Università Politecnica delle Marche
Scuola di Dottorato di Ricerca in Scienze dell'Ingegneria
Corso di Dottorato in Ingegneria Industriale

Development of a simulation tool for measurements and analysis of simulated and real data to identify ADLs and behavioral trends through statistics techniques and ML algorithms

Ph.D. Dissertation of:

Michela Pirozzi

Supervisor:

Prof. Gian Marco Revel

Ph.D. Course coordinator:

Prof. G. Di Nicola

XVIII edition

Università Politecnica delle Marche
Department of Mechanical Engineering and Mathematical Sciences
Via Brecce Bianche — 60131 - Ancona, Italy

All'amore che mi hai dato,

il piú grande che io abbia mai ricevuto.

Acknowledgments

At the end of this adventure, some acknowledgments are due. First, Professor *Gian Marco Revel*, to whom I owe the trust given to me, an element without which I would not have had the opportunity to complete this fundamental step of my life. No less obligatory are the greetings to my “personal coach” *Sara Casaccia*, who has always supported and guided me in the development of the various activities with fundamental advices and her kindness. Finally, an immense “thank you” goes to my beloved parents who have always supported me with love in every choice of my life and not only in this long journey. A real special thanks to you, *Dad*, your help has been the most important I could have in these last 8 years of university education. And a special thanks to you, *Mum*, for putting up with both of us.

Abstract

With a growing population of elderly people, the number of subjects at risk of pathology is rapidly increasing. Many research groups are studying pervasive solutions to continuously and unobtrusively monitor fragile subjects in their homes, reducing health-care costs and supporting the medical diagnosis. Anomalous behaviors while performing activities of daily living (ADLs) or variations on behavioral trends are of great importance. To measure ADLs a significant number of parameters need to be considering affecting the measurement such as sensors and environment characteristics or sensors disposition. To face the impossibility to study in the real context the best configuration of sensors able to minimize costs and maximize accuracy, simulation tools are being developed as powerful means. This thesis presents several contributions on this topic. In the following research work, a study of a measurement chain aimed to measure ADLs and represented by PIRs sensors and ML algorithm is conducted and a simulation tool in form of Web Application has been developed to generate datasets and to simulate how the measurement chain reacts varying the configuration of the sensors. Starting from eWare project results, the simulation tool has been thought to provide support for technicians, developers and installers being able to speed up analysis and monitoring times, to allow rapid identification of changes in behavioral trends, to guarantee system performance monitoring and to study the best configuration of the sensors network for a given environment. The UNIVPM Home Care Web App offers the chance to create ad hoc datasets related to ADLs and to conduct analysis thanks to statistical algorithms applied on data. To measure ADLs, machine learning algorithms have been implemented in the tool. Five different tasks have been identified. To test the validity of the developed instrument six case studies divided into two categories have been considered. To the first category belong those studies related to: 1) discover the best configuration of the sensors keeping environmental characteristics and user behavior as constants; 2) define the most performant ML algorithms. The second category aims to proof the stability of the algorithm implemented and its collapse condition by varying user habits. Noise

perturbation on data has been applied to all case studies. Results show the validity of the generated datasets. By maximizing the sensors network is it possible to minimize the ML error to 0.8%. Due to cost is a key factor in this scenario, the fourth case studied considered has shown that minimizing the configuration of the sensors it is possible to reduce drastically the cost with a more than reasonable value for the ML error around 11.8%. Results in ADLs measurement can be considered more than satisfactory.

Contents

List of Figures

List of Tables

Acronyms

AAL	Ambient Assisted Living
IoT	Internet of Things
SE	Smart Environments
SH	Smart Home
HSH	Health Smart Home
ML	Machine Learning
NB	Naïve Bayes
DT	Decision Tree
KNN	K-Nearest Neighbors
PIR	Pyroelectric InfraRed Sensor
ADL	Activity of Daily Living
AD	Activity Discovery
AR	Activity Recognition
DB	Database
DBMS	Database Management Software

CHAPTER 1

INTRODUCTION

1.1 Background and Motivation

1.1.1 The Aging of Society: global scenario

Population aging is happening more quickly than in the past. According to the United Nations Population Fund (UNFPA) projections [1], the number of people of 60 years or older was 205 million in 1950 and increased to 810 million by 2012, and will rise to 2 billion in 2050 globally. A significant proportion of elderly population suffers from age-related health issues such as dementia, Alzheimer's or cardiovascular diseases, diabetes, and different chronic diseases. This evolution of society is an important human and economic problem for the next years as current health and welfare institutions will not be sufficient to treat this percentage of elderly people. The increasing ageing population will provide many challenges for the heath care system and society, [2]. There will be an increasing number of diseases, health care costs and dependency rate of people, as a more and more growing necessity of caregivers. Alternative solutions have to be found and rapidly developed in order to supply help and independence to dependent people. The progressive decline in physical and cognitive skills coupled with these common diseases prevents elderly and dependent persons to live independently in their home and to performing basic activities of daily living (ADL), [3].

Health care quality of service and social cost are negatively affected by the aging population. A fundamental help in this context is given by the chance of obtaining qualitative information by conducting measurements and analyzes on quantitative data. The predisposition of sensors in intelligent environments capable of measuring explanatory quantities of the status of a subject together with algorithms capable of transforming data into useful information could allow facing the problem. As a matter of fact, increasing the number of care providers to handle the projected growing number of elderly population is not a realistic solution. On the contrary, “Telemedicine and Telemonitoring of elderly people is an actual challenge”, [4]. The progress in miniaturization and the low prices of sensors make smart Environments and sensor networks easiest to deploy even if this leads to new challenges for the process of the data acquired on these networks and the information extraction among them, [4, 5]. The recent advances in ambient intelligent is becoming an important research topic in recent years. Apart from adding more care providers, smart home environments expected to play a significant role to help elderly and dependent people and alleviate the burden of health care workers. Health Smart Home (HSH) represents a pervasive healthcare system in a smart environment and gained a high significance in recent years. HSH for elderly and dependent persons provides individual healthcare and social services such as nursing, rehabilitation, and health assistance in their own place, [6]. In particular, HSH systems aim to monitor and evaluate the person's health condition and their behaviour in performing daily life activities, make healthcare services more sustainable, enable elderly to live more independently and enhance quality of life at their evolving space (e.g. home, city, etc.). Specifically, these systems consider the monitoring of illness, handicap, and dependency in order to provide timely e-health services that meet the person's context and personalized needs in smart environments. The objective is to detect any deterioration regarding the person's health and prevent major complications. Moreover, the system aims to maintain the dependency level and avoid, as long as possible, the delays of recourse to healthcare institutions (e.g. nursing homes and hospitals). Thus, the system reduces medical costs, time, and facilitates the tasks of health caregivers through technology. Being able to identify the context of monitored persons by sensors is crucial to the success of proposed health monitoring systems. The context-aware paradigm in healthcare refers to the set of continuous processes that automatically acquire the person's information (e.g. behavioural, physiological, and environmental information), and are able to provide and automatically adapt the services accordingly. Context-aware assisted living systems must have a global and full visibility of the person's context. This visibility includes a good understanding of the person's lifestyle in performing the daily activities and detecting anomalies in behaviour as well the ability to predict the future health condition and anticipate risky situations.

1.1.2 Through aging well-being

A common definition of older adults does not exist. This is due to the variations in physical and mental health of the people categorized based on age. In general, the ageing process is accompanied by fragilities resulting from biological, psychological, and social degradations [3]. A person becomes fragile when his/her weakness threatens an important aspect of his/her life. According to the World Health Organisation (WHO), promoting age-friendly environments is one of the most effective approaches to respond to demographic ageing and increasing Healthy Life Year indicator. The preference of the elderly people is to "*age in place*" [7] which implies a review of care models considering that formal care costs will rise and there will be a tendency towards informal home care. A common idea to deal with this future lack of medical facilities is the development of Ambient Assisted Living (AAL) enabling to keep people at home, [8]. Age-friendly environments should empower older people to age in a better physical and mental status, promoting their social inclusion and participation and helping them to maintain their autonomy and a good quality of life. An age-friendly approach implies that living environments respect lifestyle choices, needs and preferences of people regardless of their age, enabling accessibility of all areas of community life, promoting inclusion and engagement. In this way, medical facilities could be reserved only for pathologies and emergencies.

1.2 Overview of the Research

The recent advances in ambient intelligent technologies including sensing, communications, and computing have made it possible to evaluate the human's daily behavior in smart environments, [8-10]. However, although a variety of sensor technologies are already available, existing e-health services systems do not satisfy the desired requirements and lack of selecting related data that reflect the real context. The determination of such knowledge from the huge amount of sensor data is a complex task in context-aware systems. Due to the complexity of the human's behaviour, extracting a meaningful knowledge for the context of the monitored person and detecting the health condition represent open research challenges [11-13]. Moreover, the costs to realize such intelligent environments are rather prohibitive by considering the infinite number of variables involved which could distort the measurement. In this regard it is necessary that the environment is ad hoc and specifically created, to be sure that the measure carried out can be correctly interpreted in terms of qualitative information. To this purpose many research work are dedicated to simulation tools development as a powerful instrument to reduce cost and time. Our objective is to improve the effectiveness of the e-health monitoring systems for elderly and dependent persons in a smart environment offering a system able to simulate real conditions and to convert virtual information in significant action to be taken in the real scenario. In this research work, we investigate methods for supporting the realization of smart environments in a smart way, knowing “a priori” the best configuration of the system in order to be able to extract from the measurement information on the subject behaviour. A new simulator, or better, a new service has been developed in this work. The research aims to provide:

1) A powerful instrument to simulate how a certain measurements chain, composed by PIR sensors and ML algorithms, reacts under certain conditions and in which measure the elements involved affect the accuracy of the results;
2) An instrument able to reduce time and costs of HSH implementation;
3) A real time monitoring system and a service for caregiver able to offer analysis on simulated or real data and to investigate thanks to statistical analysis and machine learning algorithms, behavioral users trends.

In order to reach the aim of this research, the following objectives are presented:

- Provide e-health services based on an automatic and homogeneous evaluation of ADLs;
- Define a parametric model able to describe the chain measurement for an e-health scenario;
- Improve the knowledge about the context of simulation in the scenario of human activity recognition;
- Determine how variation of some of the main elements of the system considered can affect the measure;
- Define the acceptable level on uncertainty in the measure that allows to minimize implementation costs;
- Offer an instrument able to optimize the real sensors network;
- Define the best ML algorithm to interpret data in the Human Activity recognition (HAR) scenario;
- Offer a service for real time analysis and monitoring able to identify and investigate user trend behavioral variation.

1.3 PhD thesis context: eWare project

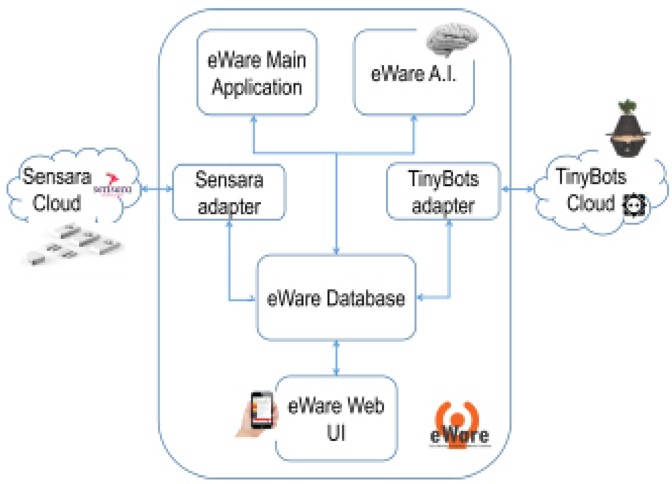
This research work is born inside the eWare scenario. The AAL project eWare “Early Warning (by lifestyle monitoring) Accompanies Robotics Excellence” is a project co-financed under the Active and Assisted Living Joint Programme of the European Commission (www.aal-europe.eu) and the National Funding Agencies of Netherlands, Italy, Norway and Switzerland. It is focused on improving the lifestyle of people with dementia and their caregivers considering the extreme impact of this disease in the world. eWare aims to develop a useful and meaningful service in co-design with human beings with main goals to:

- Reduce subjective stress of the informal carers and the patient community,

- Enhance quality of life of the informal carer and person with dementia,
- Support communication and information between formal and informal careers.

The technology and services used in eWare consist of existing lifestyle monitoring or life pattern monitoring connected and integrated with novel support robots. The eWare eco-system is characterized by the integration of the lifestyle monitoring of Sensara (motion sensors and open/close sensors) and the social robotic technology of Tinybots. To realize the eWare eco-system and integrate these technologies, three developments need to take place:

1. eWare Cloud that hosts core data
2. eWare mobile application for the caregivers
3. eWare API enables the Sensara technology and Tinybot technology to interface with the newly developed eWare eco-system.



Lifestyle monitoring is carried out by Sensara Senior Lifestyle System. Sensara is a Dutch company offering services in terms of smart technological solution for healthcare institutions and home care organizations, to be able to offer personal care to elderly people with a need for care so that they can live safely and freely in a trusted environment for as long as possible. The Sensara system consists of three PIR sensors and two open/close door sensors installed at strategic places in the home (living room, bathroom, kitchen, hallway and front door) and they automatically connect wirelessly to a gateway plugged into a standard home internet router. Sensor data are uploaded to an analytics engine “in the cloud” (Sensara Cloud) that, after two weeks, is able to recognize living patterns of the person thanks to the eWare Artificial Intelligence (eWare A.I.). When the daily behavioural patterns are known, unusual event can be detected and a warning is sent to the smartphone of the caregiver (eWare Web UI). By collecting data over a longer period in time, the system can recognize, for instance, if someone has started to walk significantly slower than two months ago. The part of social support is made by Tinybot, which is a small social robot that provide emotional support by talking, giving friendly suggestions, reminders, and playing personal music. The goal is to activate the person with dementia and prevent them from staying in a passive state. Tinybot runs a Linux-based operating system and can connect via WiFi to the Tinybots cloud backend.

The backend securely stores (amongst other data) user profiles, user's behavioural patterns, and strategies to activate the elder and it is connected to sensor data through eWare Database. Over time, the eWare A. I. learns and adapts to support the inhabitants' specific needs, giving also real time reminders and suggestions. The co-design and evaluations of the eWare eco-system will take place at four end-user sites in The Netherlands, Italy, Switzerland, and Norway with highly experienced professionals and involving a total of 300 end-users.

1.4 Thesis Contribution

The major contributions of this research work can be listed in different fields:

1. *Measurement related aspects*

   The research study has given its contribution in the development and parametrization of a measurement chain developed for a smart e-health scenario analyzing for each block which variation of parameters could affect significantly the final result. In particular, in HSH two of the biggest issues are related to the high implementation cost of a SE and the lack of datasets on which conduct studies in order to know in advance the better experimental setup for the SE itself. A solution to the problem is offered in this thesis work by considering an additional stage inside the measurement chain able to simulate how the chain itself reacts under certain conditions and variations. This helps to know in advance which could be the best configuration so as well the best analysis technique to be implemented for interpretation of results. Uncertainty related to the aforementioned aspects has been analyzed.

2. *Simulation and Virtual Reality*

   The research work offers a review of all the last approaches to simulation and virtual reality. At the same time an analysis on the simulation problem is conducted passing through motivations, advantages and weaknesses of a Virtual Environment (VE). The research work gave a contribution in this scenario by developing a simulation tool able to overcome some limits present in other models and by offering a powerful instrument to simulate the measurement chain and to conduct real time analysis on simulated dataset thanks to ML algorithms implementation. In the research work a study on different ML techniques have been conducted to define which are the most performant ones and different case studies have been implemented to test by the developed simulator the best ML algorithm to interpret such datasets.

3. *Health Service*

   An efficient e-health service has been developed in this work offering the chance to caregivers, designers and builders to dispose of an instrument able to directly and quickly analyze and discover anomalies so as well user trend behaviour deviation.

1.5 Structure of Thesis

The research work has been organized in **five** main chapters.

Chapter 1: General Introduction.

A global description of the scenario in which this thesis work gives its contribution is reported. A generic description of the aging society problem is given to show the motivation of this research.

Chapter 2: Smart Home Era

In this chapter, before to review of the state of art in the field of Ambient Assisted Living and Smart Home, the concepts of Internet of Things and Smart Environments are given. An overview of existing sensors technologies is exposed so as well a complete review in the field of Activity Daily Living is given and the Machine Learning techniques implemented in Activity Discovery and Activity Recognition are discussed.

Chapter 3: Simulation Tool Era.

This chapter reviews the most commonly used simulation tool for HAR and ADLs identification and includes a description of the two main different approaches to simulation. The state of the art on the subject is reported. Metrological aspects are deeply analysed and a parametrization of the metrological problem is given. The simulator developed in this research work is presented in this chapter and compared to those present in actual scenario.

Chapter 4: Case Studies and results

The validity of the simulator developed is presented in this chapter. Six different case studies have been conducted to test the validity of the obtained results. The measurement protocol is presented and statistical and ML results are reported.

Chapter 5: Conclusions and future works

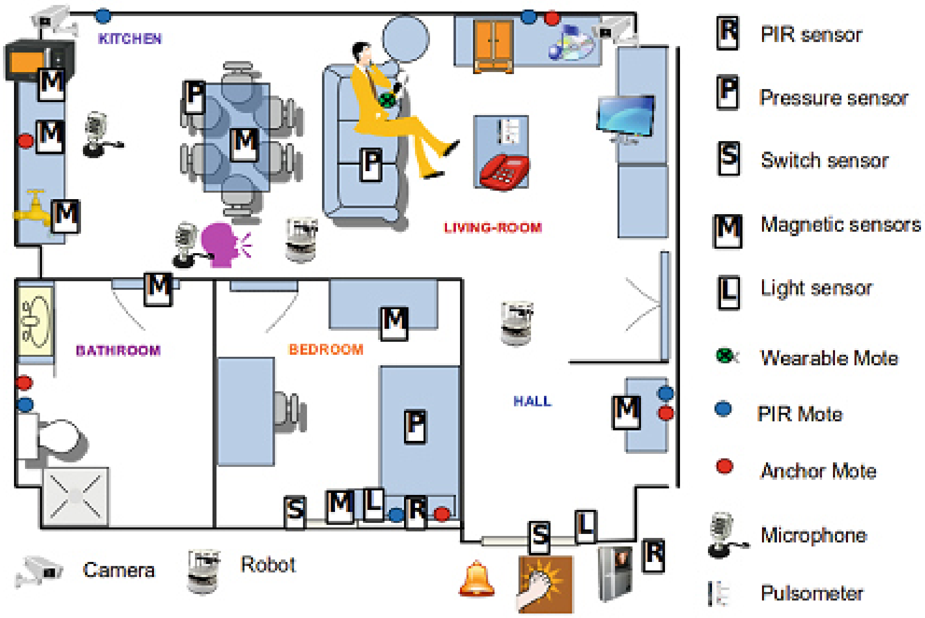
CHAPTER 2

SMART HOME ERA

The present research work born into the scenario of Smart Home technologies. To better understand the research context to which it belongs, it is necessary to provide a description of some of the main related topics.

> *"I think that everyone has some connection to people who need the technology. For me, it was not only that, but the fact that we had been designing Smart Homes in terms that this is a cool engineering advance and a machine learning application. Then I started talking to people in the psychology department and the medical field who knew that there was a need for this. Once I knew that, it was really inspiring to me to allow computer research to go from just theoretical development (which in machine learning is very common) to something that can have an impact and is needed by so many".*

(From Diane J. Cook, Regent Professor at School of Electrical Engineering and Computer Science Washington State University and director of CASAS Smart Home Project)



(Figure 1: From Giuseppe Amato et al: "*A Benchmark Dataset for Human Activity Recognition and Ambient Assisted Living* " - 2016)

- *What is a Smart Home?*
- *Why Smart Homes?*

To provide concrete answers to these questions, it is necessary to analyze the state of the art on the issue by highlighting the conditions that led to the creation of the Smart Home paradigm.

The concept of Smart Home emerges with the introduction of network enabled devices and ultramodern electronic equipment usable at home. The Internet of Things and Smart Environments have changed simple homes into smart homes, changing our traditional approach to building devices, systems, services and transforming people's lifestyles.

Elderly people who live independently and have not developed severe pathologies, need a solution that can monitor their daily life activities to ensure their safety, improve their lifestyle and reduce the costs of health services and time. To meet these requirements, it is necessary to monitor these people using data extracted from specific sensors. To achieve such a goal, different approaches based on different scientific backgrounds, different models and different considerations on the sensors that may be installed were developed.

In this chapter, therefore, a detailed analysis of the extensive literature consulted, relative to this topic and to the basic idea about smart home and its features, technologies and application areas will be produced.

First, the fundamental aspects of Ambient Assisted Living will be discussed. Secondly, the concepts of Smart Homes is developed and integrated with a technological description of the sensors and sensor networks implemented. Finally, a complete definition of Machine Learning techniques and Activity Daily Living will be given and Activity Discovery and Activity Recognition will be discussed.

2.1 Smart Homes

2.1.1 Ambient Assisted Living (AAL)

The rapid advances in the Information and Communication Technology (ICT) along with the proliferation of the broad range of sensors and actuators that are becoming readily available at low costs have facilitated the explosive growth of ***Internet of Things (IoT)*** infrastructure [14], as a subset and, at the same time, evolution of *ubiquitous computing* (or *pervasive computing*) [15].

IoT was generally defined as "*dynamic global network infrastructure with self-configuring capabilities based on standards and interoperable communication protocols; physical and virtual things in an IoT have identities and attributes and are capable of using intelligent interfaces and being integrated as an information network*" [16], [17].
One of the best result of IoT application is the development of ***Smart Environments (SE)*** systems "*able to acquire and apply knowledge about an environment and able to adapt to its inhabitants in such a way as to improve their experience in that environment*", Cook and Das [18]. They propose the goal of exploiting the large number of small computational nodes to identify and provide personalized services to the user while they interact and exchange information with the environment [19].
The context of this thesis work is exactly inserted inside a typical field of Smart Environments, the ***Ambient Assisted Living (AAL).*** An AAL can be defined as "*the use of information and communication technologies in a person's daily living and working environment to enable them to stay active longer, remain socially connected and live independently into old age*" [20], or even as "*assistant systems for the constitution of intelligent environments aiming at compensating predominantly age-related functional limitations of different target groups through technological information and communication support in everyday life. At the same time, they take charge of control and supervision of services for an independent course of life*" [21].
AAL puts together the target applications and the requirements derived from the Smart Home and Health Smart Home paradigms.
As stated in the general introduction, due to the increase of the population median age, alternatives to classical medical facilities should be found to efficiently take care of dependent people. One of the most retained solutions is to maintain in their home environment the people who do not have too severe pathologies. To this aim, it is necessary to monitor these persons using data extracted from an adapted instrumentation. To achieve such a goal, different approaches based on different scientific backgrounds, different models and different considerations on the sensor that may be installed were developed.
Safety, well-being and quality of life of every person are strongly linked to the efficiency and comfort of the house in which we live: the use of the Smart Environments technologies can provide help to enable people elderly or disabled people, often with special needs, to be autonomous, independent and continue to live in their own home for most possible time.
All these motivations are at the base of development and of always greater success of AAL techniques, precisely because they are finalized to the design of a set of technological

solutions designed to make the environment in which you live in proactive, intelligent and cooperative way to support people's independent lives by providing security, comfort and support in carrying out the activities of daily life thanks above all to the possibility of continuous and remote monitoring of environment and people who, in perspective, will also allow a abatement of the total costs of assistance, especially at home. In the most advanced forms, the AAL systems contain in particular portable and miniaturized medical devices specially designed for diagnostic and therapeutic use in a domestic or mobile environment (tele-monitoring). These can develop into individualized and person-centered forms of prevention, diagnosis, therapy and medical care.

The European Environmental Assisted Living Association (AAL) (http://www.aal-europe.eu), aimed to support and promote initiatives and projects to improving the quality of life of fragile people, has identified the main objectives that an AAL infrastructure must achieve in the chance to:

- Extend the period in which people can live in them favorite environment, increasing their autonomy and mobility;
- Maintain the health and functional capacity of the elderly;
- Promote better and healthier lifestyles for people a risk;
- Increase security, prevent social exclusion and maintain the people's relational network;
- Support operators, family members and organizations assistance;
- Improve the efficiency and productivity of resources in society that becomes older.

Several studies have been conducted on AAL projects. In [22], authors proposed a survey on real-time human wellness annotation. The work is focused on multiple tasks, such as sensor types, frameworks, data collection, processing, and analysis. In [23] a review of some smart home projects that are related to three desired services: comfort, healthcare, and security, is presented. The review also described several important components of the systems: sensors, multimedia systems, communication protocols, and algorithms. The authors of [24], provide a thorough review of environmental sensor systems which are used in many assisted living environments. In [25], authors presented an ambient-intelligence-assisted healthcare monitoring review, in which they mostly described works that are based on wireless sensor networks technologies with applications. Furthermore, they discussed several data mining techniques for ambient sensor monitoring of patients with chronic diseases and elderly people. The researchers in [3] surveyed AAL technologies for elderly care, even exploring healthcare applications that focus on algorithms for modelling elderly behaviors in smart homes. In [26] a classification of activities of elderly people in smart home scenarios is proposed, while in

[27], authors analyzed current studies related to daily activity and significant event monitoring of the elderly.

2.1.2 Smart Home concept and definition

The concept of smart homes is the most typical example of smart environments and applications of IoT in a domestic environment. As an extension of home automation, smart homes attempt to integrate different home-based smart objects to offer new or advanced functionalities to residents. The characteristic of *smart* applied to the *home* should include the capability of reasoning over the collected data from the sensors, including previous knowledge of the context, adaptability to changing situations in the environment and their inhabitants, and learning.

In literature, different definitions of smart home have been formulated. A ***Smart Home (SH)*** can be defined as

- "*a residence equipped with a communication network, linking sensors, domestic appliances, and devices, that can be remotely monitored, accessed or controlled and which provides services that respond to the needs of its inhabitants"* [28],
- *"a home which is automated through the application of the Internet of Things technologies and capable of reacting to requirements of the inhabitants, providing them comfort, security, safety and entertainment "* [29],
- "*an application of ubiquitous computing in which the home environment is monitored by ambient intelligence to provide context-aware services and facilitate remote home control*".

This last definition highlights the value of observation to provide proactive services, especially valuable in supporting the independent living of people with disabilities and elder adults.

2.1.3 Smart Home requirements and problems

The application domains and technologies used in a Smart Home give rise to a series of requirements and problems that affect the correct functioning of the infrastructure, that can be summarized in:

- *Heterogeneity* - the considered application areas are built upon devices that are extremely heterogeneous in terms of access mechanisms and data produced. In particular, devices come from various vendors and are designed for different purposes. They differ with respect to the hardware and software modules, the communication protocols, the interaction paradigms, and the data rate. Furthermore, the applications themselves produce heterogeneous data that need to be uniformed in order to be used by other applications. [30].

- *Interoperability* - an interoperable infrastructure should hide all the technical complexity concerning the access of devices in the Smart Environment, by offering simplified interfaces to devices or humans willing to access such devices [19, 31]. Interoperability is of three types: network, semantic, and syntactic.
- *Scalability* - The increasing number of connected devices gives also rise to scalability issues, such as giving communication capabilities to a large number of devices installed and dynamically added or removed to the Smart Home. For example, new functionalities should be able to be added without altering the existing components. The component needs to be developed according to standards across the solutions, which improves scalability and extensibility [30];
- *Context-awareness* - In order to support the occupants, a smart environment must be able to both detect the current state or context of the environment, and to determine what actions to take based on context information. We define context "*any information that can be used to characterize the situation of an entity, where an entity can be a person, a place, and a physical or computational object*". This information can include physical gestures, relationship between the people and objects in the environment, features of the physical environment, identity and location of people and objects in the environment. We define applications that use context to provide task-relevant information and/or services to a user to be context-aware. Context-aware systems are concerned with the acquisition of context, the abstraction and understanding of context, and application behavior based on the recognized context. As the user's activity and location are crucial for many applications, context awareness has been focused more deeply in the research fields of location awareness and activity recognition. [14, 32];
- *Security and Privacy* - Introducing Smart Environment technology in the homes results in new security and privacy challenges. Smart Environments require high level of security, because home environment contains important and private information. The modern technologies offer both opportunities and risks because the Smart Environment is highly vulnerable to attacks from the internet; if a smart device was hacked, the attacker has the potential to invade the user's privacy, steal personal information and monitor them inside the home [33, 34, 35, 36, 37, 38];

- *Usability and Accessibility* - The high technological complexity inherent in smart environments leads to problems of accessibility and usability. Accessibility is the characteristic of a device, service, resource or an environment that can be easily accessed by any type of user. The term is commonly associated with the possibility even for people with reduced or embedded sensory, motor, or psychic abilities to access and move independently in physical environments, benefiting from the resources available through the use of assistive technologies. The requirement of usability is necessary for the infrastructure to be effective, efficient, satisfying and comfortable, easy to use and safe [39];
- *Personalization* and *Adaptability* - Adaptability, understood as the ability of a system to adapt efficiently and rapidly to changes, is one of the peculiarities that a smart environment must possess. An adaptive system is therefore an open system that is able to adapt to its behavior based on the changes taking place in its environment or in parts of the system itself. The adaptability therefore allows the complete customization of the infrastructure according to the specific needs highlighted by the users, aimed at optimizing the resources and services offered [40].
- *Availability and Reliability* – In Smart Home infrastructures, one of the most challenging problems is concerned with the defining and computing of reliability and availability measures since an object or thing or device quality of service failures can lead to dangerous situations for people as well as physical infrastructures. In [41], authors propose a probability based concept for measuring the reliability and availability of the devices and things connected in Smart Home. Data reliability relying on wireless networks is subject to a number of factors. The main factors are the network coverage, device range, available power, routing protocol, and failures of the network or devices. Reliability issues have been classified into three main categories: data measurement, data communication, and data analysis [42]. Numbers of studies addresses the issues in health monitoring over wireless transmissions, for instance, using redundancy elimination [43] and data cleaning. Authors in [42] described the healthcare monitoring issues with a focus on software problems including data collection, data fusion, and data analysis. Thus, they proposed an architecture for handling data cleaning, data fusion, and context and knowledge generation for more reliable data analysis. It was observed that the number and complexity of sensors and used methods is significant if compared to traditional infrastructures. In addition, the heterogeneity of data sources adds a greater level of complexity that must be taken into account to have reliable data from any source. Existing

sensors still limited in terms of hardware and software capabilities, hence they can be subject to failure at any time.

- *Unobtrusiveness* - a Smart Home should provide innovative human-machine interactions characterized by pervasive, unobtrusive, and anticipatory communications. In this regard, devices and applications must be as much as possible unobtrusive to reach a high acceptance. It is needed to correctly balance the presence of assistive solutions in order to let the user feel safe and protected by the environment but not continuously controlled. This is particularly important in the long-term monitoring of the user and it also affects the choice of algorithms to be used.
- *Intrusiveness* - Improving the quality of life and making it easier and comfortable is one of the Smart Home challenges. Wearing and carrying sensors all the time in some proposed applications is a very cumbersome task and need more efforts to curb violations on the human lives of residents. Some studies addressed part of this problem with the mobility and portability solutions.

2.1.4 Smart Home Technologies

Smart Homes are infrastructures built around physical objects, sensors, and actuators with data communication capabilities. This section will highlight the main features of the different types of sensors installed in a smart home environment.

2.1.4.1 Sensor Systems

Data acquisition is the first step in the smart home environments in which various sources are used to gather the information related to the physical status of the person, his behavior, the environment, performed activities, and more.

The ability of a smart home system to identify and predict the behaviour of its occupants is strongly depending of the characteristics of the sensors involved, their number and predisposition so as well on artificial intelligence used to interpret data. As a consequence, the choice of the characteristics of the sensors network is a priority in a conscious smart home implementation. With an inadequate sensors environment, could not be possible to distinguish some activities from others using all data recorded, so as it could be possible for certain activities to take place in parts of the environment not covered by sensors. Nevertheless, the over-instrumentation of the environment is not a good solution to the problem: sensors can be expensive, not well accepted by occupants, and having too much data can pose algorithmic problems. The solution is generally represented by a smart definition of the measurement environment. In the coming chapter a smart and fast way to go through this direction is presented. Before, it is important to define the sensors context scenario and more, the three main categories of sensors that are typically used in smart home systems, wearable sensors, environmental sensors and multimedia devices. In this research work the focus will be on the second category.

Wearable sensors

Wearable sensors have been prominently featured in healthcare and elderly care research at home. They generally refer to sensors that are attached to human body either directly or indirectly, and which usually provide a continuous flow of information. Their small size allows them to be embedded into belts, clothes, glasses, wristwatches, shoes and mobile devices making them easier to wear. These sensors can be divided into *inertial sensors* and *vital sign sensors* (or *biosensors*).

- *Inertial Sensors*

Wearable inertial sensors can give accurate descriptive features of user's movement and body posture. *Accelerometers* are the most frequently used sensors for ambulatory activity monitoring. They can measure the value of acceleration along a sensitive axis and are particularly effective in monitoring activities related to body motion such as doing exercise, walking, standing, sitting, or walking upstairs and downstairs. Due to their small size and relatively low cost, accelerometers can be embedded into wrist bands, watches, bracelets and belts to monitor the user's activities, falls detection and wirelessly send data to mobile computing devices.

Gyroscopes use a small vibrating mass inserted into the sensor for measuring angular velocity and maintain orientation. The change of the angle compared to the initial known value can be detected over a period of time. However, inertial sensors also suffer from limitations. The placement of inertial sensors on diverse positions may result in cumbersome and uncomfortable feeling, which may lead to low acceptance by the older adults. Most wearable inertial sensors need to collect data continuously, thus the battery life and effectiveness of the device may become a great challenge. Using inertial sensors cannot in many occasions provide sufficient context information, especially when monitoring complex motions and activities that involve multiple interactions with environment objects. Inertial sensors are nowadays commonly found in recently smartphones (tablet and smartwatch). This has thus motivated researchers to use them as wearable sensors for activity recognition in general. Within the CASAS Smart Home Project (http://casas.wsu.edu/smart-homes), developed by researchers at Washington State University directed by Diane Jane Cook, was created an Android application that detects and identifies the user's daily activities using the inertial sensors embedded in the mobile device.

- *Vital Signs Sensors*

Vital signs collected from wearable biosensors such as heart rate, blood pressure and skin temperature are critical for elderly people’s health condition monitoring. There are various biosensors used to measure the wide range of vital signals: *Electroencephalography* (EEG), *Electrooculography* (EOG), *Electromyography* (EMG), *Electrocardiography* (ECG), *pressure* sensors to monitor blood pressure; *CO2 gas* sensors for the respiration; *thermal* sensors for monitoring body temperature.

These vital sign parameters can help monitor the user’s health status during the execution of activities. Based on the data collected from these biosensors, services such as further disease prediction, anomaly detection and diagnosis decision-making can be provided. Biosensors, like other wearable sensors, have as advantages their low cost, low error levels, non-intrusiveness and high accuracy. Besides, they are very sensitive to slight changes of physiological signals, and thus they can support non-invasive alternatives for continuous healthcare monitoring in smart home environments. The disadvantages of biosensors include reliability constrains and uncomfortable feeling for long time skin attaching.

Environmental sensors

Different type of sensors can be used in different areas or linked to a range of objects to monitor activities in a smart home. Most of ADLs are performed in specific locations and with specific user-object interactions. Activity can be recognized from user-object interactions combined with environment observation. It is assumed that environmental sensors data can constitute powerful information to observe the human behaviours within smart home. Simple *binary sensors*, including *state-change sensors*, *motion sensors*, *contact switches* and *pressure sensors*, may be deployed on a range of objects in smart home environments for monitoring users’ movements and locations. In real-world scenarios using a single sensor type normally cannot provide enough information for detecting activities, especially for some complex ones. Passive infrared sensors are used to detect motion in a specific area, reed switches are used to detect open/close states of doors and cupboards, and float sensors are used to measure the toilet being flushed. These low-cost, easy-to-install and long-lived binary sensors exhibit the advantages of unobtrusive user-object interaction monitoring in a privacy-preserving way. In addition, they are easy to replace and the gathered data require minimal computation resources. The main drawback is that they can only provide very limited information especially for composite and multi-user activity monitoring. *Radio-Frequency Identification* (RFID) works as a combination of environmental sensor and wearable sensor technologies. It consists of a reader worn by the user and an electronic tag attached to an object. The tag responds to a unique identifier, electronically stored in memory, when interrogated by a reader. In smart home both passive and active RFID tags may be used. A *passive* RFID tag does not contain a power source and is usually attached to an object for detecting the interaction between a user and the object. An *active* RFID tag contains a battery and is often carried by a user for personal identification throughout the house. RFID also has disadvantages, such as reader collision and tag collision. A variety of other sensors such as *light* sensors, *temperature* sensors, *humidity* sensors or *power* sensors have been also deployed and used in smart home environments to help in the detection of activities. These sensors can perform intuitive monitoring of environment and object, but on their own they only can provide very limited information for activity monitoring.

Multimedia devices

Understanding scenes and identifying activities of people from images and video streams is one of the many applications of computer vision. Recognizing activities of occupants in smart homes is thus a research problem that can be approached from this angle, especially considering how active and large the current computer vision research community is. In much the same way, recognizing activities from sound is also an approach used in smart home research. We find numerous examples where microphones are used to captured audio streams throughout the home, used as the sole data collection modality to recognize activities of occupants. In recent years, various cognitive assistants, such as the Google Home or the Amazon Echo, have been introduced in the consumer market. Such systems always include microphones, but also sometimes even cameras. As such, these cognitive assistants can, in addition to their standard uses, be used as data sources for activity recognition.

Video cameras are low-cost devices that can provide very detailed and rich context information about human actions and environmental states. Video information can provide direct and clear information about the objects within smart home, for example, the number of people. However, they face difficulties including privacy issues, high computational expense and environment dependency. On their part, *microphones* have as advantages their ability of providing accurate information about users' communications and sounds in specific locations inside smart home. However, they suffer from implementation difficulties and high computational costs associated to the audio processing algorithms necessary to distinguish different sounds, especially when there are multiple residents inside the home. Microphones, although probably to a lesser extent than video cameras, can also be perceived as privacy threats, since they can potentially record private conversations.

2.1.4.2 Pyroelectric InfraRed (PIR) Sensors

The PIR, literally Passive InfraRed sensor, is an electronic device that measures infrared rays (IR) irradiated by objects in its visual field. It is composed of a sensitive element capable of detecting a moving heat source and transforming it into an electrical signal. This feature makes it particularly useful as a motion sensor. Moreover, they are:

- Extremely sensitive and precise in the detection, even in critical environmental conditions;
- Able to withstand shock and mechanical vibrations;
- Not sensible to electromagnetic interference;
- Extremely reduced consumption.

How the PIR works

All objects with a temperature above Absolute Zero (0 Kelvin / -273.15 °C) emit heat energy in the form of infrared radiation, including human bodies. The hotter an object is, the more radiation it emits. PIR sensor is specially designed to detect such levels of infrared radiation. It basically consists of two main parts: a *Pyroelectric Sensor* (and a special convex plane lens (called Fresnel lens) which focuses the infrared signals onto the PIR sensor. A PIR actually has two rectangular slots in it made of a material that allows the infrared radiation to pass. Behind these, are two separate infrared sensor electrodes, one responsible for producing a positive output and the other a negative output. The reason for that is that we are looking for a change in IR levels and not ambient IR levels. The two electrodes are wired up so that they cancel each other out. If one half sees more or less IR radiation than the other, the output will swing high or low. When the sensor is idle, i.e. there is no movement around the sensor; both slots detect the same amount of infrared radiation, resulting in a zero output signal.
But when a warm body like a human or animal passes by; it first intercepts one half of the PIR sensor, which causes a positive differential change between the two halves. When the warm body leaves the sensing area, the reverse happens, whereby the sensor generates a

negative differential change. The corresponding pulse of signals results in the sensor setting its output pin high.

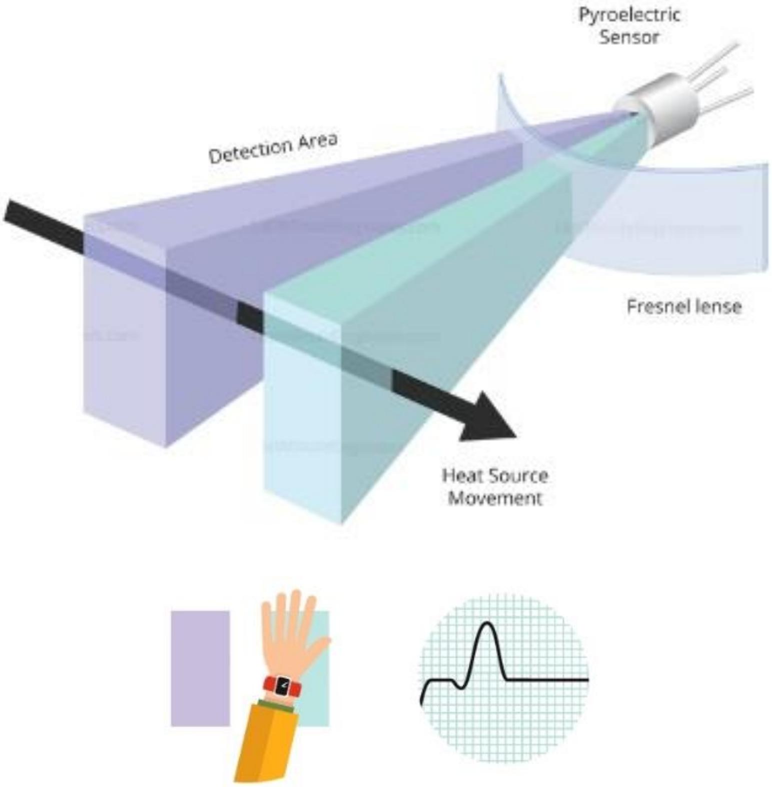


Figure 2: Detect motion with PIR sensor

Human motion detection using PIR

Numerous sensing systems have been studied using various sensors, including cameras and wearable sensors, for human movement detection and identification. Recent developments in the field of micro and nano technologies have made available wearable sensors of very small dimensions. However, moral and ethical issues inevitably arise related to questions of invasiveness and privacy. This type of sensor also requires the full cooperation of the user who in many cases could refuse or forget to wear the sensor, especially in the presence of serious diseases such as dementia and Alzheimer's. Also cameras and microphones are affected by privacy issues, being able to be considered too invasive and therefore could be rejected by users.

For these reasons, the research showed interest in approaches to behavioral analysis based on environmental sensors to obtain information on user behavior through its influence on the environment, without necessarily compromising its privacy and requesting its collaboration. Passive infrared sensors have been used to detect and localize humans because of their simplicity and less privacy concerns. By analyzing information registered by the motion sensors placed in the environment, it is possible to monitor the regularity and the level of ADLs of the inhabitant. Many of the daily activities are periodic and repetitive. An analysis of information from the environment can effectively highlight behavioral patterns relating to a person in a completely free and non-invasive way, without requiring user intervention. Moreover, other advantages of using PIR sensor in motion detection can be highlighted:

- First, they do not require any signal or device on the object to tracked;
- Second, they can work in dark environment as well, whereas vision-based system cannot;
- Third, they are cheap, easy to use and not require huge computational power.

However, they cannot provide accurate and detailed information about object within environment. Despite this, several studies have been conducted on the use of PIR sensors in the definition of human activities. Yang et al. in [44] proposed an innovative localization method for tracking human position in indoor environments based on PIR sensors using an accessibility map and A-Star algorithm, aiming at providing intelligent services. PIR sensor data were collected and sent through a Zigbee communication unit to a PC. The experiment results demonstrated the performance of method applicable for robots created to accompany elderly people living alone. In [45], authors proposed a technique to track the path of a human using PIR. In [46], an automatic identification and tracking method by combining data from PIR sensors and floor pressure sensors is presented. In [47], they proposed a region-based human tracking algorithm based on the output signals of several PIR sensors. A mathematical abstraction of a PIR sensor as a building block for the algorithm is provided. In [48], a human indoor localization system based on ceiling mounted PIR sensor nodes was proposed. In the system, five sensor nodes are utilized to form a wireless sensor network. Yun et al. in [49], present an empirical study of human movement detection and identification using PIR sensors. They have performed classification analysis with well-known machine learning algorithms, including instance-based learning and support vector machine, obtaining an accuracy of over 92% in the classification of the direction and speed of movement, of the distance interval and in the identification of the subjects. In [50], researchers introduced a wireless distributed PIR system for tracking and identify multiple humans based on their body heat radiation and gait. In [51], authors presented a person localization algorithm using an infrared ceiling sensor network for providing various ad hoc services in a office environment. The result have shown 84% accuracy on recognizing five person with support vector machine. In [52], a fall detection system is present. Luo et al. in [53] proposed a PIR-based sensing system for anomaly detection, designing a PIR sensor node able to capture the features related to space and time of human motion effectively. PIR sensor models can also be used to construct wireless sensor networks, which are intended to track and recognize multiple human targets [54]. Using only the binary information obtained by infrared sensors attached to the ceiling of a room, the human positions can be estimated, and even the number of humans in the room changes dynamically [55]. In [56], PIR sensors were deployed in a distributed sensing paradigm, which aimed at capturing the synergistic motion patterns of head, upper-limb and lower-limb. All the aforementioned studies proved the validity of the information deriving from PIR sensors. This quality together with their low cost and high level of acceptance, have made PIR the central choice in this research work.

2.1.5.3 Sensor data processing

The set of used heterogeneous sensors provides basic raw data which represent the low-level of the context. Low-level data are imperfect, uncertain and can present anomalies, errors and sensing failures. Therefore, it is required for further development to build a high-level context abstraction that can be used in providing smart home services. Raw data obtained from sensors. Through data processing, a data mining technique, it is possible modify this raw information to obtain a high-level context abstraction that can be used in providing smart home services. [26, 57, 117]. The most common data processing methods implement the following procedures are shown in Figure 3.

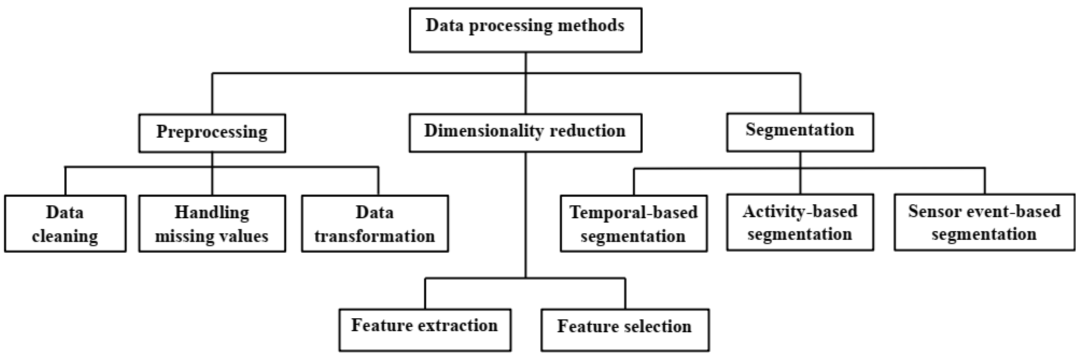


Figure 3: Sensor data processing method classification [26]

2.2 Machine Learning

All the data deriving from the aforementioned studies need to be analyzed and interpreted. The necessity to extract significant information and knowledge from this huge quantity of data is a quite impossible task to be performed if we think to do by hand.

Several ML algorithms have been developed in the recent years to study and discover ADLs. The focus in this research work is on three popular classifiers: Decision Tree, KNN and Naïve-Bayes. The reason behind this choice is given by the research study conducted on literature that is presented in the following sections. In several studies the three algorithms 1) proved to carry out good results in the ADLs identification scenario, 2) have shown quite good stability, 3) are simple to implement and interpret.

Machine learning can be considered as a set of algorithms that automatically find interesting patterns and relationships over the data. As defined in [58], Machine Learning is "*an automated process that extracts patterns from the data*". It can also be defined as "*an application of artificial intelligence that provides systems the ability to automatically learn and improve from experience without being explicitly programmed*".

Machine learning has been used extensively in human activity recognition and has applications in industry, medicine, economics, natural and technical sciences, ecology, finance, and many others. The result of the learning is the knowledge the system can use to solve new problems. An algorithm infers the properties of a given set of data and that information allows it to make predictions about other data that it might see in the future. This is possible because almost all non random data contains patterns which allows a machine to generalize. For the context of learning, the amount of data required by the algorithm is of primary importance and must be utilized to the maximum extent. The primary objective of a machine learning is to harness the predictive capabilities of the machine and hence the predictive accuracies must be as high as possible with minimal error rate.

2.2.1 Concepts, instances, attributes and patterns

Before going into the details of machine learning techniques, some concepts and common terminology must be defined. In machine learning scenario, an input is given by the sum of three elements; *concepts* (classifier), *instances* (examples) and *attributes* (features). The data provided by the classifier is referred to as an instance that is a unique example of the concept to be analyzed. Attributes may take either *numerical* or *nominal* values. Numeric attributes also called *continuous attributes* are either integer or real valued numbers. Nominal attributes also called *categorical attributes* take values from a finite set of possibilities. The aim of the learning process is to produce a distinct characterization of what the data represents in a form that the classifier can use to locate analogous traits. Therefore, the traits of each instance are judged by the attribute values which are contained in every instance. The use of the term *pattern* is common in machine learning. It is defined as a sample of data that conveys useful information that can be repeated in a recognizable way. Generally it is possible to distinguish between:

- *Numerical* pattern: are measurable values, properties or characteristics;
- *Categorical* patterns: properties and qualitative characteristics of an object that cannot be mapped numerically;
- *Sequential* patterns: data sequences of fixed or variable length.

2.2.2 Machine Learning Techniques

Machine learning can be divided into multiple sectors on the basis of the types of work to be performed. Three major areas of studies are identified in the literature: *Supervised*, *Unsupervised* and *Reinforcement method* [59].

Supervised learning is the most popular paradigm for performing machine learning operations. The algorithms are presented with a set of classified instances from which they learn a way of classifying not-seen instances. It is called supervised because the scheme works under supervision by being provided with the actual outcome for each of the training instances. It is widely used for data where there is a precise mapping between input-output data. The dataset is labeled, meaning that the algorithm identifies the features explicitly and carries out predictions or classification accordingly. As the training period progresses, the algorithm is able to identify the relationships between the two variables such that it can predict a new outcome. The success of the classification can be measured by testing the generated model with an independent set of instances for which the true classifications are known but are hidden to the classifier. Supervised learning algorithms are task-oriented. Providing it with more and more examples, it becomes able to learn more properly and provide an output more accurately. To evaluate the performance of different methods, a common practice is that to divide the set of instances into two sets: training, test and usage. The training set is used to build the classifier model, while the test set is used to measure the accuracy of the classifier, i.e, it is a measure on how well it generalizes to unseen instances. Finally, usage phase uses the model for classification on new data whose class labels are unknown. Supervised learning could be distinguished in two main categories: predictive or directed so as well divided into two branches: Classification (Support Vector Machine, Naïve Bayesan, Decision Tree, K-Nearest Neighbors, Logistic Regression) and Regression (Linear Regression, Support Vector

Regression, Ensemble Method, Neural Networks). Three of the most popular families of classifiers that have been considered in this research work: Naive-Bayes (NB), Decision Tree (DT) and K-Nearby Neighbors (KNN). Their characteristic will be deeply analyzed in 2.2.3

Unsupervised learning is the machine learning method of trying to locate hidden structure in unlabeled training data. The model is able to learn from data by finding implicit patterns. Unsupervised Learning algorithms identify the data based on their densities, structures, segments and other similar features. Unsupervised learning can be *descriptive* or *undirected* and can be divided into: *Clustering* (K. Means, Hierarchical, Hidden Markov Model, Gauss Mixture Model) and *Associations* (Apriori, FP-Growth).

Reinforcement Learning covers more area of Artificial Intelligence that allows machines to interact with their dynamic environment in order to reach their goals. With this, machines and software agents are able to evaluate the ideal behavior in a specific context. This type of learning is different from Supervised Learning in the sense that the training data in the former has output mapping provided such that the model is capable of learning the correct answer. Whereas, in the case of reinforcement learning, there is no key answer provided to the agent when they have to perform a particular task. When there is no training dataset, it learns from its own experience. Q-Learning and Monte Carlo Method are the most implemented algorithms. Figure 4 shows some of the machine learning techniques proposed in the literature.

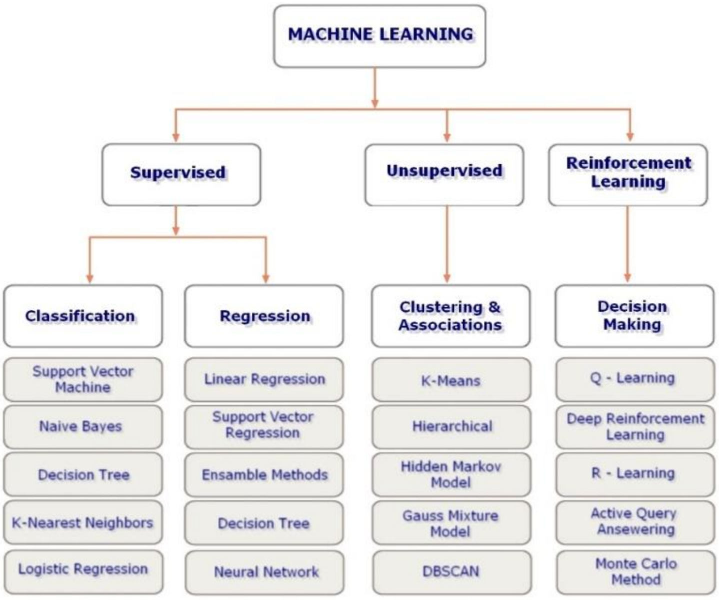


Figure 4: Machine Learning techniques

2.2.3 Classification

In machine learning, classification is a supervised learning approach in which the algorithm learns from the data input given to it and then uses this learning to classify new observation. The success of the classification can be measured by testing the generated model with an independent set of instances for which the true classifications are known but are hidden to the classifier [60]. When evaluating the performance of different methods, it is a common practice to divide the classification in three steps: *Training*, *Testing* and *Usage*. Training phase is used to build the classifier model from training instances. The classification algorithm finds the relationships between predictors and targets, then the relationships found are summarized in a model. Testing phase is used to measure the accuracy of the classifier checking the model on a test sample whose class labels are known but not used for training the model. Usage phase uses the model for classification on new data whose class labels are unknown.

A classifier refers to a mathematical function that maps input data to a category. In this section, will be discussed three popular families of classifiers that were employed in this thesis. The classifiers chosen are those most commonly used in the state of the art and collectively represent a range of different approaches.

2.2.3.1 Naïve Bayes

To Naïve Bayes classifiers family belong simple probabilistic classifiers based on applying Bayes' theorem with strong (naïve) independence assumptions between the features. [61].

Let be **V** a space of *d*-dimensional patterns and W = {w_1, w_2,… w_n} a set of *s* disjoint classes consisting of elements of **V**. For each $\mathbf{x} \in \mathbf{V}$ and for every $w_i \in W$, we denote by $\mathbf{p}(\mathbf{x}|w_i)$ the *conditional probability density* of **x** given w_i that is the probability density that the next pattern is **x** under the assumption that its class is w_i. For each $w_i \in W$, we denote by $\mathbf{P}(w_i)$ the a priori probability of w_i or the probability, regardless of observation, that the next pattern to be classified is of class w_i. For each $\mathbf{x} \in \mathbf{V}$ we denote by $\boldsymbol{p}(\mathbf{x})$ the probability density absolute of **x**, or the probability density of the next one patterns to be classified is let be **x**

$$p(\mathbf{x}) = \sum_{i=1}^{s} p(\mathbf{x}|w_i) \cdot P(w_i) \qquad \text{where} \qquad \sum_{i=1}^{s} P(w_i) = 1$$

For each $w_i \in W$ and for each $\mathbf{x} \in \mathbf{V}$ we denote by $\mathbf{P}(w_i|\mathbf{x})$ the posterior probability of w_i given **x**, or the probability that having observed the pattern **x**, the membership class is w_i. For the ***Bayes theorem***:

$$P(w_i|\mathbf{x}) = \frac{p(\mathbf{x}|w_i) \cdot P(w_i)}{p(\mathbf{x})}$$

Given a pattern $\mathbf{x}$ to be classified in one of the s classes $w_1, w_2, \ldots w_s$ of which are known the a priori probabilities $\mathbf{P}(w_1)$, $\mathbf{P}(w_2)$, .. $\mathbf{P}(w_s)$ and the conditional probability densities $\mathbf{p}(\mathbf{x}|w_1)$, $\mathbf{p}(\mathbf{x}|w_2)$ .. $\mathbf{p}(\mathbf{x}|w_s)$, the ***Bayes classification rule*** assigns $\mathbf{x}$ to the class label $b = w_s$ (for some s) for which the posterior probability is maximum:

$$\mathrm{b} = \underset{i=1,2,\ldots s}{\operatorname{argmax}} \{\mathbf{P}(\mathrm{wi}|\mathrm{x})\}$$

Maximizing the posterior probability means maximizing the conditional probability density taking into account the a priori probability of classes.

The Naive Bayes classifier combines the probabilistic model with a decision rule, known as the *maximum a posteriori* or *MAP* decision rule. Naive Bayes is one of the most efficient and effective inductive learning algorithm for machine learning and data mining. Its competitive performance in classification is surprising, because the conditional independence assumption on which it is based, is rarely true in real world applications. Several studies are dedicated to the study of the accuracy of NB Classifier. In [62], the authors have studied and compared the Naive Bayes algorithm to the Hidden Markov Model and the Conditional Random field model. Even if previous studies have shown the Markov model outdoing the Naive Bayes algorithm, the authors, performing two different types of smoothing techniques able to adjust the maximum likelihood of the classifier, obtained significant improvements in the classification accuracy of the Bayes algorithm. In [63], researchers presented various aspects of Naïve Bayes Classifier and smoothing techniques for extraction of useful data. A survey about Naïve Bayes algorithm is discussed in [64]. The authors described the concept hidden behind Naïve Bayes, text classification, traditional Naïve Bayes and machine learning technique. Some applications of Naïve Bayes and its advantages and disadvantages are discussed for a better understanding of the algorithm. In [65] the authors presented a review on application methods of Naive Bayesian Networsks in predicting disease, checking out that Naive Bayesian Networsks as the fundamental algorithm for the best performance in comparison with other algorithms. A classifier approach for detection of heart disease is proposed in [66], which shows how Naive Bayes can be used for classification purpose. In [67], researchers performed three experiments using three real-life activity datasets. To prevent estimation problems, they propose two smoothing Naive Bayes-Based Classifier for adjusting the maximum likelihood to produce more precise probability of a sensor given an activity. In [68], researchers investigate the use of a light-weight ear worn activity recognition device combined with wireless ambient sensors for identifying common activities of daily living. A two-stage Bayesian classifier that uses information from both types of sensors is presented. Detailed experimental validation is provided for datasets collected in a laboratory setting as well as in a home environment.

2.2.3.2 Decision Tree

The Decision Tree learning is one of the most used ML technique in human activity recognition scenario. Decision tree are simple to understand and interpret because they can be represented graphically. They are useful for discovering models within the data because the output of a classification can be monitored to understand the logic of the decision. In a Decision Tree, each internal node corresponds to an attribute and has borders that correspond to each possible value for that attribute. A leaf represents the expected class following the nodes from the root to the leaf. To create a Decision Tree first, the attribute to be at the root needs to be selected. Then a border is added for each possible value for that attribute, this will divide the instances set into subsets, one for each attribute value. The process is repeated recursively for each edge until all instances of a node have the same classification. The crucial step is how to choose the best attribute to be at the root of each substructure so that the depth of the final tree is small and consistent with the data. To choose the best attribute it is necessary to measure the goodness of an attribute and this is where the concept of information acquisition comes into play. The amount of information depends on previous knowledge. The algorithm also requires minimal data preparation or a feature design that can help save time and energy. Therefore, Decision Tree have been widely used in the execution of activity recognition.

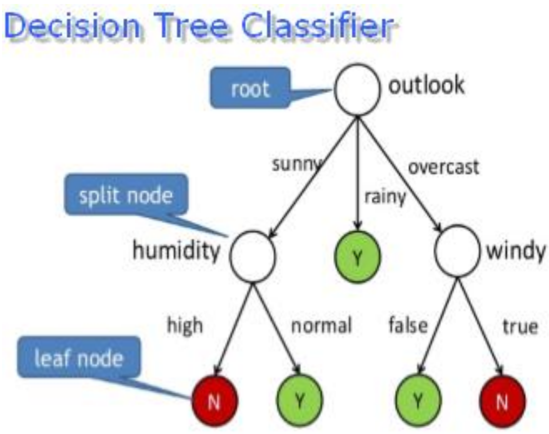


Figure 5: Decision Tree scheme

A huge quantity of works have been produced on the use of DT Classifier in ADL identification. Fan et al. [69], performed activity recognition using Decision Tree algorithms by constructing behavior and position vectors of users performing 5 different activities. The study reported high classification accuracy and less time consumption. In [70], authors propose a novel algorithm, Self-adaptive NBTree, which induces a hybrid of decision tree and Naive Bayes. The Bayes measure, which is used to construct decision tree, can directly handle continuous attributes and automatically find the most appropriate boundaries for discretization and the number of intervals. In [71] a Dt methT method is implemented for the recognition of walking, running, sitting, and standing reporting an accuracy of 98.69%. In [72], five accelerometers have been attached to hip, wrist, arm, ankle of the user and thigh in order to recognize twenty ADL and ambulatory activities. Four different classifier structures were used of which decision tree with C4.5 provided the best accuracy. Similarily, in [73] the effect of multiple sensors on the human body to evaluate the recognition accuracy has been approached. The authors investigated five classifiers (ANN, Decision tree, K-Nearest Neighbour, Naïve Bayes and SVM) to recognize five ambulatory activities using four sensors. The decision tree classifier proved to be the best algorithm. In [74], authors have developed a model that can recognize abnormal activities for assisting people living alone in a smart house environment. The idea is based on the assumption that people tend to follow a specific pattern of activities in their daily life. An open source database is used to train the decision trees classifier algorithm. Training and testing of the algorithm is performed using MATLAB. The results show an accuracy rate of 88.02% in the activity detection task. In [59], authors presents a comparative evaluation of state-of-the-art machine learning techniques to classify ADLs in the smart home domain. Forty-two synthetic datasets and two real-world datasets with multiple inhabitants are used to evaluate and compare the performance of the identified machine learning techniques. The results show significant performance differences between the evaluated techniques. Overall, neural network based techniques have shown superiority over the other tested techniques. In [75], a way to investigate the potential usefulness of an unobtrusive fall detection system, based on the use of passive infrared sensors (PIRs) and pressure mats (PMs) is presented. The system detects falls automatically by recognizing unusual activity sequences in the home environment. A decision tree based heuristic classification model is used to analyze the data and differentiate falls events from normal activities. The sensitivity, specificity and accuracy of the algorithm were around 100%, 66.67% and 90.91%, respectively, across all tested scenarios. In [76], authors presented an empirical study of human movement detection and identification using a set of PIR sensors.

They have performed classification analysis with various machine learning algorithms, including decision tree (C4.5), *k*-nearest neighbor, Naive Bayes and support vector machine. The results show that with the raw data set acquired from a single PIR sensor a precision of more than 92% was obtained in the classification of the direction and speed of movement, the distance interval and the identifying subjects. It is also possible to achieve an accuracy of over 94% by using the reduced function set extracted from two pairs of sensors. In [77], authors proposed a comparative study on activity recognition of elderly people living alone utilizing 6 classic classification algorithms: decision tree, k-nearest neighbor, support vector machine, Naive Bayes, linear discriminant analysis, and ensemble learning. These models were then adopted to recognize 10 activities of daily living. In [78] an activity aware intelligent system that supports user in his/her daily life tasks has been developed. The proposed system aims to integrate three important aspects into a smart house application (environment monitoring, user activity recognition and user-friendly interaction). The information gathered from sensors across the environment is structured as the state of the environment in a compacted form called activity frame. This specific frame is used by a predictor (based on the decision tree method), in order to recognize the activities that have been performed by the user inside his/her domestic environment.

2.2.3.4 K-Nearest Neighbors

Finally, K-Nearby Neighbors (KNN is a supervised learning algorithm, whose purpose is to predict a new instance by knowing the data points that are separated into different classes. Data points or instances are associated with each class, whose set defines the data set. Its operation is based on the *similarity of the functions*. Usually the similarity is calculated through the *Euclidean distance* **d**:

$$\boldsymbol{d}(q,t) = \sum_{i=1}^{n} (q_i - t_i)^2$$

where q and t are two instances. ***q*** is the query instance and ***t*** is an instance from the training set. A shorter distance corresponds to a greater similarity between the data point and the instance to be predicted. In addition to the distance, the algorithm expects to set an arbitrarily chosen ***parameter k***, which identifies the number of nearest data points. The algorithm evaluates the minimum distances k thus obtained. The class that gets the greatest number of these distances is chosen as a forecast.

The KNN is a ***non-parametric*** tool, so it does not make assumptions about the distribution of the data it analyzes. In other words, the structure of the model is determined by the data and is quite useful, because in the "real world", most data do not obey the typical theoretical hypotheses made. Consequently, KNN could and probably should be one of the first choices for a classification study when knowledge of the data is poor or absent.

Furthermore, it can be said that the training phase is quite fast. The lack of generalization means that KNN retains all training data. This means that all (or most) of the training data is needed during the testing phase. Depending on the problem to be solved, we can consider the use of KNN, considering the following three aspects:

- ***Type of problem***: KNN can be used both for predictive classification and regression problems. In classification problems, the KNN determines the expected class label, evaluating a type of distance. Among the most used, besides the Euclidean distance, there is the distance of Manhattan or the distance of Minkowsky:

$$\boldsymbol{d}_{Manhattan} = \sqrt[2]{\sum_{i=1}^{n} |x_i - y_i|} \quad \boldsymbol{d}_{Minkowsky} = \sqrt[p]{\sum_{i=1}^{n} |x_i - y_i|^p}$$

  After calculating the distance, the label of the majority class of the set of selected k instances is returned.

- ***Calculation time***: KNN can require a lot of memory or space to store all the data, but only performs a calculation (or learn) when a forecast is necessary (for this reason it is said that the algorithm is lazy). You can also update and maintain training requests over time to keep forecasts accurate. However, it must be said that the idea of distance or proximity can be divided into very large dimensions (many input variables) that can negatively affect the performance of the algorithm on the problem of interest. This drawback is known as the curse of dimensionality.
- ***Predictive power***: depends on the initially selected parameter k. When k is small, we are limiting the region of a given prediction and forcing our classifier to be "more blind" than the general distribution. Conversely, a large k reduces the impact of variance caused by a random error, but runs the risk of ignoring small details that could be relevant. Some authors suggest setting $\boldsymbol{k} = \sqrt{\boldsymbol{n}}$**,** where n is the number of observations in the training data set.

A fundamental aspect of the k-Nearest Neighbors algorithm is the choice of k. This task is a crucial point for the success of the prediction. Among the most used methods we find the ***Cross Validation*** method. The general idea of this method is to divide the sample of data into a defined k number of randomly extracted subsamples or segments, which will serve as a training phase. The KNN model is then applied to make predictions on the k-th segment and the error is evaluated.

As before, a small review of the most recent works adopting KNN Classifier is presented. In [79], researchers analyzed the comparison of KNN and MKNN algorithms to classify 7395 records. Comparative analysis is based on the accuracy of both algorithms. Before classification, k-Fold Cross Validation has been performed to search for the optimal data modeling resulted in data modeling on cross 2 with accuracy of 93.945%. In [80], studies on perturbing sensitive data using Gaussian noise and creating a safe KNN classifier model that provides safe mining is presented. Authors proposed an efficient approach that aims to provide better protected data mining results with minimal loss of information. In [81], a method able to estimate the indoor position of the occupants of the house by combining fingerprinting techniques with the home occupants activity pattern using K-Nearest Neighbor algorithm with Euclidean Distance is approached. In [82], with the proposed technique, 99.78% of room level classifications are correctly classified using K-nearest Neighbor (KNN). The result confirmations that the integration of KNN with Artificial Neural Network in Back Propagation techniques can give better indoor location based services In [83], the work reviews and extends the field of similarity-based classification (like k-Nearest Neighbors), presenting new analyses, algorithms, data sets, and a comprehensive set of experimental results for a rich collection of classification problems. Experiments on eight real data sets compare eight approaches and their variants to similarity-based learning. Similary, in [84], authors presented an overview of techniques for K-Nearest Neighbour classification focusing on mechanisms for assessing similarity (distance), computational issues in identifying nearest neighbours and mechanisms for reducing the dimension of the data. In [85], authors have studied the accelerometer fan android mobile phone and applied the K-Nearest neighbor algorithm to predict the activity of a single user and have received satisfactory results. In [86], a model called the Clustered KNN which is an improved KNN algorithm to detect four activities performed by four users and achieved appreciable results by utilizing limited memory and a restricted training data is created. In [87], the recognition and the differentiation between fall activities and activities of daily living was performed using the MobiFall dataset. A large database was constructed to train and validate the model. Feature selection methods were implemented to reduce dimensionality. Five different classification algorithms were implemented and evaluated based on their accuracy, sensitivity, and specificity achieved. The k-Nearest Neighbors' algorithm obtained an overall accuracy of 87.5% with a sensitivity of 90.70%, and a specificity of 83.78%. In [88] the authors, in order to improve the accuracy of recognition of activity in smart homes, reached some improvements in the data preprocessing and recognition phase even proposing a new method of sensor segmentation and a KNN algorithm modification. Results have shown the proposed method outperforming other classifiers.

2.2.3.5 Performance Metrics

Once the different type of ML techniques are defined, the final phase of a machine learning project is given by the performance metrics [118]. The correctness, efficiency and usefulness of the design and modeling process can be tested by making inference on a set of new observations called test sets. The test set incorporates information on the activities performed, so that they can be compared with the activities inferred by the predictive model using appropriate metrics. The evaluation metrics for binary classification models are: *Accuracy, Precision, Recall* and *F1 Score*.

The ***Accuracy*** is simply the percentage of instances carefully classified. It is generally the first metric that is observed when evaluating a classifier. However, when the test data is not balanced (most instances are included in one of the classes) or if it is more interested in the performance of a class, the accuracy does not really show the effectiveness of a classifier.

$$Accuracy = \frac{TP + TN}{TP + +TN + FP + FN}$$

For this reason it is useful to calculate additional metrics that collect more specific aspects of the evaluation. Before going into the details of these metrics, it is important to include the ***confusion matrix*** of the evaluation of a binary classification. Class labels in the training set can take only 2 possible values, which we will refer to as positive or negative. The positive and negative instances correctly estimated by a classifier are defined as true positive values (TP) and true negative values (TN). Similarly, instances classified incorrectly are defined as false positive values (FP) and false negative values (FN). The confusion matrix is simply a table that shows the number of instances within each of these four categories, as show in figure.

	Predicted	
	Positive	Negative
Actual True	**TP**	**FN**
Actual False	**FP**	**TN**

The Precision indicates the percentage of beginning and end of activity predicted correctly on all those actually predicted.

$$Precision = \frac{TP}{TP + FP}$$

The Recall indicates the percentage of start and end of activities predicted correctly on all the beginnings and ends of activities actually occurred.

$$Recall = \frac{TP}{TP + FN} = Sensibility$$

There is a clear compromise between precision and recall. For example, in the presence of a relatively balanced data set, a classifier that estimates above all positive instances will have a high appeal but a lower precision, since many of the negative instances will be classified incorrectly resulting in a series of false positives. Finally, the F1 Score metric, harmonic mean between precision and recall, calculated as follows

$$F1\ Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

it is used to obtain more concise information on the quality of the predictive model. it is important to note that a prediction error is considered even in the case in which the beginning and end of the activities are predicted in different time instants but very adjacent to the actual ones.

To validate a specific model it is possible to perform a ***k-fold Cross Validation*** process. This process involves partitioning the dataset in k parts of equal size. One part k is used as a test set, while the remaining k−1 are used as training sets. The process is iterated k times, where at each iteration each partition of the dataset is used exactly once as a test set. In this way it is possible to obtain an average value of the Precision, Recall and F1 values and obtain a better estimate of the predictive quality of the model.

2.3 Activity of Daily Living

As already discussed, one of most efficient way to take care of people health at home is to monitor their **Activities of Daily Living** (**ADL**). In [89] an ADL is defined as :

"Tasks performed by people in a typical day that allow independent living. Basic activities of daily living (BADL) include feeding, dressing, hygiene, and mobility. Instrumental activities of daily living (IADL) include more advanced skills such as managing personal finances, using transportation, telephoning, cooking, performing household chores, doing laundry, and shopping. The ability to perform activities of daily living may be hampered by illness or accident resulting in physical or mental disability. Health care rehabilitation workers play a significant role in teaching people to maintain or relearn these skills so that they can achieve the highest possible degree of independence".

In addition, as considered in [90], [91], and [92], activities performed by a person can be decomposed into several actions. Moreover, actions can be described as a succession of elementary moves (or motions). In healthcare, this term comprises any daily activities concerning self-care, such as feeding, bathing, dressing, grooming, work, home-making, and leisure.

ADLs can be classified into two categories, i.e., *basic* ADLs [90] (or B-ADLs) and *instrumental* ADLs (I-ADLs) [93]. B-ADLs mainly refer to self-care tasks, including bladder and bowel, control and management, feeding, dressing, personal hygiene, transferring, grooming, and sleeping. Obviously, to keep the independence in self-care is an essential point for health and safety. In contrast to B-ADLs, I-ADLs refer to a more complex scenario characterized by activities performed by a resident concerning community settings in a normal day, such as shopping, telephone usage, managing money traveling in community, preparing meals, housekeeping, and taking medications. To accomplish these ADLs, a certain mental and physical ability is necessary. Increasing inability to perform either B-ADLs or I-ADLs may result in the need for care facility placement. Both B-ADLs and I-ADLs can be seen as benchmark indicators of health and safety, and the loss of independence in the ability to personally perform ADLs is a predictor of nursing home admission, using of physician services, using of hospital services, and even mortality. Different studies have shown that the functional dependence significantly predicted later institutionalization. Normal aging changes, accident, acute illness, worsening chronic illness, and hospitalization contribute to the decline in the ability to independently perform tasks of living in the community. Besides, depression and other psychiatric illness may also lead to functional ability. Depending on situations or health status, the caused limitation may be temporary or permanent, and the patients should be promoted to the greatest degree of independence through treatment or rehabilitation. Since the physical functional decline may be the first sign of a changing health status, both the ability and the inability to perform ADLs are usually adopted to assess the person's health status, especially regarding to the older people. In order to provide objective data on the patients health status, functional assessments are usually carried out by the healthcare professionals to detect problems in performing activities of daily living and to plan according healthcare. The assessment of ADLs has been applied to the patients with various of illnesses that can result in limitation of mobility or cognitive, such as dementia, Alzheimer's disease, stroke, and neck fracture and fall-prone. What can be easily found is all of these illnesses are likely to cause increasing appearance of activity disorder or significant decline in activity level. In order to evaluate and quantify the change in ADLs, some assessments have been developed. Katz Index of ADLs is the assessment designed for assessing BADLs. A number of six functions are included in this instrument, i.e., bathing, dressing, toileting, transferring, continence, and feeding. Clients are scored yes or no for independence in each of the six functions, where the high score indicates higher independence. Katz Index is most effectively used for elderly population. Similarly, Barthel ADLs index comprises 10 items, which are similar to such as Katz Index. Specially, this index focus on recording what a patient does, rather than what a patient could do. This assessment can be used to determine a baseline level of functioning and monitor improvement in activities of daily living over time. Because I-ADLs function is usually lost before B-ADLs, the incipient physical or cognitive decline might be detected through I-ADLs assessment. Besides B-ADLs, I-ADLs are also taken into account by some assessment instruments. Bristol Activities of Daily Living Scale is designed to reveal the everyday ability of the people with memory difficulties of one form or another. A total of 20 activities are included, containing both B-ADLs and I-ADLs. And for each activity, there are five statements referring to different levels of ability. Another popular I-ADLs assessment is Lawton Instrumental Activities of Daily Living Scale (LIADLS). Competence in skills such as shopping, cooking, and managing finances are required for independent living in community. To assess more complex ADLs, Lawton et al. identified 8 items in LIADLS are scored regarding to different independence levels. The 8 items are using of telephone, shopping, preparing food, housekeeping, doing laundry, using transportation, handling, medications, and handling finances. Whereas the low score on telephone, self-medicating, and managing finances may indicate the decline in cognitive functions in community dwelling, the low score on housekeeping may more obviously point to the problems in physical function.

In literature, it is possible to distinguish between four main topics based on the ADL studies: the Activity Discovery (AD) [94], [95], the Activity Recognition (AR) [95], the Activity Prediction (AP) [96] and the Detection of Behavioural Deviation (DD) [97]. In this research work, the focus is on the AD and the AR.

2.3.1 Activity Discovery

Once ADL is modelled and the modelled is generated by learning, we use the term of **Activity Discovery** (**AD**) to indicate it as *"an unsupervised learning algorithm to discover activities in raw sensor event sequence data"* [98]. In literature, a great variety of methods using different inputs and outputs can be found. A brief review of the major methods is going to be presented in this section. What firstly it is important to say referring to the aforementioned methods is that they can be grouped by considering the semantic level of the sensors used.

The semantic level of the generated models are also highlighted to improve the comprehension of the pros and cons linked to each method.

In [99], authors model human behaviour using expert knowledge and vital sign sensors. In addition, authors skip the ADL discovery process by choosing to use ontological models instead of data learning. Therefore, the human behaviour models of this method have a very high level of semantics since they correspond to very specific situations. With such models, it is possible to directly detect dangerous situation and quickly react in case of emergency. However, those models are fully constructed using expert knowledge, and therefore subject to human mistake. In [95], the AD, called *the training stage*, initially requires a time series dataset of measured attributes from individuals performing each activity. The time series are split into time windows to apply feature extraction and thereby filtering relevant information in the raw signals. Later, learning methods are used to generate an activity recognition model from the dataset of extracted features. Likewise, data are collected during a time window, which is used to extract features. Such feature set is evaluated in the trained learning model, generating a predicted activity label. A generic data acquisition is also represented by an identified architecture for AD and AR systems. However, the need to split the data recorded during the learning period to "*individuals performing of each activity*" leads to record labels of the performed activity. In [100], authors propose a method to model, starting from a log of binary sensor events (rising and falling edges), the habits of the inhabitant. These models are extracted by sequence mining techniques and modelled by extended finite automata (EFA). The learned habits are then labelled by an expert. The expert work is thus fastidious if treating data from a big smart home. As an output, authors gives a global map of activities represented by an EFA. In [90], several machine learning using binary sensors and individuals performing of each activity as inputs have been applied. Naîve Bayes classifier, Gaussian Mixture model, hidden Markov model, decision tree, support vector machine conditional random field are considered. The computed models have semantically high information since they are trained directly with adapted and labelled data. In [101], the monitored person is asked to indicate which activity he is performing. Of course, the efficiency of this approach is confronted with the ability and the willingness of the person to declare his activity: in general, numerous reported activities errors are introduced in the database. In other works [102], the goal is the one to enrich a database by studying sensor logs or by using cameras exclusively during the learning phase. This approach is expensive, intrusive and therefore risks changing the behaviour of the patient during the learning phase. In both cases, the labelling step is difficult and unreliable. That is the reason why, in the methods proposed in this thesis, the knowledge of actually performed activities during the learning phase is not required.

2.3.2 Activity Recognition

Activity Recognition methods are model-based approaches to monitor people. The used models can be given by an expert or obtained by learning (i.e. by applying an AD method). In [90], **"***The field of* **Activity Recognition** *(***AR***) is concerned with the question of how to label activities from a sensor-based perception of the environment. The problem of AR is to map a sequence of sensor outputs onto a value from a set of predefined activity labels"*

As for AD, a great variety of methods using different inputs and outputs can be found in the literature. In this subsection, existing methods will be classified according to the type of model used to model the activities.

In [99], authors adopt the ontology-based context model to recognize the performed activity or a dangerous situation. Each context entity has attributes to describe some basic properties of the entity. Context entities are part of parent entity such as characteristics, diseases, preference, social and health ontologies are part of person ontology. Each context entity has attributes to describe some basic properties of the entity. In [91], authors make a great review on vision techniques applied to human behaviour analysis for AAL. According to the research, it can be seen that at the motion, pose and gaze estimation level, several methods achieve robust and high success rates. In [103], researchers are able to track the activity of hand washing to assist older adults with dementia. Multiple orders in the process can be correct, but not all of them. Their system is able to prompt the user if a necessary step is missing or the order of the implied actions is unacceptable. Vision is used as the only sensor in the developed system for two purposes: (1) tracking of hand location; and (2) tracking of step-specific object locations. Related to this type of activity recognition, Authors in [104] stand out in activity recognition based on object use. These authors define activities as combinations of actions and objects and intend to recognize and track objects use in order to infer human activities. Object models are acquired automatically from video, whereas object identification is based on RFID labels. At the learning phase, the user wears a RFID bracelet which reads the RFID tags attached to the surrounding objects in a home environment. Assuming that the object being moved is always the object in use an that only one object is being moved at a time, the system learns the relationship between the segmented image and the active RFID tag using a dynamic Bayesian network. As arms and hands move with the objects, skin filtering is applied beforehand. At the test phase, the system works without the RFID data as objects are recognized by detecting SIFT features within the segmented area. These key points are matched based on maximum likelihood to the previously trained SIFT points. In [105], activity recognition is approached differently. The individual silhouette is obtained at different positions of a living room. Grouped into 10–20 prototypes, each silhouette stores its center, width and height and is manually labelled with a location. A fuzzy inference method is used to estimate the most likely physical location of test silhouettes. Location estimation and previously assigned coordinates enable average speed measurement, which is used besides location in order to recognize human indoor activities. A Hierarchical Action Decision Tree (HADT) is used to classify human actions using multiple levels. At the first level, human actions are classified based on location and speed. With K-means, clustering feature patterns are obtained; and activities of daily living, like walking or visiting the bathroom, are recognized. All those presented methods are efficient in their field of appliances. Nevertheless, according to [91], at higher levels, especially at behaviour, there is still a long way to go to achieve off-the-shelf products. Still, huge advances have been made in the last ten years. But the challenge to design and develop stable and general systems still persists, as most systems only solve specific problems in very particular environments. In [106] the authors describe all inhabitant activities by only one HMM. Then, authors recognize activities by applying the well-known Viterbi algorithm. Unfortunately, the complexity of the model drastically increases with the number of activities and sensors. Furthermore, the used model has not intermediary semantic levels between activities and sensors and the precision of the recognition is not guaranteed. In [107], after converting video information to binary events traducing the human posture, the authors present a system that recognizes a set of activity modelled by HMMs. Moreover, they classify activities by a probability that allows recognizing the activity as being the one, which is represented by the most probable model. However, the previous methods can only compare models linked to the same sensor and event sets. On the contrary, in practice, activities are linked to different sensors because they are performed in different home areas and are realized by using diverse equipment in different spaces. In the majority of the presented methods, probabilistic models are used to estimate which activity is the most likely performed. For all the AR existing works, the recognition method is strongly linked with the models used to represent the activity. Therefore, if a new kind of models is used during the activity discovery, a new method, ideally based on existing ones, should be developed.

CHAPTER 3

SIMULATION ERA

Conducting measurements inside smart homes is currently one of the hottest topics in the area of the health smart environments and computer science, due to the aging of society. This condition requires long-term care to maximize the life quality of the older adult with an increasing cost and strain of healthcare resources. Intelligent Environments (IEs) such as Smart Homes (SHs) facilitate long-term monitoring of activities using sensor technology. In the same time, to be able to provide high-level intelligent solutions, algorithms to identify ADLs are necessary. An essential role is played by sensor data, a basic element for testing, classifying and connecting sensor patterns to inhabitant activities [108]. However, only few and limited data sets are currently available because of high sensor cost, availability and deployment time. Therefore, the use of simulated environments may solve these issues and facilitate the generation of such datasets. In this chapter a review of the latest developments in such scenario is presented and a discussion of motivations of the choice to develop a simulation tool as a fundamental element of the measurement chain is reported.

3.1 Simulation Tool: a key element of the measurement chain

Be able to measure physical quantities related to user life style is one of the most performant way to provide care assistance to elder people improving their quality of life and reducing health costs. The miniaturization of the sensors, their low cost and the development of algorithm able to convert quantitative information in qualitative information make the challenge interesting and feasible. In this research work, a simplified measurement chain composed by PIR sensors and ML algorithms and derived from the real scenario proposed in eWare project is analyzed and presented. With the term measurement chain, we generally refer to the set of stages of a measuring instrument which process the information detected by the physical quantity in the study, and then present a result. The measurement itself could be represented as a quantitative or a qualitative information, [109]. Different stages composing the measurement chain can be distinguished:

- Sensors network implementation, related to the measurement that needs to be performed;
- Raw data acquisition;
- Data processing;
- Uncertainty analysis;
- Application.

Inside a measurement chain the key role is played by two elements: 1) the physical quantities involved (sensors, environments, etc.); 2) the huge number of possible configuration and combination of these elements together with algorithms used to provide high level information. In smart homes and aging society scenario, the definition of the optimal configuration of the measurement chain would does not exist. It would require an enormous number of tests not feasible in terms of cost and time in the real life. In order to avoid waste, the focus is on the chance to have a simulation of the measurement chain and of its responses. The simulation tools play a key role inside this scenario. "*The basic idea is to integrate the simulation of a sensorised apartment with human behaviour modelling based on constraint-based planning that produces a sequence of daily activities*", [108]. Once the simulated environments is defined in terms of type of sensors, number, configuration, the second step is given by the chance to simulate the human behavior inside the simulated environment, which will allow to generate the desired datasets and moreover, different kinds of algorithms, generally based on ML techniques, need to be implemented to extract features and quality information. Finally, the quality information obtained thanks to the simulation have to be used to correct the parameters involved in the real environment. A general idea of the measurement chain in a global scenario is reported below.

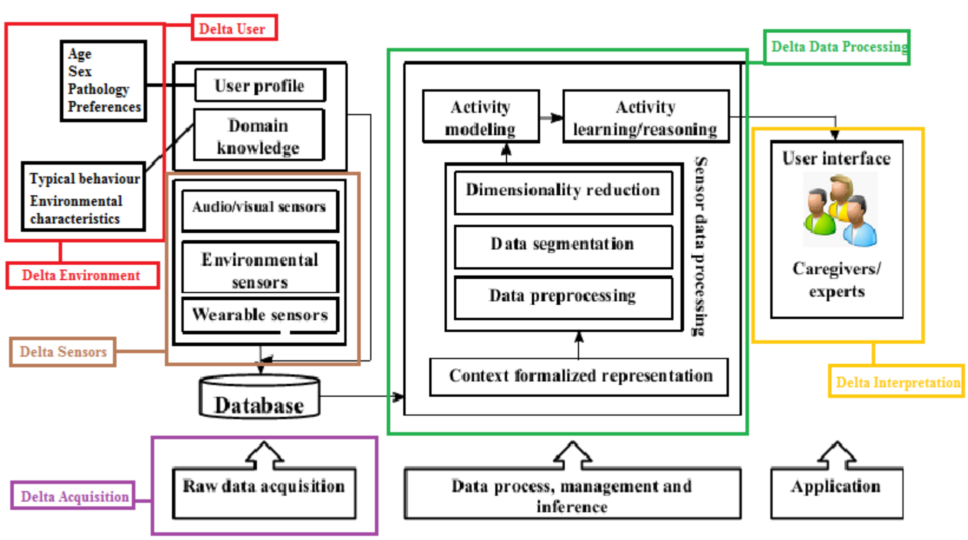


Figure 6: Measurement chain schematic view

Where the quantity delta, represents the uncertainty related to all quantities involved. It is possible to observe that result of the performed measurement will depend on a huge number of factors, such as:

1. The subject considered (sex, age, pathology, habits, culture, etc.);
2. The characteristics of the environment (number of rooms, rooms dimension, furniture, walk-on surface, etc.);
3. The sensors involved (typology, characteristics, disposition in the environment, number, etc.);
4. The acquisition phase of the signal;
5. The artificial intelligence used to analyze data;
6. The data interpretation.

Translating the previous diagram in the ad hoc case, the following structure can be obtained:

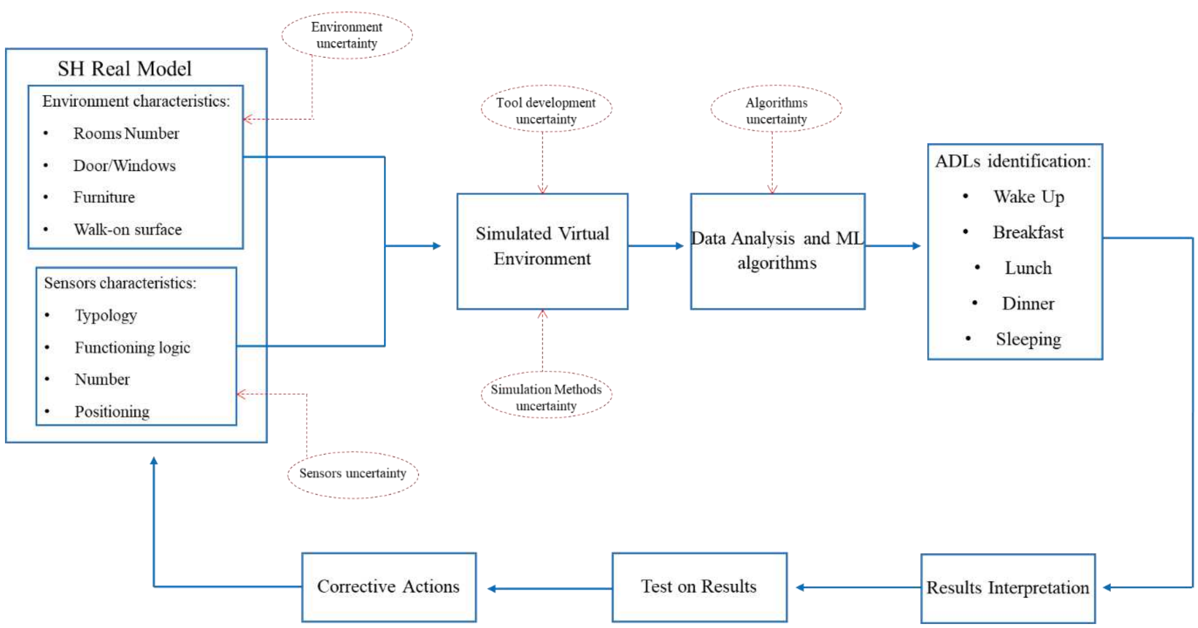


Figure 7: Measurement chain block diagram

In Paragraph 3.3 the Figure 8 is going to be analyzed into details.

Each block affects the result of the measurement according to how it is employed inside the measurement chain. Every measurement is subject to some uncertainty. A measurement result is only complete if it is accompanied by a statement of the uncertainty. This value can come from the measuring instrument, from the environment, from the operator and/or from other sources that could directly or indirectly affects the measure. At the same time it can be augmented or reduced by a different configuration of the measurement chain considered. The first step to investigate in which measure the uncertainty of each block of the chain can affects the final output is approached in this research work in paragraph 3.4.

Inside a Health Smart Home scenario, the problem related to how a certain configuration of the measurement chain could affect the measured quantity and its uncertainty it results in the impossibility of repeating the same test in a huge number of configuration without generating an enormous growth in costs and development time.

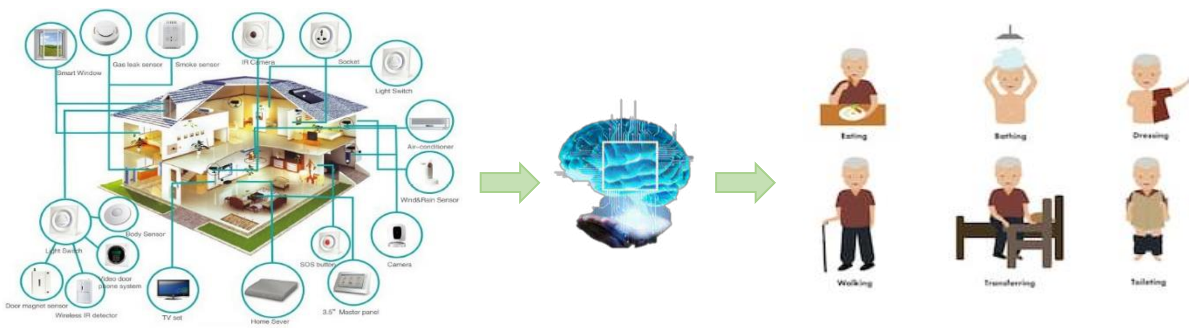


Figure 8: High cost solution

Physical IEs are expensive to implement. This is due to sensor technology costs and to the construction of the physical environments [110, 111]. About environment construction, two aspects need to be considered: first, researchers may not be aware of the ideal configuration of sensors to reach a good level of performances without pre-testing several combinations. This testing phase would require considerable time and expense and could be not feasible in real-world implementations where significant alterations may create confusion, particularly when pathological subjects are considered, [112]. Secondly, the environment construction it is a time consuming process which requires considerable groundwork before commitment to the purchase of equipment for the environment construction [112]. These environments therefore lack of flexibility [111, 113] and could have limited scalability [111]. As a consequence of these costs and constraints, not all researchers have access to datasets produced by such environments. "*Data collection from IEs is a time consuming process [110, 111] because of the nature of the monitored scenarios, which may require the collection of data for extended periods of time in order to capture events that demonstrate typical inhabitant behavior and changes in behavior over time*", [108, 110, 111, 114]. Optimal testing of new approaches generally involves data collection from different scenarios under different circumstances. This could not be possible in a real smart environment due to the difficulty in recruiting suitable participants to test the different configurations [110, 111]. Additionally, there are regulatory limitations that must be adhered to during testing on human subjects [115]. Researchers are therefore looking for alternative methods of datasets generation. As already said, one of the most popular solution is offered from generation of synthetic sensor datasets through the use of simulated IEs, which allow to accelerate research in related areas [115]. These simulated environments offer the chance to facilitate the generation of a variety of sensor datasets, even larger than those carried out from physical IEs [115]. This allows researchers to quickly test or/and evaluate new algorithms and cost effectively [110,115] providing in the same time an increased level of control over data and environment itself. "*The physical layout of environments including walls, doors and objects within the environment can be modified to test a range of use case scenarios*", [119]. The arrangement of sensors including type, number and position can be adjusted as often as required with no cost and little time and effort [116]. Researchers would have complete control over the environment and generated datasets [110], experiments could be re-run many times with small adjustments to the environment, testing algorithms under development [115]. The experiments can be restarted quickly and easily with minimal set up time [117]. The simulations allow control over environments which is not feasible in the real life, such as the manipulation of time though with easily generate months, years of datasets, [117]. Simulations may represent conceptual or already existent environments, indicating the impact of some adjustments to the environment and highlighting optimizations in sensor placement with no invasiveness or expense [112]. In this context, the simulation tool plays a key role to simulate how the measurement chain react under certain conditions and to know in advance which is the best physical configuration to perform the measurements in a real environment. The purpose to find a relationship between environments, sensors and ADLS can so be reached.

Figure 9: Simulation Tool smart solution

3.2 Related works

Several Smart Home simulators have been already developed for different purposes. Before going through the description of the implemented tool inside this research work, it is necessary to analyze the two key approaches categories to simulation. Synnott et al. (2015) distinguish two approaches between the *model-based* that facilitate the generation of data based on activity models and the *interactive* one that incorporate the use of virtual environments (VEs) and virtual sensors which respond to user interaction. A third approach, the hybrid one combining both the previous has risen recently.

3.2.1 Model based Approaches

"*Model-based simulators, such as those described by Bouchard et al. (2010) or Helal et al. (2012) use pre-defined models of the agent behaviour, describing different possible events and activities with their probability of occurrence, their order and their duration*", [114]. They enable the generation of data for extended periods, in real-time or not. These approaches are generally completely scripted so to be able to simulate several days (or more) of data, the user needs to script each day independently. This is a major drawback for the use of these simulators in e-health applications as several weeks of data are usually necessary in order to detect long-term patterns. In Bouchard et al. [118] an example of such an approach used within the SIMACT SH simulator is provided. Inside this tool the users could define the order of events, the time taken for each event and the objects involved in the event. Additionally, users could define actions associated with the completion of each step. *Mendez-Vazquez et al. [119] proved the use of Markov chains describing the order of events, combined with Poisson distribution to calculate a range of realistic activity times and probability distributions to calculate a range of sensor values to generate a simulated activity dataset, [119].* This simulated activity set contain a distribution of activities such as sleeping, walking, reading and sitting together with metrics including time and energy expenditure. Another example of a model-based approach is given by Helal et al. [120] in the PerSim simulator developed to facilitate the synthesis of data for the testing of activity recognition research. The simulator allow users to define each activity by specifying the sensors involved, the maximum and minimum typical sensor values, the order of sensor activations and activity duration. In Kormányos and Pataki [121] a simulator capable of modelling the activity of a single inhabitant within an IE is developed. The approach facilitates the modelling of individual behavior profiles, such as typical sleep amount, and the change in current state such as thirst and tiredness. Any case, generally, the quality and accuracy of the resulting datasets coming from tool model-based, relies heavily on the quality of the activity description model and associated parameters, while the construction of accurate activity models would require access to real test data describing the performance of the modelled activities.

3.2.2 Interactive Approaches

"*Interactive approaches, such as proposed by Synnott et al. (2014) or Ariani et al. (2013) assume a human controlling an avatar while it performs activities in the simulated environment"*. These approaches allow to perform more realistic and coherent behaviour given the presence of a real human involved. *''It consists of software which provides a platform for interaction with individual virtual sensors, or the use of interactive VEs combined with embedded avatars that have the potential to provide an intuitive and interactive environment simulation experience. Avatars are interactive objects that can move within VEs and passively or actively interact with the sensors contained within them, representing the behavior of real inhabitants within physical IEs*", [119]. Such models may be based on real environments or sensor specifications, or be based on conceptual environments and technology which is yet to exist. Activities can be performed in a natural manner by interaction with a virtual sensor or movement of the avatar within the VE and interaction with objects contained within the VE, [119]. This helps to create ad-hoc testing [122] recording specific activity scenarios, such as falls, interruption during activity performance, or to define the impact of a changes in object or sensor placement on the simulation result. For instance, a PIR sensor located in a certain location of the environment, the far corner of a hallway may only detect inhabitants when entering or leaving the kitchen or living room and may not detect inhabitant movement between rooms further down the hallway, [119].

3.2.3 Interactive Approaches for Context Aware Applications

Several studies have been conducted on the use of interactive VEs for the testing of context aware approaches. Even if these approaches are not used for the generation of simulated sensor datasets, they are considered as they are employed for the prototyping of solutions for use within IEs. For these studies the focus is on the illustration of the response of objects in the virtual environment based on context aware criteria more than on the output of synthetic data. Lertlakkhanakul et al. [122] described the use of a 3D virtual environment able to support interactions by multiple users simultaneously for collaborative exploration of the environment. Fu et al. [123] demonstrated an avatar-based approach for the testing of context aware applications. They provided details of a simulator which represented a virtual environment using a 2D floor plan layout, able to visualize the current state of physical sensors inside the IE, or visualizing the current state of virtual sensors thanks to the use of text boxes located next to sensor icons inside VE. The simulation of the movement inside the environment was realized by using the mouse to drag an avatar throughout the VE, generating position data. Sensors within the VE responded according to a set of context rules defined by the user in XML; however, the authors do not provide many details regarding the support for the creation of VEs and for generation of virtual sensor data. The YAMAMOTO toolkit [124] the approach is used to simulate an assistive environment by placing a virtual proximity sensor capable of responding to the location of a user controlled avatar in the virtual environment. The proximity sensor's detection range is specified by a sphere radius and this virtual sensor is capable to generate an event identical to the one of a physical real sensor. Armac and Retkowitz [125] describe the eHomeSimulator, which allow to graphically represent environments using a 2D overhead floor plan view. These environments are created in a grid format using SketchUp [126], which is then imported inside the simulator to define accessible or inaccessible areas and to add devices and avatars. Multiple user controlled avatars can be placed in the virtual environment and can be moved individually. This simulator is a good solution for testing of complex scenarios for example the case in which another person moves into the same room after the user. Other related approaches include 3DSim [127], developed to test smart devices, CASS [128], for testing home automation rules, and TATUS [129], which facilitates avatar interaction through XML commands. These studies have strongly supported advances inside IE research by facilitating rapid, low-cost testing of context aware approaches for environment automation [119].

3.2.4 Interactive Approaches for Simulated Dataset Generation

Several studies have been conducted on the use of interactive approaches to generate realistic simulated IE sensor datasets. For instance in [130] Buchmayr et al. presents a simulator for the generation and visualization of sensor data. The simulator creates the virtual environment using a 2D floor plan and facilitates the user interactions with virtual sensors through the click of the mouse on the floor plan generating sensor data output in a log file. The simulator also supports the use of different simple sensors, as, binary, contact and temperature sensors. The addition of sensors to the floor plan is supported through dragging and dropping within the 2D floor plan; however the creation of new sensor types requires development of data models, parsers and filters for each sensor not achievable from non-technical users. Several VE-based studies relating to the synthesis of IE sensor data have also used the avatar-based approach, [119]. In [131] SH Simulator is presented as a tool using a 3D virtual environment approach to the simulation of user movements in the VE with the aim to facilitate the identification of the optimal sensor configuration before doing an investment in physical sensors or real environment alterations. This simulator facilitates movement of a virtual use throughout a VE using keyboard and mouse, with the consequent generation of motion and pressure sensor data. Movements can be recorded and reused to test an alternative sensor deployments. In this approach, the creation of environments and objects is realized through the use of aseparate 3D modelling software. A similar tool was introduced by McGlinn et al. [132]. VEs can be created using a game map editing tool and users are able to configure sensors by specifying accuracy, fire rate, delay and location using the SimConfig tool. Virtual user movements into the environment are responsible of the generation of simulated data; once an avatar's position fell within a sensor's detection range, the sensor output is recorded. Krzyska [133] developed a SH simulation tool able to facilitate the creation of VEs using a 2D plan layout based on the color-coded line approach. Sensors and avatars are represented as colored dots inside the environment. The tool facilitates the placement of motion sensors with adjustable sensing radius. An avatar can be moved within the environment using click of the mouse; this generates sensor events in a log file when the avatar with his movement falls inside the sensor's detection radius. Motion sensor position and sensing radius could be configured using a Form-based approach, however the tool does not provide user interface support for the creation of additional sensor types and sensor event logging adjustment requires knowledge of the Log4J Java logging library. Ariani et al. [134] developed a simulator which helps the creation of a floor plan and the specification of a resident profile for movement speed and height. Users can specify a start time, and the end time is automatically calculated. The simulation of events is able to produce PIR and pressure mat sensor data. In Synnott et al. [135] a tool allowing the creation of 2D overhead plans of VEs is presented. The user is allowed to customize existing sensor types and to create personal ones. The tool also facilitates the recording of ADLs performances through passive and active avatar interaction with sensors placed within the environment.

The majority of these tools are any case mostly focus on the simulation itself without englobing in their structure a real time analysis on simulated datasets that would allow to evaluate the accuracy and performances of the simulation performed, the efficiency of the sensors configuration considered and the reliability of the algorithms implemented and used to interpret results.

3.3 UNIVPM HomeCare Simulator: the metrological problem

As described in the previous chapter, the VEs are a fundamental elements for SEs datasets generation. The simulator developed in this research work aims to overcome the limits present in some of the already existent tools and most of all, to offer a powerful instrument to simulate measurement on ADLs. One of the goal we want to reach is to directly evaluate how the measurement chain reacts under certain conditions to know in advance which is the best physical configuration of sensors to perform the measurements in the real environment. Before going into details through the UNIVPM HomeCare Web App it is necessary to carry out a detailed analysis on how the tool created fits and contributes to the measurement chain considered. In order to develop a suitable simulator, we started from the analysis of the parameters that characterized an ADL event. Four main aspects are considered: a) where the ADL event takes place (the SH environment); b) how it is measures (the sensors used for monitoring); c) how the ADL is performed by the user (the movements that represent the activity) and d) the artificial intelligence used to obtain high quality information.

Environment	***PIR sensor***	***Activity***	***Artificial Intelligence***
Room geometry [R(Xr,Yr)]	Position [P(Xs,Ys)]	Trajectory coordinates (Xt,,Yt)	ML algorithm used and associated formula
Rooms number (Nr)	FoV (d)	Time interval (t0, t1)	
Doors number (Nd)	Range (r)		
Obstacles (No)	Sampling frequency (Fc)		
Walking Area [A(m)]	Sensors number (Ns)		

Table 1: Problem parametrization

To precede with a parametrization of the problem it is necessary to define a function englobing all these aspects well described even in the previous block diagram. Coming back to it, it is mandatory to consider that each of the blocks is characterized by an uncertainty given by the sum of the uncertainties relating to the elements constituting the chain and that can more or less vary the result of the measurement itself.

The starting point is given by the physical environment within the test is performed. By considering a 2D model, it can be characterized and parameterized by some principal quantities as shown in the previous table. As a consequence an environmental function as the sum of the parameters involved can be defined as:

$$F_{Env}\,(x,y) = f(R(Xr,Yr),Nr,\ Nd,\ No,\ A) + \ \Delta_E(x,y)$$

Where R(Xr,Yr) takes into account of the room geometry, the quantities Nr, Nd, No refer to the number of rooms, doors and obstacles, A is the area of the walking surface expressed in meters. Given the fact that each of these parameters can affect the measure in a different way, we consider an additional quantities, the uncertainty environments, defined as Δ_E representing the variability of the aforementioned parameters. The formula describing the problem becomes:

$$F(x,y,t) = F_{Env}(x,y) + \ \Delta_E(x,y) \qquad (1)$$

However, to consider the parametrization completed, it is necessary to add the role played by the all sensors involved in the measurement. By considering PIRs sensors, each of them will have an own accuracy depending by sensor type and characteristics and the accuracy of the measure performed will depend on how such sensors are employed in the environment. The PIRs function is defined as:

$$F_{Sensors}(x,y) = n\ x\ [F_S(d,r,Fc) + \Delta_{C1}\,] + \Delta_{C2} \qquad \textit{with } n = 1,2,..$$

Where F_S is the function describing the characteristics of sensors: the filed of view, the range, the sampling frequency; Δ_{C1} is the uncertainty related to sensor characteristics while Δ_{C2} is the quantity describing the uncertainty deriving from parameter P(Xs,Ys) (see Table 1) related to the sensors configuration and disposition in the environment.

According to these considerations the (1) becomes:

$$F(x,y,t) = f(R(Xr,Yr),Nr, Nd, No, A) + \Delta_E(x,y) + n\ x\ [F_S(d,r,Fc) + \Delta_{C1}] + \Delta_{C2} = F_{Env}(x)+F_{Sensors}(x) \qquad (2)$$

Once the physical aspects (environment + sensors) are defined, we pass to analyze ADLs and MLs. The first one will mainly depend on how the activity is performed from the user. We define a function F_{ADL} as:

$$F_{ADL}(Xt,Yt,t) = ADL(Xt,Yt,t) + \Delta_{User}$$

Where ADL is the activity performed in space and time while Δ_{User} is the uncertainty related to how the ADL is performed by the user.

One of the most interesting aspects is now to define how each parameter affect the measurement. To this purpose, given the multiplicity of the variables involved, it would be impossible to think of varying each of them simultaneously in order to evaluate their consequent effect on the measure and estimate their relative uncertainty. Here the need to set constant values arises. The choice is in turn twofold: first, keeping the configuration of the sensors and the type of test performed constant, evaluating how the results vary depending on the environment considered or, secondly, given a defined environment, vary the configuration of the sensors to the same level of test performed. Given the fact one of the biggest doubts for researchers, is that of choosing the optimal configuration of sensors that allows to minimize the measurement uncertainty, in this thesis work the second of the two options considered is addressed. Test performed and environment are kept constant, configuration and number of sensors used is variable. As a consequence, the focus is on how the quantities Δ_{C1}, Δ_{C2} affect the measure.

The fundamental role is now played by the simulation tool, which allows virtually to perform tests and then predict consequently the value of Δ_{C1}, Δ_{C2}. To the purpose the contribution given by the simulator, is added to the measurement chain described up to now by the $F_{Analysis}(x)$ function which will take into account of the artificial intelligence used in the tool itself. Describing it through a function we obtain:

$$F_{Analysis} = \sum_{k=1}^{n}(F_k + \Delta_k)$$

Where the quantity $F_k(x)$ is the sum of the function of each algorithm involved in the process and $\Delta_k(x)$ is the relative algorithms uncertainty. The measurement chain will be so described by the final function:

$$F(x,y,t) = f(R(Xr,Yr),Nr,\ Nd,\ No,\ A) + \Delta_E(x,y) + n\ x\ [F_S(d,r,Fc) + \Delta_{C1}\ ] + \Delta_{C2} + ADL(Xt,Yt,t) + \Delta_{User} + \sum_{k=1}^{n}(F_k + \Delta_k\ )$$

By simplifying:

$$F(x,y,t) = F_{Env}(x,y) + F_{Sensors}\ (x,y) + F_{ADL}(Xt,Yt,t)\ + F_{Analysis}$$

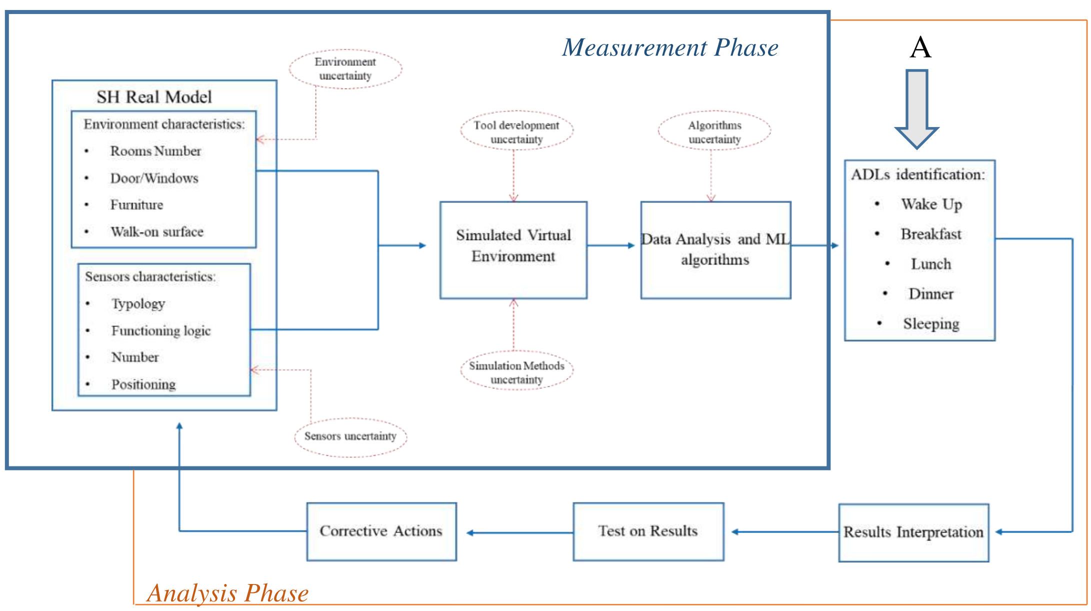


Figure 10: Measurement chain block diagram and phase evidence

The defined function is able to describe the first three main blocks of the diagram and to carry the output A described by ADLs. Once the measurements phase is completed and the behavioral trend of the user is registered and classified, the second part of the work is related to an analysis phase.

The output obtained provides an estimate of the accuracy ADLs recognition according to the chosen configuration of the sensors during the measurement. The following questions therefore arise:

- Which is the impact of a different sensors configuration on the accuracy of the measurement performed;
- To what extent, the number of sensors can influence the recognition accuracy of a given activity;
- Which would be the optimal configuration of the sensors to obtain the maximum of the accuracy;
- Which is the minimal configuration to guarantee a high level of reliability while saving on costs and time.

The chance to know the answer of these questions would have a huge impact on the feasibility of the measure. Knowing in advance the consequence of increasing or decreasing the sensors number or changings their disposition thanks to the virtual reality will allow to report this information in the real case giving the chance to realize ad hoc measure and environment.

The tool developed in this thesis work is aimed to respond to the previous questions and to this purpose.

3.4 Inside UNIVPM HomeCare Web Application

The UNIVPM Home-Care simulator is a user-friendly Web Application, based on interactive approach method, which allows the user to simulate, analyze, process and monitor data deriving from virtual PIR sensors and to generate datasets relating to ADLs. The tool is composed by three different stages:

1) The simulation stage which allow the user to simulate the physical environment, the sensors network and the virtual user behavior;
2) The analysis stage composed by statistical and ML algorithms necessary to derive qualitative information from generated datasets;

3) Monitoring stage: the tool is able to process in real time information coming from the virtual sensors network so as well information deriving from real sensors applying the logic used in the simulation to carry out significant information.

In the next section, the simulator will be presented into details.

3.4.1 A Web Application for simulator

As already mentioned, the simulator is implemented as a WEB application accessible from any device: smartphone, tablet, pc, etc. It presents as an Internet-enabled application and a browser it is sufficient to have access on it, no application needs to be downloaded on the device.

It is important to point out that a web application is not a simple optimized website but consists of a dynamic responsive site, capable of offering complex functionalities similar to those provided by the common apps installed on the device. This is possible thanks to scripting languages (both client and server side) and technologies such as AJAX, Jquery, HTML5, CSS3, Javascript that allow the creation of real applications, usable through a browser, similar to common native applications in terms of user application.
Why a web application for simulator:

- Portability: web app are multi-platform applications that allow their use on multiple devices using common technologies that can be easily adopted by developers. This generally means reduced development time and maintenance costs;
- Distribution: a definite advantage results in the deployment, as a user can access a web app, simply by entering a web address in the browser without the application being submitted to the store approval process. This determines further flexibility, in fact, a web application can be upgraded, upgraded and expanded independently by the administrator and any evolution can be used by all end users without the need for user-side updates, allowing everyone to have the same version.
- Space: a web app consists of a link to a remote application and has the advantage of not minimally affecting the storage capacity of the device and of being substantially independent of its calculation capabilities. The speed of a web app depends on the

quality of the connection to the network and on the performance of the remote server in offering the processing requested by the user.

Leak points:

- Reduced access to the device's native functionalities: it is not possible to completely access all the functions belonging to the device such as sending push notifications, the use of the compass and the accelerometer.
- User Experience: not all browsers render in the same way, so it is possible to have different experiences depending on the devices and browsers used. Moreover, it is not possible to use all the standard views as toolbars, buttons or tabs, but only those belonging to the interfaces used in the classic web pages.
- Performance: web apps still fail to support intensive CPU applications or applications containing very complex graphics such as 3D games.

The simulator is presented in the form of web applications containing web pages, digital documents through which information is made available to the end user. The page can be divided into two parts: one related to the contents and a part of text formatting (layout), graphic presentation or organization of the contents (text, digital images, etc.). The content of the pages usually consists of a document (generally HTML) and of files related to it that a web browser downloads from one or more web servers interpreting the source code to generate the visualization of the desired page.

There are two types of web pages that respectively identify two different Web programming paradigms:

- Static web page relating to the paradigm of the so-called static Web;
- Dynamic web page relating to the paradigm of the so-called dynamic Web.

In the first case, the information is usually contained in hypertexts in HTML or XHTML format which allow access to other web pages or other information content through hypertext links (links or references). In the second case, that is the one on which the implemented simulator is based, is made of scripting languages.

3.4.2 Simulator tool: Design of simulation area

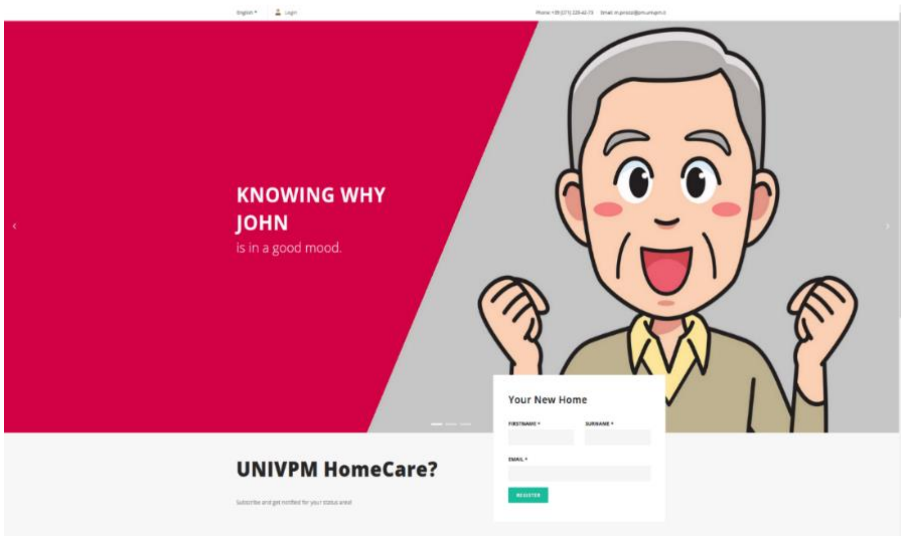


Figure 11: Univpm HomeCare Web App Main Page

The developed simulator is based on different areas. The first concerns user registration and it is the first page the user has access to, after typing the address of the Web Application on a browser. The user is asked to register his personal data and to create a username and password through which he can then access the service. Entering in the heart of the simulator, the first step to be performed before the use is the insertion of the characteristics related to the environment and the sensors network that will be used in the simulation. Through the special area "Add Sphere" the user can define the name of the simulation, import the house plan of the environment considered by defining the number of rooms and selecting the type of PIR to be used during the test. Inside the system are inserted a variety of types of PIR sensors different for characteristics, radius, angle. Each of them is identified by a unique ID that can be selected by the user. In this regard, it is necessary to anticipate that the tool is structured on two levels such as frontend and bakend (here after explained). The first is intended for the user and represented by the interface with which the user can interact, the second is related to the programmer. The user, in the frontend, can choose between different types of sensors precisely because in the backend dedicated sensor management areas are created. This is a powerful mean for expanding the tool if the chance of virtually allocating different types of sensors in it is considered (not only PIR, but also pressure, temperature, etc.) .

Figure 12: Add Sphere Web App area

After this first phase, the user proceeds with the definition of the characteristics of the environment within which the simulation is carried out. At this point in the "Tool" section, it is possible to distinguish between three phases: environmental characteristics definition, sensors positioning and trajectory simulation.

Rooms definition

Once the layout of the environment has been imported, the fundamental role is played by an extension of the HTML, the Canvas, which allows the dynamic rendering of bitmap images that can be managed through a scripting language. The user is able to interact with the loaded image and then define the area of each room, outlining its boundaries and associating it with a name. The manipulation of the size, position and rotation of these objects through the use of the mouse is made possible by the Farbricjs function which allows to change some of the attributes of such objects as their color, transparency and depth position on the page web. The main advantage of this method lies in the fact that the dimensions of the outlined rooms are calculated directly thanks to the possibility of importing a plan allocated in a grid instead of using a colors-based plan as in many simulation tools. Each point of the grid has its own x, y coordinates from which the coordinates defining the perimeter of each area are derived.

Sensor network configuration

Once the room definition is completed, the user can proceed with the positioning of the sensors in the environment. A basic question is about the level of abstraction on which the simulated sensors are handled. In general, inside simulation tool, could be possible to identify two types of sensors: those associated with particular objects which change values when a user is interacting with it (e.g. pressure sensors on a chair) and sensors associated with a particular space which change value when the user is moving through the monitored space (e.g. motion sensors). In this thesis only PIRs sensors are considered so only the second case is treated. First, the room within which the sensor will be inserted is selected so that the system prevents the sensor from being placed in a different area.

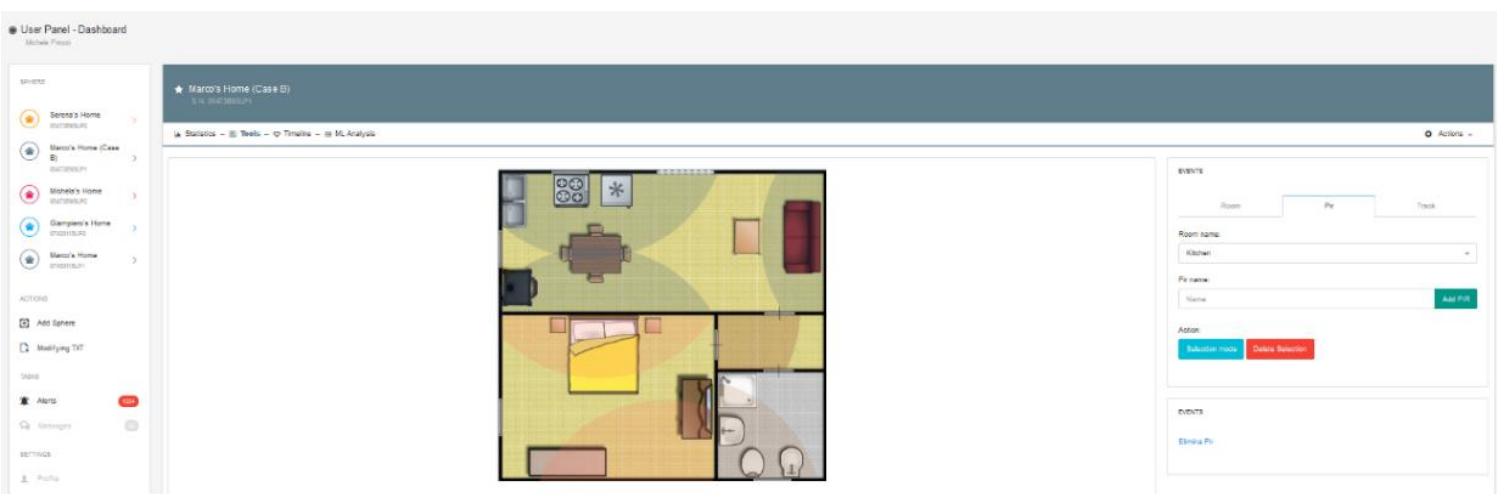


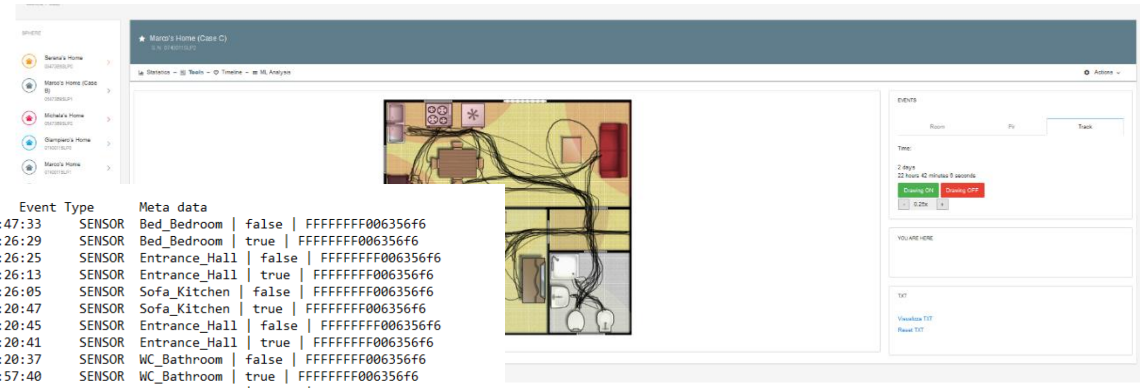
Figure 13: Tool Web App area

Sensor agents can be placed anywhere on the map. When adding a new sensor in the simulation, the real user simply needs to place it on the map and connect it to a room to be usable. The sensor radius is plotted on the map in order to facilitate the choice of the positioning through a visual analysis of the actual coverage area. Thanks to the class fabric.Rect the radius of action of each PIRs is limited inside the room it is inserted in. Since the layout is based on a coordinate grid, the positioning coordinates of the sensors are easily obtainable.

The simulated sensor reproduces how the real sensor works, as detailed as possible, which means that it synthesizes the measurement process. In the case of PIR sensors, they produce an output when the user movements falls inside the sensors detection range, passing from 0 to 1: a file .txt is produced to take into account about all the activation registered during the simulation and it will be described in the next paragraph.

Trajectory simulation

Activities take place at particular locations in the apartment, movement between those locations is part of the execution. Once sensors and environment are set, the simulation can start. The behaviour of the inhabitant agent is determined by the execution of an activity plan. Such a plan consists of a sequence of activities that are characterized by which inhabitant takes part in the activity, where the activity shall happen and which sensors are used during that activity. While executing its activities, the virtual user is in the corresponding room to perform daily and he is monitored by the rooms sensors. While performing such tasks, the user moves from one room to another realizing trajectory that can fall within the detection range of the PIR or not. If it happens, the status of the sensor from off automatically switches to on and remains there for as long as the user is intercepted by it. A geolocation system is implemented in the tool offering the chance to define the user position in the environment in real time, depending on which of the PIRs sensors is in active mode. The trajectory is simulated through the mouseMove or touchMove function (for touch devices) in javascript combined with an asynchronous call ajax for the contextual saving of the x and y coordinates in the database when the PIR is turned on and off. Through the mathematical calculation of the PIR ray with the function php sqrt it is possible to save the information only if the trace enters into the sensor radius. In order to speed up the simulation process the concept of time manipulation is introduced through the clock-speedup js library, which allows to increase



```
StartTime        Event Type    Meta data
2019-01-29 22:47:33     SENSOR  Bed_Bedroom | false | FFFFFFFF006356f6
2019-01-29 22:26:29     SENSOR  Bed_Bedroom | true | FFFFFFFF006356f6
2019-01-29 22:26:25     SENSOR  Entrance_Hall | false | FFFFFFFF006356f6
2019-01-29 22:26:13     SENSOR  Entrance_Hall | true | FFFFFFFF006356f6
2019-01-29 22:26:05     SENSOR  Sofa_Kitchen | false | FFFFFFFF006356f6
2019-01-29 21:20:47     SENSOR  Sofa_Kitchen | true | FFFFFFFF006356f6
2019-01-29 21:20:45     SENSOR  Entrance_Hall | false | FFFFFFFF006356f6
2019-01-29 21:20:41     SENSOR  Entrance_Hall | true | FFFFFFFF006356f6
2019-01-29 21:20:37     SENSOR  WC_Bathroom | false | FFFFFFFF006356f6
2019-01-29 20:57:40     SENSOR  WC_Bathroom | true | FFFFFFFF006356f6
2019-01-29 20:57:40     SENSOR  Entrance_Hall | false | FFFFFFFF006356f6
2019-01-29 20:57:28     SENSOR  Entrance_Hall | true | FFFFFFFF006356f6
2019-01-29 20:57:12     SENSOR  DressinZone2_Bedroom | false | FFFFFFFF006356f6
2019-01-29 20:47:18     SENSOR  DressinZone2_Bedroom | true | FFFFFFFF006356f6
2019-01-29 20:47:08     SENSOR  Entrance_Hall | false | FFFFFFFF006356f6
2019-01-29 20:47:02     SENSOR  Entrance_Hall | true | FFFFFFFF006356f6
2019-01-29 20:46:52     SENSOR  Fridge1_Kitchen | false | FFFFFFFF006356f6
2019-01-29 20:41:04     SENSOR  Fridge1_Kitchen | true | FFFFFFFF006356f6
2019-01-29 19:57:17     SENSOR  Fridge2_Kitchen | false | FFFFFFFF006356f6
2019-01-29 19:31:34     SENSOR  Fridge2_Kitchen | true | FFFFFFFF006356f6
2019-01-29 19:27:16     SENSOR  Fridge1_Kitchen | false | FFFFFFFF006356f6
2019-01-29 19:07:16     SENSOR  Fridge1_Kitchen | true | FFFFFFFF006356f6
2019-01-29 19:07:12     SENSOR  Fridge2_Kitchen | false | FFFFFFFF006356f6
2019-01-29 19:06:56     SENSOR  Fridge2_Kitchen | true | FFFFFFFF006356f6
2019-01-29 19:06:40     SENSOR  Entrance_Hall | false | FFFFFFFF006356f6
2019-01-29 19:06:28     SENSOR  Entrance_Hall | true | FFFFFFFF006356f6
2019-01-29 19:06:20     SENSOR  WC_Bathroom | false | FFFFFFFF006356f6
2019-01-29 18:50:30     SENSOR  WC_Bathroom | true | FFFFFFFF006356f6
2019-01-29 18:50:22     SENSOR  Entrance_Hall | false | FFFFFFFF006356f6
2019-01-29 18:50:10     SENSOR  Entrance_Hall | true | FFFFFFFF006356f6
2019-01-29 18:49:51     SENSOR  Sofa_Kitchen | false | FFFFFFFF006356f6
2019-01-29 18:10:25     SENSOR  Sofa_Kitchen | true | FFFFFFFF006356f6
```

Figure 14: Simulation trajectory and dataset extraction

Simulated activities and automatic datasets generation

Even if the details of the simulation are described in the next chapter, here we are going to provide some general information on the basic idea underlying the simulation protocol and the methodology to generate a big amount of datasets in a fast way. One week on daily activities is manually performed according to a definite procedure as in the figure.

Several PIRs activations are generally registered in the early morning, during lunch time and dinner and 5 main task are chosen to be detected:

- Wake up
- Breakfast
- Lunch
- Dinner
- Sleeping

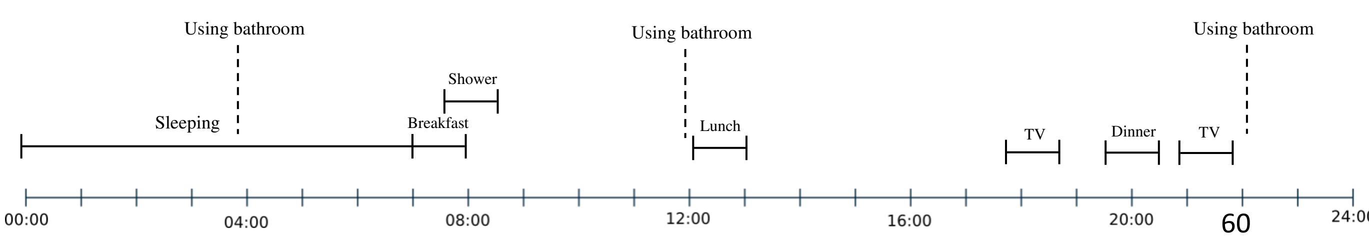


Figure 15: Daily timeline

The first week of data is manually simulated by the user through the tool. To make it faster in order to reach a big amount of data in a short time, an automatic simulation system has been implemented in the simulator. In the section “Modifying TXT” the user is allowed to import the txt file and thanks to a calendar view is possible to copy the simulation of one selected day in another one with an automatic variation of time and sensors status. In this way several months of simulation can be generated automatically just with a mouse click.

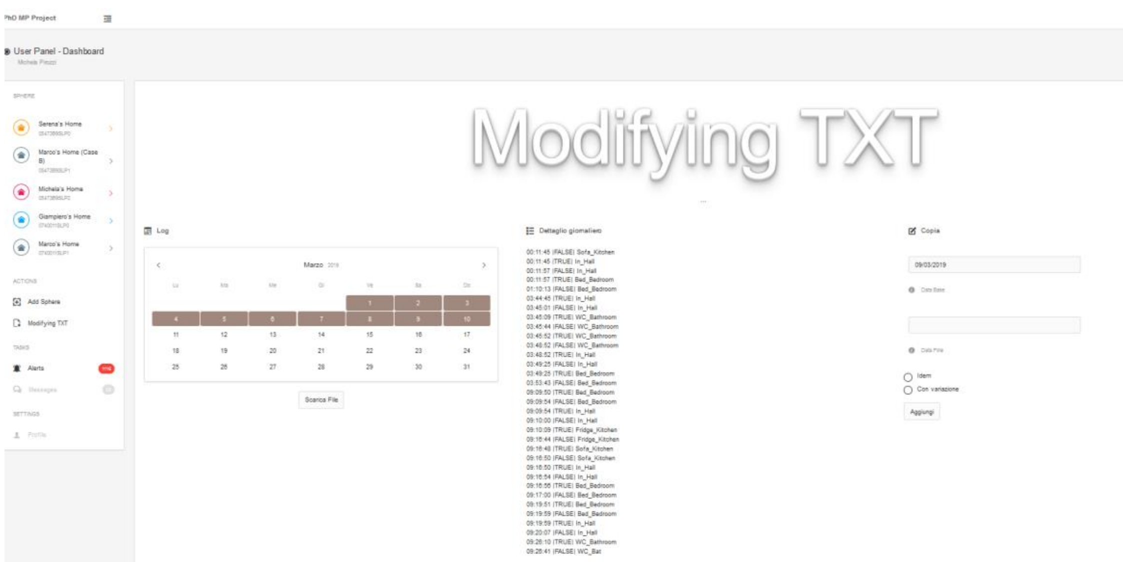


Figure 16: Automatic data generation area

Once the datasets is generated, the validity of the test performed is investigated through the Analysis Section of the simulator implemented.

3.4.3 Simulator tool: Design of analysis area

To determine the plausibility of the generated datasets, the validation of results is carried out thanks statistical and MLs algorithms.

Statistics Area

The statistic area is the main page of the simulator. It appears as a dedicated section in which different statistical algorithms are implemented in order to conduct qualitative and quantitative analysis on the simulated data. The analysis is carried out in real time, every 30 seconds a page refresh takes place looking for the presence of new data to be processed. Moreover, being the system predisposed to receive in the DB data coming from real PIRs sensors, the tool offers the chance to carry out a real-time estimate of measurements performed with real sensors. Inside the statistic section five main significant graphics are reported. The first one plots the PIRs sensors activation per day (x asses) and in time (y asses).

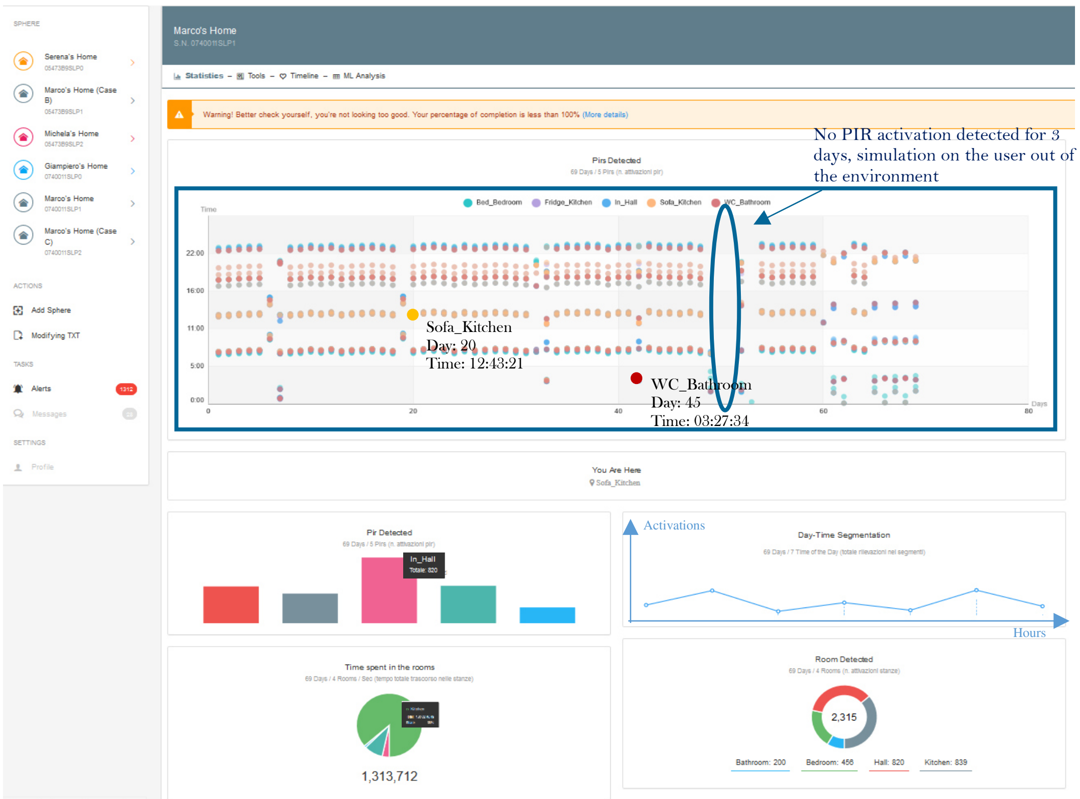


Figure 17: Statistic Web App area

It allows to quickly and visually discover the general use of the environment by the user and to establish the trend of the user behaviour.

For instance, in the figure a general PIRs activation between:

- 7 am and 8 am
- 12 pm and 1.30 pm
- 5 pm and 10 pm

Can be observed. This information on the occupancy of the environment can be used to smartly managed sensors inside a SH. It is even possible to select one of the sensors involved in the measure to analyze its behaviour singularly.

The histogram (first row, left) shows the number of activations for each PIR located in the environment. This information helps to:

- ✓ Have an idea of the most active PIRs of the environment that are necessary related to those part of the environment most used by the user;
- ✓ Easily and directly observe if a habit change occurs.

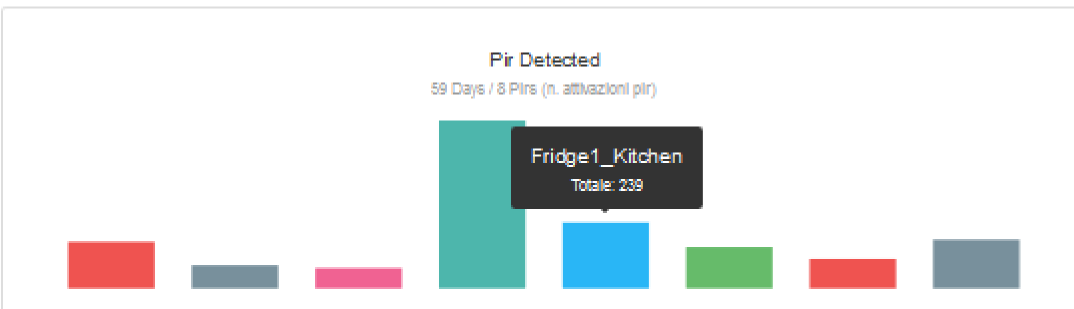


Figure 18: Histogram view of PIRs activations

The broken line (first row, right) shows the total number of PIRs activation for each range of time. The graph can be considered as a time line divided in 8 different slots: 23:00-5:59, 6:00-9:59, 10:00-11:59, 12:00-13:59, 14:00-16:59, 17:00-18:59, 19:00-20:59, 21:00-22:59. It helps to:

- ✓ Have an idea of the most active period of the day for the user: for a worker (8:00-17:00 working time) we aspect to have a broken line with points of maximum in the ranges 6:00-9:59, 19:00-20:59, 21:00-22:59 and almost a flat trend in 10:00-11:59, 14:00-16:59.
- ✓ Control if there is an overload of activation during the night, helpful to give indication on the level of stress of the patient and to monitor the sleeping phase.

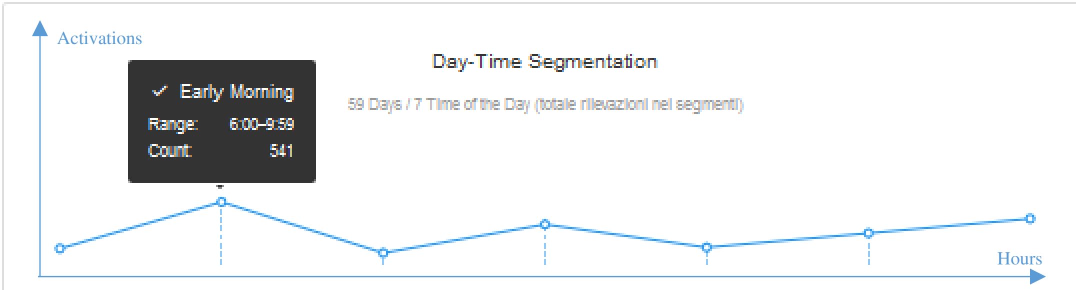


Figure 19: Cake diagram vision

The cake diagram (second row, left) shows the total time spent in each room with the relative percentage to define user environmental preferences.

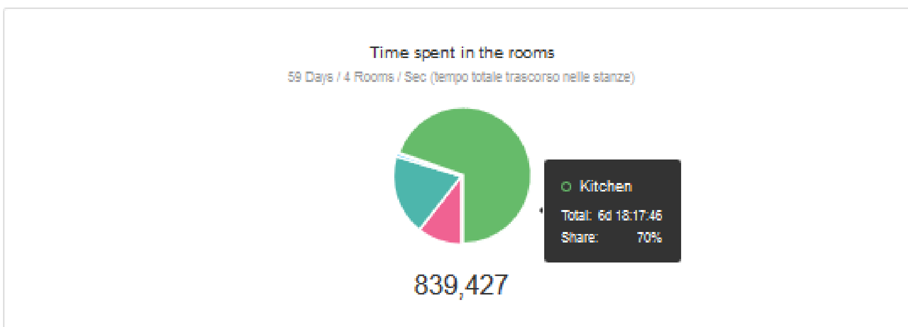


Figure 20: Cake diagram vision

The circular diagram (second row, right) shows the number of transitions between one room and another. It allows to check which is the favorite path of the user and to control that the selected path is the most convenient and logic one. For instance, supposing to consider a pathological user suffering of wandering. Suppose there are two different ways to access a room where a task must be performed, the kitchen. Suppose it can be reached directly from the entrance or through a corridor that would require a longer path. If the number of transitions between the kitchen and the corridor is high in a certain period, the caregiver could associate this information to the wandering phenomenon considering the patient is not recognizing the most logic path.

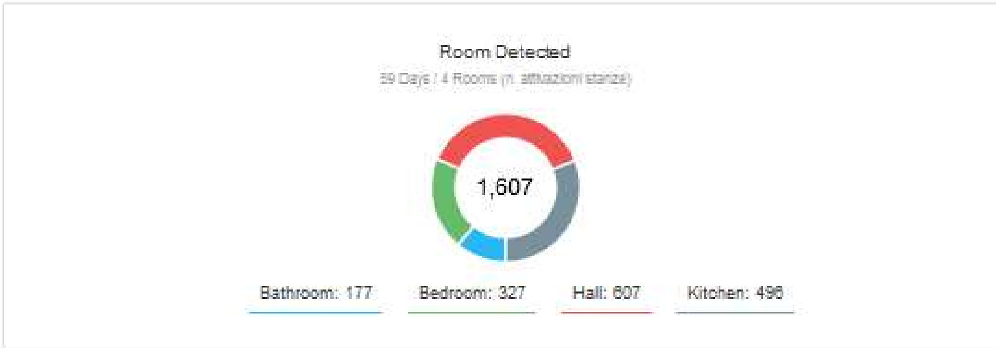


Figure 21: Circular diagram vision

Machine Learning Area

In order to provide a complete service able to carry out analyzes on the quantitative data that provide qualitative information, machine learning algorithms have been implemented in the simulator to extract ADLs from the simulated datasets in real time. As explained in 2.x machine learning algorithms are increasingly used, those on the basis of the acquired experience, allow to interpret intelligently and automatically future data and predict possible behaviors. The main objective of machine learning is that a machine is able to generalize from its experience, in other words, it is able to perform inductive reasoning. To evaluate the abilities of a machine learning algorithm, it is necessary to estimate a quantitative measure of its P performance. Often the measure of P is specific to a certain task T which the system must perform. For tasks such as classification, P is evaluated by measuring the accuracy of the model, the percentage of examples for which the model processes a correct output. It is important to study how the algorithm is able to evaluate new incoming data. Performance measurements are performed using a data set called a test set that is usually different from the set of features used to train the system (training set) so that more precise and independent evaluations can be made on the effectiveness of the training. Among the three main types, supervised, unsupervised and reinforcement learning, those implemented in this research work belong to the first class. DT, KNN and Naïve-Bayes Classifier have been studied and implemented to conduct analysis. As already seen in the dedicated literature, the reason behind this choice lays in the vastness of the use of such algorithms in this scenario due to their high level of accuracy and in the desire to define how those algorithms react under certain conditions, observing their differences in terms of level of uncertainty reachable. In the case studies considered in this thesis work, the main objective of the ML algorithms is to recognize the execution of five main tasks belonging to the ADLs class with a certain accuracy. In this regard, the data from the simulated sensors have been processed and fed to the ML algorithms in order to obtain a numerical result. Each classifier has been trained with a week of simulated data configured according to the following datasets structure:

Date	Start Time	End Time	Duration	Activity	Room
2019-01-29	22:47:33	00:48:34	2:01:01	"Gone to bed"	Bedroom
2019-01-29	07:35:10	07:37:19	00:02:10	Wake Up	Bedroom
2019-01-29	07:38:02	07:45:19	00:07:17	Breakfast	Kitchen
2019-01-29	12:15:11	13:28:42	1:13:31	Lunch	Kitchen
2019-01-29	19:34:17	20:48:37	1:14:30	Dinner	Kitchen

Activity	Time Slot	Duration Task
Wake-Up	[6:30-7:30]	[1-5] minutes
Breakfast	[7:00-8:30]	[10-30] minutes
Lunch	[12:00-13:30]	[40-120] minutes
Dinner	[19:00-21:00]	[40-120] minutes
Sleeping	[22:00-6:30]	[5-8] hours

```
//  ML statistics

    require_once '../library/vendor/autoload.php';

    $classifier = new Phpml\Classification\KNearestNeighbors(1);
    $accuracy = new \Phpml\Metric\Accuracy;


$samples = array(
                    array(0, 'Bedroom'), array(1, 'Bedroom'), array(2, 'Bedroom'), array(3, 'Bedroom'), array(4, 'Bedroom'), array(5, 'Bedroom'), array
                    array(6, 'Bedroom'),array(7, 'Bedroom'),array(8, 'Bedroom'),array(7, 'Kitchen'),array(8, 'Kitchen'),array(9, 'Kitchen'),array(10, '
                    array(14, 'Kitchen'),array(15, 'Kitchen'), array(19, 'Kitchen'), array(20, 'Kitchen'), array(21, 'Kitchen'), array(22, 'Kitchen'),
              );

$labels = array(    'Sleep', 'Sleep', 'Sleep', 'Sleep', 'Sleep', 'Sleep', 'Sleep', 'Sleep', 'WakingUp', 'WakingUp', 'WakingUp',
                    'Breakfast','Breakfast','Breakfast','Breakfast','Lunch','Lunch','Lunch','Lunch','Dinner','Dinner','Dinner','Dinner'
              );


$classifier->train($samples, $labels);

$giorno = 1;
$ar= array();

foreach ($sensor->getML($sn) as $key => $value) {

    $ml = $classifier->predict(array($value['t'], $value['r']));

    $ar[$value['d']][$ml] = $ml;


}
```

```
// Dinner

$actualLabels = array();
$predictedLabels =  array();

foreach ($ar as $key => $value) {

    if(empty($value['Dinner']) || !isset($value['Dinner'])) $value['Dinner'] = '';

    $actualLabels[] = 'Dinner';
    $predictedLabels[] = $value['Dinner'];

};

$dinner = $accuracy->score($actualLabels, $predictedLabels, true);
```

Figure 22: Examples of ML datasets and algorithms

All subsequent simulations are therefore interpreted according to the training performed. The choice to use three different types of ML lies in the desire to investigate which of the most used algorithms is the most performing in the scenario of human activity recognition. The main advantage of the ML analysis in real time is given by the chance of performing a simulation according to a certain environmental and sensor configuration and, depending on the result, testing different configurations in order to optimize the measurement chain, achieving the best compromise between identification accuracy and implementation costs.

3.4.4 Alerts Management System

Given the fact the simulator is able to work in real time, even considering data coming from real PIRs, a management alarm system has been implemented in the tool. The UNIVPM HomeCare Web App is able to manage a-typical event or to send an alert if a fundamental task has not been carried out or if a strange situation verify: the user forgot to eat, the user didn't sleep in the bedroom or he is not at home in a range of time it should supposed to be to. The simulator is able to send message or email to the caregiver and inform him about the current situation in real time. A timeline is implemented an reported in order to offer the chance to dispose of a history of a-typical events, to check their frequency, periodicity, etc. The actual limit of this section is given by the necessity to set in the programming phase the threshold to be considered for the alert. For now, the simulator is set in order to be able to send the alert when an a-typical event occurs for three consecutive times. To be efficient, there should be a section dedicated in which the caregiver defines which tasks are important according to the user habits and the logic to use to send an alarm.

3.4.5 Some technical programming aspects

LAMP and Zend framework

Depending on the technology used, the web app pages can contain graphic or multimedia elements and interactive and dynamic elements. The web page of the simulator developed is characterized by a graphic panel admin, accessible from any device, and has been developed on a LAMP server, namely Linux, Apache, MySQL and PHP. These four elements together deal with the server infrastructure that allow to generate and host dynamic web pages. The individual components are based on each other and therefore the software package is also known as LAMP Stack. The Linux operating system is the base on which the Apache server is run, but it is not able to interpret the dynamic contents, so the PHP script interpreter is responsible for this process. Therefore, the system sends the relative source code to the PHP interpreter, who has access to the MySQL database with all the information on the behavior of the website visitors. The result is then sent back to Apache and shown correctly on the user's browser.

LAMP servers have always been convenient and quickly available. It can also be defined as a set of open source libraries with the advantage to be available for free and with an accessibility of the programming code open to all. This helped to modify the code at any time and independently, as well as develop new extensions.
It is worthwhile to install a LAMP server not only because of the high flexibility and low costs, but also because the components used combine each other and allows the operation of static and dynamic web pages. Thanks to the use of the LAMP server it was possible, later, to install the Zend framework (open source framework for web application development) which favored the implementation of external classes and instances. The Zend is an object-oriented framework and thanks to the wide compatibility of its structure it can boast of being the ideal platform for the development of web applications. The collection of professional PHP packages, integrated within the framework, represents a great simplification. Zend Framework benefits from the Composer package manager, the PHPUnit Unit testing environment, a tool for continuous integration which offer the chance to automatically create and publish GitHub projects. The great advantage of Zend Framework is that all components have been programmed in such a way that they are independent of each other. Precisely this structure makes it possible to select only the modules necessary for the modular creation of a development platform desired. The flexibility of the Zend Framework is not recognizable only because of its architecture: all the components can be used without problems even on other PHP frameworks. The components of Zend Framework are based on the PHP language and also the open source software can be used without having a paid license. Even the hosting and use of a PHP web application are simple, either by working independently or by relying

on an external provider, for the use of which the PHP language is an essential element. The Zend Framework is realized with the MVC architectural pattern, which helps to improve the user experience and to make modules with different functionalities structurally independent, favoring the quality of the software. This pattern is based on the division of tasks performed by the various components of the application:

- Model: provides methods to access useful application data and also provides the Controller with a unique representation of the data requested by the user;
- View: is the data presentation level. Provides the interaction interface with the application and allows to create requests and to visualize results;
- Controller: directs the interaction flows between View and Model. Specifically, it intercepts the client's HTTP requests and translates every single request into a specific operation for the Model; later it can perform the operation itself or delegate the task to another member. Its main function is the one to call and coordinate resources to perform the action requested by the user.

From the point of view of the structure, the developed tool is a hybrid application made up of some software components that will be illustrated below.

Programming Languages

Five main programming languages have been used to develop the simulator: HTML, PHP, MySQL, CSS and JavaScript. Specifically, the use of HTML allowed both to structure and to visualize the web page and to perform purely functional tasks. The use of CSS has allowed to manage the formatting and style of the application, and to exploit the potential of the media queries. Finally, fundamental is the use of the JavaScript language, which allowed interaction with the native APIs, providing a valid tool for developing the User Interface and application logic. Regarding the PHP (Hypertext Preprocessor), it is a Scripting language, which is mainly used to create dynamic web pages and develop powerful server-side web applications, unlike JavaScript which is a client-side scripting language. The use of the PHP language allowed to quickly write complex pages generated dynamically. Among the most widespread Database Management Software (DBMS), one of the most important is mySQL. A DBMS is a software service, generally created as a server in continuous execution, which manages one or more databases. The programs that must therefore interact with a database can not do it directly but must dialogue with the DBMS which is the only one to physically access the information. This implies that the DBMS is the component that deals with all access, management, security and database optimization policies. The existing DBMSs are not all of the same type. Nowadays, for example, there is much talk about DBMS NoSQL, born to meet the needs of the most recent Web services. Yet a very large line of DBMS is that of the so-called RDBMS (Relational DBMS). Just the MySQL used in this project is an open source RDBMS that represents one of the most popular and widespread technologies in the IT. The LAMP platform incorporates MySQL for the implementation of servers to manage dynamic websites.

Frontend – Backend

Inside the developed tool it is necessary to distinguish between the frontend and the backend. The first is referred to the part of the application visible to the user, the one he can interact with and accessible through the web app URL. The second one is related to the programmer and allows the programming phase and the directory management. The directory structure is designed to be extensible towards more complex projects, providing simple subsets of folders and files for the final project. Inside the project it is possible to define use case for each directory:

- Application/: This directory contains the application itself. It hosts the system's MVC, as well as the configurations, the services used, and the bootstrap file;
- Forms /: The directory reserved for data / entry scripts.
- Layout /: This layouts directory is the base of the MVC.
- Modules /: The modules allow the developer to group a set of related controllers into a single unit;
- Controllers /, models / and views /: These directories serve as default controller, model and graphic viewer;
- Scripts /, helpers /: These directories contain action helpers;
- Bootstrap.php: This file is the entry point for applications. The purpose of the bootstrap file is to initialize the components and make them available for the application;
- Data /: This volatile directory provides a place to store application data;
- Docs /: This directory contains the documentation generated by the author.

- Library /: This directory is reserved for internal / external libraries on which the application depends. The library / folder is characterized by the presence of external libraries, in addition to the main Zend library (internal), which can be imported manually or installed via composer. For each project, composer creates a vendor folder inside the main project folder. Once the necessary library are defined, composer provides automatically to download and places them in the vendor folder.
- Public /: This directory contains all the public files for the application. index.php sets and invokes Zend_Application, which in turn calls bootstrap.php, which is then sent to the front controller. The root folder of the web server is set to this directory.
- Temp /: The temp / directory is reserved for application transient data. If the data under the temp / directory has been deleted, the application is able to continue to function with a possible decrease in performance.
- Test /: this directory contains application tests. These could be the PHPUnit tests.

Database connection

The connection to database is realized through PHP using a username, a password and the database name values at the beginning of the script code:

```
$username="your_username";
$password="your_password";
$database="your_database";
```

Then, it proceeds by replacing your_username, your_password and your_database with the username, password and MySQL database that will be used by the script. This allows you to create three variables in PHP which will then store the different details of the MySQL connection. The mysql_connect PHP function allows connection to the DB:

```
$ mysqli = new mysqli ("localhost", $ username, $ password, $ database)
```

with this line, PHP connects to the MySQL database server on localhost with the username and password provided. After establishing the connection, it is necessary to select the database to be used through the command:

$ mysqli→ select_db ($ database) or die ("Unable to select database");

through which PHP uses the MySQL connection and with it selects the database stored in the $ database variable (in our case it will select the database "your_database"). To closure of the connection with the database server it is realized through the PHP command $ mysqli→ close().

Saving occurs using asynchronous functions ajax on the client side, with data exchange in the background between web browser and server, allowing the dynamic updating of the web page without reloading the user part. The data is sent in order to not be visible to the user, through an HTTP request that the browser sends to the server (POST method). The data arrives at the controller which processes them and formats them for saving on the DB through connections and insert / update queries in the model directory. The application performs constant backups on the MySQL DB in the following tables:

- Alert: contains n alert detection lines;
- Pirs: contains the detail of the pirs with the x / y positions on the 2D map;
- Products: contains the specifications of the PIRs connected to the serial produc;
- Rooms: contains the detail of the rooms with the x / y positions on the 2D plan;
- Sensor: it contains the surveys on the tracks for the generation of the final txt;
- SerialNumber: contains the link between sphere and products, populated with production;
- Session: contains the detail of the FrontEnd side accesses;
- Sphere: contains the environments managed by the user;
- User: contains user access data.

An example of DB for PIRs activation and activation is given above. The first column refers to ID PIRs, the second column indicates the date, the location of the PIR in the environment is reported in the fourth column, the status (on or off, true or false) in the fifth.

05473B9SLP2	2019-01-01 06:51:46	SENSOR	Pir_Bed1_Bedroom	true	FFFFFFFF006356f6
05473B9SLP2	2019-01-01 06:54:22	SENSOR	Pir_Bed1_Bedroom	false	FFFFFFFF006356f6
05473B9SLP2	2019-01-01 06:54:23	SENSOR	Pir_in_Hall	true	FFFFFFFF006356f6
05473B9SLP2	2019-01-01 06:54:27	SENSOR	Pir_in_Hall	false	FFFFFFFF006356f6
05473B9SLP2	2019-01-01 06:54:27	SENSOR	Pir_Wc_Bathroom	true	FFFFFFFF006356f6
05473B9SLP2	2019-01-01 07:15:00	SENSOR	Pir_Wc_Bathroom	false	FFFFFFFF006356f6
05473B9SLP2	2019-01-01 07:15:00	SENSOR	Pir_in_Hall	true	FFFFFFFF006356f6
05473B9SLP2	2019-01-01 07:15:02	SENSOR	Pir_in_Hall	false	FFFFFFFF006356f6
05473B9SLP2	2019-01-01 07:15:06	SENSOR	Pir_Fridge1_Kitchen	true	FFFFFFFF006356f6
05473B9SLP2	2019-01-01 07:28:25	SENSOR	Pir_Fridge1_Kitchen	false	FFFFFFFF006356f6
05473B9SLP2	2019-01-01 07:28:28	SENSOR	Pir_in_Hall	true	FFFFFFFF006356f6
05473B9SLP2	2019-01-01 07:28:32	SENSOR	Pir_in_Hall	false	FFFFFFFF006356f6
05473B9SLP2	2019-01-01 12:14:55	SENSOR	Pir_in_Hall	true	FFFFFFFF006356f6
05473B9SLP2	2019-01-01 12:14:59	SENSOR	Pir_in_Hall	false	FFFFFFFF006356f6
05473B9SLP2	2019-01-01 12:14:59	SENSOR	Pir_Sofa_Kitchen	true	FFFFFFFF006356f6
05473B9SLP2	2019-01-01 12:15:09	SENSOR	Pir_Sofa_Kitchen	false	FFFFFFFF006356f6
05473B9SLP2	2019-01-01 12:15:09	SENSOR	Pir_Fridge1_Kitchen	true	FFFFFFFF006356f6
05473B9SLP2	2019-01-01 12:15:15	SENSOR	Pir_Fridge1_Kitchen	false	FFFFFFFF006356f6
05473B9SLP2	2019-01-01 12:15:16	SENSOR	Pir_Fridge2_Kitchen	true	FFFFFFFF006356f6
05473B9SLP2	2019-01-01 13:22:45	SENSOR	Pir_Fridge2_Kitchen	false	FFFFFFFF006356f6

Figure 23: DB example

CHAPTER 4

UNIVPM HomeCare Web-App Validation

4. Motivations and methods

As widely discussed in the previous sections, one of the main objectives of the simulation tool as part of the measurement chain, is to simulate and evaluate how the measurement chain reacts under certain conditions and to know in advance which is the best physical configuration to carry out the measure in a real environment reducing time and costs. In order to show the efficiency of the usage of the developed simulator, six case studies have been considered and grouped into two macro categories. The *Machine Learning Error* (also known as the out-of-sample error) has been the parameter considered and evaluated and it can be defined as a measure of how accurately an algorithm is able to predict outcome values for previously unseen data. In the first category, four case studies aim to define and show the efficiency of the simulator in determining the optimal configuration of sensors to perform the desired measurement. In the second category, two case studies are implemented to study under which conditions the developed ML algorithms collapse. Before going into the description and analysis of the developed case studies, we can summarize the objectives of the tests performed in the following list:

1) Validate the goodness and effectiveness of the developed simulator;
2) Show that the simulator is an instrument of the measurement chain that allows, through analysis on simulated data, to carry out corrective actions on the real measurement chain so that the measurement is efficient;
3) Define the optimal configuration of sensors which minimize the ML error;
4) Define the minimal configuration of sensors that minimize costs and ML error;
5) Show the tool is a real service able to conduct data analysis and showing results in real time;
6) Define the best ML algorithm for problems concerning the recognition of human activities;
7) Show the stability of the implemented algorithms;

8) Study under what considerations the aforementioned algorithms become unstable.

4.1 First case study category

4.1.1 Measurment protocol and data perturbation

To the first category belong the four case studies aimed to define 1) the optimal configuration of the sensors for the same test performed, 2) the minimal sensors configuration with a reasonable ML error and the stability of the algorithms involved. The objective is to define a relationship between sensors, environmental characteristics and user behavior, achieving the best compromise between ADLs identification ML error and cost. For this purpose, a specific and identical measurement protocol is adopted for all the four cases studies considered. Since the objective is to conduct analysis about the optimal sensor configuration, we choose to keep the environmental characteristics constant, considering to carry out the measurement in the same environment, so as well the user habits. Thus, it has been supposed to consider a healthy person, worker, with the following habits:

- Wake up in the morning between 6.30 and 7.30;
- Breakfast between 7.30-9.30;
- Lunch between 12.00-14.30 for five working days, from Monday to Friday. We consider the weekend with little variation;
- Dinner between 19.00-21.00 for five working days, from Monday to Friday. Again the weekend is considered with variation;
- Sleeping after 22.00 except in the weekend.

Five main tasks need to be identified. Once the ADLs are defined the simulation can take place. One week of manual simulation of user daily life is performed and used as training to feed ML algorithms. It is important to precise that the aforementioned structure is considered for 5 days of the simulated week, the weekend days are generated with some small differences (e.g. no breakfast, later lunch, no dinner, outside for the all day, etc.). When the first week of data is created, several weeks of simulation are automatically generated through the simulator for a total of around 3 months of data acquisition. At this point, to test the stability of the algorithms and to create simulations that are close to reality as much as possible, it has been proceeded with a dataset perturbation. The noise is added in 3 different phases with an increasing level of importance, with the purpose to see how the chain reacts when the measurement is gradually perturbated.

- The first perturbation is introduced manually in the simulation. This means the real user simulates physically the noise in the data introducing it after one month of automatic simulation. In this case the noise is presented as a random shift of 2 hours from the normal habits of the user in the performance of the main tasks;
- The second perturbation is added as a lack of data for a certain period of time plus a random simulation of non-typical behaviour of the user (random PIRs activations, different trajectories considered);
- At the third stage, one week of simulation of completely different habits of the user is added in the datasets (time shits for each task, tasks skip, etc.).

This same protocol is repeated for the four different case studies in which the only variable is given by sensors number and configuration.

4.1.2 Case Studies analysis and results

Case Study 1

Figure 24: Case study 1 PIRs distribution

The considered virtual apartment for the simulation is composed by four room: an entrance, a kitchen, a bedroom and a bathroom. The environment is set up arranging the PIRs in strategic position of the apartment, in correspondence of the furniture involved in user habits. Two PIRs sensors are placed in the living room and one is considered for bathroom, bedroom and hall. The simulation is carried out in such a way that the user performs the tasks to be identified. 2.5 months of data are simulated and disturbed with noise according to the logic described above. The generated datasets visualization is allowed thanks to the statistic area of the simulator. Considering the image below:

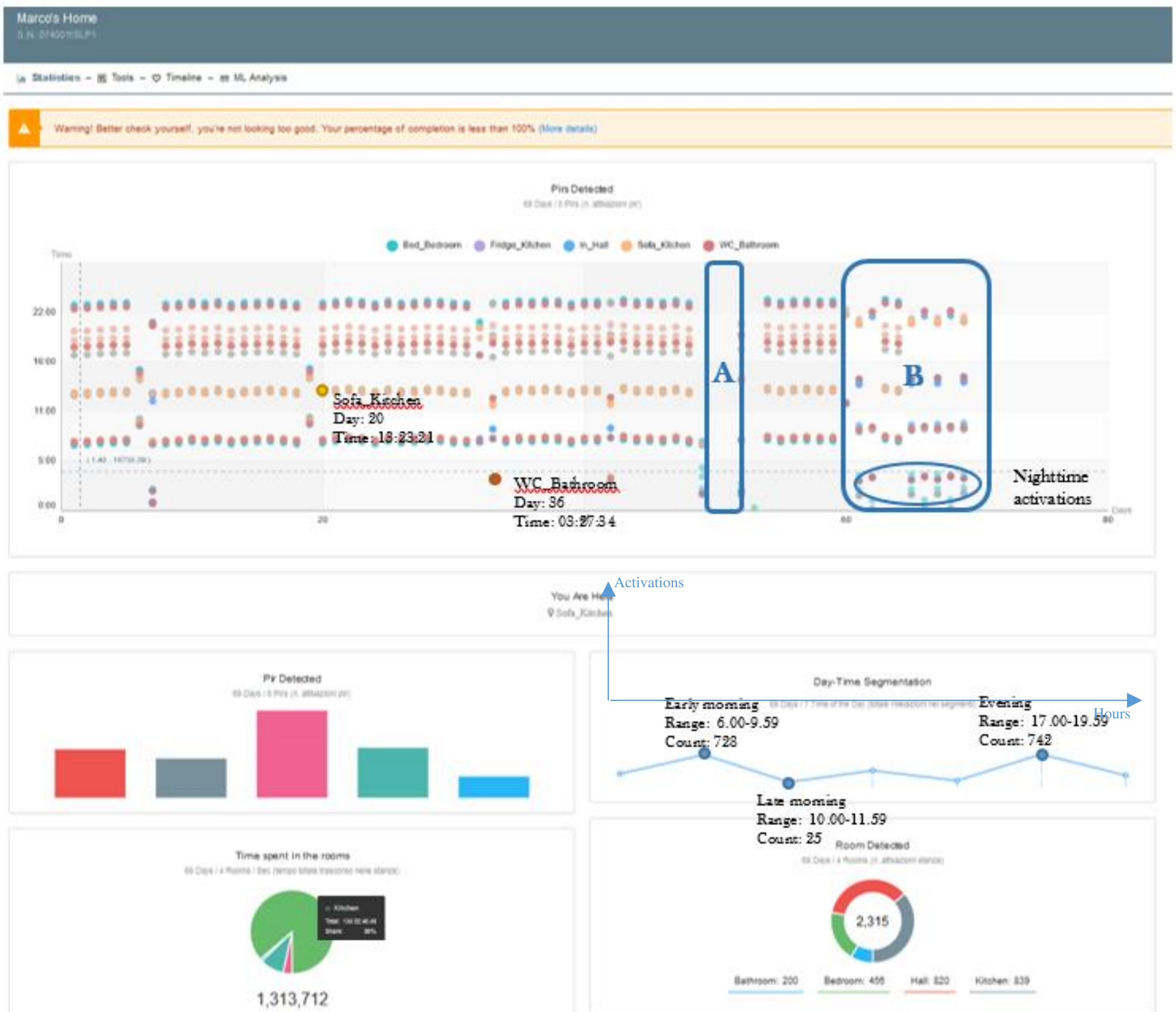


Figure 25: Case study 1 PIRs activation curves

From the first graph is it possible to see that around 70 days of data have been generated and perturbed according to the three noise stages.

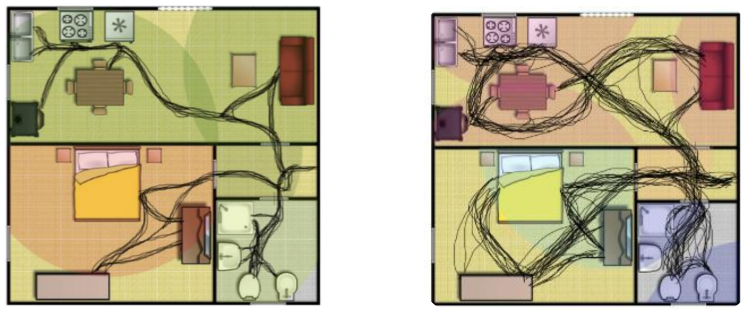


Figure 26: Trajectory examples

In particular:

- Five days of constant activities are simulated with small variation during the weekend and repeated weekly;
- A leak of information is registered (as highlighted in shape A) so as well one week of data perturbation is introduced (as in shape B);
- The broken line shows a major activity of the user in the early morning, lunch time and during the evening;
- 2315 total activation have been registered for a total of 1.313.712 hour of total sensors functioning.

As seen, the statistical analysis performed by the simulator offers the chance to have a direct and quick view on the user behaviour and to notice if a change in his habits occurs. The validity of the simulated test performed is confirmed by analyzing the ML algorithms results. Al discusses, to the purpose the parameter considered is the *Machine Learning Error*. Three main ML techniques have been implemented in the simulation tool but only one of them results with good efficiency. In the considered case study, Decision Tree Classifier offers a ML error of around 5%. It must be said that only results of the case study considering all noise sources are shown in the following pictures. Going into details:

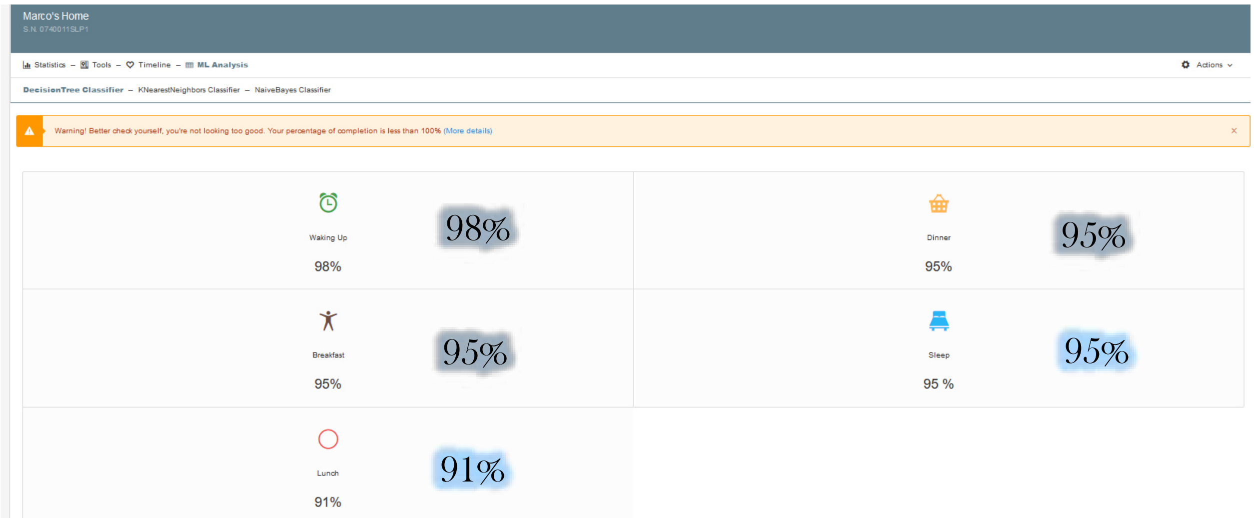


Figure 27: Case study 1 ML (DT) results

The wakeup task have been detected with a ML error of 2%, 5% for breakfast, lunch and dinner and 9% for the sleeping task. KNN and NB Classifier provides quite excellent results for some tasks and completely null for others. These results are under studies, the reason may lay in the logic of the algorithm itself.

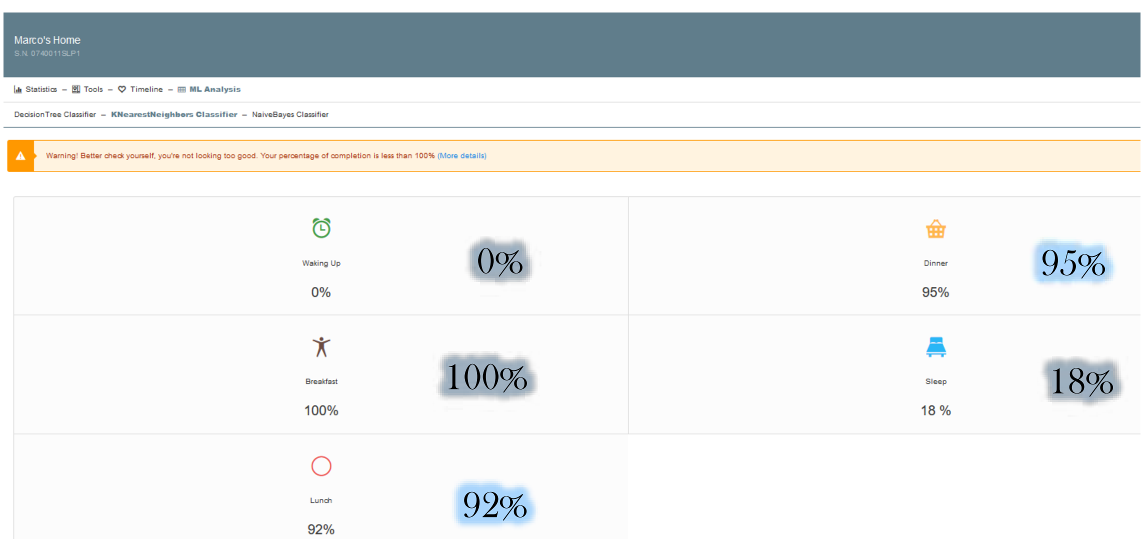


Figure 28: Case study 1 ML (KNN) results

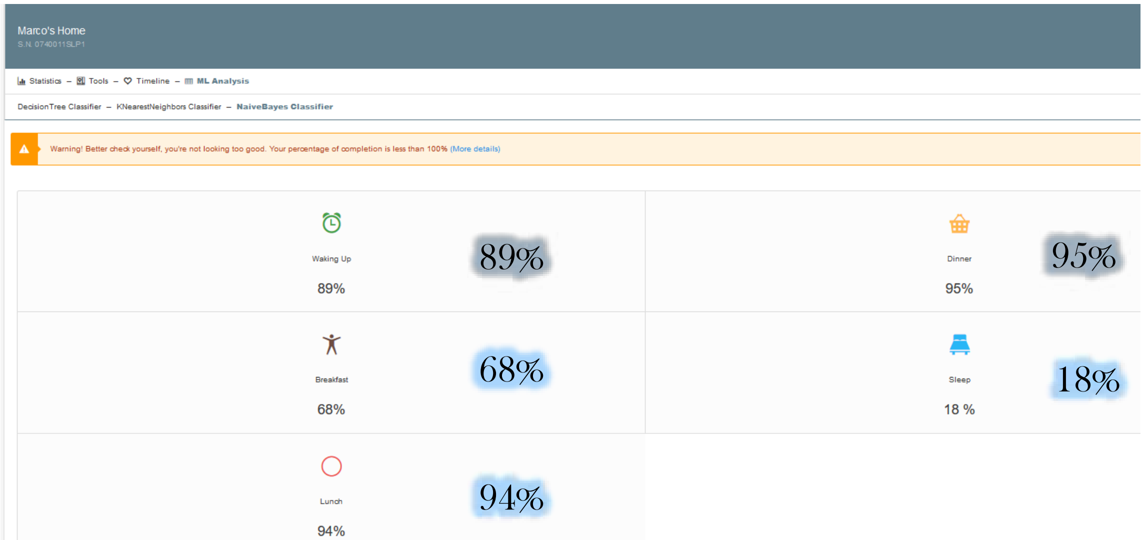


Figure 29: Case study 1 ML (NB) results

The DT Classifier is the one that better suits the case. To test the stability of the algorithm, the perturbation has been introduced in three different stages. The ML results of each test, 0 perturbation, perturbation case 1, 2 and 3 have been reported in a table to show how the algorithm reacts:

CASE A 0 (1 month no noise)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	100	100	100	96	96
KNN	0	100	100	94	8
NB	94	45	100	94	8
CASE A 1 (1 month + Noise Case 1)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	100	97	97	94	94
KNN	0	100	94	94	6
NB	94	48	100	94	6
CASE A 2 (2 month + Noise Case 2)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	98	96	98	95	96
KNN	0	100	96	95	6
NB	96	68	98	95	9
Case A 3 (3 months + Noise Case 3)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	98	95	91	95	95
KNN	0	100	92	95	18
NB	89	68	94	95	18

Table 2: ML results summary

For the case study 1, the ML results for datasets without perturbation are of 100% for wakeup, breakfast and lunch and around 94% for dinner and sleeping. Going directly to the case with all perturbation involved, it is possible to observe an increasing of ML error of the 2% for the first task and 5% for the second one, 9% for lunch and of 1% for the others. The results can be considered more than satisfying.

Case Study 2

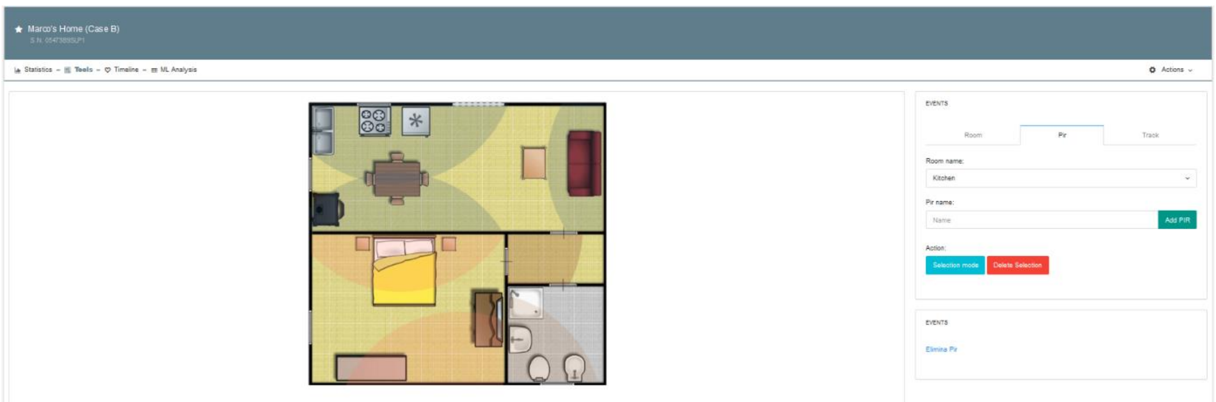


Figure 30: Case study 2 PIRs distribution

Given the fact the aim is to proof the simulator is able to define the best sensors configuration a second case study is considered. The same map has been imported but a different PIRs disposition and a different PIRs number are taken into account. To respect the first case study, three PIRs sensors are here considered in the living room, two in the bedroom and one for entrance and bathroom. Going directly to statistical analysis and ML results we can observe that:

- Around 90 days of data have been generated;
- Two lack of data have been simulated;
- 1723 total activation have been detected and 1353.513 hours of sensors functioning have been registered;
- DT classifier confirm again to most suitable for the case considered with a ML error of 1% for the tasks wakeup and breakfast, 6% for dinner and 11% for lunch and sleeping. These last two tasks record a small increase in the percentage of the ML error probably due to the different number and arrangement of PIRs sensors in the environment;

- KNN and NB Classifier increase their performances increasing the number of sensors involved in the measure even when the simulation is periodically disturbed by noise, but the results are still under studies.

Statistical an Ml results are reported below.

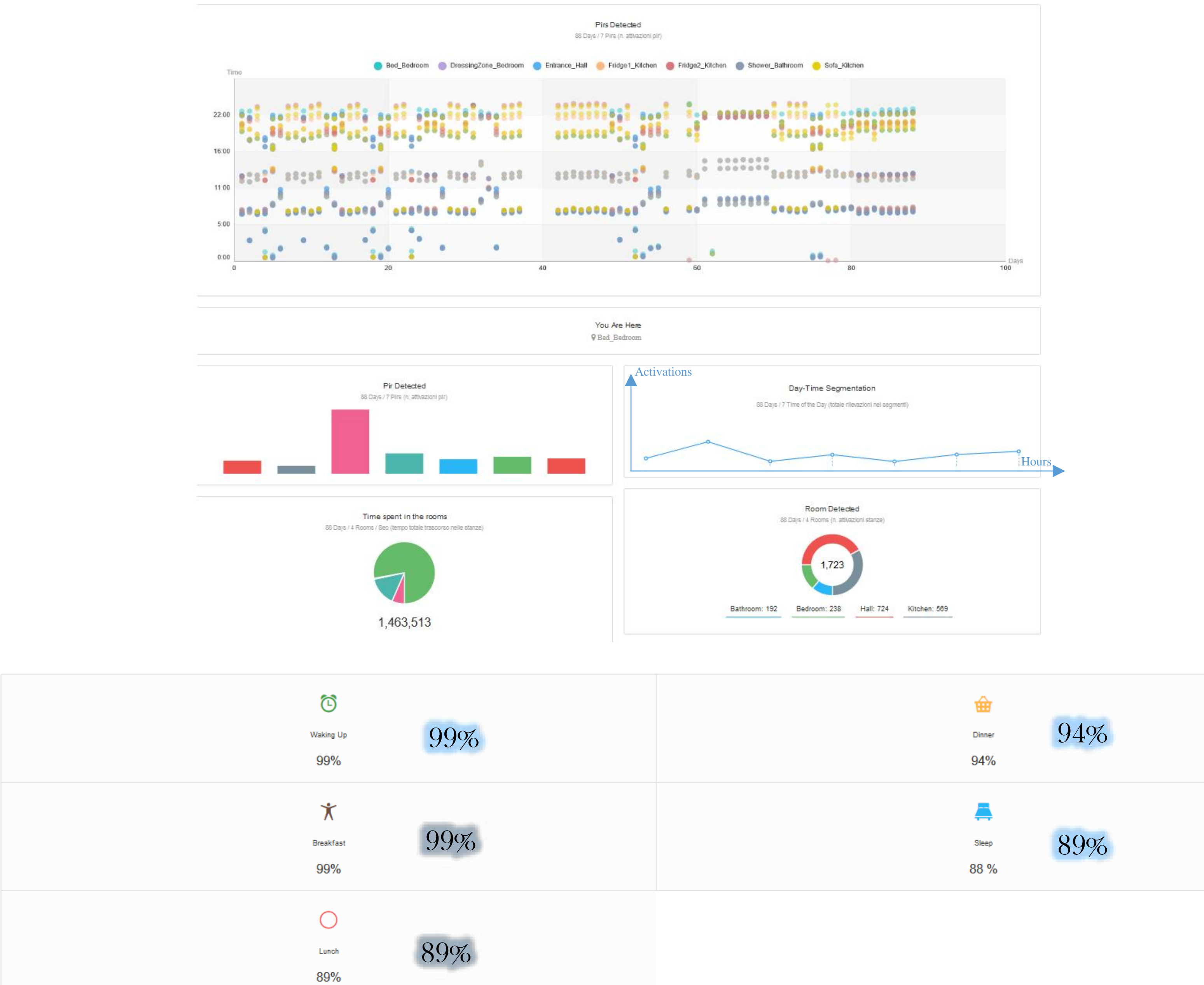


Figure 31: Case study 2 PIRs activation curves and ML results

Considering a summery table of ML results:

CASE B 0 (1 month no noise)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	100	100	90	95	100
KNN	0	100	87	100	45
NB	87	26	87	100	45

CASE B 1 (1 month + Noise Case 1)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	100	100	85	91	94
KNN	0	100	88	100	42
NB	85	80	55	100	42

CASE B 2 (2 month + Noise Case 2)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	100	100	85	92	96
KNN	0	100	87	100	40
NB	85	28	87	100	40

Case B 3 (3 months + Noise Case 3)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	99	99	89	94	90
KNN	0	99	90	99	82
NB	89	48	90	99	82

Table 3: ML results summary

Comparing the first case with zero perturbation and the last one with all perturbations involved, a ML error of 1% is registered for wakeup and breakfast tasks and 10% for sleeping is registered. The algorithm implemented confirms to be stable and results satisfying. Considering the two proofs in the middle it is possible to see an increasing of ML error for example in the task lunch from 10% to 15% and then a new decreasing with the last test: the explanation could be found in the fact that even if the Case B 3 is the one considering all the perturbations, at the same time more data are generated in a "normal" condition respect to Case B 1 and 2, where two perturbation are introduced without keeping generating standard data. This could help the ML to find again its stability.

Case Study 3 and 4

Figure 32: Case study 3 PIRs distribution

In case study 3, a different PIRs sensors configuration and number are considered: three PIRs sensors are positioned in the living room and in the bedroom while one is considered for the entrance and one for the bathroom. Going directly to statistical analysis and ML results:

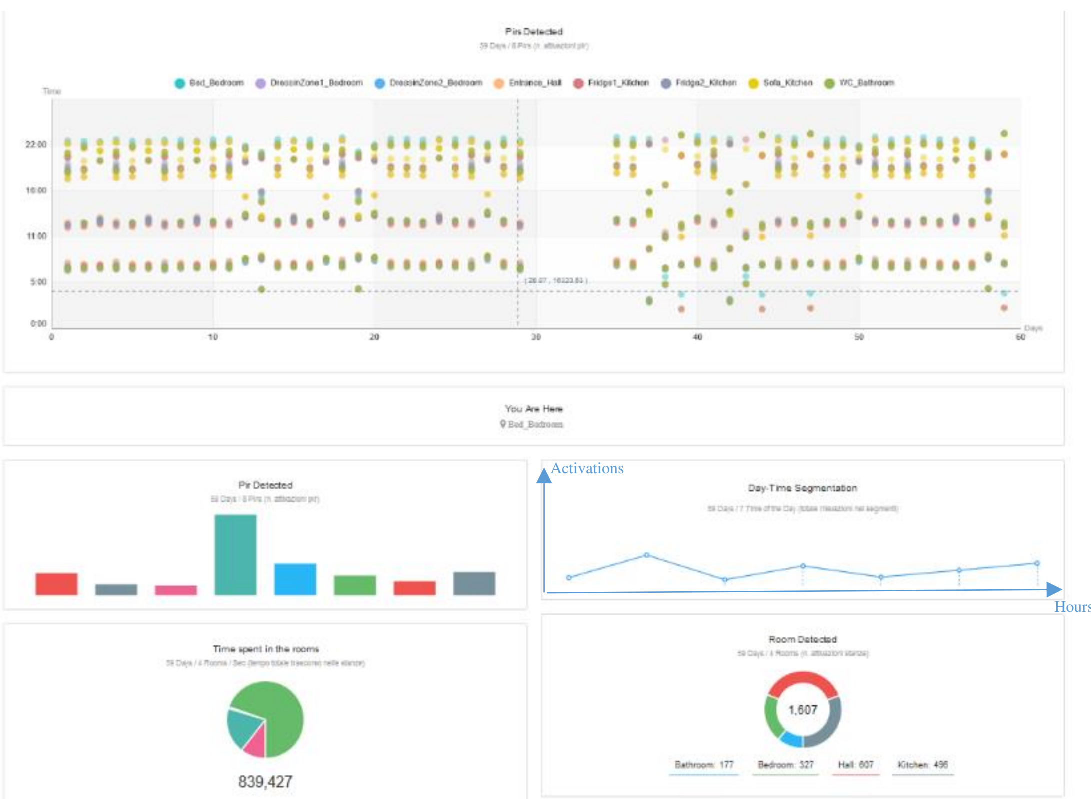


Figure 33: Case study 3 PIRs activation curves

60 days of data have been simulated for a total of 1607 total number of activation and 839.427 hours of sensors functioning. DT Classifier confirms again to be the algorithm with the best performance compared to KNN and NB. In particular, increasing the number of PIRs sensors, the stability of the algorithm is kept even in case of big perturbation of the measurement. According to results, only a ML error increase of 4% for lunch task and 2% for sleeping task have been registered compering the non-perturbed simulation with the perturbed one.

CASE C 0 (1 month no noise)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	100	100	100	100	100
KNN	0	100	100	100	7
NB	100	52	100	100	7
CASE C 1 (1 month + Noise Case 1)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	100	100	100	100	100
KNN	0	100	100	100	7
NB	100	52	100	100	7
CASE C 2 (2 month + Noise Case 2)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	100	100	98	100	96
KNN	0	100	98	100	18
NB	100	56	100	100	18
Case C 3 (2 months + Noise Case 3)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	100	100	96	100	98
KNN	0	100	96	100	20
NB	96	65	100	100	20

Table 4: ML results summary

Summarizing results for the three case studies considered:

Indicators	Marco Case A	Marco Case B	Marco Case C
PIR total Number Activation	2315	1723	1607
PIR total operation hours (h)	1313.712	1463.513	839.427
Location most active PIR	Hall Sofa and Bedroom Fridge Bathroom	Hall Fridge Bathroom Sofa	Hall Fridge Bathroom Bedroom

Table 4: user cases results summary

CASE A 0 (1 month no noise)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	100	100	100	96	96
KNN	0	100	100	94	8
NB	94	45	100	94	8

CASE A 1 (1 month + Noise Case 1)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	100	97	97	94	94
KNN	0	100	94	94	6
NB	94	48	100	94	6

CASE A 2 (2 month + Noise Case 2)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	98	96	98	95	96
KNN	0	100	96	95	6
NB	96	68	98	95	9

Case A 3 (3 months + Noise Case 3)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	98	95	91	95	95
KNN	0	100	92	95	18
NB	89	68	94	95	18

5.2% ML error DT average

CASE B 0 (1 month no noise)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	100	100	90	95	100
KNN	0	100	87	100	45
NB	87	26	87	100	45

CASE B 1 (1 month + Noise Case 1)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	100	100	85	91	94
KNN	0	100	88	100	42
NB	85	80	55	100	42

CASE B 2 (2 month + Noise Case 2)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	100	100	85	92	96
KNN	0	100	87	100	40
NB	85	28	87	100	40

Case B 3 (3 months + Noise Case 3)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	99	99	89	94	90
KNN	0	99	90	99	32
NB	89	43	90	99	32

5.8% ML error DT average

CASE C 0 (1 month no noise)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	100	100	100	100	100
KNN	0	100	100	100	7
NB	100	52	100	100	7

CASE C 1 (1 month + Noise Case 1)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	100	100	100	100	100
KNN	0	100	100	100	7
NB	100	52	100	100	7

CASE C 2 (2 month + Noise Case 2)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	100	100	98	100	96
KNN	0	100	98	100	18
NB	100	56	100	100	18

Case C 3 (2 months + Noise Case 3)	Wake Up	Breakfast	Lunch	Dinner	Sleep
DT	100	100	96	100	98
KNN	0	100	96	100	20
NB	96	65	100	100	20

2.2% ML error DT average

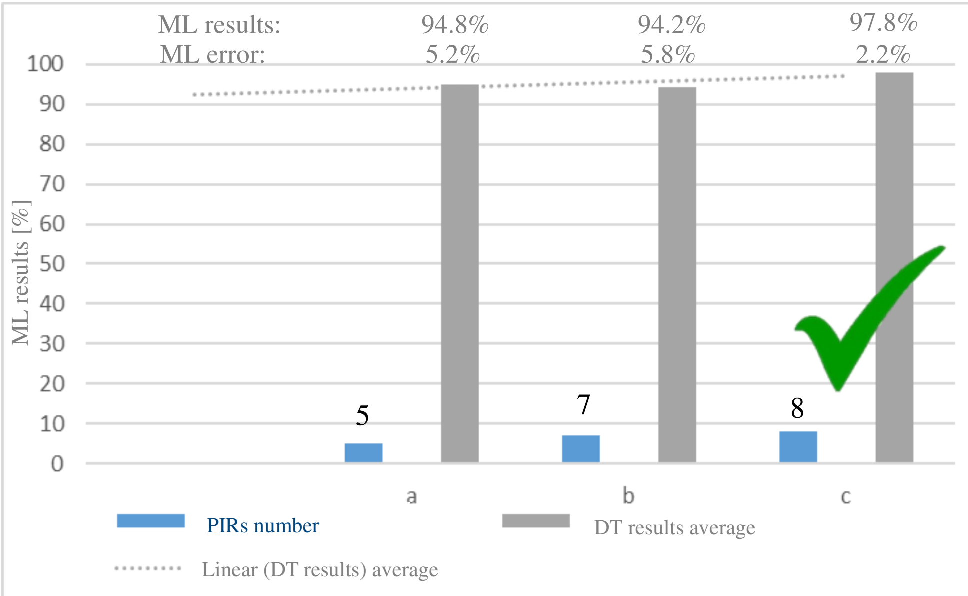


Figure 34: ML results and error respect to PIRs number used in the

As we can see from the chart, increasing the number of sensors involved in the measurement it is possible to reach a small value in the ML error for the detection of the aforementioned tasks. To this stage, the questions:

- How much is important to have a ML error of around 2% in the ADLs detection?
- How much this error costs in terms of feasibility of the measurement?
- Which could be a good compromise between a reasonable value of ML error and costs saving?

To answer to these questions, a fourth case study has been considered minimalizing the PIRs configuration to in order to reach a good compromise with costs. One sensor per room has been considered. DT classifier on 60 days of data gave results for a ML error of 5% for wakeup task, 11% for breakfast, 15% for lunch and dinner and 13% for sleeping task. The average of ML result in terms of percentage is around 88.2% and the relative ML error is around 11.2%. The compromise could be considered more than acceptable. To be more precise, it is possible to plot on a diagram on one side the curve uncertainty over the number of sensors and on the other one the total cost of the sensors. Data are retrieved from the following table. The crossing point of the two charts defines the best compromise in terms of number of sensors that minimizes ML error and overall cost.

Supposing the cost for each sensor around 20 euros:

Cost/sensors [Euros]	*Number*	*ML error [%]*	*Total cost [Euros]*
20	5	5.2	100
20	7	5	129
20	8	2.2	160
20	4	11.8	80

Table 5: Number of sensors and uncertainty related to cost

Reporting in a double chart the above data, it is possible to obtain the following structure.

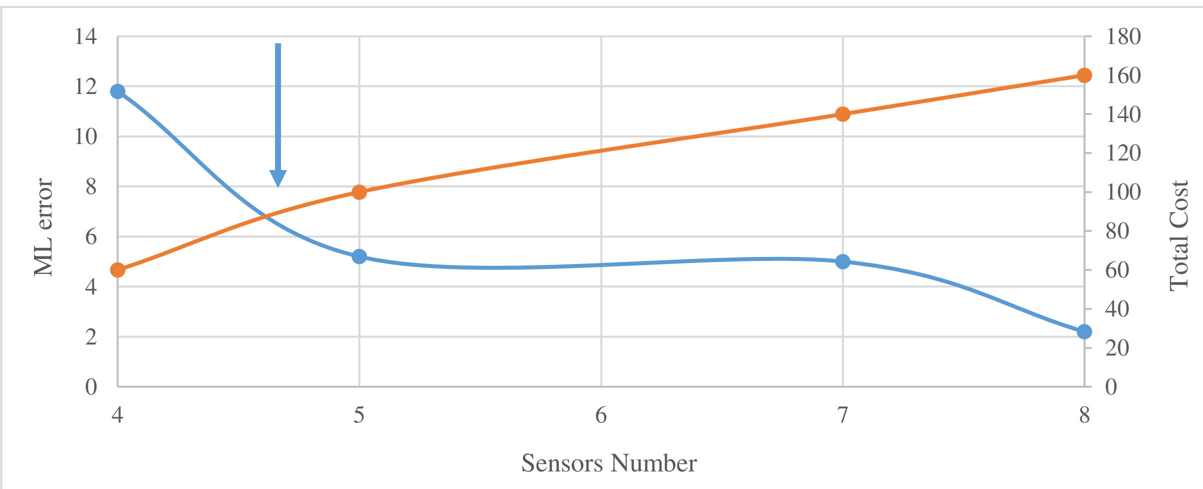


Figure 35: Double chart

From this chart is it possible to observe that the optimum of the compromise in the first category of case study is obtained for a total number of four sensors for a cost of around 80 euros with a ML error of around 6%.

4.1.3 First category case studies conclusions

According to results to this first case study category, DT classifier is the most performant algorithm to identify ADLs tasks as those described in this research work. Increasing the number of PIRs sensors is it possible to minimize the ML error even strongly perturbing the system, the value increase from 0.8% in case of non-perturbed system to 2.2% in case of perturbation. However, the increasing of sensors number has an impact in terms of costs on the measurement chain. To minimize the cost, a fourth case study has been considered to define in which measure the minimization of the configuration could affect the ML error of the task detection. Results are more than satisfactory. With a minimal configuration of just one sensor per room the ML error of the detection for the five tasks considered is around 11.8% for a total of 80 euros against a ML error of 1.8% for the double of the cost. The compromise could be considered satisfying. It is necessary to precise that this condition is verified for simple task to be identified given the fact just PIRs sensors have been considered in this work.

4.2 Second case study category

Once the stability of algorithms implemented has been proved, could be interesting to define under which conditions the aforementioned ML technique collapses. The second category of case studies aims to this purpose. Two main tests have been carried out to "stress" the algorithm in two different levels. In both simulations performed, the same environment and same PIRs number and disposition of the case study 3 have been considered. The choice is due to the high performances reached by the DT Classifier in such case study.

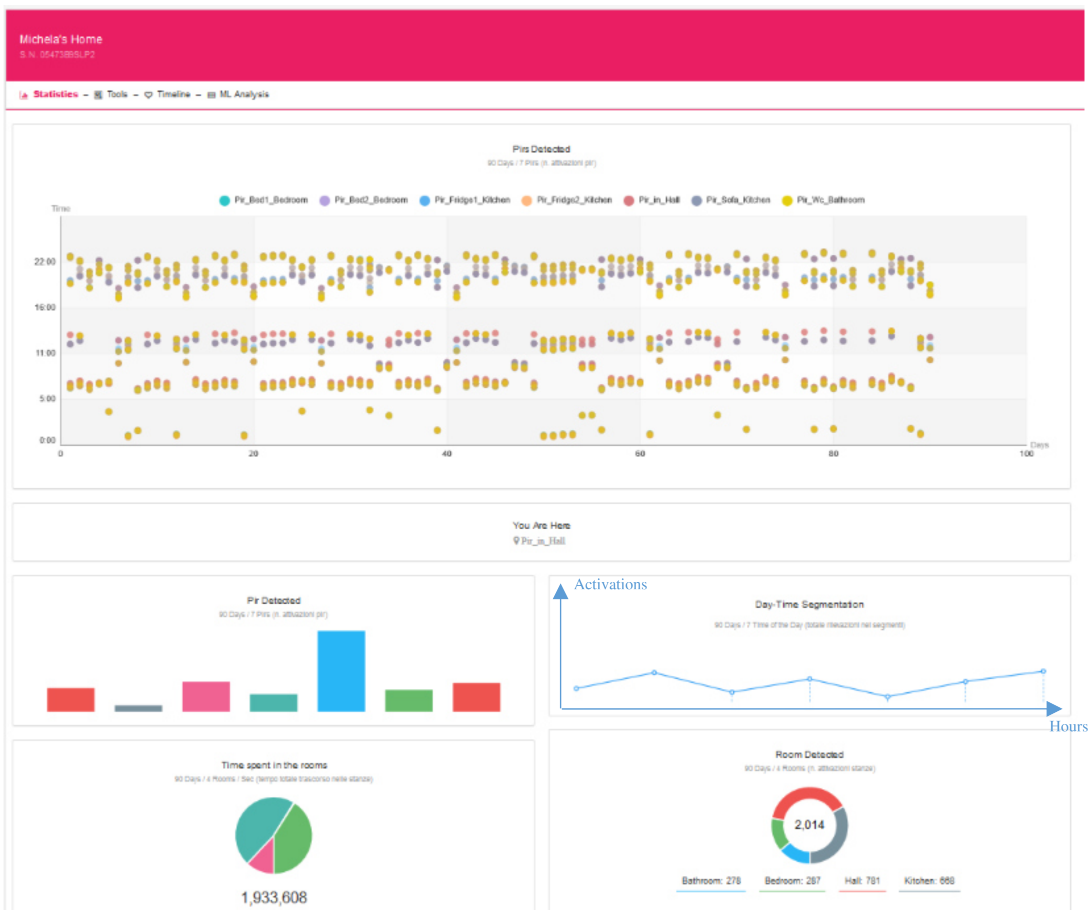
Case Study 1

In the first test, a variation of the 30% respects to the training model for ML is considered and implemented in the user habits. The virtual user is considered to perform the following tasks:

- Wake up between 7.30-8.30;
- Breakfast until 9.00;
- Lunch between 13.00-14.00;
- Dinner between 20.00-21.00;
- Sleeping after 22.30;

o Nighttime activations considered random.

Three months of simulations has been conduct.



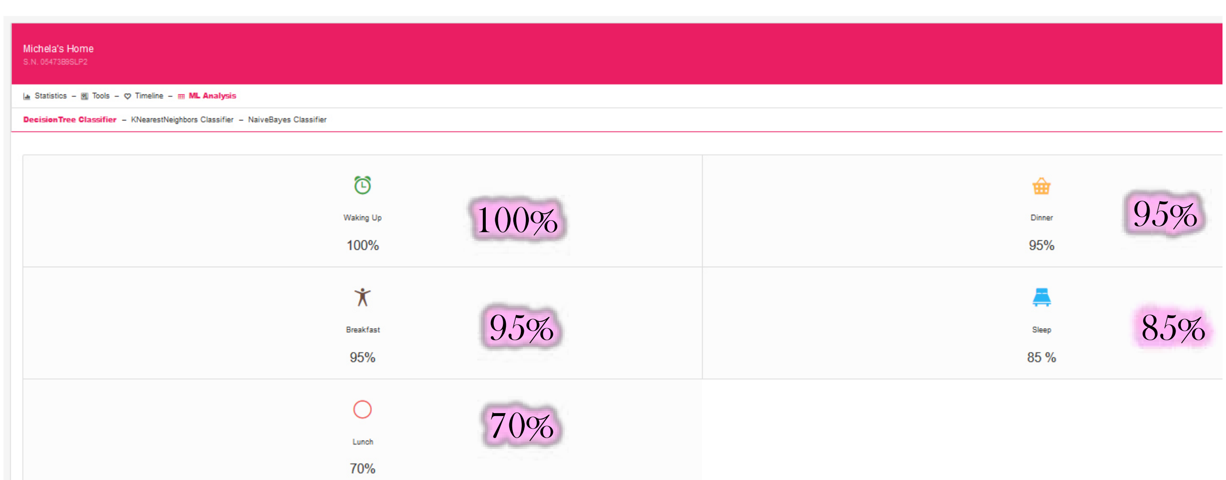


Figure 36: Second category, first case study PIRs activation and ML

As we can see from results, ML algorithm stability for DT Classifier is fully confirmed except for the task Lunch which is the one affected by the biggest difference being the slot completely shifted of one hour. So even introducing a simulation that varies the datasets for around the 30%, results are still fully satisfactory. In order to be able to define a collapse condition, a variation of the 80% from the training phase is introduced.

Case Study 2

The virtual user of the second case study is characterized by a completely different life style respect to previous one. It has been considered to simulate the habits of a user with a high level of activity in the late evening and during the night, with around 2 hours of shift in the morning and the task lunch considered as a random variable. Going into details:

- Wake up between 8.30-9.30;
- Breakfast after 9;
- Lunch always variable
- Dinner after 22.00;
- Sleeping after midnight;
- Nighttime activations considered constant.

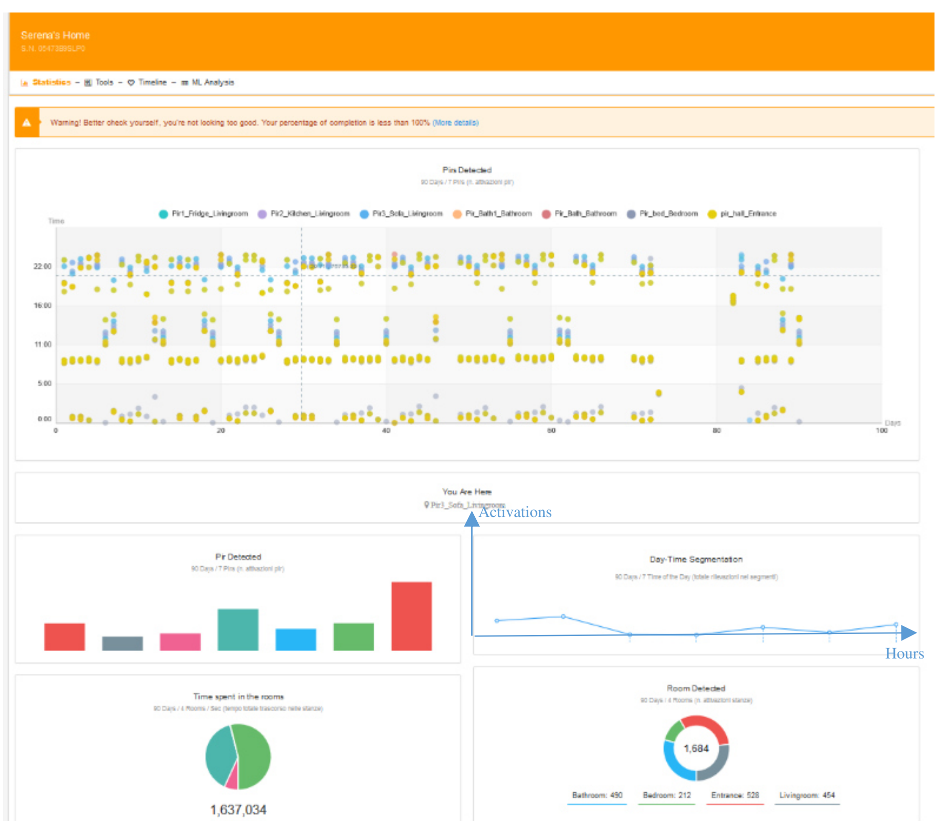
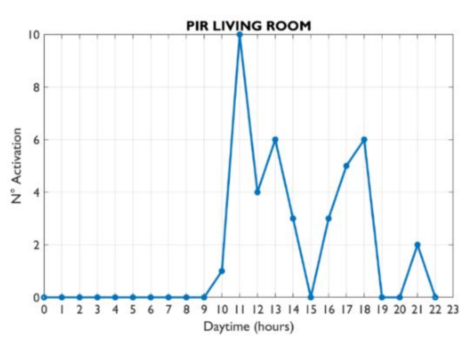


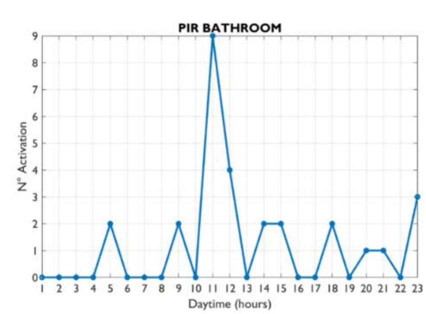
Figure 37: Second category, second case study PIRs activation

Analyzing ML results, the algorithm collapses for the intermediate tasks, breakfast, lunch and dinner, while around a 30% of ML error is identified in the Wake up and Sleeping tasks. With these case studies is it possible to confirm the stability of the algorithm for variation of habits under the 50%, while collapses conditions are reached when variation of more than 50% are introduced and simulated.

4.3 UNIVPM HomeCare Web App inside e-Ware real context

As already explained in the first chapter, this research work is born inside eWare project scenario. In the eWare project, a robot is integrated with a sensor network to measure ADLs with the aim to provide relevant suggestions to the user and reduce the caregiver influence, [136]. The eWare system is designed to support individuals with cognitive disabilities in achieving their Activity of Daily Living goals, such as having breakfast on time every day, eating dinner at appropriate times, leaving the house for activities, or going to bed at the right time. Individuals with cognitive disabilities such as dementia often have difficulties in initiating activities. eWare project introduces a novel eco-system characterized by the integration of two technologies: lifestyle monitoring and social support robotics. As already discussed, lifestyle monitoring is carried out by Sensara Senior Lifestyle System, that consists of a limited number of small sensors (a set of three passive infrared (PIR) sensors and a two open/close door sensors) installed at strategic places in the home (living room, bathroom, kitchen, hallway and front door) and they automatically connect wirelessly to a gateway plugged into a standard home internet router. An example of data we consider for the eWare project are presented below. The following data refer to different moths of monitored activity of a women of 75 years old with 3 PIR sensors placed in the kitchen, in the bedroom and in the bathroom of the apartment.





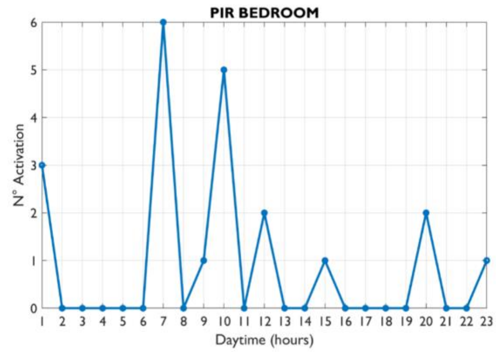
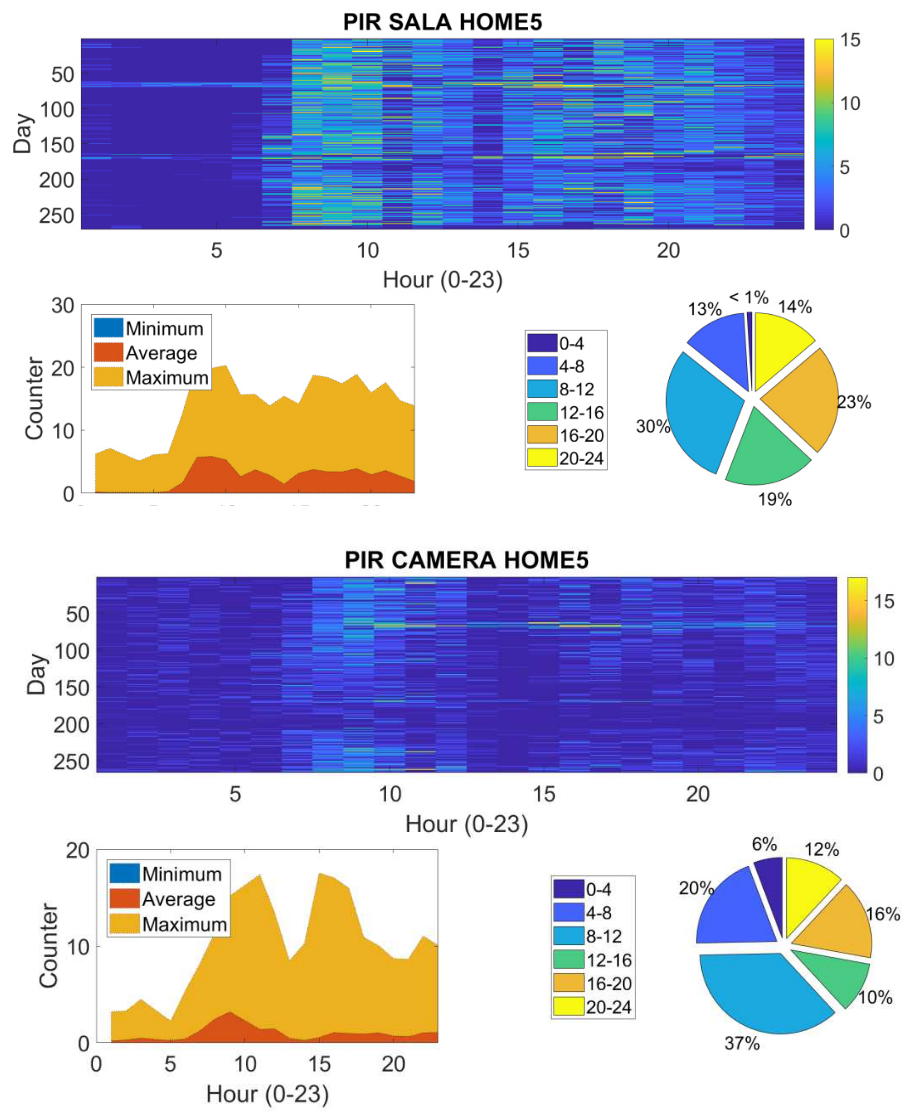


Figure 38: Real one-day PIRs activation results for eWare project



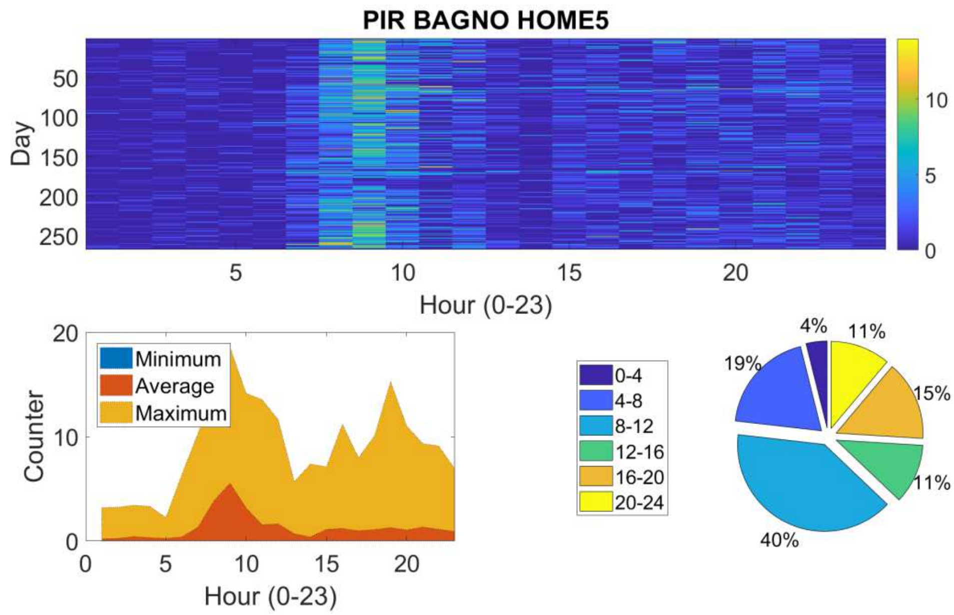


Figure 39: Real PIRs months activation results for eWare project

Sensor data are uploaded to an analytics engine "in the cloud" (Sensara Cloud) that, after two weeks, is able to recognize living patterns of the person thanks to the eWare Artificial Intelligence (eWare A.I.). When the daily behavioral patterns are known, unusual event can be detected and a warning is sent to the smartphone of the caregiver (eWare Web UI). To help these individuals, the eWare project has integrated the Sensara system and the Tessa robot to provide context relevant suggestions. These suggestions match their personal goals. The part of social support is made by Tinybot, which is a small social robot that provide emotional support for the pathological user by talking, giving friendly suggestions, reminders, and playing personal music. The goal is to activate the person with dementia and prevent them from staying in a passive state. Tinybot runs a Linux based operating system and can connect via WiFi to the Tinybots cloud backend. The backend securely stores (amongst other data) user profiles, user's behavioural patterns, and strategies to activate the elder and it is connected to sensor data through eWare Database. Over time, the eWare A. I. learns and adapts to support the inhabitants' specific needs, giving also real time reminders and suggestions. Two different kind of tests are implemented, respectively beta and alfa tests. Beta tests are real tests conducted on 300 end users (pathological subject, formal and informal caregivers). Alfa tests are preliminary tests aiming to improve the system architecture in order to: reach high level of efficiency during installation, easily solve system bugs, obtain users feedbacks. The initial participatory sessions took place in Italy, Switzerland

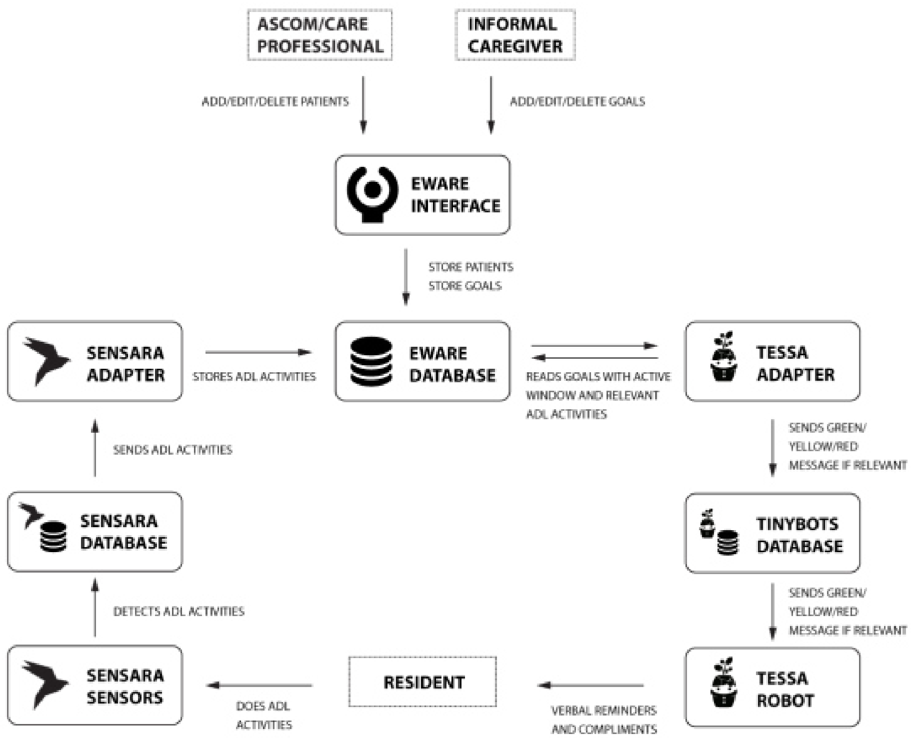


Figure 40: eWare ecosystem structure

and The Netherlands, involving the future primary, secondary and tertiary end-users of eWare. From the active involvement sessions, needs were gathered for the person with dementia in respect to support in ADLs and these were translated into system requirements for creating ADL goals with the eWare system,[136]

As described in [136], after the active involvement sessions, qualitative and quantitative information was gathered in considering nine older people with early to middle stage dementia, four informal caregivers, and one formal caregiver. Life-style monitoring sensors and Tessa were placed into the apartment of older people with mild to moderate dementia for three weeks. All participants were interviewed, except for one due to personal circumstances. The in-situ evaluation used a Wizard-of-Oz test set-up, in which the participants interacting with the robot believe it is fully autonomous, but which is operated by an unseen human being [137] and had a twofold purpose to the eWare project. First, to gain rich insights in the actual services that people with dementia would like to use in the eWare system and to see how they would respond to the interaction with the Tessa. Second, a Wizard-of-Oz experiment was also intended to see whether the sensor system would give enough information about the lifestyle for the AI to evoke meaningful insights for the robot to respond to. The experiment lasted for three weeks at four different households. People with dementia and their informal cares provided relevant input for functionalities and improvements of eWare as a response after trying out this Wizard-of-Oz version.

In this real scenario the goal of the developed tool was to provide a support instrument for technicians, developers and installers that was able to speed up analysis and monitoring times, to allow rapid identification of changes in behavioral trends, to allow system performance monitoring and to study the best configuration of the sensors network for a given environment. The tool was therefore set based on the needs of the project itself and therefore on what was happening in the apartments. This explains the absence of a test of the tool developed on a real dataset although the system is predisposed to direct acceptance of the actual data suitable for eWare and to increase the number of activities to be identified.

CHAPTER 5

CONCLUSIONS

5.1 Summary

In this research work, a simulator able to conduct real time analysis on simulated PIRs sensors measurements to discover ADLs thanks to ML algorithms has been developed and presented as a support in the eWare project scenario for technicians, developers and installers to speed up analysis and monitoring times, to allow rapid identification of changes in behavioral trends and system performance monitoring and to study the best configuration of the sensors network for a given environment. The idea behind the tool development came out from the eWare project itself. Starting from the data simulated by virtual environmental sensors, statistical algorithms and ML analysis techniques have been implemented to provide quantitative and qualitative information. As widely discussed, the development of technologies has led to the creation of intelligent environments that allow to face the growing rate of aging of the population by providing domestic care that facilitates the life of the patient and drastically reduces health costs. The construction of such environments, however, requires considerable effort in economic and time terms. In order to create effective, efficient and ad hoc environment, a huge number of tests are required to define the optimal configuration of the sensors, the analysis techniques to be implemented for the interpretation of results, etc. Such datasets are not easily available given the large amount of the variables involved both in physical (environmental and sensorial variables) and in technical implementation terms, such as the artificial intelligence adopted. As a matter of fact, one of the main problems for researchers is the one to define in which measure the variation of certain parameters could influence the goodness and reliability of the result of the measurement carried out and therefore the reliability of identification of a certain ADLs. In this research work we tried to provide a parametrization of the e-health problem and a solution able to reduce time and costs by inserting a key element in the measurement chain, represented by the simulation. Inside this thesis work a simulator of dataset related to the daily activities of virtual users has been implemented. Since the current trend is towards indirect behavioral monitoring techniques, in this first phase the tool is based on the use of non-invasive, low-cost environmental sensors, as PIRs sensors. These sensors do not provide information directly correlated to the state of health of a person but can offer behavioral information no less significant than those provided by medical sensors. The tried and tested approach is particularly innovative and promising in that, by exploiting simple and mainly targeted information for other purposes, it provides high value-added functions that can be integrated into AAL systems to increase their intelligence and make them suitable for supporting tools for monitoring well-being of people and not just environmental monitoring and management tools. The tool developed during this PhD course allows the real user to import any type of environment in the form of a map image, thus overcoming the limits provided by color-based tools, and defining both its environmental and sensorial characteristics. At present, only PIRs sensors are considered into the system even if the user can choose from different characteristics. However, the simulator is developed so that it is expandable and other types of sensors can be added in the future. The simulation can be performed according to a measurement protocol defined by the user producing an output of the virtual sensors that mimics the behavior of the real sensors. The data is specifically saved on a database and made available in txt format for subsequent processing. One of the main advantages of the developed simulator is therefore to provide the chance to generate datasets reproducing months of user activity in a completely rapid, free and automatic way. In order to provide also a service and not only an element of the measurement chain, it was decided to give intelligence to the simulator by developing algorithms able to conduct real time analysis through ML techniques and static analysis on the simulated data. Three different ML algorithms based on supervised methods have been implemented in the simulator: DT, KNN and NB Classifier. The aim is that to define which one among them is the most performing in defining some of the activity of daily living. The DT Classifier has shown much more than satisfactory results. The main advantage is to dispose of an instrument that is not only able to perform simulation but also to monitor and analyzes quantitatively and qualitatively simulations carried out. This allows therefore to carry out tests with purpose to evaluate, in real time, how variations of determined environmental parameters can influence the goodness and accuracy of the measurement carried out. The main objective of the thesis work has been that to provide a system able to simulate the response of the measurement chain according to the configuration of the chain itself in order to reach in the a real SE the best compromises between the cost of realization and goodness of the measurements performed. In this regard, in order to test the validity of the simulator developed, six different case studies have been considered and implemented. A measurement protocol has been adopted and data noise has been introduced to test the stability of the simulator in three different stages. The first four case studies aim to demonstrate the contribution of the simulator in defining the optimal configuration of the sensors which leads to minimize the ML error value in the detection of five main tasks of ADLs. The results show that by maximizing the number of sensors used in the measurement, the such value can be reduced to 0.8%. However, since in real cases the cost is one of the main factors, it was decided to conduct studies aimed to define the minimal sensors configuration that would have allowed to obtain an acceptable accuracy with a drastic costs reduction. Results were more than satisfactory and confirmed the recent trend to reduce the number of sensors. Increasing the ML error around 11.2% it is possible to reach the best compromise with costs saving. The other two other case studies are instead aimed to show the stability of the implemented ML algorithms and to define under what conditions the implemented algorithm collapses. In conclusion, the work carried out during the research work had, as a final goal, the experimentation and validation of an innovative approach aimed at the study of behavior and behavioral trends based on the processing of simulated data and to offer a powerful instrument to complete the measurement chain. It should be noted that the tool can also directly process data from real sensors.

5.2 Future works

The developed simulator offer the chance to conduct "a posteriori" analysis on the best configuration of the system to conduct the measure maximizing accuracy and reducing costs. One of the main and next future developments could be to inject a defined simulation into the system to directly obtain the optimal sensor configuration. More over, the system is now able to recognize five different main task in ADLs terms. Could be important to extend this aspect to more detailed ADLs englobing in the system other kind of sensors. For now only the first part of the measurement chain representing the problem as discussed in 3.2 has been approached. As already said, each block of the chain is characterized by a own uncertainty. In this work a study on how the uncertainty related to sensors characteristics affect results has been carried out. The next could be to define how other variables related to the system affect the measurement itself in order to carry out a function to link all the variables involved.

References

1. Majumder S., Aghayi E., Noferesti M., Memarzadeh-Tehran H., Mondal T., Pang Z., & Deen, M. (2017): "*Smart homes for elderly healthcare - Recent advances and research challenges*". Sensors, 17(11), 2496.
2. Uddin M., Khaksar W. and Torresen J. (2018). "*Ambient sensors for elderly care and independent living: a survey*". Sensors, 18(7), 2027.
3. Rashidi P. and Mihailidis A. (2013): "*A survey on ambient-assisted living tools for older adults*". IEEE journal of biomedical and health informatics, 17(3), 579-590.
4. "*Introducing knowledge in the process of supervised classification of activities of Daily Living in Health Smart Homes*", The 12th IEEE International Conference on e-Health Networking, Applications and Services, 2010.
5. Tan-Hsu Tan ; Munkhjargal Gochoo ; Shih-Chia Huang ; Yi-Hung Liu ; Shing-Hong Liu "*Multi-Resident Activity Recognition in a Smart Home Using RGB Activity Image and DCNN*", IEEE Sensors Journal, 2018.
6. Gina Sprint ; Diane Cook ; Roschelle Fritz ; Maureen Schmitter-Edgecombe: "*Detecting Health and Behavior Change by Analyzing Smart Home Sensor Data*", IEEE International Conference on Smart Computing (SMARTCOMP), 2016.
7. A. M. Burgess and C. G. Burgess: "*Aging-in-place: present realities and future directions*" in Forum on public policy, Citeseer, 2017.
8. Cornelis E., Gorus E., Beyer I., Bautmans I. and De Vriendt, P.: "*Early diagnosis of mild cognitive impairment and mild dementia through basic and instrumental activities of daily living: Development of a new evaluation tool*". PLoS medicine, 14(3), e1002250. (2017)
9. Belmonte-Fernández Ó., Puertas-Cabedo A., Torres-Sospedra J., Montoliu-Colás R. and Trilles-Oliver S.: "*An indoor positioning system based on wearables for ambient-assisted living*". Sensors, 17(1), 36. (2016)
10. Pal D., Triyason T., Funilkul S. and Chutimaskul W.: "*Smart Homes and Quality of Life for the Elderly: Perspective of Competing Models*". IEEE Access, 6, 8109-8122. (2018)
11. Pirbhulal S., Zhang H., E Alahi M. E., Ghayvat H., Mukhopadhyay S. C., Zhang Y. T. and Wu, W. (2016): "*A novel secure IoT-based smart home automation system using a wireless sensor network*". Sensors, 17(1), 69.

12. Cook D. J., Krishnan N. C. and Rashidi P. (2013): "*Activity discovery and activity recognition: A new partnership*". IEEE Transactions on Cybernetics (June 2013, Volume: 43 , Issue: 3 , pp. 820-828).

13. Lloret J., Canovas A., Sendra S. and Parra L.: "*A smart communication architecture for ambient assisted living*". IEEE Communications Magazine 2015, 53, pp. 26–33.

14. Charith Perera, Arkady Zaslavsky, Peter Christen, Dimitrios Georgakopoulos: "*Context Aware Computing for The Internet of Things: A Survey*" - IEEE Communications Surveys & Tutorials (Volume. 16, issue 1, pp 414-454 – 2014).

15. Lupiana Dennis, O'Driscoll Ciaran, Mtenzi Fredrick: "*Defining Smart Space in the Context of Ubiquitous Computing*". Special Issue on ICIT 2009 Conference - Web and Agent System.

16. Li, S., Da Xu, L. and Zhao, S."*The internet of things: a survey*". Information Systems Frontiers (2015).

17. Madakam Somayya, Ramaswamy R., Tripathi Siddharth: "*Internet of Things (IoT): A Literature Review*" - Journal of Computer and Communications, 2015.

18. D. Cook and S. Das, "*Smart Environments: Technology, Protocols and Applications"* (Wiley Series on Parallel and Distributed Computing). Wiley - Interscience, 2004.

19. Ala Al-Fuqaha, Mohsen Guizani, Mehdi Mohammadi, Mohammed Aledhari, Moussa Ayyash; "*Internet of Things: A Survey on Enabling Technologies, Protocols and Applications*" - IEEE Communication Surveys & Tutorials, vol. 17, no. 4, 2015.

20. Monekosso, D., Florez-Revuelta, F., and Remagnino, P. (2015). "*Ambient assisted living*". IEEE Intelligent Systems.

21. M. Gersch, B. Lindert, and M. Hewing, "*AAL business models. Different prospects for the successfull implementation of innovative services in the first and second healthcare market*" in Proceedings of the ALLIANCE European Conference on AAL, Malaga, Spain, pp. 11-12, 2010.

22. Khusainov, R.; Azzi, D.; Achumba, I.; Bersch, S. "*Real-Time Human Ambulation, Activity, and Physiological Monitoring: Taxonomy of Issues, Techniques, Applications, Challenges and Limitations*". Sensors 2013, 13, 12852–12902.

23. Alam, M.R.; Reaz, M.B.I.; Ali, M.A.M. "*A Review of Smart Homes—Past, Present, and Future*". IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 2012, 42, 1190–1203.

24. Al-Shaqi, R.; Mourshed, M.; Rezgui, Y. "*Progress in ambient assisted systems for independent living by the elderly*". Springer Plus 2016.

25. Salih, A.S.M.; Abraham, A. "*A review of ambient intelligence assisted healthcare monitoring*". Int. J. Comput. Inf. Syst. Ind. Manag. Appl. 2013, 5, 741–750.

26. Ni Q., Hernando A.G., de la Cruz, I. "*The Elderly's Independent Living in Smart Homes: A Characterization of Activities and Sensing Infrastructure Survey to Facilitate Services Development*". Sensors 2015, 15, 11312–11362.

27. Peetoom, K.K.B.; Lexis, M.A.S.; Joore, M.; Dirksen, C.D.; de Witte, L.P. "*Literature review on monitoring technologies and their outcomes in independently living elderly people*". Disabil. Rehabil. Assist. Technol. 2014, 10, 271–294

28. Balta-Ozkan, N., Boteler, B., and Amerighi, O. (2014). "*European smart home market development: Public views on technical and economic aspects across the united kingdom, Germany and Italy*". Energy Research & Social Science, 3:65–77.

29. Harper, R. (2003). "*Inside the smart home: Ideas, possibilities and methods"* in Inside the smart home (pp. 1-13). Springer London.

30. Michele Nitti, Virginia Pilloni, Giuseppe Colistra, Luigi: *The Virtual Object as a Major Element of the Internet of Things: a Survey* - IEEE Communications Surveys & Tutorials (Volume. 18, issue 2 - 2016).

31. Jussi K., D'Elia A., Morandi F., Hyttinen P., Takalo-Mattila J., Ylisaukko-oja A., Soininen J.P. and Salmon Cinotti T.: "*Semantic Interoperability Architecture for Pervasive Computing and Internet of Things*". IEEE Access 2 (2014): 856-873.

32. Alegre, U., Augusto, J. C., & Clark, T. (2016). "*Engineering context-aware systems and applications: A survey*" in Journal of Systems and Software, 117, 55–83.

33. Rodrigo Roman, Jianying Zhou, Javier Lopez: "*Internet of Things: Survey on Security and Privacy*" – Elsevier Computer Networks (Volume 57, Issue 10 - 2013).

34. John A. Stankovic: *Research Directions for the Internet of Things*". IEEE Internet of Things Journal (Volume: 1 , Issue: 1 - 2014).

35. Sicari S., Rizzardi A., Grieco L.A., Coen-Porisini A.: “*Security, privacy and trust in Internet of Things: The road ahead”.* Elsevier Computer Networks (Volume 76 - 2015).

36. K. Zhao and L. Ge, "*A survey on the internet of things security*" in Computational Intelligence and Security (CIS), 2013 9th International Conference on, 2013, pp. 663-667.

37. Z. Ling, "*Research on application of Internet of Things in teaching practice of security technique defense*" in Internet of Things Technologies, vol. 1, p. 029, 2013.

38. I. Ali, S. Sabir, and Z. Ullah, "*Internet of Things Security, Device Authentication and Access Control: A Review*" in International Journal of Computer Science and Information Security, vol. 14, pages 456, 2016.

39. Casacuberta J. et al. “*Evaluation of an Inclusive Smart Home Technology System*” in J. Bravo, R. Hervás, and M. Rodríguez (Eds.): IWAAL 2012, LNCS 7657, pp. 316–319, 2012.

40. T. T. Zin, P. Tin and H. Hama, "*Reliability and availability measures for Internet of Things consumer world perspectives*" *2016 IEEE 5th Global Conference on Consumer Electronics*, Kyoto, 2016, pp. 1-2. doi: 10.1109/GCCE.2016.7800446

41. Vallet T. et al. “*On Personalization in IoT*” in Conference: 2016 International Conference on Computational Science and Computational Intelligence.

42. Hyun Lee, Kyungseo Park, Byoungyong Lee, Jaesung Choi, and Ramez El- masri. “*Issues in data fusion for healthcare monitoring*”. In International Con- ference on PErvasive Technologies Related to Assistive Environments, page 3, 2008.

43. Zilong Zou, Yuequan Bao, Fodan Deng, and Hui Li. “*An approach of reliable data transmission with random redundancy for wireless sensors in structural health monitoring*”. IEEE Sensors Journal, 15(2):809818, 2015.

44. Yang D., Xu B., Rao K. and Sheng W. “*Passive Infrared (PIR) – Based Indoor Position Tracking for Smart Home Using Accessibility Maps and A-Star Algorithm*”. Sensors 2018, 18,332 – doi: 10.3390/s18020332

45. Luo, R.C.; Chen, O. “*Wireless and pyroelectric sensory fusion system for indoor human/robot localization and monitoring*”. IEEE/ASME Trans. Mechatron. 2013, 18, 845–853.

46. Al-Naimi, I.; Wong, C.B.; Moore, P.; Chen, X. "*Advanced approach for indoor identification and tracking using smart floor and pyroelectric infrared sensors*" in Proceedings of the 5th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 1–3 April 2014; pp. 1–6

47. Song, B.; Choi, H.; Lee, H.S. "*Surveillance Tracking System Using Passive Infrared Motion Sensors in Wireless Sensor Network*". In Proceedings of the International Conference on Information Networking, Busan, Korea, 23–25 January 2008; pp. 1–5

48. Xiaomu, L.; Tong, L.; Baihua, S.; Qinqun, C.; Liwen, G.; Xiaoyan, L. "*Human Indoor Localization Based on Ceiling Mounted PIR Sensor Nodes*". In Proceedings of the 13th IEEE Annual Consumer Communications & Networking Conference(CCNC), Las Vegas, NV, USA, 9–12 January 2016; pp. 868–874.

49. Yun J. and Lee S.S. "*Human Movement Detection and Identification Using Pyroelectric Infrared Sensors*". Sensors 2014, 14, 8057-8081.

50. Hao, Q.; Hu, F.; Xiao, Y. "*Multiple human tracking and identification with wireless distributed pyroelectric sensor systems*". IEEE Sens. J. 2009, 3, 428–439.

51. Tao, S.; Kudo, M.; Nonaka, H.; Toyama, J. "*Person authentication and activities analysis in an office environment using a sensor network*". In Constructing Ambient Intelligence; Springer Berlin Heidelberg: Berlin, Germany, 2012; pp. 119–127.

52. Ariani, A.; Redmond, S.J.; Chang, D.; Lovell, N.H. "*Simulated Unobtrusive Falls Detection with Multiple Persons*". IEEE Trans. Biomed. Eng. 2012, 59, 3185–3196.

53. Luo X., Tan H., Guan Q., Liu T., Zhuo H. H. and Shen B. "*Abnormal Activity Detection Using Pyroelectric Infrared Sensors*". Sensors, 2016, 16, 822.

54. Xiong, J.; Li, F.; Zhao, N.; Jiang, N. "*Tracking and recognition of multiple human targets moving in a wireless pyroelectric infrared sensor network*". Sensors 2014, 14, 7209–7228.

55. Miyazaki, T.; Kasama, Y. "*Multiple human tracking using binary infrared sensors*". Sensors 2015, 15, 13459–13476.

56. Guan, Q.; Li, C.; Guo, X.; Wang, G. "*Compressive classification of human motion using pyroelectric infrared sensors*". Pattern Recognition Lett. 2014, 49, 231–237.

57. Nef Tobias et al. "*Evaluation of Three State-of-the-Art Classifiers for Recognition of Activities of Daily Living from Smart Home Ambient Data*". Sensors (Basel, Switzerland) vol. 15,5 11725-40. 2015.

58. Kelleher, J. D., Mac Namee, B., & D'Arcy, A. (2015). "*Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*". MIT Press.

59. Alshammari Talal, Nasser O. Alshammari, Mohamed Sedky and Chris Frank Howard: "*Evaluating Machine Learning Techniques for Activity Classification in Smart Home Environments*". International Journal of Information and Communication Engineering vol 12, no:2, 2018.

60. Witten I.H., Frank E. and Hall M.A.: "*Data Mining: Practical Machine Learning Tools and Techniques*". The Morgan Kaufmann Series in Data Management Systems. Elsevier Science, 2011.

61. Zhang Harry, 2004 "*The optimality of Naïve Bayes*" in American Association for Artificial Intelligence.

62. Sarkar, A. J., Lee, Y.-K., & Lee, S. (2010). "*A smoothed naive bayes-based classifier for activity recognition*". IETE Technical Review, 27(2), 107–119.

63. Gurneet Kaur et al, "*A review article on Naive Bayes classifier with various smoothing techniques*" in International Journal of Computer Science and Mobile Computing, Vol.3 Issue.10, October- 2014, pg. 864-868

64. Pouria Kaviani, Sunita Dhotre: "*Short Survey on Naive Bayes Algorithm* " in International Journal of Advance Engineering and Research Development Volume 4, Issue 11, 2017.

65. Mostafa Langarizadeh, Fateme Moghbeli: "*Applying Naive Bayesian Networks to Disease Prediction: a Systematic Review*". Acta Inform Med. 2016; 24 (5): pp 364 – 369.

66. Dhanashree S. Medhekar, Mayur P. Bote , Shruti D. Deshmukh: "*Heart Disease Prediction System using Naive Bayes*". International Journal of Enhanced Research in Science Technology & Engineering, vol. 2 issue 3, 2013.

67. A. M. Jehad Sarkar, Young-Koo Lee & Sungyoung Lee (2010) “*A Smoothed Naive Bayes-Based Classifier for Activity Recognition*” in IETE Technical Review, 27:2, 107-119.

68. L. Atallah, B. Lo, R. Ali, R. King and G. Yang, "*Real-Time Activity Classification Using Ambient and Wearable Sensors*" in IEEE Transactions on Information Technology in Biomedicine, vol. 13, no. 6, pp. 1031-1039, Nov. 2009. doi: 10.1109/TITB.2009.2028575

69. Fan, L., Wang, Z., & Wang, H. (2013). “*Human activity recognition model based on Decision Tree*” in Advanced Cloud and Big Data, 2013 International Conference pp. 64–68.

70. Li-Min Wang, Xiao-Lin Li, Chun-Hong Cao, Sen-Miao Yuan: “*Combining decision tree and Naive Bayes for classification*” in Knowledge-Based Systems, Volume 19, Issue 7, 2006, Pages 511-515.

71. Y. Ling and H. Wang, "*Unsupervised Human Activity Segmentation Applying Smartphone Sensor for Healthcare*" pp. 1730-1734, 2015. doi: 10.1109/UIC-ATC-ScalCom-CBDCom-IoP.2015.314

72. Ling Bao and Stephen S. Intille. “*Activity recognition from user-annotated acceleration data*” in Pervasive computing, pages 1-17. Springer, 2004.

73. Lei Gao, A.K. Bourke, and John Nelson. “*Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems*”. Medical Engineering and Physics, 36(6):779-785, 2014.

74. Sanchez Veralia and Skeie Nils-Olav. (2018). “*Decision Trees for Human activity Recognition in Smart House Environments*”. Conference Paper, pages 222-229. http://doi.org/10.3384/ecp18153222.

75. A. Ariani, S. J. Redmond, D. Chang and N. H. Lovell, "*Software simulation of unobtrusive falls detection at night-time using passive infrared and pressure mat sensors*". 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, Buenos Aires, 2010, pages 2115-2118.

76. Jaeseok Yun, Sang-Shin Lee: “*Human Movement Detection and Identification Using Pyroelectric Infrared Sensors*”. Sensors **2014**, 14(5), 8057-8081.

77. Xu Z., Wang G., Guo X. (2019) “*Comparative Studies on Activity Recognition of Elderly People Living Alone*”. In Jia Y., Du J., Zhang W. (eds) Proceedings of 2019 Chinese Intelligent Systems Conference. CISC 2019. Lecture Notes in Electrical Engineering, vol 592.

78. [38] Melo Nicholas and Jaeryoung Lee. 2018. “*Environment aware ADL recognition system based on decision tree and activity frame*”. Paladyn, Journal of Behavioral Robotics. 9(1): 155-167. Retrieved 30 Sep. 2019.

79. Okfalisa, I. Gazalba, Mustakim and N. G. I. Reza, "*Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification*". 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, 2017, pp. 294-298. doi: 10.1109/ICITISEE.2017.8285514

80. R. Mynavathi, V. Bhuvaneswari, T. Karthikeyan and C. Kavina, "*K nearest neighbor classifier over secured perturbed data*". 2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave), Coimbatore, 2016, pages 1-4.

81. A. R. Winnersyah, F. Fahrianto and N. Anggraini, "*Identification and Position Estimation Method with K-Nearest Neighbour and Home Occupants Activity Pattern*". 2018 6th International Conference on Cyber and IT Service Management (CITSM), Parapat, Indonesia, 2018, pages 1-4.

82. A. Belay Adege, Y. Yayeh, G. Berie, H. Lin, L. Yen and Y. R. Li, "*Indoor localization using K-nearest neighbor and artificial neural network back propagation algorithms*". 2018 27th Wireless and Optical Communication Conference (WOCC), Hualien, 2018, pages 1-2.

83. Hu, Y.-J., Yu, M.-C., Wang, H.-A., & Ting, Z.-Y. (2015). “*A similarity-based learning algorithm using distance transformation*”. IEEE Transactions on Knowledge and Data Engineering, 27 (6), pages 1452–1464.

84. Padraig Cunningham and Delany Sarah Jane: “*k-Nearest neighbour classifiers*”. Technical Report UCD-CSI 2007.

85. Kaghyan, S., & Sarukhanyan, H. (2013). “*Accelerometer and gps sensor combination based system for human activity recognition*”. In Computer Science and Information Technologies, 2013 pages 1–9.

86. Paul, P., & George, T. (2015). *An effective approach for human activity recognition on smartphone*”. In Engineering and technology (icetech), 2015 IEEE International Conference pages. 1–3.

87. P. Vallabh, R. Malekian, N. Ye and D. C. Bogatinoska, "*Fall detection using machine learning algorithms*". 2016 24th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, 2016, pages. 1-9.

88. Wang, Z., Gao, J., Chen, R. & Wang, J.Y.: “*A modified KNN algorithm for activity recognition in smart home*”. In Proceedings of The 18th International Conference on Electronic Business, pages 53-61. 2018.

89. Farlex (2018). Activities of Daily Living definition.

90. Cook, D. J. and Krishnan, N. C. (2015). *Activity Learning: Discovering, Recognizing, and Predicting Human Behavior from Sensor Data*. John Wiley & Sons.

91. Chaaraoui, A. A., Climent-Pérez, P., and Flórez-Revuelta, F. (2012) “*A review on vision techniques applied to human behaviour analysis for ambient-assisted living*”. Expert Systems with Applications, 39(12): pages 10873–10888.

92. Fleury, A., Vacher, M., and Noury, N. (2010). “*Svm-based multimodal classification of activities of daily living in health smart homes: sensors, algorithms, and first experimental results*”. Information Technology in Biomedicine, IEEE Transactions on, 14(2): pages 274–283.

93. Lawton, M. P. and Brody, E. M. (1969). “*Assessment of older people: self-maintaining and instrumental activities of daily living*”. The gerontologist, 9 (3_Part_1):179–186.

94. Kim, E., Helal, S., and Cook, D. (2010). “*Human activity recognition and pattern discovery*”. Pervasive Computing, IEEE, 9(1): pages 48–53.

95. Lara, O. D. and Labrador, M. A. (2013). “*A survey on human activity recognition using wearable sensors*”. Communications Surveys & Tutorials, IEEE, 15(3): pages 1192–1209.

96. Mahmoud, S., Lotfi, A., and Langensiepen, C. (2013). "*Behavioural pattern identification and prediction in intelligent environments*". Applied Soft Computing, 13(4): pages 1813–1822.

97. Forkan, A. R. M., Khalil, I., Tari, Z., Foufou, S., and Bouras, A. (2015). "*A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient assisted living"*. Pattern Recognition, 48(3): pages 628–641.

98. Cook, D. J., Krishnan, N. C., and Rashidi, P. (2013). "*Activity discovery and activity recognition: A new partnership*". IEEE transactions on cybernetics, 43(3): pages 820–828.

99. Forkan, A., Khalil, I., and Tari, Z. (2014). "*Cocamaal: A cloud-oriented context-aware middleware in ambient assisted living*". Future Generation Computer Systems, 35: pages 114–127.

100. Saives, J., Pianon, C., and Faraut, G. (2015). "*Activity discovery and detection of behavioral deviations of an inhabitant from binary sensors*". Automation Science and Engineering, IEEE Transactions on, 12(4): pages 1211–1224.

101. Tapia, E. M., Intille, S. S., and Larson, K. (2004). *Activity recognition in the home using simple and ubiquitous sensors*. Springer.

102. Gaglio, S., Re, G. L., and Morana, M. (2015). "*Human activity recognition process using 3-D posture data*". IEEE Transactions on Human-Machine Systems, 45(5): pages 586–597.

103. Mihailidis, A., Carmichael, B., and Boger, J. (2004). "*The use of computer vision in an intelligent environment to support aging-in-place, safety, and independence in the home*". IEEE Transactions on information technology in biomedicine, 8(3): pages 238–247.

104. Wu, B. and Nevatia, R. (2005). "*Detection of multiple, partially occluded humans in a single image by bayesian combination of edge let part detectors*". In Computer Vision, ICCV 2005. Tenth IEEE International Conference on, volume 1, pages 90–97.

105. Zhou, Z., Chen, X., Chung, Y.-C., He, Z., Han, T. X., and Keller, J. M. (2008). "*Activity analysis, summarization, and visualization for indoor human activity monitoring*". IEEE Transactions on Circuits and Systems for Video Technology, 18(11): pages 1489–1498.

106. Van Kasteren, T., Noulas, A., Englebienne, G., and Kröse, B. (2008). "*Accurate activity recognition in a home setting*". In Proceedings of the 10th international conference on Ubiquitous computing, pages 1–9. ACM.

107. Kellokumpu, V., Pietikäinen, M., and Heikkilä, J. (2005). "*Human activity recognition using sequences of postures*". In MVA, pages 570–573.

108. Synnott, J., Nugent, C., & Jeffers, P. (2015). Simulation of smart home activity datasets. *Sensors*, *15*(6), 14162-14179.

109. Mari, L., Carbone, P., & Petri, D. (2012). Measurement fundamentals: a pragmatic view. *IEEE transactions on instrumentation and measurement*, *61*(8), 2107-2115.

110. Jon Froehlich, Eric Larson, Elliot Saba, Tim Campbell, Les Atlas, James Fogarty, and Shwetak Patel. *A longitudinal study of pressure sensing to infer real-world water usage events in the home*. In International Conference on Pervasive Computing, pages 50–69. Springer, 2011.

111. Jianxin Wu, Adebola Osuntogun, Tanzeem Choudhury, Matthai Philipose,and James M Rehg. *A scalable approach to activity recognition based on object use.* In Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, pages 1–8. IEEE, 2007.

112. Philipose, M., Fishkin, K. P., Perkowitz, M., Patterson, D. J., Fox, D., Kautz, H., & Hahnel, D. (2004). Inferring activities from interactions with objects. *IEEE pervasive computing*, (4), 50-57.

113. Patterson, D. J., Fox, D., Kautz, H., & Philipose, M. (2005, October). Fine-grained activity recognition by aggregating abstract object usage. In *Ninth IEEE International Symposium on Wearable Computers (ISWC'05)* (pp. 44-51). IEEE.

114. Renoux, J., & Klugl, F. (2018, December). Simulating daily activities in a smart home for data generation. In *2018 Winter Simulation Conference (WSC)* (pp. 798-809). IEEE.

115. Demiris, G., Hensel, B. K., Skubic, M., & Rantz, M. (2008). Senior residents' perceived need of and preferences for "smart home" sensor technologies. *International journal of technology assessment in health care*, *24*(1), 120-124.

116. Clawson, J., Pater, J. A., Miller, A. D., Mynatt, E. D., & Mamykina, L. (2015, September). No longer wearing: investigating the abandonment of personal health-

tracking technologies on craigslist. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 647-658). ACM.

117. Riboni, D., Bettini, C., Civitarese, G., Janjua, Z. H., & Bulgari, V. (2015, March). From lab to life: Fine-grained behavior monitoring in the elderly's home. In *2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)* (pp. 342-347). IEEE.

118. Bouchard, K., Ajroud, A., Bouchard, B., & Bouzouane, A. (2010). SIMACT: a 3D open source smart home simulator for activity recognition. In *Advances in Computer Science and Information Technology* (pp. 524-533). Springer, Berlin, Heidelberg.

119. Mendez-Vazquez, A., Helal, A., & Cook, D. (2009, April). Simulating events to generate synthetic data for pervasive spaces. In *Workshop on Developing Shared Home Behavior Datasets to Advance HCI and Ubiquitous Computing Research*. Citeseer.

120. Helal, S., Lee, J. W., Hossain, S., Kim, E., Hagras, H., & Cook, D. (2011, July). Persim-Simulator for human activities in pervasive spaces. In *2011 Seventh International Conference on Intelligent Environments* (pp. 192-199). IEEE.

121. Kormányos, B., & Pataki, B. (2013, July). Multilevel simulation of daily activities: Why and how?. In *2013 IEEE international conference on computational intelligence and virtual environments for measurement systems and applications (CIVEMSA)* (pp. 1-6). IEEE.

122. Lertlakkhanakul, J., Choi, J. W., & Kim, M. Y. (2008). Building data model and simulation platform for spatial interaction management in smart home. *Automation in Construction*, *17*(8), 948-957.

123. Fu, Q., Li, P., Chen, C., Qi, L., Lu, Y., & Yu, C. (2011, October). A configurable context-aware simulator for smart home systems. In *2011 6th International Conference on Pervasive Computing and Applications* (pp. 39-44). IEEE.

124. Stahl, C., & Schwartz, T. (2010, September). Modeling and simulating assistive environments in 3-D with the YAMAMOTO toolkit. In *2010 International Conference on Indoor Positioning and Indoor Navigation* (pp. 1-6). IEEE.

125. Armac, I.; Retkowitz, D. *Simulation of Smart Environments*. In Proceedings of the IEEE International Conference on Pervasive Services, Istanbul, Turkey, 15–20 July 2007; pp. 257–266.

126. Trimble Trimble SketchUp. Available online: http://www.sketchup.com/ (accessed on 12 January 2015).

127. Nazari Shirehjini, A. A., & Klar, F. (2005, October). 3DSim: rapid prototyping ambient intelligence. In *Proceedings of the 2005 joint conference on Smart objects and ambient intelligence: innovative context-aware services: usages and technologies* (pp. 303-307). ACM.

128. Park, J., Moon, M., Hwang, S., & Yeom, K. (2007, August). CASS: A context-aware simulation system for smart home. In *5th ACIS International Conference on Software Engineering Research, Management & Applications (SERA 2007)* (pp. 461-467). IEEE.

129. O'Neill, E., Klepal, M., Lewis, D., O'Donnell, T., O'Sullivan, D., & Pesch, D. (2005, February). A testbed for evaluating human interaction with ubiquitous computing environments. In *First International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities* (pp. 60-69). IEEE.

130. Buchmayr, M., Kurschl, W., & Küng, J. (2011). A simulator for generating and visualizing sensor data for ambient intelligence environments. *Procedia Computer Science*, *5*, 90-97.

131. Poland, M. P., Nugent, C. D., Wang, H., & Chen, L. (2009). Development of a smart home simulator for use as a heuristic tool for management of sensor distribution. *Technology and Health Care*, *17*(3), 171-182.

132. McGlinn, K., O'Neill, E., Gibney, A., O'Sullivan, D., & Lewis, D. (2010). SimCon: A Tool to Support Rapid Evaluation of Smart Building Application Design using Context Simulation and Virtual Reality. *J. UCS*, *16*(15), 1992-2018.

133. Krzyska, C. Smart House Simulation Tool. Master’s Thesis, Technical University of Denmark, Copenhagen, Denmark, 2006.

134. Ariani, A., Redmond, S. J., Chang, D., & Lovell, N. H. (2013, November). Simulation of a smart home environment. In *2013 3rd International Conference on Instrumentation, Communications, Information Technology and Biomedical Engineering (ICICI-BME)* (pp. 27-32). IEEE.

135. Synnott, J., Chen, L., Nugent, C. D., & Moore, G. (2014, August). The creation of simulated activity datasets using a graphical intelligent environment simulation tool. In

2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 4143-4146). IEEE.

136. Casaccia, S., Revel, G. M., Scalise, L., Bevilacqua, R., Rossi, L., Paauwe, R. A., ... & Lukkien, D. (2019, October). Social Robot and Sensor Network in Support of Activity of Daily Living for People with Dementia. In *Dementia Lab Conference* (pp. 128-135). Springer, Cham.

137. Martin B, Hanington BM (2012) Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions. Rockport Publishers, Beverly, MA