UNIVERSITÀ
DEGLI STUDI
DI CAGLIARI

Ph.D. DEGREE IN

Electronic and Computer Engineering

Cycle XXXV

TITLE OF THE Ph.D. THESIS

Sensors and Wearable-based Activity Recognition and Behaviour Analysis for Users' Quality of Life Improvement

Scientific Disciplinary Sector(s)

ING-INF/03

Ph.D. Student: Francesca Marcello

Supervisor Prof. Luigi Atzori
Co-tutor PhD Virginia Pilloni

Final exam. Academic Year 2021/2022
Thesis defence: February 2023 Session

University of Cagliari

Department of Electrical and Electronic Engineering
PhD Course in Electronic and Computer Engineering
Cycle XXXV

Ph.D. Thesis

Sensors and Wearable-based Activity Recognition and Behaviour Analysis for Users' Quality of Life Improvement

S.S.D. ING-INF/03

Candidate
Francesca Marcello

PhD Supervisor
Prof. Luigi Atzori
PhD Virginia Pilloni

PhD Coordinator
Prof. Alessandro Giua

Final examination academic year 2021/2022

Francesca Marcello gratefully acknowledges the Sardinian Regional Government for the financial support of her PhD scholarship (P.O.R. Sardegna F.S.E. - Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2014-2020 - Axis III Education and training, Thematic goal 10, Investment Priority 10ii), Specific goal 10.5".

UNIONE EUROPEA
Fondo sociale europeo

REGIONE AUTONOMA DELLA SARDEGNA

FSE 2014-2020 POR SARDEGNA

Contents

Acronyms

AI Artificial Intelligence

ANN Artificial Neural Network

APIs Application Programming Interfaces

BECM Building Energy and Comfort Management

CNN Convolutional Neural Network

DL Deep Learning

DSM Demand-Side Management

EDA electrodermal activity

GHQ General Health Questionnaire

HAR Human Activity Recognition

HMM Hidden Markov Model

HRV heart rate variability

IoT Internet of Things

JSONs JavaScript Object Notation

LSTM Long-Short-Term-Memory

ML Machine Learning

NZEB Net-Zero Energy Buildings

PSS Perceived Stress Scale

PV photovoltaic

QoE Quality of Experience

RES Renewable Energy Sources

REST Representational State Transfer

RF Radio-Frequency

RFID Radio-frequency Identification

RNN Recurrent Neural Network

SVM Support Vector Machines

WHO World Health Organization

XML eXtensible Markup Language

Abstract

The focus of the thesis is on sensors and wearable-based activity recognition and behaviour analysis for Users' quality of life improvement. The topic has been addressed considering two different fields of study, one concerning Building Energy and Comfort Management (BECM) Systems while the other is more concerning healthcare and well-being aspects. Indeed, it has been largely demonstrated how human behaviour could impact, on different levels, the quality of life. Some behaviours directly affect individuals' life, not only considering their psycho-physical health state but also as a result of the big impact that human behaviour has even on Earth's life.

Therefore, the main idea of this thesis is to analyse and monitor how human habits can have an impact on people's quality of life, inside their homes and also during their everyday activities outside the home. With this scope in mind, this thesis presents three different systems and prototypes that, thanks to different kind of sensors and devices, can understand, recognise, and learn users' habits and act accordingly to their preferences. People often undertake incorrect behaviour without realising it, so the final purpose of these systems is to provide information and advice to users in order to make them aware and correct some of these wrong behaviours or encourage positive ones.

The first system is based on a sensor network inside a Smart Home that recognises and predicts all the activities that are going on inside the home and, considering energy prices and user preferences, can schedule the appliances of the home in order to guarantee energy savings while at the same time ensure the comfort of the inhabitants.

The second system is an appliance power profiling system that analyses the power consumption data collected by smart meters, identifies which features are most relevant for the specific appliance and extracts the set of power consumption profiles that are associated with each appliance. This kind of solution is necessary so that the Smart Home system can learn and know the power consumption profiles over time of the appliances present in the building.

Finally, the third system is concentrates more on understanding how some activities, especially those related to fit activities and sleeping habits, may be reflected in how the user evaluate their own well-being and stress levels. This system makes use of simple data collected thanks to commercial wristbands and the study is fo-

cus on find a correlation between this activity data and the score obtained from a self-evaluation questionnaire about the well-being as users themselves perceive it.

The results obtained and illustrated in this thesis show how a more conscious use of energy inside buildings, achieved thanks to accurate scheduling of household appliances, can guarantee energy savings while considering the user's comfort and preferences. The study on appliance profiling made it possible to recognise different usage profiles relating to specific appliances in order to have different consumption profiles, each represented by a reference consumption profile that can be later used for making consumption predictions or for giving advice to users. The system based on common and commercial smart objects of daily use shows that a strong correlation can be found between simple data obtained from these devices about daily activities and the level of stress and well-being of users.

Introduction

0.1 Purpose

Human behaviour can affect, on different levels, the quality of life. Some examples of this statement can be found in data related to energy consumption inside buildings, that estimate how this kind of consumption constitutes more than 30% of the global energy consumption, with the 24% referring to the domestic environment and the 8% referred to public buildings [8]. Furthermore, studies on developing countries report an average annual increase in consumption of 2.2% over the last decade [8]. Apart from the economic savings that can be obtained thanks to more accurate energy management inside buildings, the discussion on the impact that such consumption has on the environment is also of great importance. Another known behaviour that can be shows as an example and that has negative impacts on environmental and human health is the overconsumption of meat, which seems to be associated with the risk of several diseases and it is a major source of greenhouse gases [36].

Furthermore, many research works are focused on better understanding how some behaviours, or changes in someone's habits, can be triggers for recognising particular illness symptoms like depressive disorder or Alzheimer's Disease [63] [62]; especially for elderly people, many works propose assistance and healthcare systems or early recognition systems for degenerative diseases to ensure a better quality of life [11] [51]. These are just a few examples, but there are many other habits and behaviours for which a correlation with psycho-physical illness has been found. People often keep on carrying on these harmful habits simply because they are not fully aware of them or because they do not know which would be a better alternative.

Therefore, monitoring, learning and understanding human habits and how they influence one person's life, is very important and useful in order to provide users with tools that can make them aware of wrong behaviours and give them some kind of suggestions to discourage these habits and consolidate virtuous ones.

The main objective of the research project proposed here is the development of a solution that knowing users' daily habits is able to define which activities have positive or negative impacts on users' quality of life and advise them to undertake one of the positive activities when a bad emotional state is encountered. To this aim, it is necessary to develop systems that using environmental monitoring sensors (e.g. motion, door, temperature, and pressure sensors) and wearable and portable devices (smartphones, smartwatches or other devices attached to the user's body provided with accelerometer, gyroscope, heart rate sensor) are able to monitor users, identify harmful and virtuous habits or abnormal behaviours, create users' profile and provide them with relevant suggestions in order to improve their quality of life, both in the short and long term. This kind of system should learn the tasks and the activities that users perform during the day and how they can impact their lives.

The subject of this thesis concerns Human Activity Recognition and Prediction. In particular, the research conducted so far has been focused on monitoring the activities carried out by users inside their homes, considering a sensor-based Smart Home scenario, and studying their habits to propose an energy efficient system. The system considers energy consumption and the user's comfort and preferences to find the best scheduling program for household appliances. More precisely, the scheduling is evaluated considering time-of-use tariffs (also dynamic) and Renewable Energy Sources (RES). Another aspect of this research work is focused on establishing whether these daily activities could be monitored using the most modern smart objects of daily use (smartphones, smartwatches, smartbands) and integrating them inside the same home scenario based on ambient sensors. User profiles will then be more accurate thanks to this new information and this data will be used not only in the energy consumption field but also to establish whether there is a correlation between individual activities and particular stress levels or emotional states that have effects on users' psycho-physical health.

The ultimate goal of this thesis is to propose an initial analysis of different solutions and methodologies that can lead to the creation of a system, as transparent and non-intrusive as possible for users, which can monitor them in their daily life, analysing their habits and behaviours, and which promotes the development of good practices that have positive effects both on their physical and mental health and the surrounding environment as a response to better use of energy sources and consumption. In order to do this, the first step is to decide how to monitor users, recognise certain activities and predict their behaviour. The creation of the first system proposed in this thesis responds to this need, suggesting a system based only on environmental sensors and particularly suitable for users less inclined to particular technologies and continuous monitoring. The proposed approach recognises the activities carried out every moment and can also predict future activities based on the user's previous knowledge and preferences. In this way, it is able to propose a scheduling program for controllable home appliances based on the user's habits, and that can also guarantee the proper use of energy, oriented towards saving and based on renewable resources, whenever it is possible to use them. A second point, which is strictly connected to the first, has also been addressed in the research work. In order to better characterise users' habits relating to the use of household appliances, the system needs to establish different consumption profiles connected to the devices. Here the primary purpose was to obtain consumption patterns directly linked to individual appliances to have differentiated profiles, which can later be easily associated with specific activities relating precisely to the use of these particular devices. Finally, the third idea comes from the willingness to answer the question of whether the use of simple commercial wearable devices could document certain habits and their effects on the psycho-physical well-being of users. With a view of not wanting to alter people's habits and wanting to realise systems that can be adapted to everyone, it was decided to investigate the usage of commercial smart bands and smartwatches, already owned by them, to monitor their activities. The data obtained from these subjects have been analysed to find correlations between certain habits and their state of stress and well-being, as perceived by themselves. The ultimate goal will be to create a unified system that can monitor users at various stages of the day, at home, thanks to the network of sensors and commercial wearable devices, and outside the house through the same wearable, and can guarantee more accurate profiles. Thanks to this added information, users' habits will be more exhaustive and can be more easily linked with information about their health. The users' preferences and profiles obtained from the home system could change automatically according to the users' inferred stress level or well-being state. In addition, by recognising from the users' attitudes and their vital signals detected from wearable if they are in a particular emotional state and knowing from their preferences what is the most suitable solution to improve this state, the system could act on the surrounding environment by intervening appropriately to improve the situation. For example, if the user is sad and the system detects this emotional state, it could turn on a particular kind of music that it is known, from the user's habits, that it makes the user feel better.

0.2 Structure of the Document

The remainder of this thesis is organised as follows: chapter 1 gives a brief description of Home Automation and Smart Buildings and an overview of some key aspects of Human Activity Recognition and Prediction systems, reviewing different applications, sensors and Machine Learning (ML) and Deep Learning (DL) approaches useful for the discussion. Section 1.4 describes the State of Art related to BECM systems, Section 1.5 illustrates some solutions for the problem of appliance profiling and describes clustering methods, while Section 1.6 addresses the descriptions of solutions for health monitoring systems based on wearable devices.

Chapter 2 presents a BECM system that, thanks to a sensors network placed inside a Smart Home, is able to learn and understand the user's daily activities and predict future activities and energy consumption of the appliances related to these activities in order to make a scheduling of the appliances that can guaranteed energy savings and at the same time take into account user's comfort thanks to the knowledge of their preferences and habits.

Chapter 3 describes a system that identifies different appliance profiles based on power consumption data gathered by smart meters. These profiles help obtain information on different consumption cycles related to the user's habits on the usage of a specific appliance so that the system can later present this information directly to the user or use it to make a more accurate evaluation of the appliances' scheduling.

Chapter 4 proposes a system that makes use of popular commercial wrist-wearable devices to find the correlation between the monitored users' activities and their stress and well-being conditions, as subjectively self-assessed by them. This Section aims to present a methodology to automatically learn which users' activities can be associated with positive and negative health conditions so that they can be later predicted as soon as the first signals are detected by wearable devices.

In chapter 5, there is a summary of the results of this thesis, with the conclusions and an outlook of possible future works.

Chapter 1

Background and State of the Art

1.1 Home Automation

The term home automation indicates an interdisciplinary science that involves engineering, information technology, safety and design and that deals with the study of technologies applied to homes or, more generally, to the buildings in which the main daily activities of a person take place, with the intent to improve the quality of their life [95]. Therefore, a home automation system needs to be able to integrate the traditional aspects of a building with new and advanced technologies, with the aim of creating a cooperative environment where it is possible to optimise the management of the building itself, improving its safety, comfort and energy efficiency. Hence, the most important aspect of this kind of system lies in the possibility that all the devices involved can communicate with each other, exchanging data and information relating to the context in which they are located.

Initially, the term home automation referred only to the field of Smart Homes. In general, a Smart Home can be defined as a home environment equipped with technologies that allow inhabitants to program and control the electrical devices to facilitate the carrying out of many activities (for example, switching on the lights, activating certain household appliances, checking the heating and air conditioning systems, etc.), to ensure safety (anti-intrusion, fire, gas leakage systems, etc.), to allow remote connection for ensuring assistance services (such as remote assistance, etc.) and energy monitoring to obtain the best levels of efficiency and savings without affecting the comfort expectations of the inhabitants [7], [97], [84]. Nowadays, however, all aspects related to the management and monitoring of buildings other than just homes, such as hospitals, schools, offices and companies, fall within the scope of home automation. The term Building Automation or Smart Building refers to this type of environment in which a large part of people's daily activities take place, where technological systems, computer networks and communication networks are managed in a coordinated and integrated way to pursue the same objectives of comfort, safety and energy efficiency, in order to improve the quality of living and working within them [94],[21].

To obtain this kind of solution, the systems must be able to perform partially autonomous functions in response to changes in the surrounding environment or based on specific user preferences, obtained as direct information or by self-learning. However, the intelligence of a Smart Home does not reside in the intelligence of the individual devices that compose it, but it is the central control that behaves intelligently, evaluating how to react with respect to the information that comes from all around.

The central aspect is that a home automation system, although it was created to improve the life that takes place inside it, is not and cannot be a closed system. Instead, it is open to other systems and the outside world. Possible users' remote management must be guaranteed, by completing the system with one or more communication systems to the outside world through, for example, warning messages, emails, web pages and applications.

1.2 Internet of Things

The interoperability between different elements and the remote management of buildings are made possible thanks to the Internet of Things (IoT), which refers to a global network composed of heterogeneous objects that collaborate and make their resources available to achieve a common goal [5]. In this kind of ubiquitous network, all objects are identifiable and acquire intelligence thanks to the exchange of information interposed between them. They are objects that take on identity and personality, able to change their behaviour based on information received from the outside and which, thanks to the network connection, can therefore assume an active role. The definition of "things" or "objects" includes elements of different categories, such as devices, equipment, tangible products and materials, and machines. All these real objects also acquire an electronic and virtual identity thanks to technologies such as Radio-frequency Identification (RFID) [91], or QR codes [71] with the possibility in this way to communicate with mobile devices such as smartphone and tablet.

Since the advent of the most advanced IoT, a series of connected equipment and devices have been included in home automation. Thanks to integrated and advanced programming, new services and opportunities with increasingly automated and intelligent functions have been included inside homes and buildings. It is clear that the IoT is a central paradigm for creating and distributing home automation solutions and is increasingly present in users' homes. In the last decade, the Smart Home has been able to experience actual exponential growth thanks to the digital evolution of recent years: houses have become truly intelligent thanks to the IoT, mobile technologies, the cloud and Artificial Intelligence (AI).

The first step for realising this complex system can be defined pre-IoT and includes the "simple" sensor network. Sensors are devices capable of obtaining information concerning the environment surrounding them according to specific application areas. Some sensors deal, for example, with the collection of data related to temperature and humidity, noise level, the presence and movement of people, and air quality. The data they perceive is then translated into a digital signal that can be measured and analysed. The actual IoT transition really occurs only when this information is shared on the network. Then a new series of phases follow this step, referring to the manipulations that are done on the data, in which the devices connected to the network are able:

1. to collect and communicate data,
2. to detect and transfer several types of different data,
3. to select data locally to transfer only those of interest that correspond to certain requirements,
4. to collect data, select it and carry out certain actions based on the indications received,
5. to collect and select data and transmit only those necessary for the projects in which they are involved, carry out actions based on indications received and carry out actions according to a local processing capacity.

It can be highlighted how the IoT is nothing more than the natural passage of the Internet from a connected network of end-user devices to a direct interconnection network of physical objects capable of communicating and cooperating without human mediation. Therefore, these connected and communicating smart objects give life to valuable and exciting solutions but can also lead to vulnerability problems that derive from putting this information into the network. The significant aspects to consider are privacy or cyber-attacks by malicious people, which would put information about people's habits and preferences or the management and control of a company at risk.

1.3 Human Activity Recognition and Prediction

The field of Human Activity Recognition (HAR) has become one of the main research topics in the last years due to the applicability of HAR solutions to many different disciplines and also due to the fact that all kinds of smart devices and sensors are increasingly common and available in everyday life. Many possible applications exist in different fields, such as health, security and surveillance, entertainment, and intelligent environments. Table 1.1 gives an overview of these applications, with some examples of the fields of use in which the HAR study is configured.

In order to create a system capable of recognising the activities performed by users, a complex process must be carried out, characterised by various tasks that could be summarised in the following points:

Categories	**Use Fields**
Security & Surveilleance	Airports, train stations, banks, parking monitoring systems
Smart Environments	Smart Home, Smart Offices, hospitals, meeting rooms, classrooms, smart automotive
Health Care	Monitoring of physical activities, living assistance, health condition assessment
Military Field	Monitoring of soldiers in battlefields
Pervasive Computing & Mobile Computing	Context-aware systems, content distribution service systems, energy efficiency

Table 1.1: Overview of Application Fields for HAR Systems

- choosing sensors and devices and their disposition;
- collect and elaborate all the data obtained from the sensors;
- create models for the activities that make the system able to do manipulation and elaboration on the data with some kind of awareness;
- choosing and implementing algorithms that can deduce the activities from sensors' readings.

Various technologies and methods can be used as a solution depending on different systems and problems; therefore, HAR can be classified according to the different choices that can be made to manage the solutions.

The first classification depends on the type of technologies and sensors that are chosen; therefore, it is possible to identify a division in vision-based or sensor-based activity recognition. In the first case, the solutions involve the use of cameras to monitor the performance of the activities and hence the data that will be analysed are images and video sequences. In this case, different computer vision techniques, such as extraction of features, segmentation of movements and tracking of actions, are used to analyse the acquired data.

Instead, in the second case, the solutions make use of a distributed sensor network for monitoring the performance of the activities. The data analysed in this case are temporal sequences of events linked to sensor state changes and/or values relating to certain parameters monitored by a specific sensor. Within this solution, there are then two other sub-categories: one relating to sensors that must be physically worn by a person to keep track of his movements (wearables, and it is also possible to use a smartphone or smartwatch for this purpose), the other relating to a sensor network arranged all around the observation environment. These sensors should be placed in strategic points or attached to specific objects to monitor the interactions of a person with the surrounding environment.

The second classification concerns the way in which the model for activity recognition is created; therefore, there are data-driven activity recognition models that are created from the analysis of the acquired data, or knowledge-driven activity recognition models, that firstly consider common sense and logic and then adapt the models to the acquired data. The first method uses the data of a pre-existing dataset and applies data extraction and ML techniques to understand the activities performed by the user and create probabilistic or statistical classification models. The second method, on the other hand, builds conceptual models of the activities of interest by making use of semantic and ontologies techniques.

The difference between the possible sensor choices basically concerns three aspects:

- the type of data that is obtained and how it is analysed, video cameras and wearable give high and medium level information while binary sensors provide lower level information;
- the level of intrusiveness, therefore, in terms of how much they can infiltrate user's privacy;
- the position with respect to the user himself.

Table 1.2 highlights these different characteristics by sensor type.

Sensor	**Position relative to Users**	**Data Semantic Level**	**Intrusiveness**
Body Sensors	Sensing on user	High	Very intrusive
Cameras	External sensing	High	Very intrusive
Wearable	Hybrid	Medium	Mildly intrusive
Microphones	External sensing	Medium	Intrusive
Motion Sensor	External sensing	Low	Non-intrusive
Magnetic Switches	External sensing	Low	Non-intrusive
Vibration Sensors	External sensing	Low	Non-intrusive
Force Sensors	External sensing	Low	Non-intrusive
Flow Sensors	External sensing	Low	Non-intrusive
Temperature, Humidity & Light Sensors	External sensing	Low	Non-intrusive

Table 1.2: Characteristics of some common sensors

The data collected from sensors are analysed using data mining and machine learning techniques to build activity models that are used as the basis of behavioural activity recognition.

With reference to modelling and classification methods, researchers have investigated the recognition of the activities using a variety of mechanisms, such as Naïve Bayes classifiers, Markov models or Support Vector Machines (SVM). Recently many studies have been addressing DL methods to recognise different types of human activities or to make predictions about their behaviour. Long-Short-Term-Memory (LSTM) Network, that are a special kind of Recurrent Neural Network (RNN), are largely used, considering home scenarios, both to the forecasting of power consumption [2] [50], and to make predictions on future activities performed by residents [24] [83].

In the next Subsections it is possible to find more details about some of these ML and DL techniques, especially those related to the rest of this discussion.

1.3.1 Machine learning

ML is a branch of AI and Data Science that, thanks to particular techniques and algorithms, can make systems capable of automatically learning, predicting and improving their performance on specific tasks over time, without or with very little human intervention, through some type of training experience. The ML algorithms are designed to emulate human intelligence by learning from the surrounding environment. Starting from a given volume of empirical or historical input data from a dataset that constitute the samples of the problem, ML algorithms can give a specific output value that aims to categorise new samples or to measure the distance from desirable outputs, thank to statistical analysis on the training data given as the input. Therefore, ML algorithms allow the systems to make decisions automatically based on received inputs.

In ML, tasks are generally classified into broader categories. These categories are based on the way in which the learning phase is approached or on how the feedback on learning is passed to the developed system. The following four different methods can therefore be identified [73]:

- supervised learning: the inputs are given to the systems as a tuple (f_i, t_i), where f_i is the set of features that described the sample i and t_i represents the target output associated with that sample. The learning algorithm is trained on this dataset and will subsequently be able to define other inputs f_i with an unknown target based on the similarities with the training data and assign the most appropriate tag output accordingly. Therefore the algorithm learns to recognise untagged elements, to catalogue them according to the tags of the tagged elements. The tags can be either labels or numerical values. In the first case, the supervised learning problem is a **classification** problem; in the second case, the learning problem is a **regression** problem;

- unsupervised learning: the input data are presented without target outputs. The aim of this kind of algorithm is to analyse the unlabelled data and find hidden patterns that allow grouping together into clusters the data based on similarities and common properties;

- semi-supervised learning: this solution is a medium between supervised and unsupervised learning. During the training, one part of the samples is made up of labelled data while another, larger part is made up of unlabelled data.It is a useful solution when the dataset does not have enough labelled data for using supervised learning algorithms. Also, this combination of labelled and unlabelled data can improve the performance of the system, making the problem more general;
- reinforcement learning: the algorithms work according to a trial and error logic, using a feedback loop of rewards and a reinforcement function that measures the degree of success of an action or decision against a predetermined goal.

Considering HAR problems, supervised learning is a common solution used for recognition based on sensors and wearable devices. In the next Subsections, a brief description of some typical ML algorithms is made, focused on solutions that have been useful for the purposes of this thesis.

Naïve Bayes Classifiers

The Bayesian classification is a statistical technique which determines the probability of an element belonging to a certain class [20]. The technique is based on Bayes' theorem, which defines the conditional probability of one event with respect to another according to the following equation:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \tag{1.1}$$

where $P(A|B)$ denotes the conditional probability of A with respect to B, $P(B|A)$ denotes the conditional probability of B with respect to A, while $P(A)$ and $P(B)$ indicate the two a-priori probabilities referred to the two events.

In the context of HAR, Naïve Bayes Classifiers is a widely used algorithm in which it is assumed that the effect of an attribute on a class is independent of the values of the other attributes [49]. This assumption, called conditional independence of classes, is intended to simplify calculations, and it is precisely for this reason that the algorithm is called "na ïve "(textit naif).
Each instance is represented by a vector of n features, i.e. independent variables, defined as $\mathbf{x} = (x_1, x_2..., x_n)$, and the class to which this instance belongs is determined by finding the maximum conditional probability of a certain class C_k with respect to the other $k-1$ classes. The conditional probability will then be:

$$P\{C_k|\mathbf{x}\} = P\{C_k|x_1, x_2, ...x_n\} = \frac{P\{x_1, x_2, ...x_n|C_k\} * P\{C_k\}}{P\{x_1, x_2, ...x_n\}} \tag{1.2}$$

however, since the independence of the features is assumed, the following equation is valid for each class C_k :

$$P\{x_1, x_2, ...x_n|C_k\} = \prod_n P\{x_n|C_k\}.$$

Finally, the maximisation affects only the numerator in 1.2, since the denominator is independent from C, so the conditional probability is given by:

$$P\{x_1, x_2, ...x_n|C_k\} = P\{C_k\} * \prod_n P\{x_n|C_k\}. \tag{1.3}$$

The main strengths of this algorithm are the following three:

- it works well in case of noisy data
- it tends to ignore irrelevant attributes
- it has a simpler training phase than other methods

The downside is represented by the assumption of the independence of the features, which may not be true in reality. However, in the case of HAR problems, good results have been found when many data are available and when working with sensors that are well differentiated according to the activity.

Hidden Markov Models

A Hidden Markov Model (HMM) is a statistical model in which the system to be modelled is assumed to be a Markov chain representing a sequence of events [26]. A Markov chain is a stochastic process with discrete state space such that the probability of transition to a new state depends only on the immediately preceding state of the system. Therefore, an HMM is composed of a finite set of hidden states, which evolve like a Markov chain, and of a set of observations generated by the states according to a certain probability distribution. The following points characterise this kind of model:

- the states evolve as a Markov chain (the new state does not depend on the whole sequence of states, but only on the predecessor);
- each generated observation, or event, depends only on the current state;
- the observations are independent of each other;
- the event is observable, the state is not.

The use of HMM models generally concerns three canonical problems:

1. knowing the parameters of the model, calculate the probability of a particular sequence of observations;

2. knowing the parameters of the model, find the most likely sequence of hidden data that generated a given sequence of outgoing observations;

3. given one or more sequences of observations, find the most probable set of probabilities of outputs and transitions. Hence, the model needs to be trained starting from data related to the sequences.

In the field of activity recognition, the hidden states correspond to the activities carried out by users, while the values read by the sensors constitute the observations. The following three parameters are obtained from the model training:

1. the initial probability of states;

2. the probability of transition from one state to another;

3. the probability of issuing an observation from a given state.

Several works have made use of hidden Markov chains to recognise human activities. For example, in [87], a model based on HMM is proposed that is able to find out what activities are carried out starting from a dataset containing the sequences of observations emitted by the sensors but without the knowledge of the actual activities performed, since they had not been labelled in the training phase.

Support Vector Machine

An SVM represents a method for classifying linear and non-linear data [47]. While the original problem can be defined in a finite dimensions space, it often happens that the sets to be distinguished are not linearly separable in that space. For this reason, the training set data is projected via a non-linear function into a space with a larger number, presumably making it easier to find a separation in this new space. Therefore, the goal is to find the best hyperplane that can separate data belonging to different classes. A good separation is represented by the distance that this hyperplane presents with respect to the closest elements (the support vectors) belonging to the two classes that we are trying to separate. Consequently to what has just been said, the best separation will be associated with that plane which is able to maximise the distance between the hyperplane and the obtained support vectors. The decision function of an SVM is defined by the following formula:

$$f(y) = \sum_{i=1}^{N} a_i \, K(x_i, y) + b \tag{1.4}$$

where y represents the unknown vector to be classified, x_i are the support vectors with their weights a_i while b is the intercept and bias term. $K(x, y)$ is a function called *kernel* that solves non-linear problems by making the implicit mapping of larger features into the space.

In [29], a method based on SVMs is proposed for the recognition of activities and the identification of the most significant sequences of behaviours, while in [47] they are used to build a behavioural model of two different inhabitants to be able to distinguish them according to the different ways in which they carry out three different activities.

1.3.2 Deep Learning

DL is considered a subset of ML and many algorithms works with artificial neural networks, that are designed to imitate the human brain and how it thinks and learns. The main aspect of DL is that, contrary to ML where the steps about features extraction and selection are essential and deeply affect the performance of the algorithms, here the process of feature engineering can be automated [48]. Similar to how the brain is made of neurons, neural networks are made of layers of nodes [25]. There are essentially three kinds of node layers: Input, Hidden and Output. The depth of the network depends on the number of Hidden layers that characterised it. The nodes inside a layer are connected to the adjacent layers and each successive layer uses the output from the previous layer as input. The information travelling between nodes is associated with corresponding weights; a node with heavier weight will have more considerable impacts on the next layer of nodes. The final layer takes into account the different weighted inputs to produce the output. The neural network tries to improve its performance by repeatedly adjusting these weights over and over again. The weights are updated through a backpropagation optimisation algorithm based on the gradient, according to the errors on the outputs. Through different cycles made of inputs-processing-outputs, networks can generalise problems and provide correct outputs associated with inputs that are not part of the training set. There are different types of neural networks, such as Artificial Neural Network (ANN) Convolutional Neural Network (CNN) and RNN; since the data typically involved in sensor-based HAR problems are strictly related to temporal dependencies and sequential tasks, the most common solution for these problems are base on RNN which allows the recognition of patterns defined by temporal distance [78].

Recurrent Neural Network

The characteristic of RNN that differentiates this kind of network from others, is that, in this case, the information can also be transmitted backwards thanks to back loops. The most basic topology is a network where the nodes in the hidden layer contain recurrent units. These back loops allow the network to create an internal state that preserves information on temporal behaviours. For this reason, RNNs can use their internal memory to process sequences of data as inputs [79]. With this configuration, the outputs of the network are dependent not only on the current state but also on previous states and computations. Considering the current hidden state h_t and the output o_t, the following equation are valid [3]:

$$h_t = f(w_h h_{t-1} + v_h x_t + b_h), \tag{1.5}$$
$$o_t = f(w_o h_t + b_o), \tag{1.6}$$

where w and v are the weights, b indicates the bias for hidden and output states, x_t is the input of the current state, and f is the activation function of the hidden layer, often chosen to be the sigmoid, tanh or sign function. These equations are recursively evaluated at every time step. Therefore, after the hidden state is obtained at time step t, it is passed back to the recurrent unit and combined with the input at time step $t+1$ in order to obtain the new hidden state at $t+1$.

LSTM networks are a modification of RNN that can remember previous data in memory for a longer time, thanks to an added cell state. The other difference between the two is how the recurrent unit works.

Long-Short-Term-Memory Network

LSTM network is an RNN designed to model temporal sequences and learn long-term dependency problems. Furthermore, LSTM networks help to resolve the problem of the exploding or vanishing gradient [78]. This is when large error gradients start accumulating and prevent the model from training with available data or when the gradients of loss functions approach zero and weights and bias of initial layers are not updated effectively [65]. The main unit in LSTM architecture is the memory block that contains the recurrently connected cell state and the three units that are the input, output and forget gates. These gates control which information is going to be kept, forgotten or passed through based on the relevance of that information, which is filtered according to the gates' set of weights [23]. The gates and the cell state are governed by the following equations:

$$\begin{aligned}
i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i),\\
f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f),\\
o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o),\\
c_t &= f_t c_{t-1} + i_t\, tanh\,(W_{xc}x_t + W_{hc}h_{t-1} + b_c),\\
h_t &= o_t\, tanh\, c_t,
\end{aligned}$$

where W are the weight matrices of the correspondent gates or cell, the terms i, f and o indicate the gates with the same initials and c represents the memory cell state, x_t is the input, σ is the sigmoid function and $tanh$ the hyperbolic tangent activation function. The forget gate controls the information that should be forgotten. The input gate identifies the elements that need to be added to the cell state. The output gate defines the data that should be immediately kept from the cell state and associated with the previous information. The information of the gates is presented through the sigmoid function; since this function can have values from 0 to 1, the values in the cell state are going to be deleted when multiplied by 0, remembered when multiplied by 1 or given partial importance when multiplied with a value between 0 and 1 [53]. The values that enter the network are regulated thanks to the hyperbolic tangent function between -1 and 1 [53]. The latest cell state c_t and the hidden state h_t are used again in the recurrent unit and the process is repeated in the next step.

The network can only interact with memory blocks through each gate, and hence the gates learn to open or close intelligently to let only relevant memory content propagate in the network and discard the irrelevant information; in this way, it is also possible to prevent the exploding or vanishing of gradients. Furthermore, thanks to this way of learning which information to keep and which information to forget, LSTM can learn the presence of time lags of 1000 time steps between inputs and target outputs, while RNN can learn only between 5 or 10 time steps[33].

1.4 Building Energy and Comfort Management System

Smart technologies can be used in all kinds of different buildings (i.e., residential, office, and retail sectors) to improve the comfort and the safety of people in their homes, concerning various topics, from healthcare and providing living assistance to environmental monitoring and ensuring energy saving. Accordingly, BECM systems have the objective of combining power consumption minimisation while preserving user comfort [90] [46] [1]. This issue has been addressed by researchers from many different perspectives.

The authors in [90], proposed an optimisation methodology in order to minimise energy consumption considering illumination, temperature and air quality as parameters for user comfort. Also in [1] the same kind of comfort parameters are considered and the reduction of power consumption is made by programming appliances usage according to a-priori set of variables, while in [76] [30] two different solutions for building management considering user preference in terms of indoor environment comfort are presented. In [72], a home management system is proposed that tries to find an optimal task plan in terms of energy consumption, while considering consumer comfort, asking the resident to define their preferred adjustments in order to schedule the activation of the tasks.

The issue of scheduling appliances according to user preferences was also addressed by [66], where Quality of Experience (QoE) is measured as a function of the interval between the preferred and proposed appliance starting time for switching controlled loads (e.g., washing machines and clothes dryers), and as a function of the interval between the preferred and proposed temperature for thermostatically controlled loads (e.g., conditioning systems and water heaters).

It is evident that user preferences and habits severely affect the results of BECM systems. For this reason, in recent years, researchers have started to observe users' behaviour in order to infer their habits and preferences. The monitoring of activities of people in their homes can be done by analysing data that can be gathered with different technologies. Different studies proposed solutions based on using cameras and wearable sensors or gathering data provided by phone accelerometer and gyroscope [60] [4]. These solutions are not very practical in home scenarios where people are often not inclined to accept those devices. Non-intrusive sensors are often preferred to monitor what activities people are performing in their house: typical devices that are installed in the environment are motion sensors, door sensors or temperature and pressure sensors, power meters [54] [70]. The data collected from sensors inside resident houses are analysed using data mining and machine learning techniques to build activity models that are used as the basis of behavioural activity recognition. As seen in Section 1.3 there are many different solutions for the recognition and prediction of resident activities, concerning the classification and recognition problem, it is possible to say that, in multiple cases, despite of its simple design and simplified assumptions, Naïve Bayes classifiers often work much better than expected, especially when a specific group of sensors can easily be identified as characteristic of a certain activity [55]. Considering the prediction problem, as highlighted in 1.3, LSTM networks are largely used to make predictions on future activities performed by residents inside their homes.

1.5 Appliances Modelling and Profiling

1.5.1 Profiling of Energy Consumption Habits

Energy efficiency is recognised as one of the most powerful tools to fight global climate change. Since the building sector is responsible for 36% of global energy consumption and 37% of CO2 emissions [43], a great effort is being spent to make buildings more efficient from an energy consumption perspective. To achieve the Paris Agreement objectives, the United Nations set the goals to build only Net-Zero Energy Buildings (NZEB) by 2030, and that the whole building stock is NZEB by 2050 at the latest [43].

Obtaining energy consumption profiles represents a key point for creating systems that can manage and monitor the energy of different Smart Buildings in the best possible way and provide support to users to ensure energy savings and a reduction in consumption. In general, consumption profiles can be derived in different ways, mainly depending on the characteristics being analysed and the attributes that are taken into account. Many research works focused on the total load curve of residential buildings under different scenarios, and the energy consumption profiling follows daily and weekly trends, seasonality and geographic areas [18], [17]. In this kind of solution, the obtained profiles are differentiated on a daily basis (i.e. time of day with greater peak) and according to the seasons (i.e. higher consumption in summer due to cooling systems). Other solutions can take into account also demographic characteristics and dwelling information [45], [85], in order to provide a different insight into how some external factors can influence the energy habits of different users. In these cases, the different profiles obtained thanks to patterns in consumption volumes and timing are also analysed by considering additional information such as residents' personal data, the number of residents, and the building itself.

Fewer research works are instead focused on energy consumption profiles that take into account the individual loads of household appliances [56], [19]. In these cases, the obtained profiles are differentiated based on the usage of appliances during the day. The profiles thus obtained take into account when household appliances are generally used but not the different energy consumption that can be associated with each of them according to the usage that users make of them. In [88], the authors proposed a method of predicting the consumption of household appliances by evaluating statistical distributions. Different consumption groups for each appliance are evaluated together, considering a-priori information about the duration of the cycles and the day of the week. On the other hand, with an unsupervised learning method, the data are grouped together based on the similarity of the different features considered without dividing the usage cycles of the appliance in advance since these cycles can be different according to the habits of different users.

1.5.2 Clustering Mechanisms

In recent years, above all, thanks to the presence and greater use of smart meters, the solutions for energy consumption profiling are based on data drive models, in which clustering techniques are the most frequently applied [86], [101], [74].

Clustering algorithms can be divided into two categories based on how the clusters are defined: hierarchical clustering and non-hierarchical clustering (or partitional clustering) [34]. Hierarchical clustering algorithms evaluate a criterion for merging or splitting clusters based on similarity and obtain the optimal nested series of partitions. Partitional clustering algorithms optimise a clustering criterion to identify the best divisions of the data [44].

Authors in [74] applied a hierarchical clustering algorithm to cumulative, normalised hourly load profiles obtaining $k = 3$ clusters differentiated in terms of the peak-period usage profiles.

Among partitional clustering algorithms, the most prevalent used is the k-means algorithm. Authors in [101] selected the k-means algorithm to cluster residential customers in different groups based on their usage patterns using half-hourly smart meter datasets and finding the best solution for $k = 6$ number of clusters. In [86] the k-means algorithm is applied to autocorrelation coefficients, also obtained from smart meter data, and obtains the best results for $k = 12$. The reasons for choosing k-means clustering are mainly related to its speed of convergence, easy implementation, and efficiency; furthermore, it is particularly suitable when working with statistical data and large datasets [85].

1.6 Health Monitoring System Using Wearable Devices

1.6.1 Health Monitoring Systems

Human behaviour can impact, on different levels, people's quality of life; some behaviours affect individuals' lives directly, especially considering their psycho-physical health state. For example, many research works are focused on better understanding how some behaviours, or changes in someone's habits, can be triggers for recognising particular illness symptoms like depressive disorder or Alzheimer's Disease[63][62]; especially for elderly people, there are many solutions that propose assistance and healthcare systems or early recognition systems for degenerative diseases to ensure a better quality of life [11] [51].

Wearable sensors are largely used to monitor users' behaviour, collecting their physical signals to describe particular actions and activities. A great deal of interest is emerging in finding solutions that, using human bio-signals detected through wearable sensors, can understand users' state, their emotions and their stress [35] [10].

There are many possible solutions for the realisation of health monitoring and sensing systems, depending on the involved devices and on observed parameters. In some cases, these systems rely on wearable devices placed in various body parts like the chest, fingers, ankle or hip[64][57]. These devices are provided with healthcare sensors to monitor some parameters related to the body condition [9] such as heart rate and heart rate variability (HRV), obtained from electrocardiogram; brain activity, monitored with electroencephalogram; electrodermal activity (EDA), computed applying a small current and measuring the resistance of the skin between two electrodes; body temperature and blood volume pulse, measured with Photoplethysmography optical technique.

Nowadays, there are many personal wearable and portable devices (smartphones, smartwatches or other devices attached to the user's body provided with accelerometer, gyroscope and heart rate sensor) that are able to monitor users, obtain context information or record vital physiological signals. Not all the parameters mentioned above can obviously be monitored through simple smartbands/smartwatches placed on the wrist, but many of the newest devices focus on incorporating more reliable sensors and monitoring even those parameters that are the most useful to characterise certain psychological states (i.e., HRV).

All these solutions allow the creation of unobtrusive systems and can be easily accepted by any type of user.

There are also many other works that realise health care monitoring systems using wireless sensor networks and cameras to obtain information about the environment other than monitoring people directly. For example, there are several fall detection systems for the elderly that use cameras, PIR sensors or sound sensors to recognise when someone has fallen down and cannot stand up [27].

The two kinds of solutions are not in opposition, in fact there are many research works that propose solutions in which both ambient monitoring and the users monitoring through wearable coexist to obtain the best results [98][41].

More recently, solutions based on Radio-Frequency (RF) signals are also becoming interesting candidates for indoor healthcare applications [32][100]. These systems can be used to monitor both users' vital signals and dangerous situations (e.g. fall detection). These kinds of solutions have the advantages of being non-intrusive, providing contactless monitoring and being less problematic in terms of privacy than video-based solutions. Still, there are also some drawbacks, mainly due to the possibility of monitoring only certain areas of interest and the necessity of considering different signal disturbances.

1.6.2 Well-Being and Stress Evaluation Systems

The term well-being is a keyword in the World Health Organization (WHO) definition of health: *a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity.* The meaning of well-being is multidimensional and an overall sense of wellness should be considered achieved when there is a balance between different elements like physical aspects, emotional and psychological feelings, social inclusion and economic wellness [59]. Higher levels of quality of life are therefore strictly correlated with perceived wellness, well-being and general healthiness.

It is possible to differentiate between two different dimensions of well-being: objective and subjective. Objective well-being is more oriented on aspects of good life concerning basic human needs and rights; on the other hand, subjective well-being, or personal well-being, is measured taking into account people's subjective evaluations of their lives [93].

In the literature, there are many different self-evaluation instruments (questionnaires) that can be used to assess subjective well-being. These surveys can be used in different fields of well-being assessment, i.e. the Patient Health Questionnaire (PHQ-9) and Mental Health Continuum-Short Form (MHC-SF) are used for detecting depressive symptoms as well as well-being [39]; others, for example, are used for clinical patients in order to validate the results of some treatments and cures [89]. Two of the most largely used questionnaires, which are the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) and the WHO-5 Well-Being Index (WHO-5), have been developed so that it is possible to monitor the mental well-being in the general population, covering all the aspects of mental well-being [81].

Also, the concept of stress is difficult to define since it is very used and presented in daily life. The term stress has been associated with many different processes, including both life stress exposure and the psychological and biological consequences of such exposures. The Perceived Stress Scale [14] is one of the most used questionnaires for measuring perceived stress levels over a given period of time.

1.7 Conclusions

In conclusion, this first part is a brief introduction to some of the main aspects relating Home Automation, sensor-based HAR problems, energy and comfort management inside buildings and health monitoring systems using wearable devices.

Sections 1.1 and 1.2 gives some definitions of Smart Homes and Smart Buildings, explaining the need to understand users' habit and preference inside their homes and how these behaviours can affect their lives. This led to the problem of HAR, which is addressed in Section 1.3, where there is a little overview of how these problems are typically addressed and a description of different classifications depending on the used technologies and sensors. Sections 1.4, 1.5 and 1.6 describe the State of the art for the three main topics addressed in this thesis, related to BECM systems, appliance profiling models and health monitoring, respectively.

Regarding the first topic, it is clear that the users' comfort and preferences with respect to household appliance usage must be considered not as a set of characteristics decided a priori since they could change over time and, above all, each user can have features different from each other. For this reason, it is important to have a system that can continuously monitor the users, learning their habits and re-learning them when needed. Hence, with respect to this subject, the main contributions provided in this thesis can be summarised as follows:

- an activity recognition algorithm used to model user profiles, which was first proposed in [58] and whose accuracy is here improved;

- an activity prediction algorithm is proposed that is based on an LSTM network. Accordingly, appliance usage is predicted for a specified time window (tests were run for a time window of 6 hours);

- user profile and activity prediction are incorporated into the user-annoyance-aware energy-cost-saving appliance scheduling algorithm proposed in [66].

Considering the second topic addressed in this thesis, different consumption profiles can be obtained regarding the energy patterns inside the house, differentiated for day hours, seasonal phases or user demographic characteristics. There are fewer works that study the consumption distinguished by individual appliances; in this case, the information about the duration of the cycles and the day of the week are considered a-priori. The main contributions of this thesis regarding this topic are:

- eight useful features are identified to recognise different consumption profiles for individual appliances;

- thanks to unsupervised clustering methods (i.e. k-means clustering), for each appliance, different consumption profiles are found that take into account the different ways of use by the user.

Lastly, considering the definitions of stress and well-being, understanding the punctual moment of stress through data gathered by a wearable is still very hard. On the other hand, it can be useful to investigate how some behaviours affect perceived well-being, even over more extended periods. In obtaining data helpful in evaluating the psycho-physical health of users, simple commercial devices of daily use can play an important role. Data gathered from these devices can be inferred to understand the activities usually performed by users and how they affect their well-being.

Therefore the main contribution regarding this aspect, presented in this thesis, is the realisation of a prototype to understand if stress and well-being can be related to data about sleep, fitness activities, and walking habits gathered thanks to commercial devices.

Chapter 2

Building Energy and Comfort Management System

2.1 Introduction

Smart buildings are characterised by the presence of sensors, actuators and smart devices that give the opportunity to monitor and control, either manually or automatically, key equipment within buildings [61]. This is the concept behind BECM systems [22] [75]. As a matter of fact, domestic electricity usage accounts for about 40% of the global energy consumption and contributes over 30% of total greenhouse gas emissions [99]. Nevertheless, user comfort is crucial when policies of Demand-Side Management (DSM) are put in place [77]. In such an intelligent scenario, one of the major goals is to provide users with tools that support cost-effective solutions to appliance management, which: i) demand the lowest effort in terms of training and management, dynamically adapting to user requirements, and ii) take into account user habits so that appliance management decisions do not conflict with them, causing a disaffection that may lead the user to turn off the system [66]. Currently, most of the literature considers user comfort as a set of hard constraints on appliance usage, which are a priori set considering general statistics [67] [69]. This approach neglects the fact that users are likely not only to have different subjective requirements with respect to others, but they also dynamically change over time. Therefore, this part of the thesis is focused on the realisation of a BECM system based on sensors deployed inside the reference building that can guarantee that user preferences and habits about appliance usage are continuously monitored, recognised and predicted. The system merges two previous studies about activity recognition [58] and appliance scheduling [66], by including the crucial activity prediction functionality. Indeed, activity prediction enables appliance scheduling by predicting which appliances are likely to be used in the following hours and scheduling them in advance so that their starting time is shifted to off-peak times when electricity tariffs are lower. This is a comprehensive system that uses sensor-based activity prediction

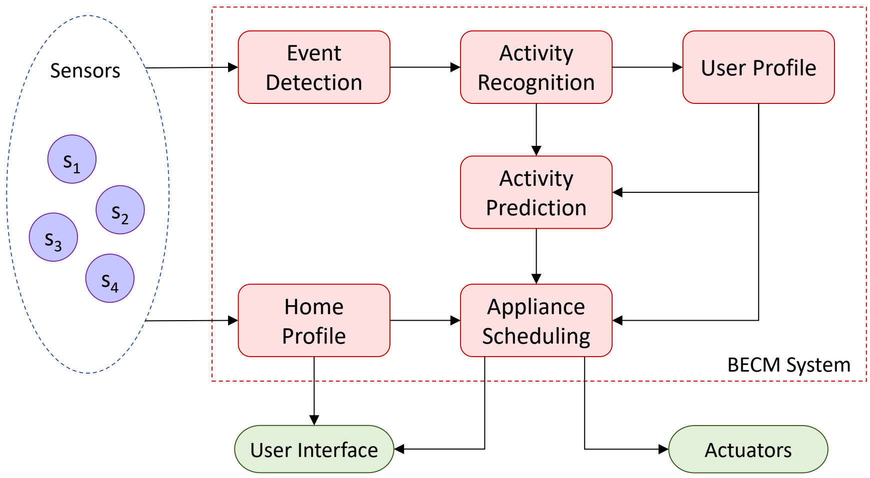


Figure 2.1: Overview of the proposed BECM system

and occupants' preference inference and integrates them into a BECM. Accordingly, based on simulations of the system on a real dataset, this Section further analyses how the proposed system affects energy-related costs.

2.2 System Model

The considered scenario is that of a BECM system, based on distributed smart home sensor networks. An overview of this system is represented in Figure 2.1. More specifically, sensors are used to make observations on users and their interactions with the surrounding environment; the combinations of these interactions, which are detected by the *Event Detection* module as events, provide meaningful information on users' activities. As described in more detail in [58], after a training period the *Activity Recognition* module can correctly recognise activities with an accuracy of more than 80% on average. Accordingly, a correlation can be observed between detected events and activities, which can be used to infer users' habits. These habits, stored in the *User Profile* module, are used with information about previously recognised activities by the *Activity Prediction* module, with the aim to predict the following activities that are expected to be carried out by the users. The information related to activity prediction, user profile and home profile is then processed by the *Appliance Scheduling* module to find a scheduling for controllable appliances that corresponds to the best trade-off between energy cost reduction and user comfort. Note that the *Home Profile* module stores home-related information collected by sensors and/or through user interfaces, such as electricity tariffs, and which appliances are installed along with their energy consumption characteristics.

In the following, more details will be provided about the core modules of the proposed BECM system, i.e. the Activity Recognition module, the Activity Prediction

Module and the Appliance Scheduling module.

2.2.1 The Activity Recognition Module

The activity recognition approach used in this system was earlier proposed in [58]. It encompasses two phases: i) *training*, during which the system learns the association between activities and their instances, i.e. sequences of detected events; ii) *running*, which uses the probabilistic model created during the training phase to associate an activity to the detected events.

a) Training phase: for each *k-th* activity instance $\mathcal{I}_{jk}$ of activity $\mathcal{A}_j$ observed during an observation time window $\mathcal{O}^A$, a feature vector

$$\mathcal{F}_{jk}(\mathcal{I}_{jk}) = [f_{1jk}, f_{2jk}, \dots, f_{ijk}, \dots]$$

is computed with the rates of detected event occurrences, that is the number of events related to one specific sensor with respect to the total number of events observed considering all the sensors within $\mathcal{O}^A$. Then, for each activity $\mathcal{A}_j$, a model vector

$$m_j = \text{mean}_k(\mathcal{F}_{jk}) = [\overline{f}_{1jk}, \overline{f}_{2jk}, \dots, \overline{f}_{ijk}]$$

is defined such that the rates of event occurrences of its sensors is the average rate for all the observed instances associated with the same activity.

b) Running phase: it relies on the use of a sensor-based windowing implementation [52], according to which sequences of detected events are divided into subsequences using an observation window $\mathcal{O}^{\mathcal{W}}(t)$ starting at time t, which contains a certain number of events equal to its size $\mathcal{W}$. Each subsequence z of events is then associated with a feature vector $\mathcal{F}_z^W$, computed analogously to m_j. Finally, the sequences of detected events are classified based on their probability of belonging to a given activity.

For further details, the reader is referred to [58].

2.2.2 The Activity Prediction Module

The *Activity Prediction* module has to provide a possible scenario ahead in the future, making predictions on which activities are most likely to be performed in the time to come. The activity prediction algorithm is based on RNN, more specifically an LSTM network has been used.

LSTM are a special kind of RNNs that are capable of learning long-short term dependencies.

A neural network usually consists of one layer of input units, one layer of output units, and multiple in-between layers called hidden layers. These hidden layers are made of hidden units that have certain weights w determined during the training phase and that define the network's final conformation. The weights are learned using the data that consists of example inputs and desired outputs obtained from those inputs. In RNN, the current output depends not only on the current input but also on the state of the system.

The main difference between LSTM and RNN networks is about the memory cell, which can keep the information for a longer time than a basic RNN. In LSTM networks, the cell state is controlled and regulated by controlling gates, that act on the signals they receive considering the importance of the information transmitted, which they filter using their own sets of weights. Therefore, the cells learn how to act on the data they received, choosing if the information should be kept or deleted, through an iterative process that allows the weights to adjust themselves during this learning phase [82] [23].

Starting from the assumption that from a sequence of previous activities $\mathcal{A}_i$ occurred at time t_i it is possible to predict the next activity most likely to be performed by the user, the module has to evaluate the probability in t, after all previous t_i, for every single activity $\mathcal{A}_j$. Considering t_0 as the current moment under investigation, the period between the current time t_0 and the time t in which the prediction is needed is split into different p time intervals of duration (Δt).

Thus, the Activity Prediction Module has two phases: one is the *training phase*, and the other is the *operational phase*. During the *training phase*, the module trains the LSTM network to learn its weights, considering a proper input vector; the output of the network consists of the predicted activity for the next moment based on the information of the input vector. Then, during the *operational phase*, the module makes use of the trained network and of the information obtained from the Activity Recognition Module, in order to predict the activities that are more likely to be carried out by the user in the future.

The vector considered as the input for the network contains information about the N previously performed activities and contextual information about the moment of the day t_0 and the day of the week W_d in which the system is at the current moment. This input vector is described as follows:

$$\mathcal{F}_{t_0} = \{\mathcal{A}_j^{t_0-N\Delta t}, \mathcal{A}_j^{t_0-(N-1)\Delta t}, \ldots, \mathcal{A}_j^{t_0-\Delta t}, t_0, W_d\}$$

where t_0 can be either the present moment or anytime in the future. The output of the network gives the probabilities for each one of the activities to be carried out at the next moment by the user, considering the information of the input vector described above. The activity related to the highest probability is considered as the most probable activity to be performed in the next moment.

Thus, during the training phase, the LSTM network learns the predictions for the activities that are going to be performed at time t_1 using the information in the input vector of time t_0, so that $t_1 = t_0 + n\Delta t$, where Δt is a predetermined time interval chosen as the most descriptive average time between two consecutive activities, according to user habits, and $n = 1$ indicates the number of time slots that separates the current moment to the moment in which the prediction is evaluated.

After the training phase is completed, the obtained network is used for the operational phase. During this second phase, using the information about the activities that have been carried out before the current moment t_0, the algorithm makes a prediction for the next activity at the next interval $t_0 + n\Delta t$ with $n = 1$. Taking into account this last prediction just gained, the algorithm uses this information to make another prediction for interval $t_0 + n\Delta t$ with $n = 2$. By repeating this behaviour, the algorithm can predict the most likely activity for every $n\Delta t$ time interval up to the time t in the future, where the prediction is needed.

These predictions are calculated every time the next Appliance Scheduling module requires a new evaluation of the activities that are most likely going to be carried out by the user so that it can elaborate a new scheduling program accordingly. Therefore, the output of the Activity Prediction module is an array containing the probabilities for each activity to be performed at time t, considering all the M activities registered in the User Profile Module, so that the Activity Scheduling module is going to make all the evaluations considering the probability vector $\{p_1, p_2, \ldots, p_M\}$.

Every activity that is recognised and predicted by the Activity Recognition and Activity Prediction modules is linked to one of the appliances of the house. In this way, the probabilities of the activities in time t, that are obtained from the predictions, are then translated into probabilities of the corresponding appliances to be used at the same time t.

These values of probabilities are then used by the Appliance Scheduling module to evaluate which appliances from the house need to be scheduled at time t in order to be switched on. A precise threshold value of probability for every appliance is considered so that the predicted probabilities obtained for that appliance at the time t must be higher than the assigned threshold to cause the system to insert the appliance into the scheduling program.

Therefore, the output from the Activity Prediction module is going to enable the scheduling algorithm to make the validation necessary for the scheduling of controllable appliances and for evaluating energy consumption.

2.2.3 The Appliance Scheduling Module

The appliance scheduling algorithm is based on the smart home energy management system proposed in [66]. By considering hourly electricity rates and power produced by Renewable Energy Sources (RES), this system dynamically shifts tasks of controlled appliances to times when it is more convenient (e.g. off-peak times), after finding a trade-off between the overall energy cost and the annoyance experienced by users as a consequence of this shift. Accordingly, appliances are subdivided into three groups:

G1: not controlled loads, i.e., small loads such as lights, and not controlled high loads such as fridges;

G2: switching controlled high loads, such as washing machines and dishwashers;

G3: thermostatically controlled high loads, i.e. appliances that are controlled by a thermostat, such as water heaters.

The energy consumption for an appliance belonging to group G1 is defined as:

$$E_l^{cons} = \int_{t=t_l^{st}}^{t_l^{end}} P_l^{cons}(t)dt \tag{2.1}$$

where $P_l^{cons}(t)$ is the power consumption of appliance l, and t_l^{st} and t_l^{end} are, respectively, the starting and ending time of its operating cycle. For switching controlled loads (i.e. G2), the operating cycle has a fixed duration, and thus the ending time only varies with the starting time: $t_l^{end} = t_l^{end}(t_l^{st}) = t_l^{st} + t_l^{exec}$, where t_l^{exec} is the execution time for appliance l. On the other hand, for thermostatically controlled ones (i.e. G3), the ending time varies with the initial internal and external temperature at the starting time, respectively $T_l^{in}(t_l^{st})$ and $T_l^{out}(t_l^{st})$, and with the temperature that is expected to be reached at t_l^{end}, i.e. T_l^{exp}. As described in more details in [66], the ending time of G3 appliances to reach a temperature T_l^{exp} is defined as

$$\begin{aligned} t_l^{end} &= t_l^{end}(t_l^{st}, T_l^{exp}) = \\ &= \begin{cases} t_l^{st} - R_l C_l \ln\left(\frac{T_l^{out}(t_l^{st}) - T_l^{exp} + R_l H_l}{T_l^{out}(t_l^{st}) - T_l^{in}(t_l^{st}) + R_l H_l}\right), & \text{if } l \text{ is ON} \\ \infty, & \text{if } l \text{ is OFF} \end{cases} \end{aligned} \tag{2.2}$$

where H_l, R_l and C_l are characteristic parameters for the appliance. More specifically, H_l is the heat rate (in Watt), R_l is the equivalent thermal resistance ($^\circ C$ /Watt) and C_l is the equivalent heat capacity (Joule/$^\circ C$).

The appliance scheduling algorithm schedules the appliances according to their related cost contribution value, which includes both the energy consumption- and user annoyance-related costs.

Based on these assumptions, the cost to shift the operating time t_l^{op} of appliance l, i.e. the cost to shift its starting time at t_l^{st} and ending time at t_l^{end}, is defined as

$$C_l(t_l^{st}, t_l^{end}) = \varphi_l\left(\Delta t_l^{op}\right) \cdot \int_{t=t_l^{st}}^{t_l^{end}} P_l^{cons}(t) \cdot \Phi(t)dt \tag{2.3}$$

where $\Phi(t)$ is the electricity tariff at time t. Let Λ be the set of G2 and G3 appliances that requested to be scheduled and have not started yet. The problem to be solved is then defined as

$$\begin{aligned} &\min_{t_l^{st}, t_l^{end}} \sum_{l \in \Lambda} C_l(t_l^{st}, t_l^{end}) \\ &s.t. \quad t_l^{end} < t^{cur} + 24h \qquad \forall l \in \Lambda \\ &\qquad\;\; P^{tot}(t) \leq P^{max} \qquad \forall t \end{aligned} \tag{2.4}$$

where the limit to the considered time span has been fixed to 24 hours after the current time t^{cur}, as defined by the first condition. The second condition ensures that the power that the grid makes available is not exceeded by the simultaneous usage of G1, G2 and G3 appliances.

The activation of an appliance occurs when the probability of the activity linked to that appliance exceeds a certain threshold, which is evaluated during the training phases of the system.

For further details about this appliance scheduling system, the reader is referred to [66].

2.3 Reference Use Case

2.3.1 Dataset Description

The experiments and simulations conducted to understand the performance of the proposed are implemented and tested using the Aruba real-world dataset from the CASAS smart environment project of the Washington State University [15]. The data were collected from one smart apartment provided with motion sensors, contact sensors in the doors or cabinets and temperature sensors, with only one resident living in the home.

Figure 2.2 shows the house plan of the apartment and the exact position of every sensor in the rooms. The description provided by the CASAS project did not give information about the specifics of the used sensor. Table 2.1 explains the number of sensors per type placed in each room. The values provided by motion and contact sensors are Boolean, whereas the ones provided by temperature sensors are numbers. There are two more sensors not listed in this table, one motion sensor and one contact sensor because they are not located in a specific room, but they are linked to the entrance of the house.

The events decoded by these sensors are significant for recording elementary actions that people are performing, for example, door sensors are easily associated with opening and closing medical cabinets, food storage, or the entrance door, while with motion sensors, it is possible to monitor the presence of the resident in one room and the proximity with a specific object or piece of furniture. The aggregation of these elementary actions defines one activity of interest. The gathered data are presented with information about the date and time of every sensor event registered, the id of the activated sensor with its value and the beginning or end of each activity that is monitored. The dataset has the structure presented in Table 2.2.

In the dataset, 10 different activities performed by the resident are noted. Table 2.3 shows the details of the number of times each activity appears in the dataset, as indicated by the user. The "Relax" activity is the one that the user has denoted as the activity occurring while staying in the living room and it involves the set of sensors arranged in that room, as shown in the house map (Figure 2.2). The "Work"

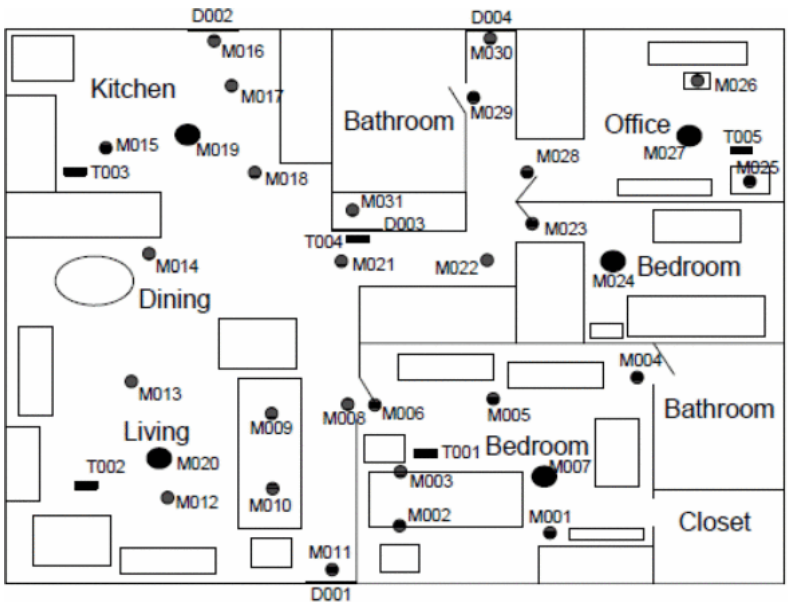


Figure 2.2: House plant of the apartment for the Aruba dataset [15].

	Motion Sensors	**Contact Sensors**	**Temperature Sensors**
Kitchen	5	1	1
Bathroom 1	2	1	-
Office	4	-	1
Dining	1	-	1
Bedroom 1	2	-	-
Living	7	1	1
Bedroom 2	7	-	1
Bathroom 2	1	-	-
Closet	-	-	-

Table 2.1: Number of sensors per type in every room of the apartment.

Date	Time	Sensor ID	Sensor Value	Activity
04/11/2010	09:56:22.785482	M018	ON	
04/11/2010	09:56:23.801652	M017	ON	
04/11/2010	09:56:26.467399	M019	ON	
04/11/2010	09:56:27.334395	M018	OFF	Meal Preparation end
04/11/2010	09:56:34.362031	M018	ON	
04/11/2010	09:56:37.729204	M020	ON	
04/11/2010	09:56:38.776094	M018	OFF	
04/11/2010	09:56:40.172391	M020	OFF	
04/11/2010	09:56:41.831135	M014	ON	Eating begin
04/11/2010	09:56:56.043362	M014	OFF	
04/11/2010	09:57:15.209217	M014	ON	
04/11/2010	09:56:16.412611	M014	OFF	

Table 2.2: Data extracted from the Aruba dataset [15].

activity is the one performed in the office room and involves the specific group of sensors placed in that area. Lastly, the "Housekeeping" activity involves a great number of all the sensors of the house due to the intrinsic dynamism of this type of activity. Sensors detect even the activities that are not registered, that correspond to "Other activity" with no label in the dataset. Since they cannot be classified accurately, this has been ignored in the proposed framework.

To evaluate the proposed system, in addition to the activities of the Aruba real-world dataset, some other activities have been simulated as performed by the same user inside this home scenario, using the same kind of sensors already installed in the house. The simulated activities are the following three activities not reported in the real dataset: using the washing machine, using the dishwasher and taking a shower, which, along with the activity of washing dishes by hand, causes the water heater to turn on. Taking a shower is supposed to be carried out by the user in the bathroom, therefore involving the motion sensors already installed close to this room and assuming that hot water is used, thus causing the water heater to switch on. The use of the dishwasher is supposed to be performed in the kitchen, involving the sensors in that area and simulating the presence of a specific cabinet containing the appropriate detergent and with a magnetic sensor to understand its opening or closing so that the activity of loading the dishwasher could be recognised

Activity	**Number of Occurrences**
Meal Preparation (MP)	1606
Relax (Rel)	2910
Eating (Eat)	257
Work	171
Sleeping (Sleep)	401
Wash Dishes (WD)	65
Bed to Toilet (BTT)	157
Enter Home (EH)	431
Leave Home (LH)	431
Housekeeping (HK)	33

Table 2.3: Activities and Statistics of the Aruba dataset.

concluded only when this cabinet had been closed. The same thing was done for the activity of using the washing machine by setting up another specific cabinet with its magnetic sensor, and placing it in a room of the house where there are no other linked activities.

The system needs a correspondence between some of the activities and the use of certain household appliances in order to predict energy consumption based on the probability of the activities to occur. Table 2.4 shows the 9 activities that have been considered during the different simulations along with their corresponding appliance owned by the user. Also, the characteristic parameters of these appliances are presented. Furthermore, the last 4 rows indicate the characteristic of two appliances/systems of group G1 that can be on regardless of the development of certain activities and the characteristic of the two types of RES considered, that are photovoltaic (PV) and a micro-wind turbine system.

The time of execution of the activities related to appliances of group G1 is not known a-priori, but the system evaluates the probability of the execution of the task and its lasting considering the user's habits and preferences, taking into account the time of the day, the day of the week and also the previous activities that the user has carried out till the current moment of evaluation. The appliances of group G2 have a mean time of execution calculated taking into account typical values of the cycles of use of the appliance itself. For G3 appliances the time of execution depends on the expected temperature obtained at the end of the operating time. There are two different target temperatures chosen as the preferred temperature for the user, one

[1]This is the maximum produced power. The power produced by RES varies during the day according to a normal distribution around the values in Fig.2.3

	Activity	Appliance	Group	Power [Wh]	Mean t_i^{exec} [min]
1	HK	Vacuum Cleaner	G1	1200	-
2	MP	Microwave Oven	G1	1000	-
3	Rel	TV	G1	30	-
4	Work	Laptop/Pc	G1	50	-
5	Laundry	Washing Machine	G2	600	130
6	Machine-Dry Clothes	Clothes Dryer	G2	1300	90
7	Wash Dishes with Dishwasher	Dishwasher	G2	400	160
8	Taking Shower (TS)	Water Heater	G3	2000	Set according to (2.2)
9	WD	Water Heater	G3	2000	Set according to (2.2)
10	Always on	Fridge/Freezer	G1	70	-
11	Always on when user is at home/not sleeping	Lighting	G1	40	-
12	NA	PV system	G4	1250[1]	-
13	NA	Wind turbine	G4	500[1]	-

Table 2.4: Correspondence between activities and home appliances and characteristic parameters of appliances [66][28]

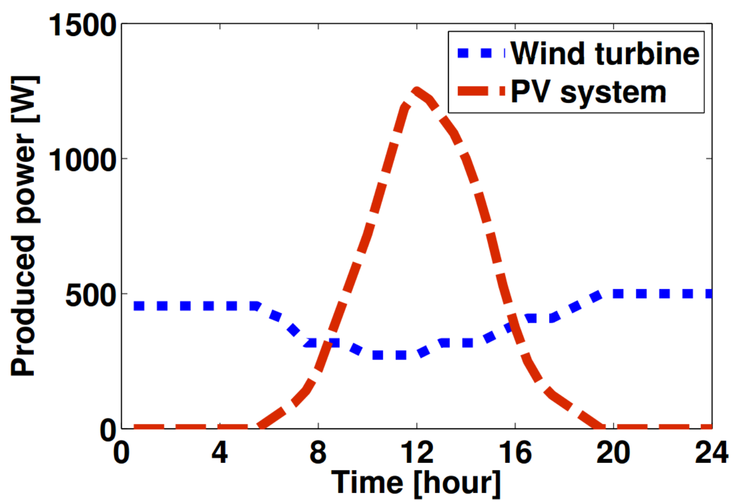


Figure 2.3: Daily power production for PV and wind turbine systems [28]

is related to the activity of "Taking a Shower", while the other one is related to the activity of "Washing Dishes" . Also the characteristic parameters that describe the dynamics of the water heater have been set with a normal distribution (with 20% deviation) around typical mean values that are [66]:

$$\begin{aligned} P_i^{heat} &= 5kW \\ R_i &= 120°C/kW \\ C_i &= 0.2kWh/°C \end{aligned} \tag{2.5}$$

The produced power of the RES has been varied according to a normal distribution (20% deviation) around the values in Fig. 2.3, up to a higher value that is consistent with those of commercial home systems [66]. In order to understand the level of discomfort experienced by the user due to the interventions of the system, the annoyance rate is used as defined in [66]; this level of annoyance is evaluated in relation to a possible shifting of appliance starting time for the controllable appliance or, with reference to the water heater, to a variation in the water temperature with respect to the user preferred temperature of use. A value of 1 of annoyance indicates that there is not any discomfort for the user in the change of time in which the appliance was turned on, while a value of 5 indicates the highest level of annoyance for the user. Annoyance levels are modelled as a normal distribution with 15% deviation.

2.3.2 Creation of User Profiles

To better understand how users' habits can have an impact in the performance of the system, three different situations are considered during the simulations. These three situations differ according to the possible grade of predictability of the activities usually performed by the user. The user's profile is then realised based on how recurring their activities are performed inside the datasets; different profiles are obtained according to how much the user routine could be considered habitual and therefore predictable. Three datasets have been used that correspond to three user's profiles:

- real-case profile: based on the real-world dataset Aruba;
- weekly habitual profile: based on a synthetic dataset created supposing that the user always has the same routine during the whole week;
- daily habitual profile: based on a synthetic dataset created supposing that the user always has the same routine every day.

To better understand the grade of predictability of the single dataset, the Levenshtein distance has been calculated between sequences of activities registered during every week and during every day. In information theory, the Levenshtein distance is a string metric for measuring the difference between two sequences. The kind of operation considered for changing one word **a** into another word **b** are the following three:

1. insertion of a symbol;

2. deletion of a symbol;

3. substitution of a symbol;

and each operation has a unitary cost. The obtained values can be normalised by dividing the sum of the costs by the maximum length between word **a** and **b**. Therefore this value can be interpreted as the degree of similarity between two sequences where 0 indicates the perfect equality while 1 is the highest inequality value. The Levenshtein distance has been calculated between sequences of activities registered during every week and during every day, hence different values of distance are evaluated for each cross comparison between different days and between different weeks. Calculating the average between all the obtained values and subtracting this result from 1, three different grades of predictability are obtained:

- real-case profile grade = 0.47

- weekly habitual profile grade = 0.73

- daily habitual profile grade = 1

considering a scale from 0 to 1, where 0 indicates an unpredictable case while 1 is the most predictable case.

Real-case Profile

For the real-case profile, the algorithm was tested using the Aruba real-word dataset from the CASAS smart environment project of the Washington State University [15] with the addition of some other activities simulated as they were performed inside the house. The detailed description is in Section 2.3.1 and all the remarks made in the Section just mentioned apply to this real-case profile. The activities considered for this simulation are those listed in table 2.4.

Weekly Habitual Profile

The second profile is realised with a synthetic dataset obtained considering a user whose habits have a higher degree of predictability. Accordingly, the dataset was created considering the first week of activities performed by the real-case user and then repeating this same week for over 6 weeks of training data. The sequences of activities are different from day to day of the same week but are the same during the same day of all the different weeks. The preferred temperatures and the level of annoyance are evaluated in the same way as the real-case profile and as explained in Section 2.3.1.

Daily Habitual Profile

The last profile is the one associated with a user who has the exact same routine for every day of the dataset. The dataset is created considering one day in which the user performs all the possible activities, and that day is repeated for the entire dataset's length. This type of user profile is not particularly realistic, but it is useful for evaluating the degree of reliability in the predictions made by the algorithm explained in the 2.2.2 Section. Also for this case, preferred temperatures and annoyance levels are considered as explained in Section 2.3.1.

2.4 Simulations Results

The performance of the overall proposed system has been evaluated considering different possible scenarios applied to the three user profiles explained in Subsection 2.3.2. Therefore, the algorithm was tested taking into consideration the following three situations:

- Scheduling Based on LSTM Network-SBLSTM: bases its scheduling evaluations on the probability of using any of the appliances at time t, calculated thanks to the LSTM network described in 2.2.2. The scheduling is programmed trying to find the best trade-off between energy savings and user annoyance due to the time change, according to the scheduling algorithm described in Section 2.2.3; This solution avoids interactions between users and the system, considering only the system's previous knowledge about user habits.

- Scheduling Based on Perfect Time-SBPT: based on a perfect knowledge of the time in which the user wants to use some of the appliances in the house. This case coincides with the possible scenario in which the user instructs the system about the exact moment they want the appliance to start and the scheduling is programmed with respect to this preferred moment, but it has the disadvantage of requiring continuous interactions between users and system;

- Without Scheduling Algorithm-WSA: describes the scenario where appliances are normally used by the resident and the scheduling is never programmed.

The two cases based on the scheduling, namely SBLSTM and SBPT, are tested both in the case the scheduling algorithm includes the user annoyance (Scheduling With Annoyance in the results' graphs) and in the case it does not (Scheduling Without Annoyance in the results' graphs).

The datasets were divided into two parts consisting of training data and test data. The training data used to learn and understand the user's behaviour consists of 6 weeks of information about the activities usually performed by the user, while the test data refers to one week. In this way, the percentage of data used for the training phase corresponds to the 70% of the total of the dataset, while the test set

Validation Accuracy for Recognition Algorithm	Validation Accuracy for Prediction Algorithm
80.3%	82.7%

Table 2.5: Validation Accuracy

is made up of the remaining 30% of data; this separation of the overall dataset is the solution generally adopted in machine learning problems [96]. The training part of the dataset has also been use for the time-series split cross-validation, starting with a train duration of 3 weeks of data and raging 7 weeks, while the validation is made on the data of the one week following the training set, each time. This subdivision resulted in 5 validations. The mean values for the validation accuracy of the recognition algorithm and the prediction algorithm are shown in Table 2.5.

Due to the fact that the shortest duration for the activities under examination is around 15 minutes, while the longest activities can elapse for several hours, the duration of the intervals (Δt) defined in Subsection 2.2.2, in which the algorithm has to recognise the ongoing activity, has been chosen equal to 15 minutes.

Therefore the system makes the operations for recognise the latest activity, predict the future activities and evaluate the appliances scheduling every $n(\Delta t)$, with n varying between 1 and 96 during all the day due to the discretisation of the time of a whole day in intervals of 15 minutes each.

2.4.1 Activity Prediction Algorithm Performance Evaluation

As explained in Subsection 2.2.2, the predictions about the activities that are going to be carried out by the user in the future are made using an LSTM Network. For this purpose, a typical LSTM model has been built, counting two hidden LSTM layers with a number of 120 units for each layer. Every layer is followed by a dropout layer to prevent overfitting. The training epoch is set to 250, the loss function is set to crossentropy, and the optimiser is Adam. The last two layers are a softmax layer and a classification layer. The softmax function is indicated for multi-class problems with mutually exclusive classes [38].

The approach here proposed monitors all the user activities in real-time. The prediction of the next activity is made taking into account three kinds of information as explained before in Subsection 2.2.2: the $N = 4$ activities previously performed by the user and recognised by the activity recognition module; the current day of the week W_d specified as a number from 1 to 7; the time of day T_0 the system is currently located, specified as a number from $n = 1$, which corresponds the midnight of every day, to $n = 96$. These numbers n derive from the discretisation explained in Section 2.4.

The choice of the number $N = 4$ of previous activities that are needed for the

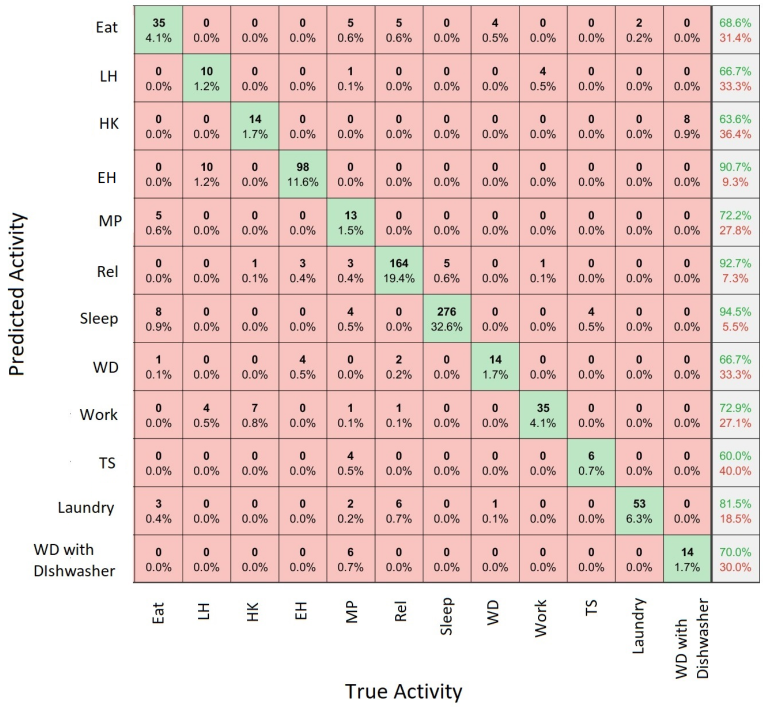


Figure 2.4: Confusion Matrix for the Prediction Algorithm for each Activity

prediction has been made considering the best results obtained according to the highest prediction accuracy. Indeed, in table 2.6, it is possible to see the results, for the real case profile, in terms of accuracy for different values of N previous activities considered, and it shows that the best result is obtained for $N = 4$. The prediction accuracy is established considering the number of times an activity is correctly predicted with respect to the number of times that this instance is incorrectly classified, predicting a label corresponding to a different activity. Therefore the accuracy is defined as:

$$Accuracy = \frac{T_l}{T_l + F_l} \tag{2.6}$$

in which T_l indicate the numbers of true positives, i.e. the number of times the activity is correctly predicted, while F_l indicates the number of false positives, therefore the number of times that the prediction is incorrectly classified. Fig, 2.4 represent the confusion matrix with the accuracy percentages of the predictions for all the activities considered during the simulations. Along the diagonal, the number of times in which the activity is correctly predicted is reported; while the column on the right shows the prediction accuracy for each class or activity. From this figure it is possible to note that most of the errors concern the activity that are less frequently performed.

The prediction of the following activities is also made for every $n(\Delta t)$ interval of duration equals to 15 minutes. More specifically, for every interval of time t_0

	N=2	**N=3**	**N=4**	**N=5**	**N=6**
Prediction Accuracy	0.851	0.854	0.864	0.859	0.857

Table 2.6: Activity Prediction Accuracy

during the testing week, the algorithm schedules appliances that are going to be used every $n\Delta t$ time intervals after t_0, according to the predictions of the activities that the Activity Prediction module has made for intervals t after t_0 and up to 9 hours forward in the future. Every 15 minutes then, the scheduling algorithm has to re-evaluate its scheduling consideration based on the new information about the previously performed and recognised activities and the new prediction obtained from the LSTM network, with always the aim of finding the best compromise between energy savings and user's comfort. In order to evaluate the performance of the system in terms of the capability to predict the next activities correctly so that the appliance scheduling can be made successfully, the first tests were run for the three different profiles described in Section 2.3.2, considering only the SBLSTM case, i.e. the case that makes use of the LSTM network. The results in terms of scheduling errors due to errors in predictions are shown in table 2.7. Three types of error can be considered:

- the number of missed schedules that is calculated considering the total number of times each appliance i has been scheduled during the week N_{sched}^{week}, and counting the number of missed schedules, that is the times when the user would have liked the appliance to be switched on but that did not happen N_{sched}^{miss}. The appliance is considered unscheduled when the wait to turn it on reaches the maximum degree of annoyance for the user.

  Therefore the error for appliance i is obtained using the following equation:

$$E_i^{miss} = \frac{N_{sched}^{miss}}{N_{sched}^{week}} \tag{2.7}$$

- the number of additional schedules, calculated considering the total number of times each appliance i has been scheduled during the week N_{sched}^{week}, and then counting the times when the algorithm has predicted that the user would have liked the appliance to be switched on, but that was not true, and it is indicated as N_{sched}^{add}.

  Therefore this error is calculated as:

$$E_i^{add} = \frac{N_{sched}^{add}}{N_{sched}^{week}} \tag{2.8}$$

	Weekly Habitual Profile			Real Case Profile		
	$\mathbf{E_i^{miss}}$	$\mathbf{E_i^{interval}}$	$\mathbf{E_i^{add}}$	$\mathbf{E_i^{miss}}$	$\mathbf{E_i^{interval}}$	$\mathbf{E_i^{add}}$
Dishwasher	0	0.01	0.20	0	0.01	0
Washing Machine	0	0.03	0	0	0.04	0.14
Water Heater	0	0.04	0.45	0	0.06	0.47

Table 2.7: Scheduling Errors due to Prediction Algorithm

- the difference in terms of time intervals between the one predicted as the user's preferred time for switching on the appliances and the ground truth value of that time.

  This error is considered taking into account the total amount of intervals I of 15 minutes inside the scheduling windows of 6 hours, that has been considered during the test. The choice of this duration for the scheduling window has been done taking into account the best compromise between computational complexity and scheduling results. Indeed, to be sure that the time to correct any prediction errors is enough,a fairly large window is needed. Therefore, this error represents the times in which the algorithm correctly predicts that the user would like to use one of the appliances i, but there is a miscalculation on the exact time interval of activation. This difference in the time interval is calculated by counting the number of n intervals that separate the true time interval from the one predicted by the algorithm and then comparing it to I.

  Therefore this error is calculated as:

$$E_i^{interval} = \frac{n}{I} \tag{2.9}$$

The first two errors are reported in the first and in the third column for each profile of Table 2.7; while the third error reported in the list is shown in the second column for each profile.

The results in Table 2.7 and only those relating to two of the three profiles described in subsection 2.3.2, specifically the results about the weekly habitual and the real case profiles are reported. That is because in the situation of a user with a daily habitual profile, the errors are always 0. Indeed, the predictions are always right since the training data for the LSTM network and the test data are exactly the same.

When considering the other two profiles, there are some errors in the additional number of schedules, especially in the use of the water heater. This is due to the fact that for this appliance the value of the threshold needed to establish its scheduling has been chosen very low because, in relation to the user's comfort, it is still better to be sure that the appliance is switched on whenever it is needed, and even for

Weekends, holidays and everyday from 21:00 to 8:00	**Everyday from 8:00 to 21:00**
0.00089 €/kWh	0.0012 €/kWh

Table 2.8: Energy Pricing

the times that are predicted by mistake, then to have the water heater off when the user really wants to use it. On the other hand, it is possible to see that, in any of the considered profile scenarios, there are no cases in which the algorithm misses the scheduling of the appliances, therefore making consistently accurate predictions on the fact that the user would have carried out certain activities and that the activation of the corresponding household appliance has occurred correctly. The errors in terms of time intervals in which the appliance is scheduled are always very low. Indeed, for the most part, the algorithm predicts the activities connected to the appliances one interval of time before the right time. Anyway, as it will be presented in the following set of results, this condition is further evaluated in how much this difference from the preferred operating time annoyed the user.

2.4.2 Scheduling Algorithm Performance Evaluation

The proposed scheduling system has been then assessed in terms of energy savings and user's comfort, which represent the main objective of the system as a whole. The evaluation of the energy costs during the test week has been made considering the tariffs listed in Table 2.8, based on some typical Italian time-of-use tariffs set by the company ENEL during the year 2021.

The results in Fig.2.5 show the cost savings obtained during the test week thanks to the scheduling of the appliances. The graph shows the percentage savings obtained in the scenarios in which the scheduling of household appliances takes place compared to the case where no scheduling program is evaluated.

These results are obtained taking into account the fact that cost savings are coming from a scheduling of switching controlled high loads to hours where the energy has lower prices or to hours where there is a production of energy thanks to the Renewable Energy Sources and considering that there is a reduction in energy consumption due to a better optimisation in the usage of the water heater, which is switched on only at times of interest for the user and not every time the temperature drops below a certain value.

The two scenarios, in the Scheduling Without Annoyance case, base their evaluation scheduling only considering energy savings and not taking into account the possible discomfort of the user due to the scheduling. This is reflected in an even lower cost of energy at the end of the week as reported by the higher savings values shown by the graph series Scheduling Without Annoyance compared to the graph series Scheduling With Annoyance in Fig.2.5. Even if the algorithm does not take into

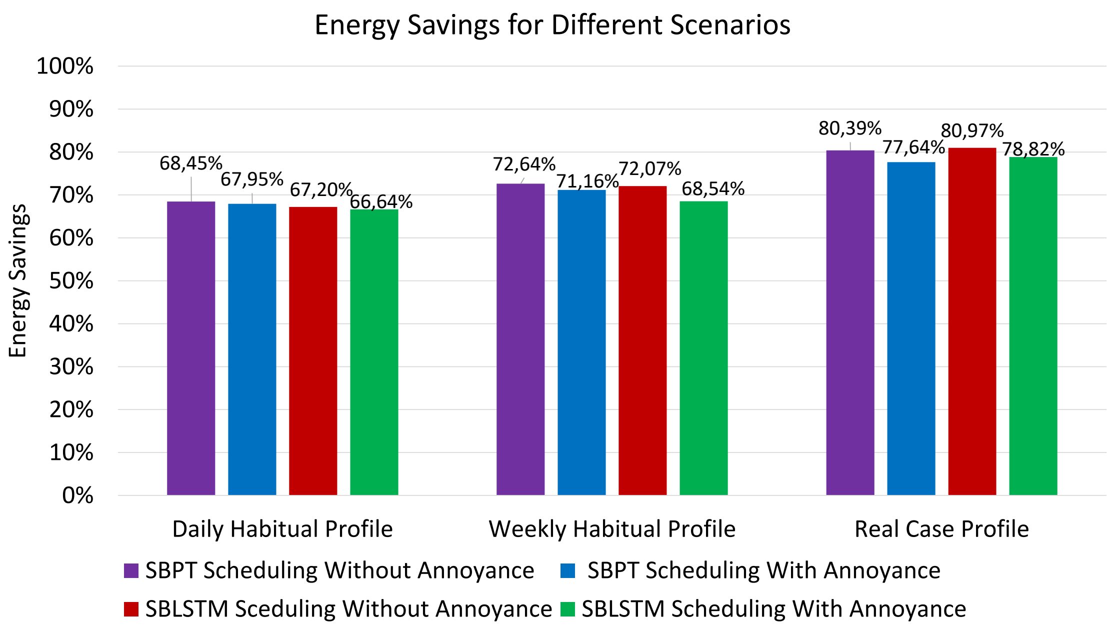


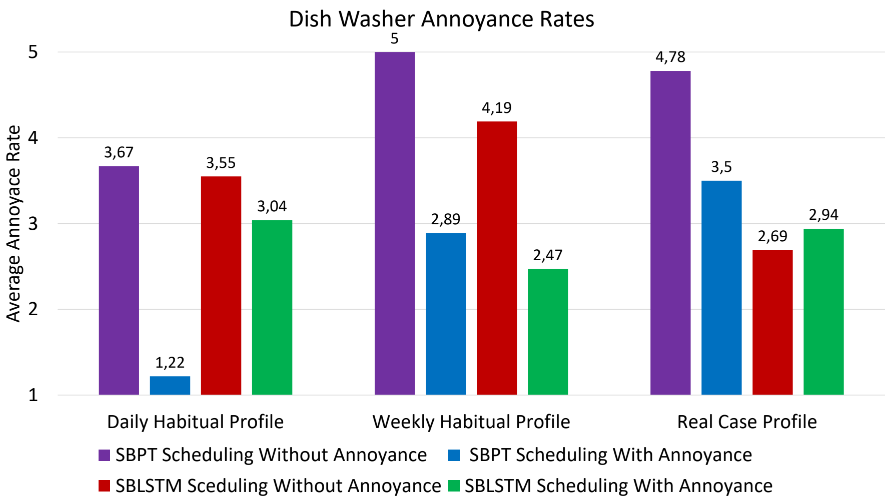
Figure 2.5: Energy Savings with respect to the WSA Scenario

consideration the annoyance of the user, the scheduling of the controlled appliances is anyway programmed in a range of ± 3 hours from the preferred time.

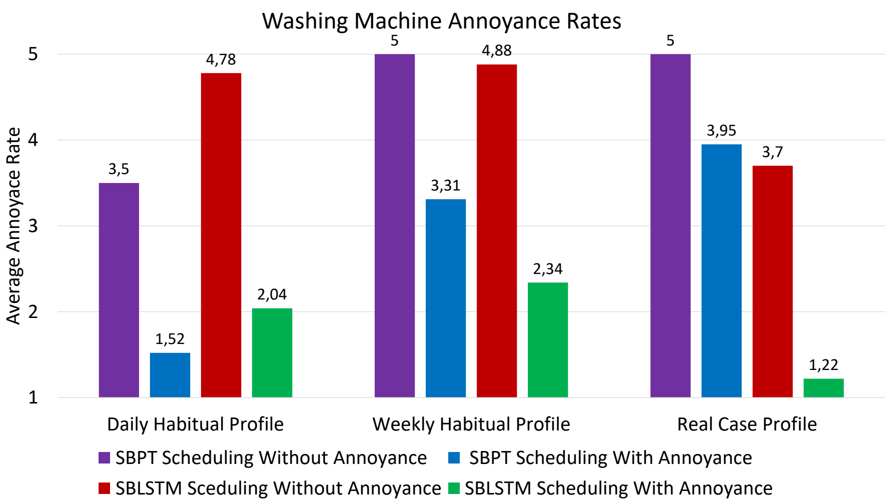
With respect to user annoyance, the system evaluates the user's discomfort due to the time shifts of the scheduling, considering a scale from 1 to 5. As explained in Subsection 2.3.1, level 1 of annoyance indicates that the user did not feel any discomfort in the time change, while a level of annoyance equal to 5 indicates that the user had the highest discomfort due to this time alteration.

Fig. 2.6 shows the mean level of annoyance obtained for each appliance, considering the three profiles and the combination of the different scenarios that presuppose the scheduling of the appliances. Each figures is referred to one of the three controllable appliances in the house, i.e. the washing machine, the dishwasher and the water heater; the comparisons concern the two scenarios SBPT and SBLSTM, that are considered in the two cases of Scheduling With Annoyance or Scheduling Without Annoyance.

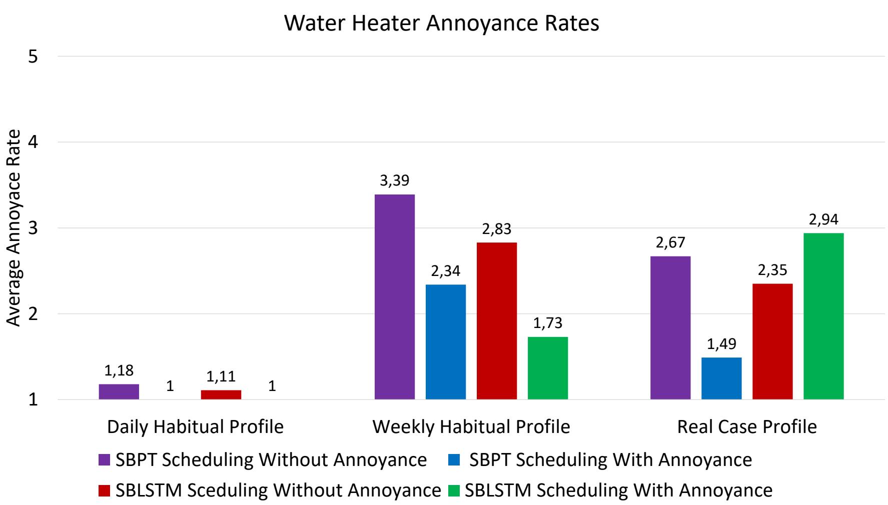
In most cases where the annoyance of the user is considered before taking the decision of the better time for the scheduling algorithm, the annoyance rate is lower than value 3. There are some exceptions that can be explained both in the possible errors while predicting the right preferred moment for the user to switch on the appliances and sometimes in the randomness of certain user preferred moments that are very distant from the hours of greatest energy saving. In every situation, it is anyway possible to note how the annoyance rate for the user in the scenarios concerning the Scheduling Without Annoyance is always higher than in the scenarios



((a)) Dishwasher



((b)) Washing Machine



((c)) Water Heater

Figure 2.6: Average annoyance rate for each appliance considering SBPT and SBLSTM Scenarios and Scheduling With or Without Annoyance

that take into account the user's discomfort. Therefore, against a slight improvement in energy savings over the week, there is a significant increase of the annoyance rates. The only exception that seems to differ from the other results is referred to the use of the water heater in Fig. 2.6(c). In this case, the annoyance rate is always very low, even for the scenarios of Scheduling Without Annoyance. This is due to the fact that in this case the discomfort is evaluated considering the water temperature and not the hour in which the appliance is switched on, therefore if the temperature is considered acceptable by the user regardless of the time, the annoyance rate does not increase.

Then, the scheduling system here proposed, which considers both energy savings and user annoyance, is capable of ensuring energy savings making better use of appliances and resources while preserving the user's preferences.

2.5 Conclusion and Future Works

This part of the thesis focuses on a solution for energy and comfort management inside buildings, with the purpose of reducing energy waste thanks to a proper control over appliances while at the same time ensuring the well-being of users. To this aim, a BECM system is proposed that integrates a solution for two different problems: the first one concerns the need for such a system to be able to know users' behaviour and preferences and to predict usual activities; the second is about the necessity to manage appliances with respect of that behaviours and preferences and with respect of energy consumption.

The system has been tested in a real scenario, evaluating if the predictions were correct and proposing a coherent scheduling that could guarantee energy savings. The obtained results show that, as expected, the scheduling of the appliances can guarante energy savings, reducing consumption over a week of at least 66.64% in comparison with the classic use of energy and appliances. The prediction module permitted a quite accurate scheduling based on probabilities, even if some of the activities have given some problems because the statistical data about them were based on a few instances. Furthermore, it was possible to guarantee that the annoyance rate was never too high, thus respecting user comfort.

The main limitations of this work concern the fact that the system was tested on a single dataset with a single inhabitant within the reference scenario. Some activities are performed infrequently, so more data needs to be obtained. Furthermore, the data on the consumption of controllable household appliances are based on presumed data and take into account the "worst case" (in terms of cycle duration and consumption). In contrast, a study on the usage habits of the specific appliance, based on historical consumption, could improve the algorithm scheduling program. Improving the algorithm for the recognition of the activities and their prediction, integrating the system with other technologies (i.e. data from wearable, information obtained from the analysis of Wi-Fi signals), would lead to a more accurate knowledge of the activities and make both recognition and prediction more reliable.

Hence, future works will investigate the proposed system's adaptability to different real-case scenarios, trying to improve the prediction module considering a larger training phase and more instances of the activities and corresponding use of appliances. Furthermore, it will be evaluated how different consumption cycles for the use of the appliances can generate different predictions in the energy consumption analysis and improve these predictions.

Chapter 3

Energy Consumption Profiling of Appliances inside Smart Buildings

3.1 Introduction

As discussed in the previous Section, it has been demonstrated that building occupants' behaviour has a major impact on energy efficiency, not only considering energy consumption per se but also on energy cost [40][16]. Indeed, energy efficiency encompasses all the measures to either reduce the total energy consumption, e.g. by reducing the appliance usage, or shift the consumption when it is more convenient, e.g. off-peak hours of time-of-use rate plans or, in the case RES are installed, when RES power is produced. To this aim, two of the main actions to reduce the impact of buildings on greenhouse gas emissions are: raising awareness among building occupants about their energy consumption habits, and scheduling appliances to more convenient times. Both actions require to have knowledge of the power consumption profiles of the building appliances, which varies significantly over time, with peaks and spikes that have very different values as compared to the nominal consumption power.

Therefore, the main idea is to develop a system that identifies appliance profiles based on power consumption data gathered by smart meters. The system continuously monitors the power consumption of the appliances connected to the smart meters to extract the features that best characterise their consumption profiles. Such features are later used by a clustering algorithm to determine the clusters that better represent the energy consumption profiles that can be associated with each appliance. Accordingly, the parameters associated with each cluster can be used to convey relevant information about occupants' energy consumption habits, either to the occupants themselves or to a Smart Building energy and comfort management system.

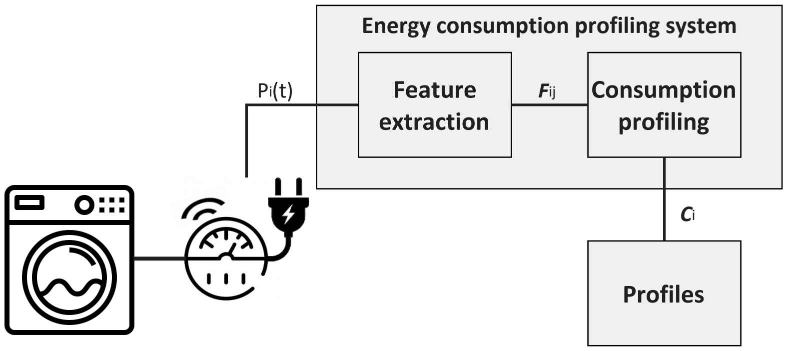


Figure 3.1: Scheme of the reference scenario and proposed system

3.2 System Overview

The reference scenario considered in this part of the thesis is that of a Smart Building where appliances are connected to smart meters so that their energy consumption can be monitored and profiled, as shown in Figure 3.1.

Each appliance i is connected to a smart meter s_i, which continuously monitors its power consumption and sends its values $P_i(t)$ to the *energy consumption profiling system*. Such a system is made of two components: the *feature extraction* and the *consumption profiling* blocks.

The feature extraction block collects all the $P_i(t)$ values and analyses them. More specifically, whenever appliance i starts consuming an amount of power enough higher than 0, the feature extraction block associates the following $P_i(t)$ values to a specific *consumption cycle* j. Once the consumption cycle is considered finished, i.e. when the measurements of the power reach a value lower than a threshold and this value is maintained for quite a long time , the feature extraction block computes the set of parameters that characterise the consumption cycle, i.e. its set of features $\mathcal{F}_{ij}$.

The set of features is then sent to the consumption profiling block. Once a sufficient number of sets of features is available for a specific appliance i, the consumption profiling block starts the clustering process on all the sets of features associated with appliance i. The output of this process is given by the set of clusters $\mathcal{C}_i = \{\mathcal{C}_{mi}\}$ identified by the clustering process for appliance i, each of which corresponding to the set of consumption cycles that are grouped in the cluster. For each cluster $\{\mathcal{C}_{mi}\}$, a *reference consumption cycle* $\mathcal{P}^*_{mi}$ is then computed.

3.3 Energy Consumption Profiling System

As described in the previous Section, every appliance in the Smart Building is associated with different energy consumption cycles that depend on the users' habits for that appliance. The data on the power absorbed by each load is recorded and then grouped according to different consumption cycles. This information is used to recognise similar patterns automatically, thanks to clustering algorithms. The approaches used to design the blocks that compose the energy consumption profiling system outlined in the previous Section are described in detail in the following Subsections.

3.3.1 Feature Extraction Block

The feature extraction block is the module of the system in charge of extrapolating meaningful information from raw data in order to obtain a feature vector representative of the different appliance consumption cycles.

In order to separate the different cycles, first, the raw data have been manipulated to eliminate some errors and noise. All the power values $P_i(t) \leqslant 1W$ have been considered measure errors of the smart meter and cleared. Then, the beginning of each cycle is considered whenever there is a time t in which $P_i(t) > 1W$, while the end of it is considered for time t when $P_i(t) < 1W$ for at least 15 consecutive minutes. This first processing only removes the major errors due to the smart meter's accuracy, but it is not sufficient to determine the duration of the consumption cycle precisely. Therefore, the data are again manipulated to find more defined consumption cycles. From the observation of the consumption cycles obtained, it has been found that the consumption peaks convey interesting information about the consumption cycle. Therefore, for each cycle obtained from the first division, the first and the last peaks $P_i(f), P_i(l)$ are taken into consideration. The important information is the time in which these peaks occur and their values. The starting operational time is then considered as the time t where $P_i(t) \geqslant \sigma_f \cdot P_i(f)$, while the ending time is the time t for which $P_i(t) \leqslant \sigma_l{}_i(l)$. The values of σ_f and σ_l have been set to 0.1. Indeed, the power values read from the smart meters tend to drop really fast whenever they concern only fluctuations in the network and not real data about the use of the appliance, so a value 10% lower than the peaks can be considered as an error and discarded.

Once the cycle is obtained, it has to be described by different representative features which take into account the information extrapolated from the raw data. For each cycle, the feature extraction block computes the feature vector $\mathcal{F}_{ij}$ based on the information relating to consumption, power, duration and waveform characteristics. More specifically, it has been observed that the most representative feature vector consists of the following 8 features: the energy consumption at the end of the cycle; the mean value of power; the max and min values of power; the duration from start time to end time; the number of peaks and the average distance between them and lastly the waveform shape factor.

Therefore, for each consumption cycle, a feature vector is created that describes the different usage that the user can engage over the same appliance i, and that is then compared and grouped with other similar feature vectors. The set of all the vectors is then used as input for the consumption profiling block.

3.3.2 Consumption Profiling Block

The consumption profiling block finds similar patterns between feature vectors in order to obtain different consumption groups for each appliance. For that, a clustering algorithm is needed in the process to divide the data based on their attributes and organise them into meaningful structures.

Given the numerical and statistical nature of the considered features, among the different clustering algorithms and methods, the k-means algorithm has been chosen in order to cluster the data into similar groups [12]. In fact, this algorithm iteratively searches for the centroid that represents the cluster and at every step each data point is assigned to its nearest centroid, based on the squared Euclidean distance. The centroid is recomputed taking the mean of all data points assigned to that centroid's cluster.

In this case, the k-means algorithm is applied separately on each appliance features dataset so that the obtained set of clusters $\mathcal{C}_i = \{\mathcal{C}_{mi}\}$ is already differentiated for every appliance i. Therefore, each of these clusters corresponds to one energy consumption group.

Different consumption cycles j are associated with different energy consumption groups as the algorithm clusters them; the purpose of the consumption profile block is then to find the *reference consumption cycle* $\mathcal{P}^*_{mi}$ that is representative for the entire cluster.

Therefore, the reference consumption cycle is computed taking into account the mean power consumption, minute by minute, of all the consumption cycles j that belong to the same cluster $\mathcal{C}_{mi}$:

$$\mathcal{P}^*_{mi} = \text{mean}_j(\mathcal{P}_{ij}(t) : \mathcal{P}_{ij}(t) \in \mathcal{C}_{mi})$$

Since the cycles belonging to the same cluster have different duration of time, this reference consumption cycle has the same duration as the most extended cycle in the same group.

The information that can be extrapolated from it is used as representative of the entire cluster.

3.4 Performance Evaluation

The performance evaluation of the proposed system is done using the DEDDIAG dataset [92], which is a real-world domestic electricity demand dataset. The dataset contains data from 15 different homes about power measurements of a total of 50 appliances that have been recorded at a frequency of $1Hz$. The appliances that are reported in the entire dataset are dishwashers, washing machines, dryers but also refrigerators, coffee machines and TVs.

At this early stage, the main focus of the system is on monitoring and profiling the consumption of high-consuming controllable appliances, therefore the only data

	Dishwasher	Washing Machine	Dryer
Cluster 1	38.5%	35.6%	7.8%
Cluster 2	7.3%	1.7%	18.6%
Cluster 3	1.2%	12.1%	24.5%
Cluster 4	27.8%	45.4%	1.0%
Cluster 5	25.2%	5.2%	48.1%

Table 3.1: Percentage Cluster Population

that are considered are those about the dishwasher, the washing machine and the dryer. Furthermore, as a preliminary assessment of the system, the inspected data entirely comes from one single house among the 15 reported in the dataset, i.e. House number 12.

For this home, the monitoring over the appliances was performed for 320 days, whit a rate of missing days of recording around 3%. The measurements are recorded so that only value changes are stored; therefore, in order to be more easily processed, the data have been padded to fill the gaps in the incomplete datetime series so that a power consumption sample is available for every minute of the recording period time.

From this data, the different consumption cycles are elaborated as explained in Section 3.3.

The dishwasher's data have produced 238 consumption cycles, while for the washing machine and the dryer there are only 190 and 107 cycles, respectively. The datasets composed of the feature vectors obtained by the feature extraction block from the consumption cycles mentioned above are used to obtain the clusters for each appliance that best represent the consumption cycle groups. In order to understand the best k number of clusters, different tests have been performed, changing the number of k from $k = 2$ to $k = 15$ and then evaluating the optimal cluster number result. This evaluation has been made using Calinski-Harabasz criterion values [12]. The Calinski-Harabasz criterion gives the highest Calinski-Harabasz index value to the solution that have a large between-cluster variance and a small within-cluster variance with respect to the number k.

During this simulation, the optimal number of clusters obtained for all the three considered appliances was $k = 5$. For each of these clusters, the reference consumption cycle is obtained as explained in Section 3.3.2. In order to ensure that, within the cluster, the number of outliers was not too high and that they could not negatively influence the approximation of the reference cycle, some of these consumption cycles j were dropped from the cluster. In particular, the cycles defined as outliers were all the elements of $\mathcal{C}_{mi}$ with power values greater than three standard deviations from the mean for every minute. In this way, the total number of cycles decreased

	Energy Consumption	**Mean Power**	**Max Power**	**Min Power**	**Duration**	**Peaks Nr**	**Peaks Distance**	**Shape Factor**
Cluster 1	52777	463.0	1571.9	0.05	114	32	3.4	1.6
Cluster 2	4505	21.1	2230.3	3.8	214	62	4.0	2.5
Cluster 3	53619	646.0	1571.1	0.9	83	34	3.4	1.5
Cluster 4	60183	573.2	1570.3	0.1	105	26	3.2	1.4
Cluster 6	47449	481.9	1562.4	0.4	98.5	33	4.0	1.5

Table 3.2: Dishwasher Features for Each Reference Cycle

	Energy Consumption	**Mean Power**	**Max Power**	**Min Power**	**Duration**	**Peaks Nr**	**Peaks Distance**	**Shape Factor**
Cluster 1	27131	130.4	2150.1	0.2	208	69	3.6	4.1
Cluster 2	31767	392.2	1927	0.2	81	25	5.3	3.9
Cluster 3	58217	383.0	2039.1	0.1	152	43	3.3	1.9
Cluster 4	19690	180.6	2179.9	0.2	109	31	4.3	4.8
Cluster 5	45739	304.9	2183.2	0.9	150	48	3.3	3.0

Table 3.3: Washing Machine Features for Each Reference Cycle

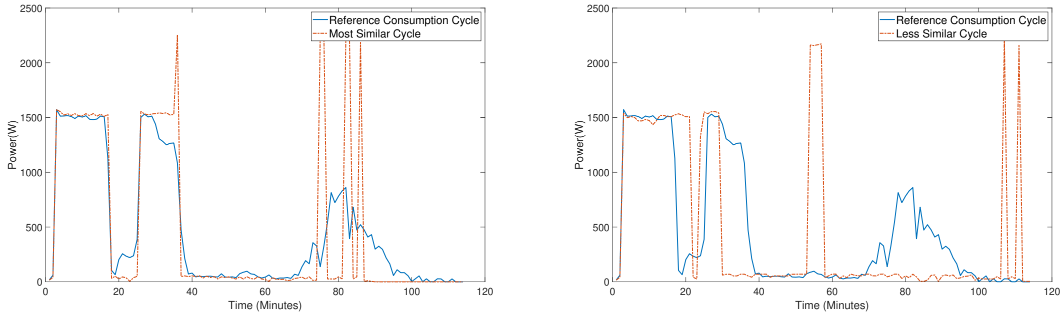
to 234 for the dishwasher, to 174 for the washing machine and to 102 for the dryer. Table 3.1 shows the percentage of elements that belong to each cluster for every appliance.

Tables 3.2, 3.3 and 3.4 show the values for the 8 features of the reference consumption cycles for each cluster of the dishwasher, the washing machine and the dryer, respectively.

In order to understand how the consumption cycles of the same cluster could differ from the reference one, every one of them has been considered based on their distance from the cluster centroid. Therefore the one that is less similar to it is far away from the centroid, while the one more similar to it is considered the one closer to the centroid.

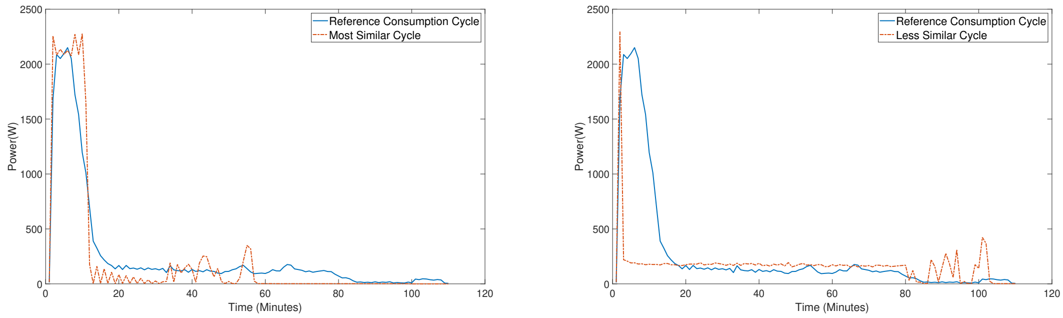
	Energy Consumption	**Mean Power**	**Max Power**	**Min Power**	**Duration**	**Peaks Nr**	**Peaks Distance**	**Shape Factor**
Cluster 1	91153	1688.9	2669.8	0.5	54	16	4	2.5
Cluster 2	411590	1470.0	2626.9	0.2	280	88	4.1	2.6
Cluster 3	79944	1211.3	2457.8	0.4	66	20	3.9	1.6
Cluster 4	71296	1584.4	2630.3	0.2	45	15	3.5	2.1
Cluster 5	128310	1379.7	2677.1	1.1	93	31	3.3	2.5

Table 3.4: Dryer Features for Each Reference Cycle



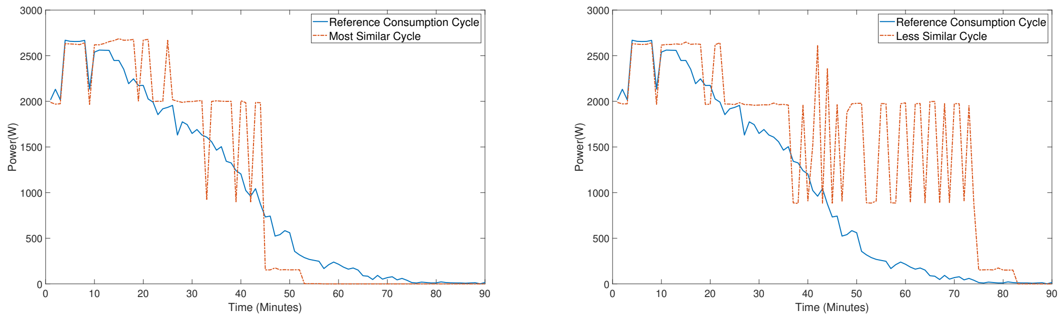
((a)) ((b))

Figure 3.2: Comparison between the most similar cycle (**a**) and the less similar cycle (**b**) for Cluster 1 of the Dishwasher with respect to the reference consumption cycle.



((a)) ((b))

Figure 3.3: Comparison between the most similar cycle (**a**) and the less similar cycle (**b**) for Cluster 4 of the Washing Machine with respect to the reference consumption cycle.



((a)) ((b))

Figure 3.4: Comparison between the most similar cycle (**a**) and the less similar cycle (**b**) for Cluster 5 of the Dryer with respect to the reference consumption cycle.

	Dishwasher		**Washing Machine**		**Dryer**	
	Estimated Energy Value	**Percentage Error**	**Estimated Energy Value**	**Percentage Error**	**Estimated Energy Value**	**Percentage Error**
Cluster 1	52777	1.1	27131	4.7	91153	0.4
Cluster 2	4505	0.4	31767	0.2	411590	2.0
Cluster 3	53619	3.2	58217	0.9	79944	2.5
Cluster 4	60183	0.7	19690	9.4	71296	0
Cluster 5	47449	0.8	45739	0.4	128310	4.1

Table 3.5: Total Energy Consumption

As an example, fig. 3.2 shows the two graphs that compare the reference consumption cycle of the dishwasher's cluster 1 with the two extreme cycles selected as explained above. Fig. 3.3 shows the same results for cluster number 4 of the washing machine, while fig. 3.4 shows the graphs for cluster number 5 of the dryer. The clusters chosen to be represented here are those that have the higher number of elements in them, as shown in table 3.1.

From these images, it is possible to see that, especially for the most similar consumption cycles, the reference one is a good approximation of the entire waveform, also for the peaks of the signals. The main problems of the less similar consumption cycles can be found in the different duration, as it is shown in the case of the washing machine 3.3(b) and the dryer 3.4(b), or in the displacement of power peaks for the dishwasher 3.2(b).

It has then been investigated how much the approximation of using the reference consumption cycle in place of the real consumption cycle, e.g. if it was used to predict the expected consumption of an appliance in an energy and comfort management system, would affect the performance of the system. To this end, the results of Tables 3.5 and 3.6 show the errors in the energy consumption obtained at the end of the entire cycle operations and on the values of power measurements, time interval by time interval, with respect to the average power value of the reference cycle.

The errors considering the overall energy consumption are shown in Table 3.5. The table presents the errors regarding the percentage difference between the overall estimated energy value obtained from the reference consumption cycle and the energy values of the other cycles of the same cluster, considering each cluster and every appliance. The errors equal to 0 refer to clusters that had a single element within them.

From the table rows, it is possible to see that, in general, the percentage error is always less than the 10% in the total consumption of the cycle. The best results are obtained for the dishwasher, for which the cycles generally have a trend with a lower number of peaks and are, therefore, more efficiently represented by a reference cycle which is the average of their values. For all the appliances, it is also possible to note

	Dishwasher		**Washing Machine**		**Dryer**	
	Mean Power	**Error Interval**	**Mean Power**	**Error Interval**	**Mean Power**	**Error Interval**
Cluster 1	463.0	±79.0	130.4	±25.2	1688.0	±219.4
Cluster 2	21.1	± 2.0	392.2	±89.1	1470.0	±277.2
Cluster 3	646.0	±101.0	383.0	±80.3	1211.3	±249.3
Cluster 4	573.2	±81.0	180.6	±20.1	1584.4	±0.0
Cluster 5	481.9	±77.6	304.9	±29.5	1379.7	±253.2

Table 3.6: Minute-wise power errors in $W \cdot min$

that the errors are always very low even for the clusters that were less populated, i.e. cluster 3 for the dishwasher, cluster 2 for the washing machine and cluster 1 for the dryer, not considering in this last case the cluster number 4 that have only 1 element inside.

The other type of error considered is the difference in the values of power estimated for each interval of time. In this case, the considered time is a one-minute interval and therefore, the error is computed minute by minute until the end of the reference consumption cycle. Table 3.6 shows, in column 1, the average power value that occurs for each minute when considering the reference consumption cycle of each cluster, while column 2 shows how much the individual cycle values can deviate from the average for each interval. In this case, it is possible to see that the errors are quite low if compared to the mean value, and again the worst errors are found for the clusters that have fewer data and are harder to define with a single average representative cycle. Also for this table, the rows with 0 values refer to clusters composed of only one element.

3.5 Conclusion and Future Works

This Chapter presents an energy consumption profiling system which, based on the power consumption of appliances monitored by smart meters, infers the consumption profiles that best describe their most common usage characteristics. The system first detects the most relevant features that describe the consumption cycles of appliances. Then, based on such features, it clusters the sets of consumption cycles associated with each appliance using the k-means algorithm. Simulation results on a real-case dataset show that it is possible to find different consumption profiles according to users' habits of appliance usage. From the clusters obtained it is possible to use a single reference consumption cycle that can be used as the representative for all the cycles of the cluster with good approximation and acceptable errors, both in the overall energy consideration and in the power instantaneous values referred to specific intervals.

The significant limitations encountered in this work concern using a single dataset referring to a single house, with data relating to some household appliances. The presented approach is based on the average of the various cycles, while simulations using other metrics, such as the barycenter, could give different insights for the reference cycle. Another interesting aspect is the study of additional features depending on the appliance considered, while in this work, the same group of features is evaluated for all devices. The reference profiles presented here take into account only the total energy consumption of the cycle; by recognising particular features relating to specific appliances, profiles can be defined considering other characteristics (time of use, temperatures, days of the week).

Future works will be focused on analysing if different sets of features have different results depending on the appliance considered. Furthermore, diverse types of appliances such as HVAC, microwave ovens and laptops will be investigated. Finally, the system will be integrated into the energy and comfort management System presented in Chapter 2. In this way, the prediction about the activities linked to the appliances will be accompanied by forecasts on energy consumption differentiated by different usage profiles, making the system more complete and improving the forecasts.

Chapter 4

Daily Activities Monitoring for Well-Being and Stress Correlation

4.1 Introduction

The market of wearable devices is constantly growing and expanding, enabling sensors, actuators and communication interfaces to be embedded into wrist-worn devices, garments, and jewellery [31]. As such devices are becoming more and more common, also the so-called techno-sceptics are starting to accept their everyday use, not only for fitness tracking, but also to monitor their own health. Accordingly, many systems have been developed to recognise simple repetitive body motions such as running, walking, sleeping, sitting, getting up or going up and down the stairs based on the data collected by accelerometers, gyroscopes, magnetic sensors, and other wearable sensors [68][6]. A step ahead can be taken considering that these data provide relevant information also about their owners' health conditions and quality of life [35] [10].

Indeed, it has been amply demonstrated that individuals' habits, either positive or negative, affect their psycho-physical health. The literature presents several studies that prove the correlation of lousy sleep and lack of physical activity with stress and illness conditions [80]. Accordingly, by early recognising or even predicting the triggers that cause bad health conditions, it can be possible to warn people in order to discourage negative behaviours and correct their habits. On the other hand, by understanding which activities correspond to an improvement in people's health, it is possible to encourage such behaviours, thus enhancing their quality of life.

The system proposed in this Chapter makes use of common commercial wrist-wearable devices to collect data about users' activities and habits, and find their correlation with self-assessed stress and illness conditions. The objective is to take a first step towards the development of the criteria and methodology that enable automatic recognition of the parameters that affect users' health conditions, so that such conditions can be later predicted as soon as the first signals are detected by

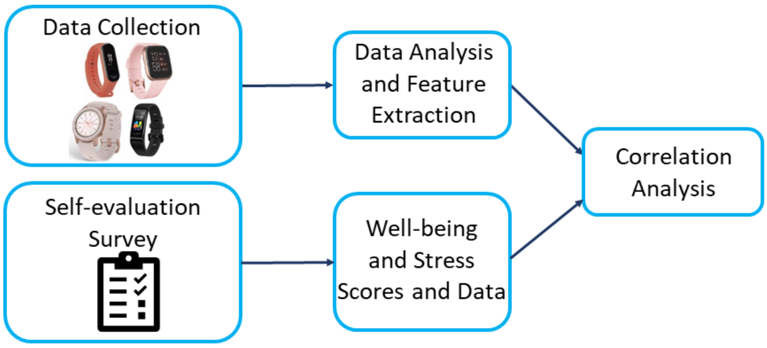


Figure 4.1: Overview of the proposed system

wearable devices. A first prototype has been realised to determine if the data about sleep, fitness activities and walking habits, collected by the wearable devices of some volunteers, have a correlation with their stress levels and illness conditions as self-assessed by them. The results show that it is possible to find a correlation between data about users' activities and night sleep and the level of well-being perceived by users themselves, while monitoring them with simple commercial smartbands already owned by the users.

4.2 Proposed Methodology and Design of Preliminary Prototype

The system proposed in this Section makes use of the data collected by common commercial wrist-wearable devices to understand their correlation with low or high stress levels and illness conditions. In the following Subsections, the proposed methodology and its implementation in a preliminary prototype will be described in detail. More precisely, how data are collected from wearable devices of different brands and models and appropriately prepared to be later processed is described in Subsection 4.2.1. The questionnaire that has been given to volunteers so that they could self-assess their stress levels and illness conditions is defined in Subsection 4.2.2, whereas Subsection 4.2.3 describes the logic and metrics that have been considered to find the correlation between the collected data and the questionnaires' results. Figure 4.1 shows an overview of the proposed system and its main components.

4.2.1 Data Collection from Wearable Devices

In order to collect information on user habits and on the activities generally performed by them, typical wearable devices such as smartwatches and smartbands are used. In order to reduce the burden of designing specific interfaces to collect data from devices of different brands and models, the Google Fit platform is used. Indeed, Google Fit is an open ecosystem that allows the synchronisation of data collected with a multitude of different wearable devices. Such data are tied to the user's Google account and can be easily shared with other apps. Furthermore, specific Android and Representational State Transfer (REST) Application Programming Interfaces (APIs) are provided to help app developers collect and use fitness-related sensor data in their applications. In particular, REST requests return structured data, generally under the form of JavaScript Object Notation (JSONs) or eXtensible Markup Language (XML), named resources.

Data collected by the Google Fit APIs are then appropriately saved by a Google App Script in datasheets, to be later processed. The Google App Script is a cloud-based JavaScript scripting language that allows to extend many of the features of Google Apps and to create lightweight cloud-based applications. The app script created to collect and store wearable data creates different buckets obtained daily and divided according to the type of activity that was carried out. A bucket represents a set of data aggregated over a certain time interval and according to specific requests. Buckets can be: time buckets, which collect all the required information for specific hours, days or time intervals; session buckets, which describe sessions, i.e. time spans in which specific metadata are recorded, such as precise training sessions or continuous activities (e.g. sleep); activity type bucket, which contain all the information associated with the type of activity carried out; activity segment bucket, where a segment consists of a continuous sample of the same activity over time. The most appropriate bucket type for the proposed solution is the activity segment bucket, since it is more representative of real-time requests that could be made any time of the day.

In the Google Fit storage, data are associated with the sensor that has collected them, namely the data source. By using the POST method, the elements are created by making requests for each of these data sources. Aggregate data is obtained from each of them, in this case, according to each activity segment.

Access to user data is allowed only with prior authorisation from the user; indeed, Google APIs make use of the OAuth 2.0 authorisation and authentication protocol. Therefore, in order to receive the user data, the application first needs to obtain an access token that gives it permission to access the API. Requests made by the application require an authentication step in which users log in with their Google account. After logging in, they are asked if they want to grant one or more permissions required by the application. This process is called user consent. The requested authorisations provide the following permissions to Google Fit scopes: read data about the carried out activities; read information about the user's physical characteristics; receive sensor data about the user position; receive sensor data about the heartbeat; read the data related to the *sleep* activity.

The script is automated to transcribe data day by day into Google spreadsheets. Such data cannot be immediately used, but require further modification and processing steps beforehand. To this end, a Python code based on the NumPy and Pandas libraries was developed, along with Google Sheet APIs, which enable data manipulation of Google spreadsheets. More specifically, reading/writing values from/to

Google spreadsheets is enabled by the spreadsheets and the spreadsheet values collections.

Once the data are retrieved and imported into the Python program, they are appropriately processed. Each of the datasets obtained from the spreadsheets of the users is linked to an anonymous ID known by the users. This ID is the same one that identifies the users when they have to answer some questions about their well-being. The data are then elaborated for every day, joining together the different segments or sessions that belong to the same performed activity detected by the wearable devices. For every activity there is then one single row every day that takes into account the information about the number of steps, burned calories, covered distance and how long that activity was performed.

4.2.2 Self-Evaluation of Well-being and Stress

The focus of this study is to assess the well-being of generic users which do not present pathological conditions, and to measure a level of stress that is considered as the normal daily stress correlated with minor life problems, such as work stress and similar. For this reason, the WHO-5 Well-Being Index questionnaire described in Section 1.6.2 has been selected among the possible ones due to its applicability to general population and its briefness; in order to include the stress assessment, one question from the Perceived Stress Scale (PSS) has been added to the questionnaire, i.e.: "how often have you felt nervous and stressed?". In the survey there are also three other questions selected from the General Health Questionnaire (GHQ) [37]. There are four different versions of the GHQ; in this case the GHQ-28 has been considered, which provides four scores measuring physical symptoms, anxiety and insomnia, social dysfunction, and severe depression. The selected questions are related only to the physical items, which can be easily linked to a generic physical state of health.

In order to answer the questions, the users had to complete a survey sent to them by email. In the first question they had to indicate their anonymous ID so that, later on, it was possible to link their score, obtained from the survey, to the GoogleSheet containing the data of their performed activities, obtained from the wearable.

There are two different surveys, one is related to users' feelings after one entire month of monitoring, and the other one is related to a single week. Other than the questions extracted from WHO-5, PSS and GHQ-28 surveys, in the monthly questionnaire there are two other questions that specifically investigate whether users have had any general health problems (cough, allergies) or fever and for how many days.

Therefore, the questions presented to users are the following: "I have felt cheerful in good spirits", "I have felt calm and relaxed", "I have felt active and vigorous", "I woke up feeling fresh and rested", "How often have you felt nervous and stressed?", "Have you been feeling perfectly well and in good health?", "Have you been getting any pains in your head?", "Have you had difficulty in staying asleep once you are off?", "Have you been ill (cough, allergies, cold)?", "Have you had a fever?".

The second questionnaire has the same questions as the previous one, with one additional question that requires users to indicate the approximate hours spent outdoors ("How many hours have you spent outdoors this week?"). This last question has been inserted in order to try to understand if there is some kind of correlation between the time generally spent outdoors and the sensation of well-being perceived by users. These results can be used to decide if it could be helpful to monitor this aspect directly through wearable devices during a second phase of experimentation. The same question is not asked in the first survey because it is very unlike that users can answer a specific question like that with accuracy for an entire month.

4.2.3 Metrics to Assess the Correlation Between Wearable-Collected Data and Self-Evaluation Results

In order to understand how users' typical behaviour in terms of daily activities and daily sleep has an impact on their perceived well-being and stress, it is necessary to measure the correlation between the data obtained through the monitoring, using smartwatches and smartbands, and the survey scores.

There are different statistical methods that can be used to assess the correlation between two or more variables. Pearson Correlation [35], Spearman Correlation [42], t-test [10] are some typical formulas that allow establishing the level of correlation of the features under consideration. For example, in [35] the authors make use of Pearson Correlation to understand what information taken from smartwatches sensors about body sensing and environmental light are useful to determinate the mood state of users (happiness and activation); in [13] features extracted from smartwatches are used to find some correlation with respect to the stress condition reported by users, through the Mutual Information evaluations, that is a measure of mutual dependence between the two variables.

The aim of this study is to find a correlation between the information concerning users' daily activities and daily sleep, that are obtained from different types of smartwatches or smart bands, and the scores about perceived well-being and level of healthiness deduced from the survey reported in the above Section 4.2.2. For this purpose, Spearman's rank correlation coeffiecients among involved variables are used.

Spearman's correlation coefficients provide a measure of the strength and direction of association between two ranked variables and it is a measure of a monotone association. Values of ± 1 indicate a strong positive correlation or a strong negative correlation, so that as one variable increases, the other variable also tends to increase (positive), or decrease (negative). A value equal to 0 indicates that the association between the two ranks is not monotonic. Spearman rank correlation can be calculated with the following formula:

$$r_{sr} = \frac{\sum(R_{ix} - \overline{R}_x)(R_{iy} - \overline{R}_y)}{\sqrt{\sum(R_{ix} - \overline{R}_x)^2 \cdot \sum(R_{iy} - \overline{R}_y)^2}} \tag{4.1}$$

where R_{ix} and R_{iy} indicate the ranks of the $i-th$ values of variables X and Y; $\overline{R}_x$ and $\overline{R}_y$ are the means of R_{ix} and R_{iy} values.

4.3 Preliminary Performance Evaluation

The objective of this first prototype was to assess the soundness of the proposed methodology and have a first estimation of the correlation between the observed parameters and users' health. For this reason, even though the number of involved volunteers is quite low, it is still enough to draw some preliminary conclusions. The monitoring of users' behaviour is done *in-the-wild*, therefore in uncontrolled real-life environments, counting 4 people, 2 females and 2 males, aged between 28 and 36 years. The subjects were all informed about the type of data that the system was going to collect and the purpose of their collaboration; they all used their personal smart devices already possessed and were instructed to perform their activities how they usually do and use the devices like they always have. Indeed, this is necessary to be sure that the results are affected as least as possible by the so-called *observer effect*, i.e. the disturbance caused by the awareness of being observed. The wearable devices that have been used to collect data are: one Xiaomi Mi Band 3, two Xiaomi Mi Band 5 and one Watch Amazfit GTR.

Data were collected for 2 months, recording information concerning the number of steps, the distance travelled on foot, the number of calories burned and the average heart rate. This information is linked to one precise activity recorded in Google Fit and obtained from the wearable devices. The information of the starting and ending time and the duration is also recorded for every detected activity.

The data are then aggregated by day, so that for every day it is possible to consider the amount of time dedicated to moving/fitness activities and the total sums for steps, calories and distances. For the monitoring of sleep activities, the only data that are considered are the burned calories and the duration of the daily night sleep.

The evaluation of the well-being and stress levels are done considering two different time periods. Firstly, the subjects were asked to answer questions related to the first month of the experiments. The questions of this survey are the ones from the WHO-5, PSS and GHQ-28 that have been explained in 4.2.2. On the other hand, during the second month, the subjects had to answer the same questions of the ones in the first survey but related to one week at time, at the start of each of the three weeks.

Therefore, for each user, 4 different scores from the survey can be considered. The first one related to the whole first month, and the other three scores related

	Mean Daily Steps	Mean Daily Calories	Mean Daily Distance	Mean Daily Duration	Ill Days	Fever	Survey Score
Mean Daily Steps	1	0.738	1	0.643	-0.546	-0.094	**0.856**
Mean Daily Calories	0.738	1	0.738	0.929	-0.081	0.514	**0.723**
Mean Daily Distance	1	0.738	1	0.643	-0.546	-0.094	**0.856**
Mean Daily Duration	0.643	0.929	0.643	1	-0.081	0.514	**0.724**
Ill Days	-0.546	-0.081	-0.546	-0.081	1	0.589	-0.193
Fever	-0.094	0.514	-0.094	0.514	0.589	1	0.016
Survey Score	**0.856**	**0.723**	**0.856**	**0.724**	-0.193	0.016	1

Table 4.1: Spearman's Rank Correlation Coefficients for Fitness Activities

each to one of the three weeks of the second month. Points are calculated so that greater final score values are related to a greater level of well-being experienced by users. For every question there are 5 possible answers useful to rate how well each of the statements applies to the user; each of these answers can give a point between 0 to 5. Therefore, the maximum score for the surveys is 40 considering the first 8 questions. The remaining answers are evaluated as the amount number of ill days or days with fever with respect to the total number of days considered in the referred period.

The obtained scores are then related to the following features calculated from the data of the wearable devices:

- the daily mean values of the duration of the activities (Mean_Daily_Duration) and the total time spent performing the activities during the month or the week; fitness activities are considered all together to obtain an estimation of the time spent doing physical movement over the considered period,
- the mean daily steps number (Mean_Daily_Steps) and the total number of steps counted during the month or the week,
- the mean daily burned calories (Mean_Daily_Calories) and the total calories expended during the month or the week,
- the mean value of the travelled distances (Mean_Daily_Distance) per day and the travelled distance of the month or the week.

Table 4.1 shows the results in terms of the Spearman correlation coefficients. These results refer to the totality of the data and scores obtained in the two months of experimentation. Only the most significant ones are shown, i.e., those that can be significantly related to the survey scores. For example, there is obviously also a

	Mean Daily Duration	Ill Days	Fever	Survey Score
Mean Daily Duration	1	-0.500	**-0.866**	-0.112
Ill Days	-0.500	1	0.125	**-1**
Fever	**-0.866**	0.125	1	0.022
Survey Score	-0.112	**-1**	0.022	1

Table 4.2: Spearman's Rank Correlation Coefficients for Sleep Activities

strong correlation between the total steps and the travelled distance, but it is not useful information. High values of correlation can be considered those that have an absolute value greater than 0.7.

Therefore, it is possible to notice that there is a strong correlation between the mean values of the information related to the daily activities over the time period and the correspondent well-being and stress evaluation reported with the scores of the surveys. The correlation is positive, thus indicating that physical activities bring a greater sense of well-being.

Table 4.2 shows the Spearman correlation coefficients referred to the daily sleep of the users. From these results, it is possible to see a strong correlation between the fact the the users had a fever during the reference period and the mean duration of users' night sleep; also, there is a strong correlation between the number of ill days and the general score obtained from the survey. Both the correlations are negative, thus indicating how not feeling in good health affects both the users' night rest and the users' final results of the evaluation of their well-being and perceived stress.

Since the questions of the second survey, sent every week, have more details, the correlations between data and scores are also evaluated week by week. The results in table 4.3 show a similar trend to the general results already presented, with a strong correlation between data about activities and the total score of the surveys. The only exception can be found in the correlation between the mean time spent performing the activities and the total score, where the values are always under 0.5. This is due to the fact that the duration spent on physical activities by each user is always very different from each other and thus is not very discriminating over one week. Regarding the analysis considering each user alone for the overall period of the experimentation, the results show similar significant correlation coefficients for activity data and survey scores comparable to those in the table 4.1.

The additional question of the weekly questionnaire about the time spent outdoors by users gave one suggestion of a strong negative correlation with the mean duration of the activities. Indeed, the correlation coefficients are -0.811, -0.716 and -0.732 for each week. From this result it is possible to understand that the

	Week1 Survey Score	Week2 Survey Score	Week3 Survey Score
Mean Daily Steps	0.810	0.800	0.800
Mean Daily Calories	0.721	0.712	0.800
Mean Daily Distance	0.786	0.800	0.800
Mean Daily Duration	0.452	0.400	0.406

Table 4.3: Spearman's Rank Correlation Coefficient by Week

users performed most of their activities inside their home. However, this result could be influenced by the particular situation in which we find ourselves globally due to COVID-19 restrictions, for which very often it is not allowed to spend a lot of time outside one's own homes. On the other hand, a strong positive correlation was also found with the mean daily steps and the mean daily distance, as it is expected.

4.4 Conclusion and Future Works

This part of the thesis presents a first approach to the design of a methodology to early recognise and predict stress and illness conditions, based on their correlation with data collected through common commercial wearable devices. The first prototype presented in this thesis assesses the correlation between some specific parameters describing sleep, fitness activities and walking habits, with stress levels and well-being state as self-evaluated by the users of the wearable devices. Preliminary results prove that there is a strong correlation between simple data collected through smartbands and the well-being and state of the healthiness of the users assessed using self-evaluation surveys.

The main limitation of this study is undoubtedly linked to the small number of participants involved and the fact that it was possible to collect only averaged data for the investment period. Furthermore, obtaining only some information relating to user activities and parameters was possible. Moreover, there are more parameters that, if investigated, could lead to more significant results. The study in question is an initial evaluation of the methodology, while the essential part will be to establish how to predict psycho-physical developments from previously collected data.

Therefore, future works will be focused on expanding and validating the proposed methodology with more volunteers monitored for a longer period of time. Furthermore, other parameters, such as heart rate variability, will be included. The system will be later integrated with data gathered by ambient sensors inside users' homes, when present, which will improve the context-awareness of the system by incorporating other activities' recognition. It may be interesting, for instance, to evaluate if the time devoted to cooking or eating has a correlation with users' weight.

Chapter 5

Concluding Remarks and Future Works

The main idea of this research work was to establish how people's habits and behaviours can affect certain aspects of their lives. Indeed, it is known that some actions lead to consequences that can be negative, both on the psycho-physical health of the user and on the surrounding environment. Good habits, on the other hand, when recognised as such, should be encouraged and highlighted. Therefore, the thesis focused on developing systems that could help users in these terms, acting to improve their quality of living from different points of view. In particular, an analysis has been presented, within the Smart Home scenario, to improve the use of energy and, at the same time, guarantee user comfort; on the other hand, it has also been analysed how it is possible to investigate the way in which physical activity and sleep quality can affect the psycho-physical health of the user.

In chapter 1, the problem of recognising and predicting activities carried out by users is addressed. This Section starts with the study of the different works from the literature and the presentation of different possible approaches depending on the sought solution, with an insight about the approaches that were useful for the discussion made in the next Sections. In particular, solutions using a network of environmental and non-intrusive sensors arranged inside a Smart Home and solutions based on wearable for health monitoring are considered.

chapter 2 addresses in detail the realisation of a BECM system that is focused on monitoring and controlling key equipment inside smart buildings and understanding users' behaviours, with the aim of providing users with tools that support cost-effective solutions in appliance management. The main objective of this system, described in Section 2.2, is to learn users' daily behaviour and to be able to predict their future activities based on statistical data about the activities they usually perform. The system can then execute a scheduling algorithm of the appliances based on the expected energy consumption and user annoyance related to shifting the appliance starting time from their preferred one. Experimental results demonstrate that thanks to the scheduling algorithm, energy cost can be reduced by at least 66.64% depending on different users' routine predictability profiles, just by shifting the use of the appliance to time periods of non-peak hours. Scheduling based on probability evaluation of the preferred time of usage of the appliances allows obtaining evident energy savings, even considering some errors in predicted activities.

In chapter 3, Section 3.3 proposes an appliance power profiling system that analyses different consumption profiles that are associated with each appliance. Preliminary simulation results show that each profile can be approximated with a single reference consumption cycle representative for the entire cluster, with errors that are always lower than 10%, considering both total energy consumption and power values time interval by time interval.

Finally, the last chapter 4 focuses on a different research area that is about health monitoring systems. In Section 4.2, a system that makes use of popular commercial wrist-wearable devices to find the correlation between the monitored users' activities and their stress and well-being conditions, as subjectively self-assessed by them, is proposed. This Section aims to present a methodology to automatically learn which users' activities can be associated with positive and negative health conditions so that they can be later predicted as soon as the first signals are detected by wearable devices. The implementation and preliminary results of a first prototype of the proposed system, which monitors users' sleep and activity and assesses the correlation with stress levels and illness conditions, have been presented in Sections 4.2 and 4.3.

The major limitations of all the works presented concern the difficulty in finding data and volunteers, which has led to being able to evaluate the proposed solutions only in small groups (single houses, few users). It will therefore be necessary to consider all the proposed solutions, even on a larger scale. Finally, at the moment, the three systems are independent of one another. The final system will instead be a single complex entity which, thanks to the combined use of sensors and other non-intrusive systems to monitor the surrounding environment and to the use of simple wearable devices, will have a complete understanding of all the habits and activities typically performed by the users and will also monitor their vital parameters. Therefore, it will be able to act by giving targeted advice or modifying the surrounding environment to make it more suitable considering users' particular state of mind based on prior knowledge of their habits and preferences.

Hence, future works will primarily focus on improving the appliances' energy consumption profiling system, analysing different sets of features for different appliances and adding new appliances into the simulations. Furthermore, this system will be integrated into the BECM system of chapter 2. Indeed, the main purpose is to be able to connect the prediction of the different appliance energy consumption profiles to the prediction of certain activities associated with that particular household appliance. The other important aspect concerns the health monitoring system. For this research activity, future efforts will involve expanding the proposed system, including more users and analysing more parameters. Also in this case, the final purpose will be to integrate this type of monitoring based on commercial wearable devices within the BECM system for a more accurate and complete knowledge of the user's habits.

Chapter 6

List of publications related to the Thesis

- Marcello, Francesca & Pilloni, Virginia.
  "Daily Activities Monitoring of Users for Well-Being and Stress Correlation Using Wearable Devices"
  2021 IEEE Global Communications Conference (GLOBECOM), 2021
- Marcello, Francesca & Pilloni, Virginia.
  "Smart Building Energy and Comfort Management Based on Sensor Activity Recognition and Prediction"
  CEUR Workshop Proceedings, 2020

List of Figures

List of Tables

Bibliography

[1] A. Ahmad and J. Y. Khan. Real-time load scheduling, energy storage control and comfort management for grid-connected solar integrated smart buildings. *Applied Energy*, 259:114208, 2020.

[2] R. E. Alden, H. Gong, C. Ababei, and D. M. Ionel. Lstm forecasts for smart home electricity usage. In *2020 9th International Conference on Renewable Energy Research and Application (ICRERA)*, pages 434–438. IEEE, 2020.

[3] M. Alessandrini, G. Biagetti, P. Crippa, L. Falaschetti, and C. Turchetti. Recurrent neural network for human activity recognition in embedded systems using ppg and accelerometer data. *Electronics*, 10(14):1715, 2021.

[4] Y. Asim, M. A. Azam, M. Ehatisham-ul Haq, U. Naeem, and A. Khalid. Context-aware human activity recognition (cahar) in-the-wild using smartphone accelerometer. *IEEE Sensors Journal*, 20(8):4361–4371, 2020.

[5] L. Atzori, A. Iera, and G. Morabito. The internet of things: A survey. *Computer networks*, 54(15):2787–2805, 2010.

[6] S. Balli, E. A. Sağbaş, and M. Peker. Human activity recognition from smart watch sensor data using a hybrid of principal component analysis and random forest algorithm. *Measurement and Control*, 52(1-2):37–45, 2019.

[7] N. Balta-Ozkan, R. Davidson, M. Bicket, and L. Whitmarsh. Social barriers to the adoption of smart homes. *Energy Policy*, 63:363–374, 2013.

[8] U. Berardi. A cross-country comparison of the building energy consumptions and their trends. *Resources, Conservation and Recycling*, 123:230–241, 2017.

[9] Y. S. Can, B. Arnrich, and C. Ersoy. Stress detection in daily life scenarios using smart phones and wearable sensors: A survey. *Journal of biomedical informatics*, 92:103139, 2019.

[10] Y. S. Can, H. Iles-Smith, N. Chalabianloo, D. Ekiz, J. Fernández-Álvarez, C. Repetto, G. Riva, and C. Ersoy. How to relax in stressful situations: a smart stress reduction system. In *Healthcare*, volume 8, page 100. Multidisciplinary Digital Publishing Institute, 2020.

[11] E. Carmeli, B. Imam, and J. Merrick. Assistive technology and older adults. In *Health care for people with intellectual and developmental disabilities across the lifespan*, pages 1465–1471. Springer, 2016.

[12] K. T. Chui, K. F. Tsang, S. H. Chung, and L. F. Yeung. Appliance signature identification solution using k-means clustering. In *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*, pages 8420–8425, 2013. doi: 10.1109/IECON.2013.6700545.

[13] L. Ciabattoni et al. Real-time mental stress detection based on smartwatch. In *2017 IEEE International Conference on Consumer Electronics (ICCE)*, pages 110–111, 2017. doi: 10.1109/ICCE.2017.7889247.

[14] S. Cohen, T. Kamarck, R. Mermelstein, et al. Perceived stress scale. *Measuring stress: A guide for health and social scientists*, 10:1–2, 1994.

[15] D. J. Cook. Learning setting-generalized activity models for smart spaces. *IEEE intelligent systems*, 2010(99):1, 2010.

[16] T. Csoknyai, J. Legardeur, A. Abi Akle, and M. Horváth. Analysis of energy consumption profiles in residential buildings and impact assessment of a serious game on occupants' behavior. *Energy and Buildings*, 196:1–20, 2019.

[17] T. Csoknyai, J. Legardeur, A. A. Akle, and M. Horváth. Analysis of energy consumption profiles in residential buildings and impact assessment of a serious game on occupants' behavior. *Energy and Buildings*, 196:1–20, 2019. ISSN 0378-7788. doi: https://doi.org/10.1016/j.enbuild.2019.05.009. URL `https://www.sciencedirect.com/science/article/pii/S0378778818334790`.

[18] L. Czétány, V. Vámos, M. Horváth, Z. Szalay, A. Mota-Babiloni, Z. Deme-Bélafi, and T. Csoknyai. Development of electricity consumption profiles of residential buildings based on smart meter data clustering. *Energy and Buildings*, 252:111376, 2021. ISSN 0378-7788. doi: https://doi.org/10.1016/j.enbuild.2021.111376. URL `https://www.sciencedirect.com/science/article/pii/S0378778821006605`.

[19] J. Dahihande, A. Jaiswal, A. A. Pagar, A. Thakare, M. Eirinaki, and I. Varlamis. Reducing energy waste in households through real-time recommendations. In *Fourteenth ACM Conference on Recommender Systems*, pages 545–550, 2020.

[20] L. Devroye, L. Györfi, and G. Lugosi. *A probabilistic theory of pattern recognition*, volume 31. Springer Science & Business Media, 2013.

[21] P. Domingues, P. Carreira, R. Vieira, and W. Kastner. Building automation systems: Concepts and technology review. *Computer Standards & Interfaces*, 45:1–12, 2016.

[22] A. I. Dounis and C. Caraiscos. Advanced control systems engineering for energy and comfort management in a building environment—a review. *Renewable and Sustainable Energy Reviews*, 13(6-7):1246–1261, 2009.

[23] Y. Du, Y. Lim, and Y. Tan. Activity prediction using lstm in smart home. In *2019 IEEE 8th Global Conference on Consumer Electronics (GCCE)*, pages 918–919, 2019. doi: 10.1109/GCCE46687.2019.9015492.

[24] Y. Du, Y. Lim, and Y. Tan. Activity prediction using lstm in smart home. In *2019 IEEE 8th Global Conference on Consumer Electronics (GCCE)*, pages 918–919. IEEE, 2019.

[25] G. Duwarah and D. I. Islam. Human activity recognition using lstm model.

[26] S. R. Eddy. What is a hidden markov model? *Nature biotechnology*, 22(10): 1315, 2004.

[27] P. V. Er and K. K. Tan. Non-intrusive fall detection monitoring for the elderly based on fuzzy logic. *Measurement*, 124:91–102, 2018.

[28] M. Eroglu, E. Dursun, S. Sevencan, J. Song, S. Yazici, and O. Kilic. A mobile renewable house using PV/wind/fuel cell hybrid power system. *Int. journal of hydrogen energy*, 36(13):7985–7992, 2011.

[29] I. Fatima, M. Fahim, Y.-K. Lee, and S. Lee. A unified framework for activity recognition-based behavior analysis and action prediction in smart homes. *Sensors*, 13(2):2682–2699, 2013.

[30] M. Fayaz and D. Kim. Energy consumption optimization and user comfort management in residential buildings using a bat algorithm and fuzzy logic. *Energies*, 11(1):161, 2018.

[31] T. M. Fernández-Caramés and P. Fraga-Lamas. Towards the internet of smart clothing: A review on iot wearables and garments for creating intelligent connected e-textiles. *Electronics*, 7(12):405, 2018.

[32] F. Fioranelli, J. Le Kernec, and S. A. Shah. Radar for health care: Recognizing human activities and monitoring vital signs. *IEEE Potentials*, 38(4):16–23, 2019. doi: 10.1109/MPOT.2019.2906977.

[33] F. A. Gers, J. Schmidhuber, and F. Cummins. Learning to forget: Continual prediction with lstm. *Neural computation*, 12(10):2451–2471, 2000.

[34] P. Gianniou, X. Liu, A. Heller, P. S. Nielsen, and C. Rode. Clustering-based analysis for residential district heating data. *Energy Conversion and Management*, 165:840–850, 2018. ISSN 0196-8904. doi: https://doi.org/10.1016/j.enconman.2018.03.015. URL `https://www.sciencedirect.com/science/article/pii/S019689041830236X`.

[35] P. A. Gloor, A. F. Colladon, F. Grippa, P. Budner, and J. Eirich. Aristotle said "happiness is a state of activity"—predicting mood through body sensing with smartwatches. *Journal of Systems Science and Systems Engineering*, 27 (5):586–612, 2018.

[36] H. C. J. Godfray, P. Aveyard, T. Garnett, J. W. Hall, T. J. Key, J. Lorimer, R. T. Pierrehumbert, P. Scarborough, M. Springmann, and S. A. Jebb. Meat consumption, health, and the environment. *Science*, 361(6399):eaam5324, 2018.

[37] D. P. Goldberg, R. Gater, N. Sartorius, T. B. Ustun, M. Piccinelli, O. Gureje, C. Rutter, et al. The validity of two versions of the ghq in the who study of mental illness in general health care. *Psychological medicine*, 27(1):191–197, 1997.

[38] I. Goodfellow, Y. Bengio, and A. Courville. Deep learning (adaptive computation and machine learning series). *Cambridge Massachusetts*, pages 321–359, 2017.

[39] F. Grant, C. Guille, and S. Sen. Well-being and the risk of depression under stress. *PLoS one*, 8(7):e67395, 2013.

[40] O. Guerra-Santin, H. Bosch, P. Budde, T. Konstantinou, S. Boess, T. Klein, and S. Silvester. Considering user profiles and occupants' behaviour on a zero energy renovation strategy for multi-family housing in the netherlands. *Energy Efficiency*, 11(7):1847–1870, 2018.

[41] M. Haghi, S. Neubert, A. Geissler, H. Fleischer, N. Stoll, R. Stoll, and K. Thurow. A flexible and pervasive iot-based healthcare platform for physiological and environmental parameters monitoring. *IEEE Internet of Things Journal*, 7(6):5628–5647, 2020. doi: 10.1109/JIOT.2020.2980432.

[42] C. Heine, C. H. Gong, and C. Browning. Dual sensory loss, mental health, and wellbeing of older adults living in china. *Frontiers in public health*, 7:92, 2019.

[43] IEA. 2021 global status report for buildings and construction. *United Nations Environment Programme*, 2021.

[44] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3):264–323, 1999.

[45] H. C. Jeong, M. Jang, T. Kim, and S.-K. Joo. Clustering of load profiles of residential customers using extreme points and demographic characteristics. *Electronics*, 10(3):290, 2021.

[46] W. Jin, I. Ullah, S. Ahmad, and D. Kim. Occupant comfort management based on energy optimization using an environment prediction model in smart homes. *Sustainability*, 11(4):997, 2019.

[47] R. Kadouche, B. Chikhaoui, and B. Abdulrazak. User's behavior study for smart houses occupant prediction. *annals of telecommunications-annales des télécommunications*, 65(9-10):539–543, 2010.

[48] R. Kanjilal and I. Uysal. The future of human activity recognition: deep learning or feature engineering? *Neural Processing Letters*, 53(1):561–579, 2021.

[49] G. Kaur and E. N. Oberai. A review article on naive bayes classifier with various smoothing techniques. *International Journal of Computer Science and Mobile Computing*, 3(10):864–868, 2014.

[50] N. Kim, M. Kim, and J. K. Choi. Lstm based short-term electricity consumption forecast with daily load profile sequences. In *2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)*, pages 136–137. IEEE, 2018.

[51] B. Kon, A. Lam, and J. Chan. Evolution of smart homes for the elderly. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 1095–1101, 2017.

[52] N. C. Krishnan and D. J. Cook. Activity recognition on streaming sensor data. *Pervasive and mobile computing*, 10:138–154, 2014.

[53] F. Landi, L. Baraldi, M. Cornia, and R. Cucchiara. Working memory connections for lstm. *Neural Networks*, 144:334–341, 2021.

[54] Y. Liu, L. Nie, L. Liu, and D. S. Rosenblum. From action to activity: sensor-based activity recognition. *Neurocomputing*, 181:108–115, 2016.

[55] L. Lu, C. Qing-Ling, and Z. Yi-Ju. Activity recognition in smart homes. *Multimedia Tools and Applications*, 76(22):24203–24220, 2017.

[56] F. Luo, G. Ranzi, W. Kong, G. Liang, and Z. Y. Dong. Personalized residential energy usage recommendation system based on load monitoring and collaborative filtering. *IEEE Transactions on Industrial Informatics*, 17(2): 1253–1262, 2021. doi: 10.1109/TII.2020.2983212.

[57] S. Majumder, T. Mondal, and M. J. Deen. Wearable sensors for remote health monitoring. *Sensors*, 17(1):130, 2017.

[58] F. Marcello, V. Pilloni, and D. Giusto. Sensor-based early activity recognition inside buildings to support energy and comfort management systems. *Energies*, 12(13):2631, 2019.

[59] H. L. Meiselman. Quality of life, well-being and wellness: Measuring subjective health for foods and other products. *Food Quality and Preference*, 54:101–109, 2016. ISSN 0950-3293. doi: https://doi.org/10.1016/j.foodqual.2016.05.009. URL `https://www.sciencedirect.com/science/article/pii/S0950329316301069`.

[60] A. E. Minarno, W. A. Kusuma, H. Wibowo, D. R. Akbi, and N. Jawas. Single triaxial accelerometer-gyroscope classification for human activity recognition. In *2020 8th International Conference on Information and Communication Technology (ICoICT)*, pages 1–5. IEEE, 2020.

[61] D. Minoli, K. Sohraby, and B. Occhiogrosso. Iot considerations, requirements, and architectures for smart buildings—energy optimization and next-generation building management systems. *IEEE Internet of Things Journal*, 4(1):269–283, 2017.

[62] J. K. Morris, R. A. Honea, E. D. Vidoni, R. H. Swerdlow, and J. M. Burns. Is alzheimer's disease a systemic disease? *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1842(9):1340–1349, 2014.

[63] C. W. M. Ng, C. H. How, and Y. P. Ng. Major depression in primary care: making the diagnosis. *Singapore medical journal*, 57(11):591, 2016.

[64] D. Norhafizah, B. Pg, T. H. Lim, N. S. Binti, M. Arifin, et al. Non-intrusive wearable health monitoring systems for emotion detection. In *2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pages 985–989. IEEE, 2017.

[65] S. W. Pienaar and R. Malekian. Human activity recognition using lstm-rnn deep neural network architecture. In *2019 IEEE 2nd wireless africa conference (WAC)*, pages 1–5. IEEE, 2019.

[66] V. Pilloni, A. Floris, et al. Smart home energy management including renewable sources: A qoe-driven approach. *IEEE Transactions on Smart Grid*, 9 (3):2006–2018, 2016.

[67] Z. Pooranian, J. Abawajy, et al. Scheduling distributed energy resource operation and daily power consumption for a smart building to optimize economic and environmental parameters. *Energies*, 11(6):1348, 2018.

[68] W. Qi, H. Su, C. Yang, G. Ferrigno, E. De Momi, and A. Aliverti. A fast and robust deep convolutional neural networks for complex human activity recognition using smartphone. *Sensors*, 19(17):3731, 2019.

[69] M. Rasheed, N. Javaid, A. Ahmad, Z. Khan, U. Qasim, and N. Alrajeh. An efficient power scheduling scheme for residential load management in smart homes. *Applied Sciences*, 5(4):1134–1163, 2015.

[70] D. Riboni and M. Murtas. Sensor-based activity recognition: One picture is worth a thousand words. *Future Generation Computer Systems*, 101:709–722, 2019. ISSN 0167-739X.

[71] J. Rouillard. Contextual qr codes. In *2008 The Third International Multi-Conference on Computing in the Global Information Technology (iccgi 2008)*, pages 50–55. IEEE, 2008.

[72] A. Samadi, H. Saidi, M. A. Latify, and M. Mahdavi. Home energy management system based on task classification and the resident's requirements. *International Journal of Electrical Power & Energy Systems*, 118:105815, 2020.

[73] I. H. Sarker. Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, 2(3):1–21, 2021.

[74] A. Satre Meloy, M. Diakonova, and P. Grunewald. What makes you peak? cluster analysis of household activities and electricity demand. European Council for an Energy Efficient Economy, 2019.

[75] M. Shafie-Khah and P. Siano. A stochastic home energy management system considering satisfaction cost and response fatigue. *IEEE Transactions on Industrial Informatics*, 14(2):629–638, 2017.

[76] P. H. Shaikh, N. B. M. Nor, et al. Intelligent multi-objective optimization for building energy and comfort management. *Journal of King Saud University-Engineering Sciences*, 30(2):195–204, 2018.

[77] M. H. Shoreh, P. Siano, et al. A survey of industrial applications of demand response. *Electric Power Systems Research*, 141:31–49, 2016.

[78] D. Singh, E. Merdivan, I. Psychoula, J. Kropf, S. Hanke, M. Geist, and A. Holzinger. Human activity recognition using recurrent neural networks. In *International cross-domain conference for machine learning and knowledge extraction*, pages 267–274. Springer, 2017.

[79] N. K. Singh and K. S. Suprabhath. Har using bi-directional lstm with rnn. In *2021 International Conference on Emerging Techniques in Computational Intelligence (ICETCI)*, pages 153–158, 2021. doi: 10.1109/ICETCI51973.2021.9574073.

[80] R. Stanton, Q. G. To, S. Khalesi, S. L. Williams, S. J. Alley, T. L. Thwaite, A. S. Fenning, and C. Vandelanotte. Depression, anxiety and stress during covid-19: associations with changes in physical activity, sleep, tobacco and alcohol use in australian adults. *International journal of environmental research and public health*, 17(11):4065, 2020.

[81] N. Tamminen, T. Kettunen, T. Martelin, J. Reinikainen, and P. Solin. Living alone and positive mental health: a systematic review. *Systematic reviews*, 8 (1):1–8, 2019.

[82] N. Tax. Human activity prediction in smart home environments with lstm neural networks. In *2018 14th International Conference on Intelligent Environments (IE)*, pages 40–47, 2018. doi: 10.1109/IE.2018.00014.

[83] N. Tax. Human activity prediction in smart home environments with lstm neural networks. In *2018 14th International Conference on Intelligent Environments (IE)*, pages 40–47. IEEE, 2018.

[84] G. M. Toschi, L. B. Campos, and C. E. Cugnasca. Home automation networks: A survey. *Computer Standards & Interfaces*, 50:42–54, 2017.

[85] G. Trotta. An empirical analysis of domestic electricity load profiles: Who consumes how much and when? *Applied Energy*, 275:115399, 2020. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2020.115399. URL `https://www.sciencedirect.com/science/article/pii/S0306261920309119`.

[86] A. Tureczek, P. S. Nielsen, and H. Madsen. Electricity consumption clustering using smart meter data. *Energies*, 11(4), 2018. ISSN 1996-1073. doi: 10.3390/en11040859. URL `https://www.mdpi.com/1996-1073/11/4/859`.

[87] K. Viard, M. P. Fanti, G. Faraut, and J.-J. Lesage. An event-based approach for discovering activities of daily living by hidden markov models. In *2016 15th International Conference on Ubiquitous Computing and Communications and 2016 International Symposium on Cyberspace and Security (IUCC-CSS)*, pages 85–92. IEEE, 2016.

[88] D. Villanueva, D. San-Facundo, E. Miguez-García, and A. Fernández-Otero. Modeling and simulation of household appliances power consumption. *Applied Sciences*, 12(7):3689, 2022.

[89] R. R. Villar, S. P. Fernández, C. C. Garea, M. Pillado, V. B. Barreiro, and C. G. Martín. Quality of life and anxiety in women with breast cancer before and after treatment. *Revista latino-americana de enfermagem*, 25, 2017.

[90] F. Wahid, L. H. Ismail, R. Ghazali, and M. Aamir. An efficient artificial intelligence hybrid approach for energy management in intelligent buildings. *KSII Transactions on Internet & Information Systems*, 13(12), 2019.

[91] R. Want. An introduction to rfid technology. *IEEE pervasive computing*, (1): 25–33, 2006.

[92] M. Wenninger, A. Maier, and J. Schmidt. Deddiag, a domestic electricity demand dataset of individual appliances in germany. *Scientific Data*, 8(1): 1–15, 2021.

[93] M. Western and W. Tomaszewski. Subjective wellbeing, objective wellbeing and inequality in australia. *PloS one*, 11(10):e0163345, 2016.

[94] M. Wigginton and J. Harris. *Intelligent skins*. Routledge, 2013.

[95] C. Wilson, T. Hargreaves, and R. Hauxwell-Baldwin. Smart homes and their users: a systematic analysis and key challenges. *Personal and Ubiquitous Computing*, 19(2):463–476, 2015.

[96] N. C. Wong, C. Lam, L. Patterson, and B. Shayegan. Use of machine learning to predict early biochemical recurrence after robot-assisted prostatectomy. *BJU international*, 123(1):51–57, 2019.

[97] G. Wood and M. Newborough. Energy-use information transfer for intelligent homes: Enabling energy conservation with central and local displays. *Energy and buildings*, 39(4):495–503, 2007.

[98] F. Wu, T. Wu, and M. R. Yuce. An internet-of-things (iot) network system for connected safety and health monitoring applications. *Sensors*, 19(1):21, 2019.

[99] L. Yang, H. Yan, and J. C. Lam. Thermal comfort and building energy consumption implications–a review. *Applied energy*, 115:164–173, 2014.

[100] X. Yang, S. A. Shah, A. Ren, N. Zhao, J. Zhao, F. Hu, Z. Zhang, W. Zhao, M. U. Rehman, and A. Alomainy. Monitoring of patients suffering from rem sleep behavior disorder. *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, 2(2):138–143, 2018. doi: 10.1109/JERM.2018.2827705.

[101] J. Zethmayr and R. S. Makhija. Six unique load shapes: A segmentation analysis of illinois residential electricity consumers. *The Electricity Journal*, 32(9):106643, 2019. ISSN 1040-6190. doi: https://doi.org/10.1016/j.tej.2019.106643. URL `https://www.sciencedirect.com/science/article/pii/S1040619019301800`.