16 hard falls, 17 soft falls and 5 normal situations. The overall accuracy of the two approaches was similar. However, the proposed approach could better identify soft falls. On the other hand, DEN produced more false alarms. Authors proposed a solution to that, by introducing localization in the house. This helped on classifying the areas as zones that have a higher movement probability, thus a higher fall chance. This localization, according to the authors, could reduce false fall alarms.

Ameva Algorithm

Álvarez de la Concepción et al. (2017) presented a mobile phone-based activity recognition system. They have gathered data from volunteers, between 19-48 years old, with a smartphone attached to their waist. The generated set was split randomly to training (70%) and testing (30%). Additionally three more datasets were employed for evaluation purposes all containing data gathered from mobile phones and wearables (Shoaib, Scholten, & Havinga, 2013; Weiss & Lockhart, 2012; Zhang & Sawchuk, 2012). Although all datasets were using younger adults as subjects, the system was also evaluated using people wearing accessories that would simulate the movement of elders. Of all detected activities the higher performance was achieved at fall detection with 98% accuracy, followed by cycling with 97.91%. Lowest accuracy was observed on driving (93.63%) and walking (93.5%). Driving was mainly confused with being idle, as the small movements generate a similar acceleration profile. Aforementioned results were obtained from a test case, where a young person was simulating an elder's movement and the average accuracy of their system was 95%.

Apart from evaluating the classifier, authors performed experiments regarding the energy consumption of their solution. Since the recognition was performed offline, it was crucial to assess the power needs. As optimizations were done on battery usage,

their application could run for approximately 18 hours. Comparing their energy needs with activity recognition systems proposed in the literature that perform an offline recognition, showed that the Ameva algorithm approach was the most energy efficient method.

Threshold based methods

The threshold-based method, presented by Mighali et al. (2017) was evaluated under laboratory conditions and among young and elderly people. Subjects performed targeted activities for one minute and occasionally they had small breaks standing still. Of the collected data 60% were used for training while the rest formed a validation set. The binary classification between standing still and moving achieved high accuracy, 97.2% for the former and 97.9% for the later. Validating the BLE beacon localization module, was achieved by using two scenarios. One with beacons placed on opposing walls of two rooms and one within a 5-meter proximity and visual contact between them (close to the separating door). First scenario had 100% correct localization probability, while the second one, depending on the distance between the two beacons (varying from 0.5 to 3 meters) had a probability between ~90% and 100%.

Santiago et al. (2017) developed an application running on android device, in order to evaluate their threshold-based fall detection system. The pendant was equipped with a microprocessor an accelerometer and gyroscope as well as Bluetooth capabilities. Experiments were done on different fall scenarios, such as backwards, right/left side and right/left diagonal. Front fall scenarios were not tested as they were considered dangerous. Overall accuracy was 86.6% with each fall's average accuracy shown on Table 10.

Type of fall	Number of tests	Accuracy
Backwards	11	92%
Right side	10	83%
Left side	11	92%
Diagonal left	10	83%
Diagonal right	10	83%

Table 10. Accuracy & number of tests per type of fall.

Online Daily Habit Modeling & Anomaly Detection

Meng et al. (2017) evaluated their work against two datasets: a fall detection dataset with accelerometer and gyroscope data (Ojetola et al., 2015) and the opportunity activity recognition dataset (Sagha et al., 2011) containing data from both wearables and ambient sensors for six daily activities. Apart from the precision metric, authors used the false alarm rate (11) and miss detection rate (12), with $n_{detected}$, $n_{true}, n_{false}, n_{activity}$ being the number of detected activities, correct detections, false detections and total activities respectively.

$$FA_{Rate} = \frac{n_{false}}{n_{detected}} = 1 - \frac{n_{true}}{n_{detected}} \quad (11)$$

$$MD_{Rate} = 1 - \frac{n_{true}}{n_{activity}} \qquad (12)$$

Firstly, the OAR model was evaluated using the fall detection dataset. Authors stated that the performance was very good, but false alarms did occur. Using all dimensions of the 3D accelerometer and gyroscope the performance greatly increased, recognizing 13 out of 14 falls and no false alarms were generated. Performance wise OAR achieved an 87.45% precision with 12.5% false alarm rate and 8.9% miss detection rate, outperforming other systems proposed for the same dataset. Using the opportunity dataset, the precision was 78.4%, with false alarm and miss detection rate 21.5% and 11.6% respectively. On opportunity dataset OAR was inferior to information theoretic score approach (Chavarriaga, Sagha, & Del R. Millán, 2011), on precision and false alarm, but it was the best on miss detection metric.

Additionally, authors conducted performance experiments, comparing their method with other methods proposed on the same datasets. They have implemented all solutions on Matlab, and run them under the same conditions. Their approach was found to be faster (delay for fall detection: 0.86 seconds), compared to Decision Trees (delay 1.84) and Hidden Markov Models (delay 1.32 seconds).

Probabilistic Behavior Modeling

Aran et al. (2016) used a dataset gathered anonymously covering 104 days and 45 different subjects with average age 84.3 years old. Annotations were achieved by asking subjects to report the activities' start and finishing time. For the location inference engine, the bathroom usage data were used for testing. The system was able to correctly classify 81/94 bathroom events and misclassified 46 other events as bathroom. The precision and recall of the location inference engine were found to be 86% and 64% respectively. For outing detection results were better with 86% precision and 94% recall.

After validating the location and outing detection engines, authors evaluated the anomaly detection system. Two different type of anomalies were considered: a sensor malfunction and behavior changes. Sensor malfunction could be observed in data, as there were days the subject was constantly reported in one room. Anomalies were annotated manually. Comparing the detected anomalies with the annotated ones, was done by thresholding the detected anomalies and check if an hourly detection score was higher than the threshold, thus marking the day as anomaly. If the marked day was also manually annotated as anomaly, then it was a correct detection otherwise a false one. In total 104 days of 3 subjects were used for testing. Two approaches were considered, one that evaluates all data at the end of the day and one that does it hourly. The second method had slightly better performance with 72% TPR and 38% FPR compared to 66% TPR and 29% FPR of the daily evaluation.

6   Qualitative analysis

Summarizing the results discussed on the previous section, a qualitative analysis can be performed. Since each work was evaluated on different dataset and a different set of activities was recognized, an empirical comparison was made and no experiments were

conducted. Table 11, shows each system's classification method, metric used, number of activities recognized (N) and their performance. Results obtained from Alcalá et al., (2017) were omitted, since the metrics used (belief and plausibility) could not provide results directly comparable with the rest of the reviewed papers.

Authors	Classification	Metric	N	Results (avg)		
Aguilar et al.	DEN	Accuracy	2	95%		
Álvarez de la Concepción et al.	Ameva	Accuracy	9	95%		
Aran et al.	Probabilistic behavior model	TPR, FPR	4	TPR	72%	
				FPR	38%	
Arifoglou & Bouchachia	RNN	Accuracy, TPR, FPR	9	Vanilla	Dataset A	95.5%
					Dataset B	87.9%
					Dataset C	86.7%
				LSTM	Dataset A	96.7%
					Dataset B	87.2%
					Dataset C	87.4%
					TPR	91.43%
					FPR	40.96%
				GRU	Dataset A	96.1%
					Dataset B	87%
					Dataset C	86.6%
Capela et al.	Decision Trees	F-score	6	Able bodied	55%	
				Stroke	47%	
Fan et al.	HMM & DT	Accuracy	-	93%		
G. Hu et al.	HMM	F-score	8	94%		
Hu et al.	SVM	Accuracy	2	Ambient	Labeled	74.58%
					Unlabeled	25.2%
				Ambient + Fitbit	Labeled	82.04%
					Unlabeled	26,12%
Kang et al.	CNN & Fuzzy logic	F-score	4	Simple	Walking	0.9944
					Standing	0.9922
				Transition	Turn left	0.8019
					Turn right	0.9772
Meng et al.	ODHMAD	Precision, FA /MD rate	11	Fall dataset	Precision	87.4%
					FA rate	12.55%
					MD rate	8.9%
				Opportunity dataset	Precision	78.49%
					FA rate	21.5%
					MD rate	11.64%
Mighali et al.	Threshold method	Accuracy	2	97.55%		
Riboni et al.	FABER	F-score	3	93%		

Authors	Classification	Metric	N	Results (avg)		
Sansrimahachai & Toahchoodee	C4.5 Trees	Accuracy	7	93.5%		
Santiago et al.	Threshold	Accuracy	2	86.6%		
Sztyler et al.	RF	F-score	8	Position recognition	89%	
				Subject specific	84%	
				Cross subject	1 device	79%
					2 devices	82%
Yao et al.	Dictionary	F-score	23	96%		
Zambrana et al.	Rule based	F-score	2	96%		

Table 11. Aggregate results. (-) = Not applicable, (N) = Number of activities

Usage of already trained models to different persons, whose data were not part of the training set, is a common issue when performing activity recognition. That problem was only addressed by Sztyler et al., (2017) with their cross subject recognition approach. The ability to use their system without further retraining, gives a significant advantage when moving to a real-life scenario.

Another aspect important on HAR systems is energy consumption. This is important on deployments performing offline recognition (i.e. smartphones, portable computers etc.). Álvarez de la Concepción et al., (2017), performed experiments regarding battery consumption and optimized their approach to minimize battery usage. Energy efficiency has to be further investigated in the literature as the majority of the state-of-the-art systems do not take it into consideration.

Obtrusiveness is also an important characteristic of the presented activity recognition systems. Smartphones and wearables are ambiguous regarding their intrusiveness. Since most of those devices are equipped with a GPS, camera and microphone, they could be potentially exploited to intrude a person's privacy. Additionally, the need to carry them or wear them, could be considered as a certain type of intrusion. All papers presented in this review that exploit smartphones/wearables for sensing, (Álvarez de la Concepción et al., 2017; Capela et al., 2016; G. Hu et al., 2017; Mighali et al., 2017; Sansrimahachai & Toahchoodee, 2017; Santiago et al., 2017; Sztyler et al., 2017) took advantage of the built in accelerometer and gyroscope of those devices without using any functionalities that could be considered intrusive.

Ambient sensors and RFID tags on the other hand provide the lowest intrusiveness. More specifically works exploiting ambient sensors and RFID (Aran et al., 2016; Arifoglu & Bouchachia, 2017; Fan et al., 2017; R. Hu et al., 2017; Meng et al., 2017; Riboni, Bettini, Civitarese, Janjua, & Bulgari, 2015; Sebestyen et al., 2016; Zambrana et al., 2016) provided low intrusion in a person's privacy as they only had to be installed in the house. Lowest obtrusiveness was achieved in the work presented by Alcalá et al., (2017). In their work disaggregated power data were used to recognize performed activities. Power disaggregation could be achieved using power data coming from a single device, thus no special installations or further care of the house residents is required (Nalmpantis & Vrakas, 2018).

The number of activities recognized and the different type of activities is an additional characteristic that has to be taken into consideration when designed or implementing an HAR system. Table 12 presents the different type of activities identified on each paper. A set of activities, as shown on Table 10, is identified by most state-of-the-art systems. Those actions are related to movement, walking, sitting, standing and stair climbing. Meng et al., (2017) proposed solution is the one that recognized the most activities of all the presented papers. Their work was evaluated against two different datasets, with the opportunity dataset (Sagha et al., 2011) providing data for more performed activities. Higher number of identified actions though, could result on lower performance score as the classification task becomes more complex.

Activities are considered to have distinct boundaries by most proposed HAR systems. Kang et al. (2018) recognized transition activities in their work with high F-score. This is important, as actions in real-life scenarios "blend" with each other. Another important characteristic of their solution was its scalability. According to authors, their method could potentially be trained (with small changes) to recognize any set of a simple and/or transition activities.

Authors / Activities	Aguilar et al	Alcalá et al.	Álvarez de	Aran et al.	Arifoglou et	Capela et al.	G. Hu et al.	Hu et al.	Kang et al.	Meng et al.	Mighali et	Riboni et al.	Sansrimaha-chai et al.	Santiago et al.	Sebestyen et al.	Sztyler et al.	Yao et al.	Zambrana et al.
Walking	-	✓	-	-	-	✓	✓	-	✓	-	✓	-	-	-	✓	✓	✓	-
Run	-	✓	-	-	-	-	-	-	-	-	✓	-	-	-	✓	✓	-	-
Jumping	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	-	-
Sitting	-	-	-	-	-	✓	-	-	-	-	✓	-	-	-	✓	✓	✓	-
Standing	-	✓	-	-	-	✓	✓	-	✓	-	✓	-	-	-	✓	✓	✓	-
Stairs	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	✓	-	-
Falling	✓	✓	-	-	-	-	-	-	-	✓	-	-	-	✓	-		✓	-
Lying	-	-	-	-	-	✓	-	-	-	-	✓	-	-	-	-	✓	-	-
Sleeping	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	✓
Immobility	-	-	-	-	-	✓	-	-	-	-	-	-	✓	-	-	-	-	-
Entry/exit bed	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-
Small movements	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	✓	-
Cycle	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Driving	-	✓	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-
localization	-	-	-	✓	-	-	-	-	✓	-	✓	-	-	-	-	-	-	-
Outing	-	-	-	✓	✓	-	-	-	✓	-	-	-	-	-	-	-	-	-
Appliance usage	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Authors / Activities	Aguilar et al	Alcalá et al.	Álvarez de	Aran et al.	Arifoglou et	Capela et al.	G. Hu et al.	Hu et al.	Kang et al.	Meng et al.	Mighali et	Riboni et al.	Sansrimaha-chai et al.	Santiago et al.	Sebestyen et al.	Sztyler et al.	Yao et al.	Zambrana et al.
Use toilet	-	-	-	-	√	-	-	-	-	-	-	-	-	-	-	-	-	-
Brush teeth	-	-	-	-	√	-	-	-	-	-	-	-	-	-	-	-	-	-
Prepare /eat food	-	-	-	-	√	-	-	-	-	√	-	√	-	-	√	-	-	-
Get drink	-	-	-	-	√	-	-	-	-		-		-	-	√	-	-	-
Doors usage	-	-	-	-	-	-	-	-	-	√	-	-	-	-	-	-	-	-
Drawer usage	-	-	-	-	-	-	-	-	-	√	-	√	-	-	-	-	-	-
Open close fridge/ dish-washer	-	-	-	-	-	-	-	-	-	√	-	-	-	-	-	-	-	-
Interaction with objects	-	-	-	-	-	-	-	-	-	√	-	√	-	-	-	-	-	-

Table 12: Activities recognized on each paper.

Another conclusion that can be drawn by observing table 12 is that more complex activities, such as interaction with objects and furniture, cooking and eating could be recognized using data from ambient sensors and RFID tags. Those sensing deployments could expose more information about smaller movements and object usage. On the other hand accelerometer and gyroscope provide information only about the trunk movement. Riboni et al., (2015) for example, deployed a sensing network using RFID and ambient sensors. The sensing network allowed them to identify usage of food containers, interaction with medicines and furniture. Those activities could provide more information when detecting abnormal behaviors since complex behaviors could be modeled more accurately.

6.1 HAR Taxonomy

Dividing the current literature on activity recognition and abnormal behavior detection focused on elderly people, could be achieved using a variety of criteria. The taxonomy used on this review, firstly divides the proposed systems based on whether activity recognition or behavioral modeling / anomaly detection is the goal. Another criterion used is the type of sensors employed (i.e. smartphones/wearables, ambient sensors etc.). Lastly, the literature is classified based on the type of recognition, i.e. online or offline. A summary of the design choices made by each author can be seen on Table 13.

Authors	System	Sensing	Data pro-cessing
Alcalá et al.	ABD	Non-intrusive load monitoring	Online
Álvarez de la Concepción et al.	HAR	Smartphone	Offline

Authors	System	Sensing	Data processing
Aran et al.	ABD	Ambient sensors	Online
Arifoglou & Bouchachia	ABD	Ambient sensors	Online
Capela et al.	HAR	smartphone	Offline
Fan et al.	ABD	Ambient sensors/wearables/smartphones	Online
G. Hu et al.	HAR	RFID/accelerometer	Online
Hu et al.	HAR	Ambient sensors/wearable	Online
Meng et al.	ABD	Ambient sensors	Online
Mighali et al.	HAR	Wearables/BLE	Online
Riboni et al.	ABD	Ambient Sensors /RFID	Online
Sansrimahachai & Toahchoodee	HAR	smartphone	Online
Santiago et al.	HAR	Smartphone/wearable	Offline
Sebestyen et al.	HAR	Ambient sensors	Online
Sztyler et al.	HAR	Smartphone/wearable	Offline
Yao et al.	HAR	RFID	Online
Zambrana et al.	ABD	Ambient sensors	Online

Table 13. Literature divided based on proposed taxonomy. Human Activity Recognition=HAR, Abnormal Behavior Detection=ABD.

As seen on the table above, offline processing has been chosen only by architectures exploiting the smartphone/wearable devices. The built-in accelerometer, gyroscope and magnetometer, as well as the presence of a processor and local memory, makes them a solid choice. The absence of the aforementioned capabilities on ambient sensors is what makes online recognition a one-way solution when the system relies on them for data gathering. It is worth mentioning that energy optimization in order to reduce power consumption, thus extending the recognition active time was only addressed by Álvarez de la Concepción et al. (2017).

Another characteristic extracted from systems' classification based on design criteria, is that systems performing behavior modeling and anomaly detection rely on ambient sensors for data gathering. This is due to the fact that information gained from those sensors could provide insights not only about the actions of the elderly but also their interaction with devices and objects. Additionally, wearables and smartphones require older adults to carry them in order to identify their actions, thus preventing constant behavior analysis, i.e. waking at the middle of the night, taking a shower etc.

On the other hand, the majority of systems performing only activity recognition, employ smartphones and wearables on their work. Accelerometer, magnetometer and gyroscope provide valuable data regarding acceleration, orientation with respect to magnetic north/south and the angular velocity with respect to the body axis respectively. Those features could characterize most physical activities, while stable values denote inactivity. Ambient sensors for HAR have been chosen instead of mobile/wearables for their low intrusiveness and information returned that could describe more complex activities.

7 Issues & Challenges

The main challenge at the moment when performing activity recognition, is the use of the same system without retraining on different subjects. Sztyler et al. (2017) tried to achieve cross subject recognition using subject grouping with physical characteristics and the results were promising. Using subject grouping, though requires a large and diverse amount of training data, in order to create sets with enough examples to train each classifier. In future, cross subject recognition has to be addressed, as no need for retraining, could make HAR systems accessible to a larger number of households.

It is important for HAR systems to be able to identify transition activities. Kang et al. (2018) despite the fact that they only recognized two activities in their work (walking, standing), they addressed the problem of transition activities. In uncontrolled environments the boundaries between different actions are not distinct. Additionally, multiple activities could be performed at the same time, thus a system should be able to identify more than one activity. Feature research could focus on architectures that can correctly identify transitions or multiple activities. This could be achieved using either fuzzy logic or multilabel classification algorithms. Data sets that expose such transitions are also needed, since available data have distinct activities.

Another issue when performing HAR on elderly people, is the absence of a pure dataset gathered directly from seniors. As people age, their motor pattern changes and differs from people of a younger age. This is mainly a result of changes on the body and potential health issues (e.g. osteoarthritis). Another challenge is convincing seniors to help with data annotation to provide a ground truth. There is work done on that field though, with more datasets becoming available, either gathered from elderly volunteers, or by using kits that restrict a younger person's movements, simulating an elder.

Complex sensor arrays used on activity recognition are rarely measured regarding their power and computational needs. Data gathering from heterogenous sensors, filtering, preprocessing, feature extraction and classification may need high computational power. An approach on that problem could be the separation of those tasks on different processing units.

An additional challenge is that the majority of the HAR systems are evaluated on laboratory conditions. Moreover, the subjects used on evaluation are rarely elders. This could provide false evaluation results about the system. Activities performed by youngers on controlled conditions are not close to actions from elders on real-life scenarios. Although, testing with elderly people seems as an obvious solution, there are ethical and practical complications on using them as part of an experiment.

The absence of a universal framework for evaluating and comparing HAR and abnormal behavior detection systems was also one of the conclusions in this review. At the moment, comparison with already published approaches, for evaluation purposes, is only doable by using the same dataset or implementing proposed algorithms again and retrain them with a common dataset. Creation of such a framework will provide a more solid evaluation tool. Still there are a few considerations that make that task dif-

ficult such as the sensors used, offline or online recognition etc. Since sensors and architecture of the system vary, common criteria are needed, that would set the requirements of the evaluation frameworks.

Activities with similar sensor profile is also an open challenge on HAR domain. While most actions provide distinct sensor readings, there are activities that could not be easily identified. An example is elevator usage and standing or driving and walking. A possible solution is the incorporation of sensors that could potentially provide more valuable information regarding those activities. Driving and walking, for example, could be misclassified when using a smart phone on the waist to gather data. Adding a smart band or smartwatch on the system, will allow the capture of upper body acceleration, thus distinguish those actions.

Another issue that has to be addressed, regarding evaluation, is the absence of a common metric. Although there are many metrics available for scoring HAR systems the number of recognized activities is taken into account by none. Using the available metrics, systems that identify a larger set of actions are prone to lower average scores. A new metric needs to be defined that would consider the size of the action set. Such a metric would provide a reliable scoring technique allowing the performance comparison of HAR frameworks that are now hard to compare.

Regarding abnormal behavior detection, TPR and FPR metrics provide a solid choice. An improvement that could be done when calculating the score for an anomaly detection system would be the consideration of other characteristics, apart from correct and false alarms. System's response time should be considered and further investigated as a factor that could affect the overall score. Real time behavior analysis and notifications are crucial, especially when anomalies are critical (e.g. no medicine taken, inactivity for a long time during a usually active hour etc.).

In terms of classification, deep learning and RNN have to be investigated further with non-intrusive activity recognition. Work done on that field showed promising results for both HAR and behavior analysis domains. Especially for behavior anomalies detection, LSTM neural networks were proven to be on par with state of art methods. As future work, further usage of those techniques is suggested to explore potential advantages on the field.

8   Conclusions

Human activity recognition and abnormal behavior detection, are domains with increased scientific interest. This review, presented recent advantages in those fields extensively. In total seventeen approaches were discussed, while performing an analysis on their results and categorizing them based on design choices and purpose. Reviewed material, allowed us to detect open issues and challenges on the area. Investigating each system's architecture, led to useful conclusions and feature work needed on the domain.

One of the areas we would like to focus is the continuous recognition of activities, including transitions activities. Additionally, a system that could generalize well without further re-training should be examined. As already mentioned, there are activities that have a similar motor pattern. In order to address that challenge feature work could

be focused on solutions that exploit multiple classification techniques. Each classifier could recognize a specific subset of activities and the individual results could be merged, (e.g. using fuzzification) or reported separately.

9 Acknowledgments

10 References

Abdallah, Z. S., Gaber, M. M., Srinivasan, B., & Krishnaswamy, S. (2015). Adaptive mobile activity recognition system with evolving data streams. *Neurocomputing*, *150*(Part A), 304–317. https://doi.org/10.1016/j.neucom.2014.09.074

Aguilar, P. A. C., Boudy, J., Istrate, D., Dorizzi, B., & Mota, J. C. M. (2014). A dynamic evidential network for fall detection. *IEEE Journal of Biomedical and Health Informatics*. https://doi.org/10.1109/JBHI.2013.2283055

Alcalá, J., Ureña, J., & Hernández, Á. (2015). Activity supervision tool using Non-Intrusive Load Monitoring Systems. In *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA* (Vol. 2015–Octob). https://doi.org/10.1109/ETFA.2015.7301622

Alcalá, J., Ureña, J., Hernández, Á., & Gualda, D. (2017). Assessing Human Activity in Elderly People Using Non-Intrusive Load Monitoring. *Sensors*, *17*(2), 351. https://doi.org/10.3390/s17020351

Álvarez de la Concepción, M. Á., Soria Morillo, L. M., Álvarez García, J. A., & González-Abril, L. (2017). Mobile activity recognition and fall detection system for elderly people using Ameva algorithm. *Pervasive and Mobile Computing*, *34*, 3–13. https://doi.org/10.1016/j.pmcj.2016.05.002

Amiribesheli, M., & Bouchachia, A. (2015). Smart Homes Design for People with Dementia. *Proceedings - 2015 International Conference on Intelligent Environments, IE 2015*, 156–159. https://doi.org/10.1109/IE.2015.33

Aran, O., Sanchez-Cortes, D., Do, M. T., & Gatica-Perez, D. (2016). Anomaly detection in elderly daily behavior in ambient sensing environments. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 9997 LNCS, pp. 51–67). https://doi.org/10.1007/978-3-319-46843-3_4

Arifoglu, D., & Bouchachia, A. (2017). Activity Recognition and Abnormal Behaviour Detection with Recurrent Neural Networks. *Procedia Computer Science*, *110*, 86–93. https://doi.org/10.1016/j.procs.2017.06.121

Baldinger, J.-L., Boudy, J., Dorizzi, B., Levrey, J.-P., Andreao, R., Perpère, C., … Lacombe, A. (2004). Tele-surveillance System for Patient at Home : The MEDIVILLE System. In

Verlag Berlin Heidelberg 2004. https://doi.org/10.1007/978-3-540-27817-7_59

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Capela, N. A., Lemaire, E. D., Baddour, N., Rudolf, M., Goljar, N., & Burger, H. (2016). Evaluation of a smartphone human activity recognition application with able-bodied and stroke participants. *Journal of NeuroEngineering and Rehabilitation*, *13*(1), 5. https://doi.org/10.1186/s12984-016-0114-0

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Comput. Surv.*, *41*(3), 1–58. https://doi.org/10.1145/1541880.1541882

Chavarriaga, R., Sagha, H., & Del R. Millán, J. (2011). Ensemble creation and reconfiguration for activity recognition: An information theoretic approach. In *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics* (pp. 2761–2766). https://doi.org/10.1109/ICSMC.2011.6084090

Cho, K., Merrienboer, B. Van, Bahdanau, D., & Bengio, Y. (2014). On the Properties of Neural Machine Translation : Encoder – Decoder Approaches. *Ssst-2014*, 103–111. https://doi.org/10.3115/v1/W14-4012

Damaševicius, R., Vasiljevas, M., Šalkevicius, J., & Wozniak, M. (2016). Human Activity Recognition in AAL Environments Using Random Projections. *Computational and Mathematical Methods in Medicine*, *2016*, 1–17. https://doi.org/10.1155/2016/4073584

Del Rosario, M. B., Wang, K., Wang, J., Liu, Y., Brodie, M., Delbaere, K., … Redmond, S. J. (2014). A comparison of activity classification in younger and older cohorts using a smartphone. *Physiological Measurement*, *35*(11), 2269–2286. https://doi.org/10.1088/0967-3334/35/11/2269

Do, T. M., Loke, S. W., & Liu, F. (2013). HealthyLife: An Activity Recognition System with Smartphone Using Logic-Based Stream Reasoning. In *International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services* (pp. 188–199). https://doi.org/10.1007/978-3-642-40238-8_16

Durand, V. M., & Barlow, D. H. (2003). Essentials of abnormal psychology (3rd ed.). *Essentials of Abnormal Psychology (3rd Ed.).* Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc4&NEWS=N&AN=2003-06650-000

F., S., & S., B. (2005). Telesurveillance of geriatric patients in a hospital using passive infra-red sensors. *Journal on Information Technology in Healthcare*.

Fan, X., Xie, Q., Li, X., Huang, H., Wang, J., Chen, S., … Chen, J. (2017). Activity Recognition as a Service for Smart Home: Ambient Assisted Living Application via Sensing Home. *2017 IEEE International Conference on AI & Mobile Services (AIMS)*, 54–61. https://doi.org/10.1109/AIMS.2017.29

Forman, G., & Scholz, M. (2009). Apples-to-Apples in Cross-Validation Studies: Pitfalls in Classifier Performance Measurement. *HP Labs*, *12*(1), 49–57. https://doi.org/10.1145/1882471.1882479

Garcia-Valverde, T., Garcia-Sola, A., & Botia, J. A. (2010). Improving RFID's location based services by means of Hidden Markov Models. In *Frontiers in Artificial Intelligence and Applications* (Vol. 215, pp. 1045–1046). https://doi.org/10.3233/978-1-60750-606-5-1045

Gonzalez-Abril, L., Cuberos, F. J., Velasco, F., & Ortega, J. A. (2009). Ameva: An autonomous discretization algorithm. *Expert Systems with Applications*, *36*(3 PART 1), 5327–5332. https://doi.org/10.1016/j.eswa.2008.06.063

Hall, M. (1999). Correlation-based Feature Selection for Machine Learning. *Methodology*, *21i195-i20*(April), 1–5. https://doi.org/10.1.1.149.3848

Hochreiter, S., & Urgen Schmidhuber, J. (1997). LONG SHORT-TERM MEMORY. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hong, X., Nugent, C., Mulvenna, M., McClean, S., Scotney, B., & Devlin, S. (2009). Evidential fusion of sensor data for activity recognition in smart homes. *Pervasive and Mobile Computing*. https://doi.org/10.1016/j.pmcj.2008.05.002

Hu, G., Qiu, X., & Meng, L. (2016). RTagCare: Deep human activity recognition powered by passive computational RFID sensors. *18th Asia-Pacific Network Operations and Management Symposium, APNOMS 2016: Management of Softwarized Infrastructure - Proceedings*. https://doi.org/10.1109/APNOMS.2016.7737258

Hu, G., Qiu, X., & Meng, L. (2017). Human Activity Recognition Based on Hidden Markov Models Using Computational RFID. In *2017 4th International Conference on Systems and Informatics (ICSAI)* (pp. 813–818). Hangzhou, China.

Hu, R., Pham, H., Buluschek, P., & Gatica-perez, D. (2017). Elderly People Living Alone : Detecting Home Visits with Ambient and Wearable Sensing. https://doi.org/10.1145/3132635.3132649

Huynh, T., & Schiele, B. (2005). Analyzing features for activity recognition. In *Proceedings of the 2005 joint conference on Smart objects and ambient intelligence innovative context-aware services: usages and technologies - sOc-EUSAI '05*. https://doi.org/10.1145/1107548.1107591

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*(2), 201–211. https://doi.org/10.3758/BF03212378

Kang, J., Kim, J., Lee, S., & Sohn, M. (2018). Transition activity recognition using fuzzy logic and overlapped sliding window-based convolutional neural networks. *Journal of Supercomputing*. https://doi.org/10.1007/s11227-018-2470-y

Kelly, J., & Knottenbelt, W. (2015). The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Scientific Data*, *2*. https://doi.org/10.1038/sdata.2015.7

Kornowski, R., Zeeli, D., Averbuch, M., Finkelstein, A., Schwartz, D., Moshkovitz, M., … Pines, A. (1995). Intensive home-care surveillance prevents hospitalization and improves morbidity rates among elderly patients with severe congestive heart failure. *American Heart Journal*, *129*(4), 762–766. https://doi.org/10.1016/0002-8703(95)90327-5

Kumari, P., Mathew, L., & Syal, P. (2017). Increasing trend of wearables and multimodal interface for human activity monitoring: A review. *Biosensors and Bioelectronics*. https://doi.org/10.1016/j.bios.2016.12.001

Labrador, M. A., & Lara, O. D. (2014). Human Activity Recognition using Wearable Sensors, (6), 2–7.

Lara, O. D., & Labrador, M. A. (2013). A Survey on Human Activity Recognition using Wearable Sensors. *IEEE Communications Surveys & Tutorials*, *15*(3), 1192–1209. https://doi.org/10.1109/SURV.2012.110112.00192

Mainetti, L., Patrono, L., & Rametta, P. (2015). Capturing Behavioral Changes of Elderly people through Unobtruisive Sensing Technologies, 1–3.

Meng, L., Miao, C., & Leung, C. (2017). Towards online and personalized daily activity recognition, habit modeling, and anomaly detection for the solitary elderly through unobtrusive sensing. *Multimedia Tools and Applications*, *76*(8), 10779–10799. https://doi.org/10.1007/s11042-016-3267-8

Mighali, V., Patrono, L., Stefanizzi, M. L., Solic, P., & Rodrigues, J. P. C. (2017). A Smart Remote Elderly Monitoring System based on IoT Technologies, 43–48.

Mukhopadhyay, S. C. (2014). Wearable sensors for human activity monitoring: A review. *IEEE Sensors Journal*, *15*(3), 1321–1330. https://doi.org/10.1109/JSEN.2014.2370945

Nalmpantis, C., & Vrakas, D. (2018). Machine learning approaches for non-intrusive load monitoring: from qualitative to quantitative comparation. *Artificial Intelligence Review*, pp. 1–27. https://doi.org/10.1007/s10462-018-9613-7

Nef, T., Urwyler, P., Büchler, M., Tarnanas, I., Stucki, R., Cazzoli, D., … Mosimann, U. (2015). Evaluation of three state-of-the-art classifiers for recognition of activities of daily living from smart home ambient data. *Sensors (Switzerland)*, *15*(5), 11725–11740. https://doi.org/10.3390/s150511725

Ojetola, O., Gaura, E., & Brusey, J. (2015). Data set for fall events and daily activities from inertial sensors. In *Proceedings of the 6th ACM Multimedia Systems Conference on - MMSys '15* (pp. 243–248). https://doi.org/10.1145/2713168.2713198

Paul, S. S., Tiedemann, A., Hassett, L. M., Ramsay, E., Kirkham, C., Chagpar, S., & Sherrington, C. (2015). Validity of the Fitbit activity tracker for measuring steps in community-dwelling older adults. *BMJ Open Sport & Exercise Medicine*, *1*(1). https://doi.org/10.1136/bmjsem-2015-000013

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol*, *56*(3), 303–308. https://doi.org/10.1001/archneur.56.3.303

Ramasso, E., Rombaut, M., & Pellerin, D. (2006). A Temporal Belief Filter Improving Human Action Recognition in Videos. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*. https://doi.org/10.1109/ICASSP.2006.1660299

Rashidi, P., & Mihailidis, A. (2013). A survey on ambient-assisted living tools for older adults. *IEEE Journal of Biomedical and Health Informatics*, *17*(3), 579–590. https://doi.org/10.1109/JBHI.2012.2234129

Riboni, D., Bettini, C., Civitarese, G., Janjua, Z. H., & Bulgari, V. (2015). From lab to life: Fine-grained behavior monitoring in the elderly's home. *2015 IEEE International Conference on Pervasive Computing and Communication Workshops, PerCom Workshops 2015*, 342–347. https://doi.org/10.1109/PERCOMW.2015.7134060

Riboni, D., Bettini, C., Civitarese, G., Janjua, Z. H., & Helaoui, R. (2015). Fine-grained Recognition of Abnormal Behaviors for Early Detection of Mild Cognitive Impairment, (Mci), 149–154. Retrieved from http://arxiv.org/abs/1501.05581

Robnik-Šikonja, M., & Kononenko, I. (2003). Theoretical and Empirical Analysis of ReliefF and RReliefF. *Machine Learning*, *53*(1–2), 23–69. https://doi.org/10.1023/A:1025667309714

Ronao, C. A., & Cho, S.-B. (2016). Human activity recognition with smartphone sensors using

deep learning neural networks. *Expert Systems with Applications*, *59*, 235–244. https://doi.org/10.1016/j.eswa.2016.04.032

Rosenhan, D., & Seligman, M. (1984). *Abnormal Psychology*. W W Norton & Co Ltd.

Ruan, W. (2016). Unobtrusive human localization and activity recognition for supporting independent living of the elderly. *2016 IEEE International Conference on Pervasive Computing and Communication Workshops, PerCom Workshops 2016*, 16–18. https://doi.org/10.1109/PERCOMW.2016.7457085

Sagha, H., Digumarti, S. T., Millán, J. D. R., Chavarriaga, R., Calatroni, A., Roggen, D., & Tröster, G. (2011). Benchmarking classification techniques using the opportunity human activity dataset. In *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics* (pp. 36–40). https://doi.org/10.1109/ICSMC.2011.6083628

Sansrimahachai, W., & Toahchoodee, M. (2017). Mobile-phone based immobility tracking system for elderly care. *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, 3550–3553. https://doi.org/10.1109/TENCON.2016.7848718

Santiago, J., Cotto, E., Jaimes, L. G., & Vergara-Laurens, I. (2017). Fall detection system for the elderly. *2017 IEEE 7th Annual Computing and Communication Workshop and Conference, CCWC 2017*, 1–4. https://doi.org/10.1109/CCWC.2017.7868363

Sebestyen, G., Stoica, I., & Hangan, A. (2016). Human activity recognition and monitoring for elderly people. *2016 IEEE 12th International Conference on Intelligent Computer Communication and Processing (ICCP)*, 341–347. https://doi.org/10.1109/ICCP.2016.7737171

Shinmoto Torres, R. L., Ranasinghe, D. C., Shi, Q., & Sample, A. P. (2013). Sensor enabled wearable RFID technology for mitigating the risk of falls near beds. In *2013 IEEE International Conference on RFID, RFID 2013*. https://doi.org/10.1109/RFID.2013.6548154

Shoaib, M., Scholten, H., & Havinga, P. J. M. (2013). Towards Physical Activity Recognition Using Smartphone Sensors. In *2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing* (pp. 80–87). https://doi.org/10.1109/UIC-ATC.2013.43

Simon, C., & Weber, P. (2009). Evidential networks for reliability analysis and performance evaluation of systems with imprecise knowledge. *IEEE Transactions on Reliability*. https://doi.org/10.1109/TR.2008.2011868

Singh, A., & Misra, N. (2009). Loneliness, depression and sociability in old age. *Industrial Psychiatry Journal*, *18*(1), 51. https://doi.org/10.4103/0972-6748.57861

Sztyler, T., & Stuckenschmidt, H. (2016). On-body localization of wearable devices: An investigation of position-aware activity recognition. In *2016 IEEE International Conference on Pervasive Computing and Communications, PerCom 2016*. https://doi.org/10.1109/PERCOM.2016.7456521

Sztyler, T., Stuckenschmidt, H., & Petrich, W. (2017). Position-aware activity recognition with wearable devices. *Pervasive and Mobile Computing*, *38*, 281–295. https://doi.org/10.1016/j.pmcj.2017.01.008

Thies, W., & Bleiler, L. (2013). 2013 Alzheimer's disease facts and figures. *Alzheimer's & Dementia : The Journal of the Alzheimer's Association*, *9*(2), 208–245. https://doi.org/10.1016/j.jalz.2013.02.003

Tran, T., Sutton, C., Cocci, R., Nie, Y., Diao, Y., & Shenoy, P. (2009). Probabilistic inference over RFID streams in mobile environments. In *Proceedings - International Conference on Data Engineering* (pp. 1096–1107). https://doi.org/10.1109/ICDE.2009.33

United Nations, Department of Economic and Social Affairs, P. D. (2015). *World Population Ageing 2015*. *World Population Ageing 2015*. https://doi.org/ST/ESA/SER.A/390

United Nations, Department of Economic and Social Affairs, P. D. (2017). World Population Prospects The 2017 Revision Key Findings and Advance Tables. https://doi.org/10.1017/CBO9781107415324.004

Van Kasteren, T. L. M., Englebienne, G., & Kröse1, B. J. A. (2010). Human Activity Recognition from Wireless Sensor Network Data : Benchmark and Software. *Activity Recognition in Pervasive Intelligent Environments*, 165–186. https://doi.org/10.2991/978-94-91216-05-3_8

Weiss, G. M., & Lockhart, J. W. (2012). The Impact of Personalization on Smartphone-Based Activity Recognition. *AAAI Workshop on Activity Context Representation: Techniques and Languages*, (October 2016), 98–104. Retrieved from http://www.aaai.org/ocs/index.php/WS/AAAIW12/paper/download/5203/5564

Witten, I. H., Frank, E., & Hall, M. a. (2011). *Data Mining: Practical Machine Learning Tools and Techniques, Third Edition*. *Annals of Physics* (Vol. 54). https://doi.org/10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C

World Health Organization. (2015). World report on ageing and health. 2015. *Luxembourg, Luxembourg*, 1–260.

World Health Organization. (2017). *Mental health of older adults: Fact Sheet*.

Yao, L., Sheng, Q. Z., Li, X., Gu, T., Tan, M., Wang, X., & Zou, W. (2017). Compressive Representation for Device-Free Activity Recognition with Passive RFID Signal Strength. *IEEE Transactions on Mobile Computing*, *X*(X). https://doi.org/10.1109/TMC.2017.2706282

Yusuf, B., Woo, J., Botzheim, J., Kubota, N., & Tudjarov, B. (2017). Robot Partner Technology Based on Information Support System for Elderly People and Their Family. *Proceedings - 2016 3rd International Conference on Computing Measurement Control and Sensor Network, CMCSN 2016*. https://doi.org/10.1109/CMCSN.2016.35

Zambrana, C., Rafael-Palou, X., & Vargiu, E. (2016). Sleeping recognition to assist elderly people at home. *Artificial Intelligence Research*, *5*(2). https://doi.org/10.5430/air.v5n2p64

Zhang, M., & Sawchuk, A. A. (2012). USC-HAD: A Daily Activity Dataset for Ubiquitous Activity Recognition UsingWearable Sensors. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12* (p. 1036). https://doi.org/10.1145/2370216.2370438

Zimmermann, J.-P., Evans, M., Lineham, T., Griggs, J., Surveys, G., Harding, L., … Roberts, P. (2012). Household Electricity Survey: A study of domestic electrical product usage. *Intertek*, 600. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/208097/10043_R66141HouseholdElectricitySurveyFinalReportissue4.pdf