Early diagnosis of disorders based on behavioural shifts and biomedical signals

Ane Alberdi Aramendi

October 2017

Electronics and Computing Department
Faculty of Engineering
Mondragon University

Early diagnosis of disorders based on behavioural shifts and biomedical signals

A thesis submitted
in fulfillment of the requirements for the degree of
Doctor of Philosophy

Ane Alberdi Aramendi

Supervised by:

Dr. Asier Aztiria Goenaga
and
Dr. Adrian Basarab

Committee:
Chair: Dr. Diego López de Ipiña González de Artaza (Universidad de Deusto)
Member: Dr. José Javier Yanguas Lezáun (Fundación INGEMA)
Member: Dr. Javier Mar Medina (Hospital Alto Deba)
Member: Dr. Maria Goreti Carvalho Marreiros (Polytechnic Institute of Porto)
Secretary: Dr. Maitane Barrenechea Carrasco (Mondragon Unibertsitatea)

October 2017

Electronics and Computing Department
Faculty of Engineering
Mondragon University

Atte ta amai,

"Education is a progressive discovery of our own ignorance."
-William James Durant. 1965.

Originality Statement

I declare that I am the sole author of this work. This is a true copy of the final document, including any revision which may have been ordered by my examiners. I understand that my work may be available to the public, either at the university library or in electronic format.

Abstract

There are many disorders that directly affect people's behaviour. The people that are suffering from such a disorder are not aware of their situation, and too often the disorders are identified by relatives or co-workers because they notice behavioural shifts. However, when these changes become noticeable, it is often too late and irreversible damages have already been produced. Early detection is the key to prevent severe health-related damages and healthcare costs, as well as to improve people's quality of life.

Nowadays, in full swing of ubiquitous computing paradigm, users' behaviour patterns can be unobtrusively monitored by means of interactions with many electronic devices. The application of this technology for the problem at hand would lead to the development of systems that are able to monitor disorders' onset and progress in an ubiquitous and unobtrusive way, thus enabling their early detection. Some attempts for the detection of specific disorders based on these technologies have been proposed, but a global methodology that could be useful for the early detection of a wide range of disorders is still missing.

This thesis aims to fill that gap by presenting as main contribution a global screening methodology for the early detection of disorders based on unobtrusive monitoring of physiological and behavioural data. The proposed methodology is the result of a cross-case analysis between two individual validation scenarios: stress in the workplace and Alzheimer's Disease (AD) at home, from which conclusions that contribute to each of the two research fields have been drawn. The analysis of similarities and differences between the two case studies has led to a complete and generalized definition of the steps to be taken for the detection of a new disorder based on ubiquitous computing.

Laburpena

Jendearen portaeran eragin zuzena duten gaixotasun ugari daude. Hala ere, askotan, gaixotasuna pairatzen duten pertsonak ez dira euren egoerataz ohartzen, eta familiarteko edo lankideek identifikatu ohi dute berau jokabide aldaketetaz ohartzean. Portaera aldaketa hauek nabarmentzean, ordea, beranduegi izan ohi da eta atzerazeinak diren kalteak eraginda egon ohi dira. Osasun kalte larriak eta gehiegizko kostuak ekiditeko eta gaixoen bizi kalitatea hobetzeko gakoa, gaixotasuna garaiz detektatzea da.

Gaur egun, etengabe zabaltzen ari den Nonahiko Konputazioaren paradigmari esker, erabiltzaileen portaera ereduak era diskretu batean monitorizatu daitezke, gailu teknologikoekin izandako interakzioari esker. Eskuartean dugun arazoari konponbidea emateko teknologi hau erabiltzeak gaixotasunen sorrera eta aurrerapena nonahi eta era diskretu batean monitorizatzeko gai diren sistemak garatzea ekarriko luke, hauek garaiz hautematea ahalbidetuz. Gaixotasun konkretu batzuentzat soluzioak proposatu izan dira teknologi honetan oinarrituz, baina metodologia orokor bat, gaixotasun sorta zabal baten detekzio goiztiarrerako erabilgarria izango dena, oraindik ez da aurkeztu.

Tesi honek hutsune hori betetzea du helburu, mota honetako gaixotasunak garaiz hautemateko, era diskretu batean atzitutako datu fisiologiko eta konportamentalen erabileran oinarritzen den behaketa sistema orokor bat proposatuz. Proposatutako metodologia bi balidazio egoera desberdinen arteko analisi gurutzatu baten emaitza da: estresa lantokian eta Alzheimerra etxean, balidazio egoera bakoitzari dagozkion ekarpenak ere ondorioztatu ahal izan direlarik. Bi kasuen arteko antzekotasun eta desberdintasunen analisiak, gaixotasun berri bat nonahiko konputazioan oinarrituta detektatzeko jarraitu beharreko pausoak bere osotasunean eta era orokor batean definitzea ahalbidetu du.

Acknowledgements

Atzo balitz bezela akordatze naiz tesi honi ekin niyon egunetaz. Ilusiyoz beteta, laneako gogoz ta ikasteko desiatzen. Harrezkeo, daneako denborie eon da: momentu hobiek eta txarrauek bizitzeko, lan batzun emaitzak jasotzeko ta emaitzaik eman eztebenan alde ona billatzen sayatzeko, hankak lurrien jartzeko ta lorpen txiki bakoitze baloratzen ikasteko...

Ta konturatzeako, pasa da. Esker hitz hauek idatziz eman nahi niyoke tesi honi bukaerie. Izene, asko izen die bide hontan lagundu diben pertsonak, bai maila teknikuen, baita maila pertsonaliene, ta jende honi danai zor diyot lan honen emaitze. Zuek gabe, etzalako hau posible izengo.

Lehenik eta behin, nere eskerrik haundiyena, nere zuzendari izen dan Asier Aztiriai, aukera hau eman izenaatik, hasiera hasieratik nigan jarritako konfiantziatik eta hiru urte hauetan ziher emandako babes, laguntza ta aholku danaatik. Elkarlanien jarraitzeko esperantzaz, lagun bat dakezu Azpeitin ber dezuneako! De plus, j'aimerais remercier mon co-directeur, Adrian Basarab, d'avoir accepté de participer à ce processus sans hésiter et sans rien attendre en retour et de m'avoir aidée avec ses avis et conseils au cours de ces années. Eskerrak baitare, Mondragon Unibertsitateai, tesi hau burutzeko baliabidiek eskeintzearren.

Bigarrenik, egunerokotasunien hainbeste lagundu dieben lankide ta lagunek aipa nahi nituzke. Zuekin konpartiutako kafetxuk ta elkarrizketak izen dielako momentu askotan aurrea jarraitzeko inderra eman diebenak... Aitor A., Aitor L., Enaitz, Mikel I., Alain, Dani, Unai B., Mikel M., Urtzi, Miren, Iñaki G., Unai I., Miriam, Oscar, Pablo, Raúl... ta gainontzeko departamentu-kidiei. Bihotz bihotzez, millesker ekipo, momentuoro eman diezuen babesaatik eta egunero lanea etortzie atsegin bihurtzeatik! Eskerrak baitare, azken aste hauetan departamentu berri hontan hain goxo hartu ta etxin bezela sentiarazi nauzuen lankidioi.

Third, I'd like to mention the colleagues from Washington State University. My sincerest thanks for your great contribution to Maureen Schmitter-Edgecombe, Alyssa Weakley and specially, to Diane J. Cook, to which I owe the opportunity to live an incredible experience in the USA. Thank you so much to the CASAS team: Samaneh, Bryan, Beiyu, Yuchen, Sal and others, for your warm welcome and treatment during my stay, et merci à toi, Milli, pour être ma copine de route et faire partie de mon aventure Américaine! *Go Cougs!*

Azkenik, nere aipamena zuei, nere pentsamentue lanetik apartau ta biziko gauza garrantzitsutaz gozatzea eamate nazuenoi: atte, ama ta Enaitzi, kuadrillakuei ta Jokini, zuentzat nere eskerrik berona, hasieratik nigan sinisteatik ta beti-beti hor eoteatik. Eztakezue meritu makala...

Contents

List of Figures

List of Tables

List of Acronyms

AD Alzheimer's Disease

ADL Activities of Daily Living

AI Artificial Intelligence

AmI Ambient Intelligence

ANN Artificial Neural Networks

ANS Autonomic Nervous System

BOLD Blood Oxygen Level-Dependent

BP Blood Pressure

BVP Blood Volume Pulse

CAAB Clinical Assessment using Activity Behaviour

CAD Computer-Aided Diagnosis

CBF Cerebral Blood Flow

CBFS Correlation Based Feature Selection

CSF Cerebrospinal Fluid

CV Cross Validation

DTI Diffusion Tensor Imaging

ECG Electrocardiogram

EDA Electrodermal Activity

EEG Electroencephalogram

EMG Electromyogram

FA Fractional Anisotropy

fMRI Functional Magnetic Resonance Imaging

GDS Geriatric Depression Scale

GM Grey Matter

HR Heart Rate

HRV Heart Rate Variability

kNN k Nearest Neighbors

LDA Linear Discriminant Analysis

LR Linear Regression

LZC Lempel-Ziv Complexity

MAE Mean Absolute Error

MCI Mild Cognitive Impairment

MD Mean-Diffusivity

MEG Magnetoencephalogram

MLP Multilayer Perceptron

MMSE Mini Mental State Examination

MRI Magnetic Resonance Imaging

MRSI Magnetic Resonance Spectroscopic Imaging

MTL Medial Temporal Lobe

NasaTLX NASA Task Load Index

PC Principal Component

PCA Principal Component Analysis

PD Pupil Diameter

PET Positron Emission Tomography

PRMQ Prospective and Retrospective Memory Questionnaire

RBANS Repeatable Battery for the Assessment of Neuropsychological Status

RBF Radial Basis Function

RCI Reliable Change Index

RF Random Forests

RQ Research Question

RSME Rating Scale Mental Effort

SAM Self Assessment Manikin

SBS Sequential Backward Selection

SCL Skin Conductance Level

SCR Skin Conductance Response

SFS Sequential Forward Selection

sMRI Structural Magnetic Resonance Imaging

SPECT Single Photon Emission Computed Tomography

ST Skin Temperature

SVM Support Vector Machines

SVR Support Vector Regression

TCD Transcranial Doppler

TI Thermal Imaging

TUG Timed Up and Go

WM White Matter

1

Introduction

Contents

This chapter introduces the work carried out in this PhD thesis which deals with the aspect of the early diagnosis of disorders based on behavioural shifts and biomedical signals. In this chapter, we first present the reasons and motivations for this thesis, followed by a statement of the main hypotheses and objectives. Then, the research methodology followed for the conclusion of this work is explained. Finally, the organization of this dissertation is exposed.

1.1 Motivation

There are many disorders (*e.g.* diseases such as Alzheimer's Disease (AD), stress, Attention Deficit Hyperactivity Disorder, Parkinson's Disease, eating disorders, depression, autism...) that directly affect users' behaviours. Usually, people suffering from such a disorder are not aware of their situation, and too often those disorders are identified by relatives or co-workers because they notice behavioural shifts. However, when these changes become noticeable, it is often too late, when the disorder has progressed too much and irreversible damages have been caused. Early detection of these disorders allows one to take measures to prevent them completely or to slow down their progress, reducing the consequent health-related damages and economic costs, as well as improving people's quality of life.

In the last decade, thanks to the emergence and continuous improvement of smart electronic devices, smart environments are becoming increasingly known and frequent around us. Smart environments are surroundings that have been enhanced with sensors integrated in the environment and in specific devices and objects. These embedded sensors, acquire knowledge about the environment and its inhabitants, so that the gathered information can be applied in order to improve users' experience [1]. Therefore, smart environments provide an excellent infrastructure to monitor users' activity and behaviour in a completely transparent and ubiquitous way. Based on this monitoring, systems that detect users' frequent behaviours and shifts have been developed, and validated, for example, in work environments (in an office [2]) and in home environments with patients with AD [3].

Nonetheless, these behavioural shifts are yet to be mapped to the disorders that might be provoking them. From the point of view of healthcare and biomedical fields, a great opportunity is being seen in the use of these smart environments to, among other things, detect disorders in time. For that purpose, it is necessary to develop smart environments that are capable of detecting changes in the behaviour of people indicating the onset of a disease.

Towards this goal, there is much work to be done in several research and organizational areas. Some of these include:

1. Research on the development of new smart devices, environments and cities, towards their massive use and expansion
2. Research on the Big Data area, for the development and/or improvement of the current technical infrastructure to store, transfer and process the big amounts of data that the smart environments can collect.
3. Research on the many individual disorders to understand the behavioural shifts that they provoke and, therefore, be able to implement systems that recognize these symptoms when they arise.
4. Perform a reorganization of the current health-care system to adopt and normalize the use of new technologies and smart environments as support assistance in health-related applications.
5. Begin a sensitization process to increase users' acceptance towards these emerging technologies.

This PhD thesis aims at contributing to the third and fourth point by proposing a multidomain methodology for the early detection of disorders, which defines the procedure to follow for the

research and implementation of smart environments able to detect specific disorders. The use of this methodology can ease the implementation of smart environments providing decision support to the health-specialists, hence, help in introducing these technologies to the current health-care system.

1.1.1 Trends in this research area

Awareness of the advantages that the monitoring of behavioural biomarkers involves for the early detection of disorders (most importantly the possibility to monitor them in an unobtrusive and ubiquitous way) is increasing, and therefore, more and more effort is being done on the research aimed at discovering new behavioural biomarkers for specific disorders. In this sense, changes in in-home activity patterns of the elderly have been associated to dementia [4], changes in computer [5] and smartphone [6, 7] use patterns to stress, specific traits in gait to Parkinson's Disease [8, 9], and certain patterns in the interactions with smart tablets to autism [10]. All these research suggest that changes in behaviour might be the earliest observable symptoms for many disorders and are therefore important to monitor and understand. Some European research projects are also aligned with this hypothesis, *e.g.* i-Prognosis [11], ICT4Life [12] or Brainview [13], which aim at predicting Parkinson's Disease, Alzheimer's Disease and other dementias, as well as autism and attention deficit hyperactivity disorder from behavioural traits.

Nonetheless, all existing works focus on the detection of a single or few specific disorders, without taking into account the similarities and analogies that exist between them allowing to define a multidomain methodology. Currently, the actively used methodology for the early detection of multiple disorders is the screening method, defined by the World Health Organization in 1968 [14]. This methodology is not up to date with the technological advances that have taken place in recent years. We believe that a methodology for the early detection of multiple disorders that provoke behavioral shifts could be defined making use of technologies such as Smart Environments.

The widespread use of such a methodology for different disorders can highly contribute to the discovery and understanding of the correlations and/or causalities that may exist between the behavioural and physiological shifts and the status of the disorders. Consequently, this can lead to increase our current knowledge level about the disorders, to find out new behavioural biomarkers for an earlier diagnosis and to monitor their progress unobtrusively. This would lead to developing ubiquitous and unobtrusive monitoring systems for specific symptoms of disorders, with the resulting early detection.

1.1.2 Scenarios

The development of this PhD thesis will be held focusing on two disorders: workplace stress and AD in home environments. The in-depth analysis of these individual scenarios will allow to focus on the steps to follow in order to give a solution to each one of the two cases, while allowing to infer a global solution by comparing them.

The more individual scenarios analysed, the better and the more robust global solution could be inferred. Nonetheless, for having limited time and space, the work has had to be limited to two case studies. The reasons why such different scenarios have been chosen are the following:

- The workplace and the home environments are the places where people spend most of their time.
- The implication of such different scenarios allows to validate the non-specific methodology that will be useful for a wide variety of disorders.
- Data of stress in the workplace are much more accessible so that they can be used for acquiring experience and developing the first algorithms, in order to minimize the problems that can arise when working with AD patients.

1.2 Hypotheses and Objectives

This section presents the objectives and hypotheses that compose the foundation of the upcoming research work. In order to address this question, we take into account the problem first presented in Section 1.1 and also the trends in current research exposed in Chapter 2.

The central hypothesis of this work is:

"Correlations can be found between physiological and behavioural shifts and the status of the disorders, so that a multidomain methodology for the early detection of disorders can be defined based on shifts detected on unobtrusively collected physiological and/or behavioral data".

In order to validate the hypothesis, the general goal of this research work is:

$\Rrightarrow$ To define a mutidomain methodology for the early detection of disorders based on unobtrusively collected physiological and behavioural data.

This objective can be divided into several sub-objectives:

- Research the current state of the art in stress detection and AD diagnosis, especially the types of symptoms that are measured and how they are used for the recognition of the aforementioned disorders.
- Select the behavioural and physiological signals for each of the two considered scenarios, and develop the required processing algorithms for their characterization and further use.
- Define the characteristics of the required data for the research and collect the data or obtain access to existing datasets of such characteristics.
- Analyse the data using Artificial Intelligence (AI) and machine learning techniques in order to search for correlations and patterns between the state of the disorder and the physiological and behavioural data.
- Integrate and validate the knowledge extracted from the data analysis process in the two scenarios.
- Deduce a global methodology to find useful correlations and markers for the early diagnosis of several disorders.

1.3 Methodology

The research methodology selected for this PhD thesis is the Case Study Research. This strategy can be defined as "An empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident" [15].

Such a study is characterized by the following traits [16]:

- One or few instances are investigated in depth.
- The instances are examined in their natural setting.

- It is a holistic study, where the researcher focuses on the complexity of relationships and processes and how they are interconnected and inter-related, rather than isolating individual factors.
- Multiple data sources and methods are used.

The methodology has been slightly modified from its original version [15] in order to adapt it to the needs of the current research, leading to a more iterative strategy.

Figure 1.1 shows the resulting research method followed in this work, with details about the tasks carried out in each stage and the planning of the publications. As shown in the figure, after a first literature review about the early diagnosis of disorders, specific case-studies have been chosen. Later, a thorough review of the selected cases has been done, and the data sources to be analysed have been identified. Then, individual analyses have been performed for each case, validating the proposed data analysis solution and drawing case-specific conclusions. Finally, a general methodology for the early detection of disorders has been presented generalizing the knowledge acquired on the analyses of individual case-studies.

1.4 Thesis Outline

This dissertation is divided into six chapters. In this section, we provide a short summary of the contents of each of the chapters.

Chapter 1 is this introduction.

Chapter 2 reviews the literature on the early detection of the two disorders under study (*i.e.* stress and AD) from a multimodal point of view, as well as the current strategies for the early detection of disorders, and highlights the gaps of the State of the Art in this area.

Chapter 3 presents the work performed for the first case-scenario on stress detection in office environments.

Chapter 4 explains the work on the early AD detection from smart home data performed for the second case-scenario.

Chapter 5 exposes the multidomain methodology defined for the early detection of disorders based on unobtrusive physiological and behavioural measurements resulting from a cross-case analysis of the two case-studies presented in the previous chapters.

Chapter 6 concludes the thesis by summarizing the main findings and contributions of this dissertation, and provides some possible lines for further work.

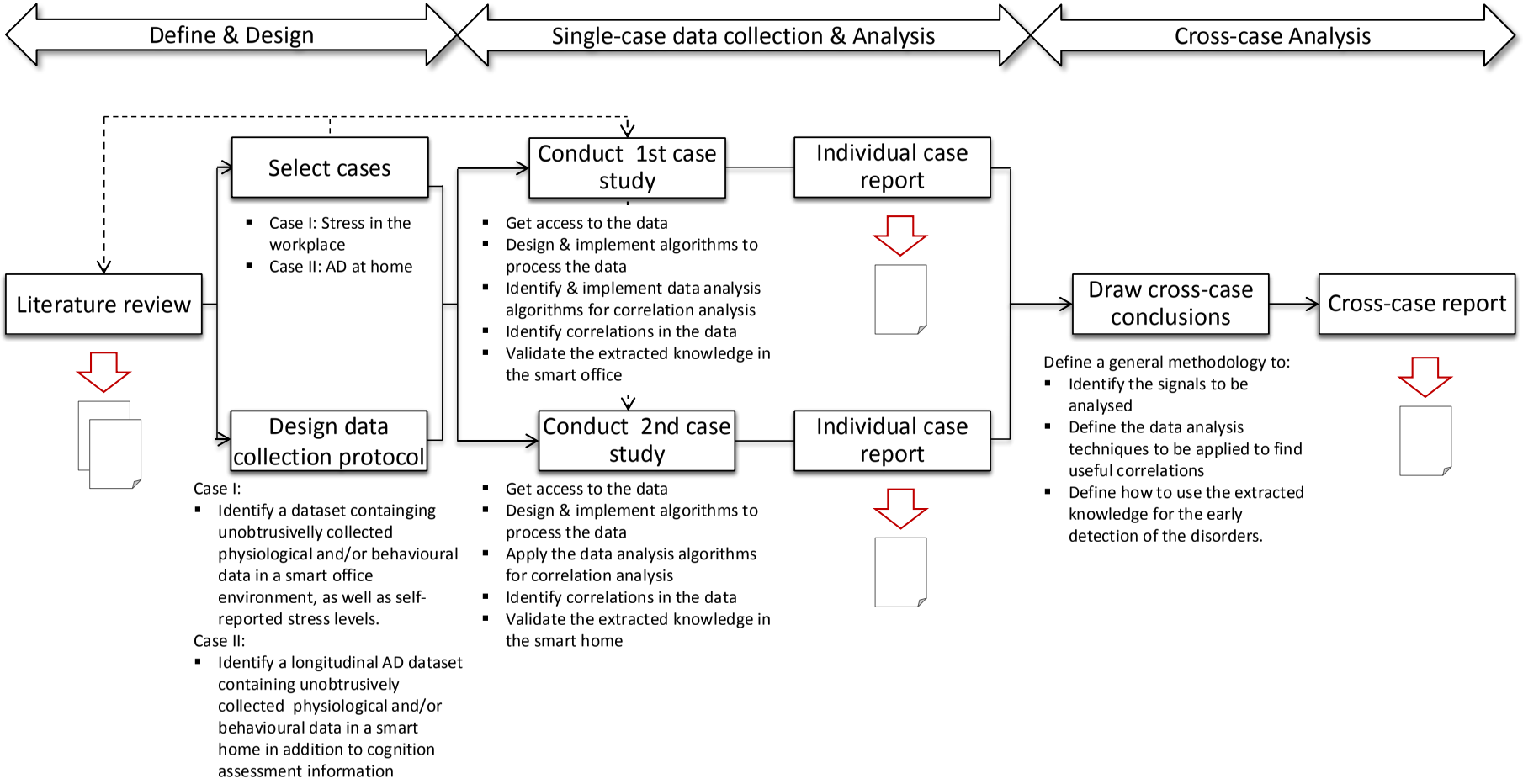


Figure 1.1: Case Study Method used in this thesis project

2

State of the Art

Contents

The literature review presented in this chapter is structured into six main parts. First, a brief section giving the definitions of some basic concepts to understand the following work is exposed. The following two sections explain the current state of the art on the detection of the two disorders under study: stress in the workplace and Alzheimer's Disease (AD) at home. Later, a section about the data mining process and the steps followed to model the disorders is provided. Then, a section explaining the current methodology for the early detection of disorders is exposed. Finally, the open research areas are listed.

2.1 Definitions

This section aims at introducing and clarifying the definitions of some of the basic concepts that will be used throughout this document. These terms could be understood in multiple ways, and therefore, this section tries to clarify the definition that has been taken into account during the development of this thesis to avoid any kind of ambiguity.

"Multidomain" and *"multimodal"* terms are extensively used throughout the document, thus, it is important to clearly define them.

In this thesis document, the term *"multidomain"* is defined as:

Definition 2.1.: A system involving or providing a solution for several domains or disorders, which could be stress, AD, depression, bipolar disorder, and so on.

Whereas the term *"multimodal"* is defined as:

Definition 2.2.: A system involving information sources from several modalities, which are psychological, physiological and behavioural.

The terms *"psychological"*, *"physiological"* and *"behavioural"* are used to present the taxonomy of the symptomatology of the disorders under study in this thesis project. Below, a definition of each one of the modalities is given.

"Psychological" symptoms are defined as:

Definition 2.3.: "Of or relating to the mind or mental activity" [17] and they do not involve the execution of an action.

"Physiological" responses are defined as:

Definition 2.4.: Part of the normal functioning of a living organism or bodily part [18], therefore, they are non-voluntary actions or responses, and very hard or impossible to notice by external observation.

"Behavioural" is interpreted as:

Definition 2.5.: "The manner of conducting oneself" [19], so that, unlike physiological responses, they involve an action that could be controlled or relatively easily observed externally (*i.e.* by colleagues, by family members).

2.2 Stress detection

This section aims at responding the following Research Question (RQ)s:

What is stress? What are its symptoms? How can these symptoms be measured in order to detect stress?

Introduction

Stress was defined for the first time by Hans Selye as "the non-specific response of the body to any demand for change" [20]. When these demands come from job-related sources, we are talking about "occupational stress", which has more specifically been defined as "the emotional cognitive, behavioural and physiological reaction to aversive and noxious aspects of work, work environments and work organizations. It is a state characterized by high levels of arousal and distress and often by feelings of not coping" [21].

Three levels of stress can be distinguished depending on the time of exposure to stressors:

- Acute stress: It is the innate "flight-or-fight" response in face of stressors, and it is not considered harmful [22]. It is provoked by punctual stressors, for instance, a job interview. The sympathetic nervous system, which is part of the human Autonomic Nervous System (ANS) [23], is responsible for activating the glands and organs for defending the body from threat, *i.e.* for the stress response [24].
- Episodic stress: It is said when acute stress episodes go from being punctual, to be more frequent.
- Chronic stress: It appears when stressors persist for a long-time, causing physical and mental problems on the worker. Hypertension, musculoskeletal disorders, depression, sleep problems, and suicide attempts are only some examples of the health-related problems that it can provoke.

The Sympathetic Nervous System provokes the stress response in humans [24], provoking psychological, physiological and behavioural symptoms [25]. Examples of these modalities include:

- Psychological symptoms of stress comprise the increase of strong negative emotions, such as anger, anxiety, irritation or depression [26].
- Examples of physiological responses include changes in the hormonal levels of the body, sweat production, increased Heart Rate (HR) and muscle activation [27]. Respiration becomes faster and the Blood Pressure (BP) increases [28]. Skin Temperature (ST) and Heart Rate Variability (HRV) vary [29], as well as the Pupil Diameter (PD).
- Behavioural symptoms include changes in facial expressions, head movements [30] and General Somatic Activity. In an office environment, interaction patterns with the computer can also be considered.

Figure 2.1 shows the multimodal nature of stress. It can be seen that stress is affected by the context, *i.e.* by the personal characteristics of the subject and circumstances that are not subject-dependent like events, places or moments. Stress responses are evidenced, at least, in the three aforementioned domains.

Currently, stress is detected measuring some of these symptoms. Whereas psychological questionnaires and salivary cortisol measurements have been considered as the gold standard for a long time, in the recent years, the search field has greatly expanded.

The following of this section reviews the signals that are being used to measure stress levels following the presented taxonomy:

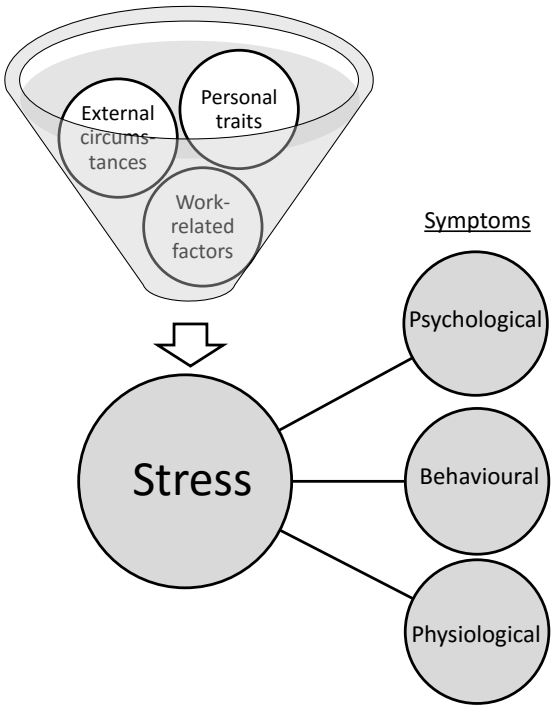


Figure 2.1: The multimodal nature of stress. The figure is inspired from [30].

2.2.1 Psychological evaluation

Psychological evaluation of stress can be carried out by means of self-report questionnaires or by being interviewed by a psychologist. The former is one of the most widely used ways to measure stress levels in humans and it is considered a reliable method. Examples of these tests are: The Stress Self Rating Scale [31], the Perceived Stress Scale and the Stress Response Inventory.

Strengths and Weaknesses of Psychological Evaluation to Detect Stress

The advantage of using psychological questionnaires to detect stress is their reliability. They are the result of years of work by a whole branch of psychology, and their validity has been proven throughout this time. Despite this, their weaknesses are more remarkable when the goal is the early detection of stress.

On one hand, these questionnaires do not provide information about the stressors nor about the evolution of the stress levels. On the other hand, these tests can be taken from time to time, but this might not be enough to detect the subtle changes which could indicate an early stage of a major problem. Actually, these tests are only taken when the affected or their colleagues realize about the situation, and this is too late in the vast majority of the cases. Furthermore, questionnaires are subjective and require users' full attention.

2.2.2 Physiological measurements

Stress can be evaluated by a wide variety of physiological measurements. While some of them are widely accepted for this purpose, the validity of others is still a research topic. This section presents the most relevant physiological signals used for stress detection, which are shown in Figure 2.2.

Hormone levels

The stress response changes the endocrine and immune systems by releasing adrenaline and cortisol hormones [32] from the adrenal cortex and the adrenal medulla, respectively. Under stress, the ability to regulate cortisol levels decreases [33] and consequently, people suffering from chronic stress have elevated cortisol levels.

Cortisol levels are considered a reliable biomarker of psychological stress [33] and can be measured in blood, urine or saliva, being the latter the preferred one due to its non-invasive nature [34].

Electrocardiogram (ECG)

The ECG is the recording of the electrical activity of the heart, which reflects directly the mechanical activity of the heart function [35].

ECG is one of the most used signals in stress detection research because the activity of the heart is clearly affected by ANS changes [36]. An ECG can be easily recorded placing some electrodes on specific places of the body and measuring the potential difference. The number of electrodes and their positions can vary, but one of the most simple and effective ways is the Lead-II configuration, which consists of placing three electrodes: one on the right arm, one on the left arm and the last one on the left leg. The most common features computed with an ECG are the HRV and its variants.

Effectiveness of ECG and HRV features for stress detection has been shown by several authors [27, 37, 38].

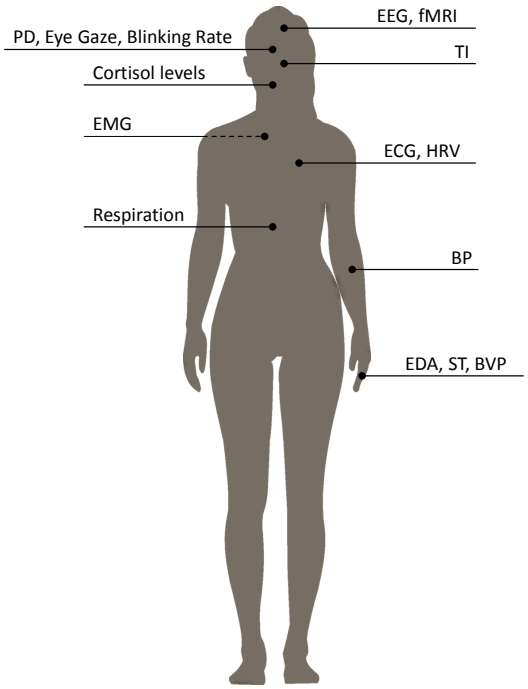


Figure 2.2: Physiological measurements for stress detection

Electroencephalogram (EEG)

An EEG is a test that measures the electrical activity of the brain. It is monitored by placing an array of electrodes on the subject's scalp so that the electrical fluctuations are recorded. The number of

electrodes depends on the application. EEG signals can be divided into four main frequency bands: Alpha (8-13 Hz), Beta (13-30 Hz), Delta (0.1-4Hz) and Theta (4-8 Hz).

Alpha waves reflect a calm, open and balanced psychological state, so Alpha activity decreases in stress situations [33]. Besides, Beta activity reflects cognitive and emotional processes [39] so it increases with mental workload and thus with stress. Stress has also been related to changes in Right Frontal Activity [40] provoking frontal asymmetry.

EEG has not only been used for stress detection [33] but also for emotion recognition based on arousal and valance [40].

Electrodermal Activity (EDA)

The EDA, also known as Skin Conductance or Galvanic Skin Response, is defined as a change in the electrical properties of the skin [41]. Under emotional arousal, increased cognitive workload or physical activity, the level of sweating increases, changing the skin properties, *i.e.* increasing conductance and decreasing resistance [36, 42]. EDA can be measured placing two electrodes on the skin surface next to each other and applying a weak electrical current between them. EDA can be decomposed into Skin Conductance Level (SCL) and Skin Conductance Response (SCR) signals (also called EDL and EDR), which are the background tonic and the rapid phasic components respectively. SCR appear as a small wave superimposed on the SCL [43]. SCRs can be both Event Related or Non-Specific, depending on whether they have been a response to a specific stimuli or not.

EDA measures activity in endocrine sweat glands, which are innervated by the sympathetic sudomotor nerves [44], and is therefore one of the best real-time correlates of stress [45]. It is linearly related to arousal [46] and it has been widely used in stress and emotion detection [24, 45, 47, 48].

Blood Pressure (BP)

BP is the pressure of the blood against the inner walls of the blood vessels and it can be measured using a stethoscope and a sphyngomanometer [49].

It has been proven that stress increases BP [25] over time. Nevertheless, it might not be suitable for detecting subtle stress responses in real-time [23]. This could be explained by the fact that unlike HRV, which is regulated by the "central command", BP is regulated peripherally and is influenced by local conditions in working muscles which could mask the changes of mental workloads.

Skin Temperature (ST)

Physiological variations in the ST mainly come from localized changes in blood flow caused by vascular resistance or arterial BP, which in turn are influenced by the ANS activity [50], suggesting that stress level changes ST. ST can be easily measured placing a temperature sensor in contact with the skin, usually on the hands.

The way ST responds under stress has been analysed in stress and emotion detection researches [48, 50, 51, 52]. However, ambiguous results have been reported. Whereas some affirm that ST in hands rises with stress [25] others have stated the opposite [53].

Facial skin temperature is also starting to be measured by emerging non-invasive technologies such as Thermal Imaging (TI).

Thermal Imaging (TI)

Several existing studies have affirmed that stress can be measured from thermal images due to the temperature changes suffered from stressed individuals [54, 55, 56]. Facial temperature can be easily measured using an Infrared camera, which is a completely unobtrusive method, making it interesting for office-place applications. In the past few years, this technique has been included in the set

of stress measuring methods. Warming of the corrugator muscle has been reported as the best indicator [55, 57]

The promising results obtained with TI, have led other researches to analyse facial blood flow under stress situations with even more sophisticated methods such as Hyperspectral Imaging [58].

Electromyogram (EMG)

An EMG measures the electrical activity of the muscles by using electrodes placed over the muscle of interest, which is usually the Trapezius. As it is known that stress elevates muscle tone, many researches have been done to analyse the potential of EMG for stress level assessment.

These works have concluded that a significant increase in Trapezius muscle activity is suffered during mental stress, provoking involuntary reactions [38]. This increase is translated into an increased EMG amplitude and a decrease in the amount of gaps, *i.e.* short periods of relaxation. Low frequency contents also increase significantly under stress situations, so EMG signals give a useful information for detecting mental stress.

Respiration

In 1973 researchers from the department of psychology of Peking University discovered that when the stress level changes, the speed and depth of respiration system also change [59]. Due to this finding, respiration has been measured in many stress-related researches [27, 45, 59, 60] together with other physiological signals. Respiration can be measured with a pneumotachometer (or pneumotachograph).

Unfortunately, the literature suggests that respiration monitoring is not as worth as other physiological signals. Healey et al. [45] found out that contribution of respiration signals to stress detection was far from being as evident as EDA or HRV's contribution. Wei [59] also qualified respiration signals as less effective for stress classification than EMG signals.

Blood Volume Pulse (BVP)

BVP is the measure of the volume of blood that passes over a photopletismographic sensor with each pulse [61]. Photopletismography, consists of measuring blood volume in skin capillary beds in the finger, relying on the capability of blood for absorbing light.

BVP allows to measure information of Heart Rate Variability (HRV) non-intrusively [36] so that it can be used instead of Electrocardiogram (ECG)s.

Pupil Diameter (PD), Eye Gaze and Blinking Rate

PD, eye gaze and blink rates can be measured with infrared eye tracking systems or with Image Processing techniques applied to visual spectrum images of the eyes.

Pupil dilations and constrictions are governed by the Autonomic Nervous System (ANS) [36]. Thus, PD exhibits changes under stress situations [62] and literature suggests that it can positively contribute to stress detection [52, 63, 64]. Eye gaze has also been measured in different situation such as in reading [65], in driving [66] and in working with a computer [63]. Eye gaze spatial distribution has been found to be positively correlated with stress levels. Regarding blinking rate, an increase under stress has also been reported in several studies [57, 67].

Functional Magnetic Resonance Imaging (fMRI)

fMRI is a non-invasive imaging technique for the functional analysis of the brain. It consists of measuring the oxygen concentration of the different brain areas when the subject is developing different tasks or when he is at the rest state for evaluating the default mode network. A Blood

Oxygen Level-Dependent (BOLD) image contrast that provides an indirect measure of neuronal activity is achieved [68, 69]. This is also done with a Magnetic Resonance Imaging (MRI) scanner. Therefore, fMRI can be used to produce activation maps showing the parts of the brain that are involved in a particular mental process.

The number of research papers on functional brain activities associated with emotional stress using fMRI has increased in the recent years for several reasons. On one hand, it is non-invasive and it does not involve radiation, making it safer. On the other hand, it is easy to use and it has a good spatial resolution. The downsides are its lack of temporal resolution and high cost. That is why the use of EEG has been preferred [70]. Furthermore, this method is restrictive by nature and it does not allow monitoring in the workplace [55].

Table 2.1 summarizes the aforementioned physiological signals and features present in the literature.

Strengths and Weaknesses of Physiological Measurements to Detect Stress

A wide variety of physiological measurements has been reviewed in this section. Even if all the methods rely on different principles, they share most common strength and weaknesses.

In fact, most of the aforementioned techniques need a specific device to be placed in direct contact with the user, making them obtrusive methods, as well as costly, due to the need of extra equipment. Wearables are being developed so as to overcome the obtrusivity of these methods, but the continued use of most of them is not realistic yet. Despite these drawbacks, such methods are the preferred ones nowadays to detect mental stress due to their high reliability.

2.2.3 Behavioural responses

Behavioural responses may also be used to identify stress. Advances in technology increasingly allow to measure the behavioural changes shown by people under stress, and although they are not as present as the physiological signals, they have also been considered in the literature. This section presents the most relevant behavioural signals used for stress detection, which are shown in Figure 2.3.

Keystroke and mouse dynamics

Keystroke dynamics refers to the study of the unique characteristics that are present in an individual's typing rhythm when using a keyboard or keypad [81]. The same way, mouse dynamics are affected by the subject's characteristics when moving the computer's mouse or clicking on its buttons.

Keystroke and mouse dynamics have been widely analysed in the security area for authentication of people [82,83]. Zimmermann et al. [84] first mentioned the possibility of using mouse and keyboard dynamics information to measure the affective state of the user and thenceforth they have also been widely used for emotion recognition [85] as Kolakowska et al. explain in their recently published review [86]. Stress detection, although to a lesser extent, has also been the objective of some researches based on keystroke and mouse dynamics. Small variations in these patterns have been attributed to stress.

The biggest advantages of using a keyboard and a mouse for this purpose is that the developed technique is not intrusive and there is no need of any special hardware.

Table 2.1: Physiological features of the stress detection literature

Signal	Domain	Features
ECG	Amplitude	μ, SD, power and energy [38]
	HR	μ [71], min, max [62, 72]
	HRV	μ [71], LF/HF [23, 70, 73], HF/All [29], (LF+MF)/HF [45], tone, energy, complexity [29, 74, 75], kurtosis, skewness, κ_{2-4} [38], SDNN, SDANN, RMSSD, pnn50, SDSD, triangular index (TINN), ULF, VLF, LF, HF [76] and total power [29, 65, 76]
EEG	Amplitude	μ, fractal dimension [77, 78], higher order crossings [78], Hjorth parameters. [77]
	Event Related Potentials	Mean amplitude [70]
	Spectrum	Power of the absolute and relative α_{eeg}, β_{eeg}, δ_{eeg}, θ_{eeg} bands [77].
EDA	SCL	μ [71], SD, min, max, RMS [24, 36, 52, 65, 72], kurtosis, skewness, 1st derivative [79], difference of the max and min from baseline, positions of the max and min, zero crossings [6],
	SCR	Onsets, peaks, durations, magnitudes, n°of peaks [6], peak height [24, 47], mean magnitudes, mean durations [45], latency, rising time, recovery time, sum of magnitudes, sum of durations, amplitude & duration percentiles, sum of the estimated areas under the responses, the average area under the rising half of SCRs [79]
BP	Systolic	Mean and SD [25]
	Dyastolic	Mean and SD [25]
ST		μ [52], min, max, SD [38, 62]
EMG	Amplitude	μ, median, SD, min, max [45], range, minRatio, maxRatio, 1st and 2nd derivatives [27, 59, 80]
	Contraction signal	μ, static, median and peak loads, gaps/min, time between gaps [27, 59, 80]
	Spectrum	Mean and median frequency [27, 59, 80]
Respiration	Amplitude	μ, SD, 1st and 2nd differential, median, min, max [27, 50], range, maxRatio [59], rate
	Spectrum	Power of 0-0.1Hz, 0.1-0.2Hz, 0.2-0.3Hz, 0.3-0.4Hz bands [45]
BVP	Amplitude [52]	
	Inter-beat Interval	LF/HF, μ, SD
	HR	See ECG
	HRV	See ECG
Eye dynamics	PD	μ [52], SD [65], % dilation, ratio of variation [63]
	Eye position	Gaze distribution, % saccades [63], μ, SD, distance, n°of forwards and backwards tracking fixations, and time fixed [65]
	Blinks	Blink rate, eye closure speed [63]
TI	Temperature of facial ROIs' temperature	μ, SD, kurtosis, skewness, IQR, min and max [54, 56, 57, 77]
fMRI	Activation map	Activation of ROIs [31]

Posture

It has been shown that posture is a good indicator of the feelings of the worker towards the tasks they are carrying out [87]. Thus, individuals' postural behaviour may also provide important information about their stress levels.

Anrich et al. [88] measured the posture of office workers using a pressure distribution measuring system installed in their chairs and affirmed that the amount of fast movement increases during stress tests compared to control tests. Others [47] have analysed the posture using visual techniques. Specifically, a Kinect has been used to detect the interest levels of the office workers on the tasks they were involved in. Using techniques such as depth information and skeletal tracking, the inclination of the person and consequently an indicator of the workers' motivation was deduced.

Posture can be analysed in a transparent way for the user, but the use of imaging techniques might lead to privacy problems.

Facial expressions

Automatic recognition of facial expressions has been the subject of many researches [66,89,90]. They can be estimated with computer vision techniques [47].

An optical computer recognition technique to detect facial expressions related to stress induced by workload was tested [89], where eyebrow and mouth movements' measurement were found to be very useful. Facial expressions have also been used for emotion recognition [47], even in combination with head movements [91].

Speech analysis

Many researchers agree with the fact that stress changes human vocal production [92, 93, 94]. It has been found out that changes in pitch (fundamental frequency) and in the speaking rate are usual under stress situations, together with variations in features related to the energy and spectral characteristics of the glottal pulse [93]. Speech analysis has caused interest, principally, because it can be easily measured in a completely unobtrusive way. Furthermore, results of some research suggest that Electrodermal Activity (EDA) results can be overcame by speech features [24].

Nevertheless, voice-based stress analysis can be ineffective both in quiet and noisy spaces [95], due to the lack of speech recordings and to the excessive noise, respectively. Most of the research done in stress recognition from voice, has been carried out in laboratories [24] or in quiet environments [96]. As an exception, the research carried out by Lu et al. [93], aimed at detecting stress both in indoor and outdoor acoustic environments, using mobile phones, with quite satisfactory results. Nonetheless, the use of speech recordings might directly affect users' privacy.

Smartphone use patterns

Nowadays, a huge amount of information related to users' behaviour can be extracted from smartphones. Call logs, SMS, e-mails, internet browsing, app's usage, location data and many other knowledge can be easily obtained without the user even noticing it.

Recently, the usefulness of this unobtrusive information collecting method for stress detection has been affirmed [7]. For example, Sano et al. [6] discovered a change in the amount and length of both sent and received SMS, as well as in their screen on/off patterns.

[1] Figure adapted from "ergonomics-at-work" by Lab Science Career used under CC BY 4.0

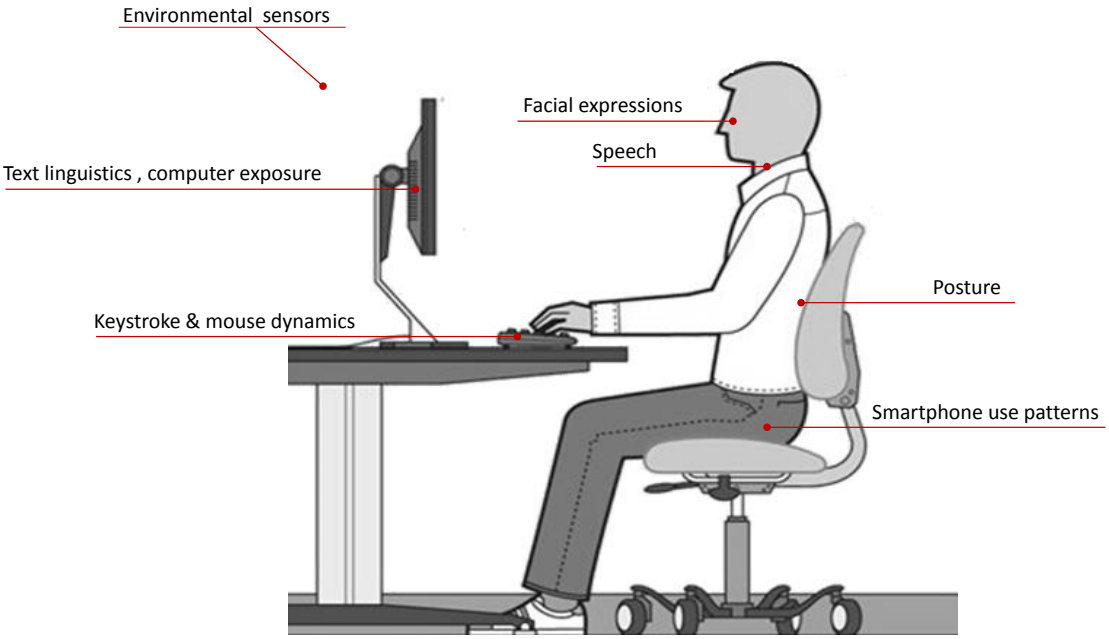


Figure 2.3: Behavioural measurements for stress detection [1]

Computer exposure

It is natural to think that computer exposure of workers changes under high stress levels because high workload is one of the reasons for individuals to be stressed, and this could lead workers to spend more time in front of the computer. Eijckelhof et al. [5] have studied whether stress levels affect the overall human-computer interactions within a day, *i.e.* computer exposure times and breaks' frequency and lengths, using a specific software for this purpose. They affirmed that workers suffering from individually oriented stressors, *i.e.* overcommitment and high perceived stress levels, spend more time in front of the computer during the day, while the workday itself is not extended. In addition, they suggested that these stressors do not affect on their break patterns. Besides, they concluded that workers with high levels of organizationally oriented stressors, *i.e.* effort and reward, tended to have fewer short (30s-5min long) computer breaks and slightly longer breaks (more than 15 min).

Environmental sensors

Cook et al. define smart environments as "one that is able to acquire and apply knowledge about the environment and its inhabitants in order to improve their experience in that environment" [1]. For this purpose, smart environments are surroundings that have been enhanced with sensors integrated in the environment and in specific devices and objects. These environments have been used to detect people's behaviour patterns [3, 97].

Suryadevara et al. [98] were the first ones in carrying the emotion detection problem, including stress, to a Smart House, suggesting that an initial change in regular daily activities can mean changes in health, and hence, in stress levels too. They created a two part monitoring system, which included on one hand, physiological information obtained from Heart Rate (HR), Skin Temperature (ST) and EDA signals in order to determine the emotion of the person. On the other hand, a smart house with wireless sensors to monitor house appliances' using patterns in order to detect abnormal behaviours. Unfortunately, both parts of the system as described in the article were completely independent, and

a method for integrating their results was not even mentioned.

Smart offices are the scenarios where workers' stress levels can be best observed. Thus, some research work have been lately placed in office environments [2, 99, 100, 101]. Nonetheless, most of the reported conclusions do not result from real office-work settings but from experiments under artificial conditions where participants were not performing their usual work and/or stress was elicited with atypical stressors for an office worker.

To overcome this obstacle and get to know stress in the most natural environment possible, Saskia et al. [71] recently analysed the possibility of detecting stressful situations and estimating mental states from unobtrusively collected smart office physiological and behavioural data. More specifically, they built a smart office environment where workers' ECG, EDA, facial expressions, posture and body movements and computer use patterns were recorded all together. These data were collected in an experiment where the participants performed real office-work and were being stressed with common real office-related stressors such as time pressure and e-mail interruptions. They succeeded in accepting their hypothesis and built both stress and mental workload prediction models from the smart office data. They also analysed the importance of building individual stress detection models instead of generic models, concluding that specialized models for particular groups of people with similar characteristics might be much more effective on this task. However, as most of the existing literature does, these authors ignored the temporal nature of stress and only considered the use of instantaneous values of the physiological and behavioural data to create the prediction models.

Text Linguistics

People's writing patterns can vary depending on their stress levels. On one hand, some pressure can enhance the writing abilities of a person, making writings of better quality, using a more diverse lexicon, etc. On the other hand, mood can be directly reflected on the text being written, especially, in free texts. Therefore, analysing text linguistics can be an added value for a stress recognition system.

Currently, there exist many tools that allow to automatically analyse linguistic features, as, for example, LIWC [102], Harvard General Inquirer [103], Semantria [104], SentiStrength [105], Synesketch [106]. These tools can be used both to measure writing performance in users by means of lexical diversity measures, or to directly analyse the "feelings" of the text, which is their main purpose. They count the word-type rate (such as the self-reference rate, or article rate), as well as their polarity, *i.e.* their positivity or negativity, and strength (the degree in which they are positive or negative) ([102, 105]). There is a whole scientific branch dedicated to the Sentiment Analysis of texts, which could be considered the neighbour of stress detection. We refer the user to the review of Taboada et al. [107] for further information.

Whereas Sentiment Analysis techniques have been widely used for analysing, for example, depressive moods [108, 109], only a few studies have focused on inferring stress levels from texts: this is the case of Saleem et al. [110] and Vizer et al. [111]. The former used this technology to analyse online forum posts and detect user stress levels from them.

The latter analysed timing, keystroke and linguistic features in free texts, in order to distinguish between physical stress, cognitive stress and no stress situations. An improvement on lexical performance under both types of stress was found, reducing the number of mistakes, increasing lexical and content diversity and decreasing pause lengths.

Table 2.2 shows a summary of the behavioural measurements and features used in the state of the art.

Strengths and Weaknesses of Behavioural Responses to Detect Stress

Behavioural measurements for stress detection are much less in number than the physiological ones. Nonetheless, the rapid advances in technology are allowing to broaden the field. Behaviour can be monitored unobtrusively and most of the time in a completely transparent way for the user. Furthermore, for certain measurements such as computer usage patters, no extra equipment is needed, which makes them more affordable. Nonetheless, sometimes, privacy issues can arise, for instance, in the case of methods that require the use of vision techniques or mobile phones.

Table 2.2: Behavioural features for stress detection used in the literature

Signal	Domain	Features
Keyboard use	Keystroke	(Average) duration between keystrokes, (average) dwell times (the time a key is pressed), number of keystrokes, typing speed, use of particular keys, pause rate, time between two consecutive keystrokes, duration of digraphs and trigraphs, n°of events [71, 112, 113]
	Pressure [46]	
Mouse use	Movement	[x,y] coordinates, total distance, stillness, horizontal, vertical and total velocity, acceleration, jerk, angular velocity and average speed against the movement direction [46, 71, 112, 113]
	Distance	Covered distances, euclidean distances, difference between covered and euclidean distances and time elapsed between the following events: two button presses, a button press and the following button release, two button releases, a button release and the following button press [112]
	Clicks	n°of clicks , menu and toolbar clicks [112]
	Wheel use	n°of wheel events [46, 112, 113]
Posture	Chair Pressure	μ of several frequency bands [88]
	Lean direction	Gradient front-to-back, gradient side-to-side [47], distance [71]
	Joint coordinates & angles	μ and SD of joint angles, μ and SD of bone orientations [71]
Facial Expressions	Points of Interest	Mean smile intensity, mean eyebrow activity, mean mouth activity [47, 66, 89]
	Facial information and emotions	Head orientation, facial movements, action units, emotion [71]
Speech	Amplitude	Speaking rate, voiced and unvoiced speech [24, 92]
	Intensity	μ, range and variability [93]
	Pitch (f0)	μ, min, max, SD, median, jitter, range, 1st derivation [24, 93, 96]
	Spectrum	Spectral centroid, smoothed energy, energy $>$ 500Hz
	MFCC [2]	μ, σ^2, min, max of the first 12 cepstral, δ and δ-δ coefficients [24, 92]

[2] Mel Frequency Cepstral Coefficient

Signal	Domain	Features
	TEO-CB-AutoEnv features [3]	f0 and harmonic related parameters. [93]
	RASTA-PLP [4]	μ, σ^2, min, max
Smartphone use	Calls	n°of calls, time spent in calls, μ, σ^2 and median of call duration [7], incoming/outgoing calls, μ, SD and median of unanswered calls and n°of people contacted by call
	SMS	μ, SD and median of time spent writing SMS, μ, SD and median of the SMS' length, n°SMS, received/sent SMS, n°of people contacted by SMS [6]
	Screen use	μ, SD and median of time spent with screen on and n°of times the screen went on [6]
	Contacts list	Changes of the n°of contacts, n°of phone numbers, and n°of mail addresses
	Battery	Ratio between time not charging and charging
Computer exposure	Interaction	Total interaction, (n°of short breaks, n°of medium breaks, n°of long breaks) / workday [5], use of applications (n°of app and tab changes) [71]
	Log on/off	Length of workday
Environmental sensors	Activity	Frequency and time spent in ADLs [114]
	Inactivity times	Wellness function [98]
Text linguistics	Free text [111]	Lexical & content diversity, average word length, average sentence length, polarity, strength, rates of: nouns, verbs, function words,conjunctions, cognition operations, emotive words, modifiers, adjectives, intensity words, positive and negative affect, sensory information, passive tense, other references, modal verbs, negations, group references, self-references, generalizing terms [110]

Summary

- Stress' symptoms are evidenced psychologically, physiologically and on people's behaviour.
- Psychological symptoms are assessed by interviews or self-reported tests.
- A wide variety of physiological symptoms exist and their measurement vary depending on their nature. In general they are obtrusive methods and might be costly, but they are the most reliable.
- Behavioural symptoms can be measured in a completely unobtrusive and transparent way but their effectiveness must be further researched. Moreover, care must be taken to always ensure users' privacy.

[3] Teager Energy Operator based non-linear transformation
[4] Relative Spectral Transform - Perceptual Linear Perception

2.3 Alzheimer's Disease (AD) diagnosis

This section aims at responding the following Research Question (RQ)s:

What is AD? What are its symptoms? How can these symptoms be measured in order to detect AD?

Introduction

AD is a progressive, degenerative disorder that attacks brain's nerve cells, or neurons, resulting in loss of memory, thinking and language skills, and in behavioural changes [115]. It mostly affects people over 65 years old and its incidence rate grows exponentially with age [116]. It is the most common form of dementia [117] and despite what some people may think, AD is not a normal part of ageing.

People developing AD undergo three different stages:

- Preclinical AD: Changes in the brain may start happening, but the patient does not show any symptoms [118]. Therefore, nowadays, this phase can not still be detected. In fact, it is believed that this stage can start 20 years before any symptom is evidenced.
- Mild Cognitive Impairment (MCI): In this stage, symptoms related to thinking ability may start to be noticeable for the patients themselves and for the nearest family members, but they do not affect their daily life [118]. Not all the people diagnosed with MCI develop AD, but only an estimated %10-%15 of them every year [119, 120] and the reason why some people develop dementia and others do not, remains still unknown. Two different types of MCI are distinguished: amnesic MCI and non-amnesic MCI [121]. The former refers to patients who have impairment in the memory domain, and the latter to patients who have impairment in one or more non-memory domains of cognition, as, for example, attention or language processing. It is believed that subjects suffering from amnesic MCI are more likely to develop AD [122].
- Dementia due to AD: Memory, thinking and behavioural symptoms are already evident and affect the patient's ability to function in daily life.

People suffering from AD, show symptoms of several types and in different degrees, depending on the progression level of the dementia. These symptoms can be distinguished into three main modalities, which are physiology, psychology and behaviour. Figure 2.4 shows the multimodal nature of AD.

- Psychological symptoms include changes in mood and personality. Depression is the commonest symptom, but apathy, irritability, agitation, euphoria, disinhibition, delusions and hallucinations are also part of AD symptomatology [123]. These patients can also become suspicious, confused, fearful or anxious [124].
- The physiological symptoms of AD include accumulations of big amounts of Amyloid plaques (Aβ) and Neurofibrillary Tangles (aggregates of hyperphosphorylated tau protein) [124], neuronal death [118] with its consequent cortical and hippocampal atrophies [121] and cerebral hypoperfusion. Reduction of the volume of the hippocampus is probably the most common pronounced change [125], being a symptom which is already evidenced in the mild stage (reductions of about 15-30%) and which worsens over time [126].

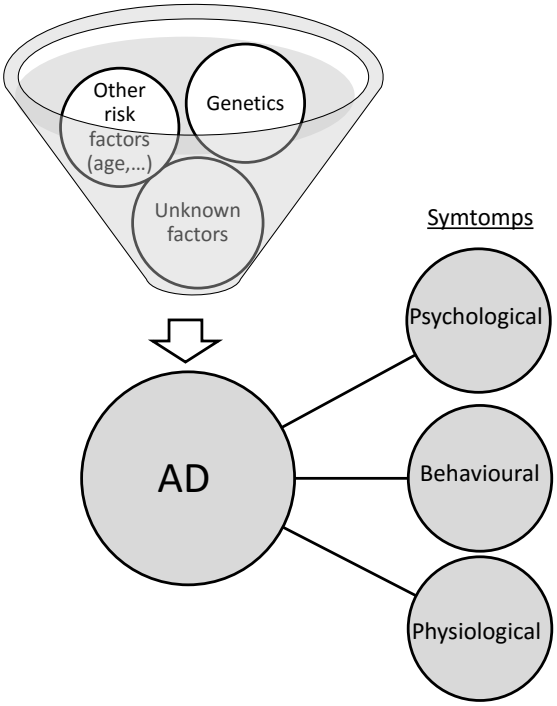


Figure 2.4: The multimodal nature of AD

- Behavioural symptoms include the cognitive changes and their direct consequences. The clinical hallmark and earliest manifestation of AD is episodic memory impairment [127]. People in early AD stages may also have difficulties solving daily problems, for example, with number-related tasks as managing finances. As the disease progresses, this cognitive symptoms become much worse, and the patients start having troubles recognizing people nearby, including family members [124, 128]. Progressive deterioration of cognition leads to incoherent behaviour and limits the patient's capacity to perform his tasks of everyday life, who may take much more time than before performing daily activities, start being much less sociable due to communication difficulties, or have problems in driving due to deteriorated vision. Gait and balance dysfunction [123] might also arise.

Currently, AD is detected measuring some of these symptoms. Whereas questionnaires about the cognitive impairment such as the Mini Mental State Examination (MMSE) have been widely used, protein levels in Cerebrospinal Fluid (CSF) have been measured for a more objective diagnosis. In the recent years, neuroimaging modalities are gaining strength and Positron Emission Tomography (PET) and MRI are already considered essential for a reliable AD diagnosis.

The following section reviews the signals/images and features that are being researched for the early diagnosis of AD.

2.3.1 Psychological evaluation

As depression is one of the most frequent non-cognitive symptoms in AD, psychological evaluation is mainly focused on depression measurement.

Tests like the Montgomery and Åsperg Depression Scale, the Geriatric Depression Scale (GDS), the Cornell Scale for Depression in Dementia and the Nurses Observation Scale for Geriatric Patients are some possibilities to assess depression levels in AD patients [129].

Strengths and Weaknesses of Psychological Measurements to Detect AD

Psychological questionnaires are not the commonest methods used to evaluate the presence of dementia in the patients. Even if they can reliably measure depression levels, this is just a typical symptom of AD, but its incidence does not always mean that there is a cognitive impairment problem. Furthermore, as it happens with other questionnaires and measurements, this are not realistic for a continuous monitoring of the patients. Therefore, they are not the most suitable method for the early detection of dementia.

2.3.2 Physiological measurements

The use of measurements and imaging modalities that reflect the physiological state of AD patients is probably the most reliable method to verify the real state of the disease. While some of these methods are already accepted for the clinical use, the validity of others is still a research topic. In this section, the current use and state of the physiological measurements that have been considered for AD research is introduced (See Figure 2.5).

Cerebrospinal Fluid (CSF)

Reliable and valid biomarkers of AD have been found in biofluids, including CSF [130]. CSF is a clear fluid that surrounds the brain and spinal cord mainly for protection, and it must be obtained by lumbar puncture [131]. CSF "is the only body fluid in direct contact with the extracellular space of the brain and thus biochemical changes due to pathological brain processes are more probable to be reflected in CSF than in other body fluids" [132].

Decreased Aβ42 values have been found in the CSF of AD patients compared to healthy subjects [133], as well as increased total tau and some specific tau epitopes (p-tau231, p-tau181 and p-tau199) [133,134]. Other chemical components like CSF Isoprostanes have been found to be increased in AD patients even in the preclinical stage [135]. The amount of CSF in the Hippocampus region is also related to AD [136], probably due to the decreased size of the Hippocampus, which leaves space for more CSF.

Blood tests

Blood samples can be obtained in a less intrusive and less costly way [131] and more frequently than CSF samples, and thus, AD biomarkers on blood have also been searched.

Features extracted from blood samples are similar to the ones extracted from CSF, but, so far, it is no clear if blood samples could help in discriminating AD and healthy patients, neither if they could serve as a predictor. Blood-based biomarkers of AD have not been still accepted due to the "failure to replicate findings" [131] and to the ambiguous results obtained in different studies.

Structural Magnetic Resonance Imaging (sMRI)

sMRI is a non-invasive imaging technique for structural analysis. Shortly, it consists of applying strong magnetic fields to the area that is wanted to image (*e.g.* the brain) while the different tissues are distinguished thanks to their particular relaxation responses, *i.e.* the radiofrequency signal emitted by the protons of each tissue, [69] when the magnetization stops. This is done with an MRI scanner.

sMRI is currently used for the diagnosis of AD [137]. This technique can help diagnosing AD in two ways: on one hand, it allows to measure Medial Temporal Lobe (MTL)'s atrophy, which is closely related to cognition and memory, with very high definition [138,139] and on the other hand, it

enables changes on tissue characteristics due to vascular damage to be detected [69]. MTL's atrophy is earliest evidenced in the Hippocampus and the Entorhinal Cortex [137, 140, 141, 142].

Making use of image processing techniques applied to sMRI, Computer-Aided Diagnosis (CAD) systems have been developed, achieving satisfactory results distinguishing AD patients from normal controls [125, 137, 140, 141, 143, 144, 145, 146] and less convincing (yet promising) results in distinguishing MCI from normal controls [125, 137, 140, 141, 145].

Currently, MRI's role is quite blurry in the early disease stages [147]. Atrophy of the Hippocampus can be differenced clearly in AD patients compared to healthy people, but, unfortunately, it may not be so obvious at the early stages, hindering the use of sMRI for early detection. Furthermore, brain atrophy is not specific to AD [69]. Despite all this, MRI scanners are highly available nowadays and the results that are being achieved are encouraging, so further research is worth.

Functional Magnetic Resonance Imaging (fMRI)

The use of fMRI in AD research has contributed to brain activity-related discoveries. AD patients have reduced activity in the MTL [148], particularly in the Hippocampus [148, 149, 150, 151, 152], but also in the Entorhinal Cortex [148], while an increased activation has been reported in the Prefrontal Cortex, probably, due to a compensation mechanism [153, 154]. Deactivation in Posteromedial Cortical areas such as the Posterior Cingulate Cortex and the Medial Parietal Cortex has also been found to be anomalous in AD patients [155, 156].

Nevertheless, these anomalies are much less evident in MCI patients, which could difficult the use of fMRI as an early detection component.

Some few examples of automatic analysis of fMRI images in AD detection can be found in the literature. These include the works of Khazaee et al. [157] and Tripoliti et al. [158, 159].

The biggest advantages of fMRI are probably its noninvasive and no radioactive nature because as they can be used safely in a repetitive manner [68], they facilitate the longitudinal studies. Furthermore, it offers a relatively high spatial and temporal resolution [121] of the activation map of the brain.

Magnetic Resonance Spectroscopic Imaging (MRSI)

MRSI, is a non-invasive imaging method that can be performed in a standard MRI scanner. It can detect biochemistry by using signals from organic molecules, allowing in-vivo detection and measure of concentration of some low molecular weight metabolites [68, 160, 161]. Based on the phenomenon of "chemical shift" [162], MRSI provides a spectra in which each peak represents a metabolite or group of metabolites. The area under the peak is related to the concentration of the metabolite. These metabolites include Myo-Inositol (mI), Choline (Cho), N-acetyl Aspartate (NAA), Creatine (Cr), glutamate and Glutamine (Glu) [163].

MRSI has allowed to find in AD patients metabolite abnormalities like decreased NAA or NAA/Cr levels [164, 165, 166, 167, 168, 169, 170], elevated mI/Cr ratio [165, 168], increased or decreased Cho/Cr ratio levels depending on the stage of the disease [171] and decreased Glu levels [168, 169, 170] in the Grey Matter (GM). NAA/mI ratio has been found to be useful for distinguishing between AD patients and healthy subjects. In fact, some affirm [172, 173, 174] that this is the most robust marker of the disease. MRSI could also help in the prediction from MCI to dementia. Some studies have reported lower NAA/Cr [175, 176, 177, 178, 179] and higher Cho/Cr [180] levels in several brain regions in MCI patients who developed dementia than stable MCI subjects. Nevertheless, some disagree with these findings [181, 182] so further research is needed to verify MRSI's predictability from MCI to AD. Some researches have also affirmed the potential of MRSI to help in distinguishing different types of dementia from AD, such as Frontotemporal Dementia [183].

Nevertheless, MRS has some drawbacks. The low signal-to-noise ratio of the images, the long acquisition times [161], which in turn makes this system sensitive to motion artifacts [160], and their low spatial resolution are some of them. Consequently, MRS "is little used in the clinical evaluation of subjects with dementia" [184]. Furthermore, for the best of our knowledge, CAD systems based on MRSI have not been reported up to date.

Transcranial Doppler (TCD) Ultrasonography

TCD ultrasonography is an imaging technology that has been used to image direction and speed of Cerebral Blood Flow (CBF) [185]. It is based on the Doppler effect, and it is executed with an ultrasound probe [186].

An increased carotid intima-media thickness in AD patients compared to healthy subject has been found [187] by this method, as well as a higher degree of carotid atherosclerosis [188]. The total CBF [189, 190], the cerebrovascular reserve capacity and the mean flow velocity [185, 191, 192, 193] are decreased whereas the pulsatility index has shown increased values [185].

Vascular impairment can be detected by several imaging methods like PET or Single Photon Emission Computed Tomography (SPECT), but ultrasonography is a non-invasive and cheaper alternative. Unfortunately, nowadays, it can just serve to monitor the vascular system's state for AD prevention and it is no clear if this technique could really serve as an AD diagnosis method.

Diffusion Tensor Imaging (DTI)

DTI is a MRI technique that can provide information about brain tissue microstructure. It can be obtained non-invasively using an ordinary MRI scanner. It takes advantage of the Brownian motion phenomenon suffered by water molecules in human tissues, who affirms microstructure of the human tissues can be inferred from the water molecules' diffusion patterns [194]. In other words, DTI identifies indirectly "the microscopic aspects that provide measures reflecting the patterns in size, orientation and organization of tissue which are supposed precursors to the final stage of macroscopic tissue atrophy" [195, 196].

DTI can provide relevant information about a person's cognitive state, being Mean-Diffusivity (MD) and Fractional Anisotropy (FA) the most important measures used for it [68]. FA has shown significant differences in the cingulum, splenium of the Corpus Callosum, uncinate fasciculus, superior longitudinal fasciculus and frontal lobes between AD patients and healthy controls, and MD in the Hippocampus, splenium of the Corpus Callosum, parietal lobes and temporal lobes. MD has been found to increase with cognitive performance decline, especially in the temporal structures while FA decreases [197]. The Hippocampus area, the Posterior Cingulate Cortex and the Corpus Callosum have also shown moderate early cognitive dysfunction evidence in DTI images [195, 198], which could allow early detection of AD. For this purpose, DTI has shown superior effect sizes compared to volumetric MTL measurements [199].

Machine learning methods have been applied to DTI images in several researches both for automatic MCI and AD diagnosis. O'Dwyer et al. [120] and Dyrba et al. [200] made use of DTI images whereas Wee et al. [201] combined both DTI and fMRI images.

DTI has shown to be a very potential tool in the early diagnosis of AD, because it can detect alterations that cannot be detected, for example, by conventional sMRI [202]. It still presents some drawbacks, because there is still uncertainty about the best choice of diffusion parameters and about the methods to use to manage crossing fibers [68]. Nevertheless, efforts to overcome these obstacles are being done, so, DTI could be soon accepted as a clinical diagnosis tool.

Positron Emission Tomography (PET) Scans

PET imaging is a nuclear imaging technique that provides three-dimensional images of a brain at the molecular and cellular level [203]. It consists of injecting or making inhale a substance, called radiotracer, that contains a positron emitter to the patients, detecting the emitted radiation by a scanner and computing a digital image that represents the distribution of the radiotracer in the body [204]. This distribution might show anomalies in the brain tissues. Depending on the chosen radiotracer, different kinds of PET scans can be done. PET Scans are done with PET scanners and cyclotrons are used for the preparation of the radiotracers.

In AD diagnosis, many different radiotracers have been used for four main purposes: mainly the ^{11}C-PIB to image the accumulation of the Aβ plaques on the brain, the ^{18}F-fluorodeoxyglucose to image the glucose consumption of the brain, ^{11}C-PMP, ^{11}C-MP4A, ^{11}C-MP4B, ^{11}C-Nicotine and others to image the neurotransmitter systems of the brain and finally ^{11}C-(R)-PK11195 to image the inflammation in the central nervous system which can cause neuronal death [205]. The glucose consumption imaging is based on the idea that as brain mainly uses glucose for energy production, glucose metabolism is closely related to neuronal function, both at rest and during functional activation [206, 207].

CAD systems have been developed to try to automatically diagnose AD and MCI with PET scans [208] even combining both ^{18}F-fluorodeoxyglucose and PiB PET scans [209]. A group of investigators of the University of Granada has published several important works proposing automatic PET based AD diagnosis tools [210, 211, 212], reporting high accuracies.

The advantage of PET is that it has the ability to display very mild symptoms [209]. Unfortunately, while theoretically is not a high risk for the patients, it involves exposure to radiation, and, therefore, it is a method that should better be avoided. Furthermore, it is an expensive method and is not highly available, although this fact is changing in recent years [69]. These reasons lead us to believe that PET imaging is not the best-suited method for massive monitoring of the population.

Single Photon Emission Computed Tomography (SPECT)

SPECT or perfusion SPECT is also a nuclear imaging method that tracks CBF and measures brain activity [213]. Its principle is very similar to PET, as both consist on introducing short-lived radionuclides into an amyloid binding molecule, being different the radionuclides used for the two techniques: while PET uses emitting positrons, SPECT uses photons [214]. The two most common radiotracers used for SPECT are ^{99m}Tc-hexamethylpropyleneamine Oxime (HMPAO) and ^{99m}Tc-ethylcysteine dimer (ECD) [215].

SPECT has shown to be a valuable aid for the early diagnosis of AD [216], because it allows to image the hypoperfusion suffered by AD patients. A correlation between the progression of AD and the loss of cortical CBF in various brain regions [217] has been found with SPECT. A significant correlation was also found between the total tau and phosphorylated tau concentrations in CSF and perfusion in the left parietal cortex [218]. Nevertheless, it is not yet clear in which brain areas this hypoperfusion is most evidenced and thus which one would be the most accurate one for AD diagnosis. Temporo-parietal region has been considered practical for the early detection of AD [219], but its sensitivity and specificity is still questioned [216]. Some suggest that Posterior Cingulate Cortex and precunei regions could be more useful [220] while MTL and Hippocampus regions can not be analysed due to the depth to which they are located [221].

CAD systems have been developed using SPECT images and machine-learning techniques [210, 212, 219, 222, 223].

SPECT shows lower resolution and higher variability [224] than PET, but its radiotracers are cheaper and easier to acquire [225], being probably better suited for longitudinal repetitive studies. Furthermore, SPECT can be carried out by means of a Gamma camera, a device that is already

available in most of the greater hospitals [226]. SPECT has also shown the potential to aid distinguishing between AD and other dementias, namely, Frontotemporal Dementia, Vascular Dementia and Lewy Body Dementia, as well as between AD patients and healthy controls [215, 226]. Nevertheless, the heterogeneity of the results suggest that it should be combined with other methods. Weih et al. suggested in their review [226] that SPECT could be better used to rule out AD instead of for diagnosing it, as it presents a much higher specificity than accuracy both in distinguishing AD patients from healthy controls and in predicting progression from MCI to AD. The results reported above encourage SPECT-based AD diagnosis research. Nonetheless, this can be questioned due to its invasive nature provoked by the use of radiotracers.

Electroencephalogram (EEG)

EEGs are the recordings of the electric field of the scalp caused by the electrical signals exchanged between neurons [227]. Thus, they reflect the communication activity between nerve cells, which is of great importance in neurological diseases like AD.

Studies have shown that Electroencephalogram (EEG) may have the potential for an early AD detection. It has been widely accepted that at least 3 types of changes occur in AD patients' EEG signals: the power of low frequencies is found to be increased while the power of the high frequencies is decreased; their complexity, which is the measure of the number of different patterns in the signal [227], is reduced and synchrony or correlation between EEG signals of the different parts of the brain is reduced [227, 228].

Despite these findings, only a few examples of CAD techniques which rely only on EEG signals can be found in the literature [229, 230, 231, 232].

Many researchers [227, 233] support the use of EEG for a longitudinal monitoring of changes in the brain, due to the cheap and non-invasive nature of this method and because of the ease with which anybody can take samples without the need of going to a medical facility each time. It is a "simple, relatively inexpensive and potentially mobile brain imaging technology" [234] but further research is needed for EEG to be included in a clinical AD diagnosis.

Magnetoencephalogram (MEG)

MEG is a non-invasive medical imaging technology. MEG identifies the brain activity by measuring the magnetic field created by the electric current flowing within the neurons. Thus, measurements follow a similar principle to the ones obtained by EEG because both measure the same sources of brain activity. A MEG scanner is needed for this imaging purpose.

MEG findings related to AD are similar to those of EEG. Increased delta and theta activity [235, 236, 237, 238] in frontal and central areas [239] and decreased alpha activity in posterior and temporal regions [239], *i.e.* slower signals, has been reported in several pieces of research. A generalized loss of functional interactions (*i.e.* decreased synchrony) has also been found [235, 240, 241]. People with MCI have also been investigated with MEG, verifying that their symptoms are somewhere between those of AD and controls [236, 242, 243].

Even if MEG has probably been less studied than EEG, its potential for positively contributing to a computer-aided AD recognition system has been proven. Gómez et al. [244, 245, 246, 247, 248] have highly contributed to the use of these signals on AD diagnosis, concluding that MEG really has the potential to discriminate between AD and normal controls. No results were provided for the MCI case.

MEG can be done without placing uncomfortable electrodes and it is less affected by conductivity issues related to the skull and scalp [249], they do not require a reference, they are less affected by volume conduction, and furthermore, they can obtain more sensitive measurements of the cortical activity than scalp EEG. The disadvantage of MEG is the interference that Earth's magnetic field or

the electrical devices can introduce, so measurements must be done in a heavily shielded room with all the electrical devices around switched off, which complicates its use as part of a global routine monitoring system.

Eye dynamics

It has been hypothesized that "the pattern of AD-specific neurodegeneration may affect neural circuitry of the eye movement system in a unique manner that allows the clinical differentiation of AD from other cognitive disorders" [250]. In order to verify this hypothesis, eye movements of AD patients have been compared to those of healthy subjects in many studies and effectively, it has been proven that AD patients suffer from changes in oculomotor and pupillary functions [251]. Specifically, changes in saccades, smooth pursuit function and in the pupillary response have been found by some researchers. Saccades are "rapid, conjugate movements of the eyes, which serve to orient the high acuity foveal region of the retina onto a specific region of visual space" [252]. It is thought that saccades are of particular interest because they are very related to attention and thus, they are likely to be disturbed by cognitive impairments associated with neurodegenerative disorders such as AD, as well as by dysfunctions related purely to oculomotor execution [253]. All these behaviours can be measured easily and in a non-invasive way in a laboratory, making the patients perform specific tasks like the reflexive paradigm, the memory-guided paradigm [253], the gap/overlap task or the antisaccade task [254] while their eyes are being tracked by cameras or infrared systems [255] and image processing techniques.

Despite the promising AD detection power of eye dynamics, only few works (*e.g.* [255]) based on this method are present the literature.

The problem with these measurements is that these anomalies are not always present in AD patients and furthermore, they are not unique to them [253]. Moreover, it is not clear if MCI patients also show signs of these anomalies because while some refuse this fact [256], others have found some evidences [254, 257]. Crutcher et al. [258] have also found that some MCI preferred the new images the same as the control patients, demonstrating the variability between patients' patterns. Consequently, further research is needed to verify if they might be used both as AD biomarkers and as predictors in the early stages.

Strengths and Weaknesses of Physiological Measurements to Detect AD

A wide variety of physiological measurements are being researched for AD and MCI diagnosis, due mainly to the development of new imaging modalities. A review of the measurements can be found in Table 2.3.

Despite all the methods rely on different principles, they share most common strength and weaknesses. Nearly all the reviewed technologies are obtrusive methods that require the use of specific devices. Thus, they are mainly expensive methods, and not practical because the patient must move to the hospital each time a test should be done. Therefore, they are not realistic for a longitudinal and massive monitoring of the elderly. Nonetheless, they are the most objective methods and where we can find the most reliable biomarkers because they directly show the state of the brain.

Table 2.3: Physiological features for AD detection used in the literature

Signal	Domain	Features
sMRI	Intensity	μ and SD [146, 158, 159] of GM intensity
	Volume	GM voxel locations [146], GM, White Matter (WM) and total eigenbrains, GM, WM and total Partial Least Squares -brains [137], GM ICA basis functions [141], hippocampal volume [145], CSF volume [125], relative GM volume [158, 159], medial temporal lobe's intensity and volume change estimation [143], cortical thickness within ROIs [145], structural thickness, contour area, volume and structural curvature of ROIs [140]
	Manifold-based learning features	Coordinates [145]
	Atrophic voxels	Mean intensity, Jacobian of ROIs [145]
DTI	Intensity	(mean [120, 201]) FA [200], (mean [120]) MD [200, 201], (mean [120]) Axial Diffusion, (mean [120]) Radial Diffusion [201], fiber connectivity network [201]
PET	All brain voxels or voxels of interest	Eigenbrains (PCA) [210, 211, 212], ICA based feat. [211], Linear Discriminant Analysis projections [210], intensity of ROIs [208, 209]
fMRI	Head motion	Length of the path [158, 159] (see [259])
	Activation patterns	N°of activated voxels, max z-score, size of the cluster where the max z-score belongs to, n°of significant clusters, % of activated regions that belong to a ROI, total activation of ROIs [158, 159], atlas based ROIs' clustering coeffs. of functional connectivity networks of several f sub-bands [201]
	BOLD response	Amplitude, undershoot and transit time, and amplitude of the regional CBF, of the venous volume, of the vascular signal and of the deoxyhemoglobin signal [158, 159]
	Brain network graph	Measures of functional segregation (clustering coefficient, (normalized) local efficiency), of functional integration (characteristic path length and global efficiency), local nodal measures (degree, participation coefficient, betweenness centrality) and network small-worldness. [157]
Eye dynamics	PD	SD/μ during tests and outside them, pupil dilatation [255]
	Fixations	Median duration, mean re-fixation depth, total duration [258], total n°of fixations, total time of fixations, novelty preference [255, 258]
	Saccades	Orientation [255]
SPECT	Histogram	μ, σ^2, entropy [222]
	Co-ocurrence matrix	Angular second moment, contrast, inverse difference moment, entropy, ρ_{xy} [222]
	Voxels of interest	Eigenbrains (PCA) [210, 212], normalized MSE
MEG	Temporal signals	SampEn, Lempel-Ziv Complexity (LZC) [245, 260], ApEn, MSFD[5] [260], HFD[6] [260, 261], Cross-ApEn [244]
	Spectrum	Shannon Spectral Entropy [260]

[5]Maragos and Sun's fractal dimension
[6]Higuchi's Fractal Dimension

Signal	Domain	Features
Speech	Hesitation and Puzzlement	Question rate, confusion rate, no answer count, rate of pauses in utterances, filler sounds [262]
	Words	Verb, noun, pronoun, adverb, adjective, particle, and conjunction rates, unintelligible word rate [262]
	Complexity	Phonemes per word, words per recording, standarized word entropy, phone entropy [262]
	Fluency	Length of voice segment, length of pause, short time energy and spectral centroid, average of the voiced and unvoiced segments, % voiced/voiceless and spontaneous speech evolution along the time [127, 263], maximum and minimum voiced and unvoiced segments [127]
	Emotional temperature	Average, SD, max and min of pitch and intensity, mean and SD of the period, and RMS, shimmer, local jitter, NHR[7], HNR[8] and ρ_{xx}, fraction of locally unvoiced frames, degree of voice breaks [127]
	HFD	Max, min, σ^2, SD [127]
EEG	δ, θ, α, β and γ bands' spectrum	Power densities [229, 264], total spectral power [264], specific spectral power ratios (see [264]), coherence between several combinations of pairs of electrodes [231], peak α band [264], spectral peaks of biauricular references, spectral peaks of bipolar references [231], median frequency, spectral entropy [264]
	Temporal signals of ROIs	Connection weights derived from IFAST methodology [230], Hjorth parameters, SampEn, LZC [264]
	First derivative	Total spectral power, peak α band frequency, median frequency, spectral entropy, SampEn, LZC [264]

2.3.3 Behavioural responses

Methods to measure the behavioural symtomps that can be evidenced in AD have been searched, in order to track the cognitive impairment of the patients. Some of these methods rely on tests and specific tasks that might be tested from time to time, whereas other methods are much more recent and rely on unobtrusive and ubiquitous technologies. This section presents the behavioural signals that could be used for AD diagnosis, which are shown in Figure 2.5.

General behaviour assessment tests

Behavioural changes suffered by AD patients might be measured by means of questionnaires or tests. These tools may help the patient himself and his relatives to take conscience of the real state of the disease. Tests like the Behavioural Pathology in Alzheimer's Disease Rating Scale (BEHAVE-AD) [265], the Brief Psychiatric Rating Scale (BSRS) [266], the Behaviour Rating Scale for Dementia of the Consortium to Establish a Registry for Alzheimer's Disease [267], Neuropsychiatric Inventory (NPI) [268] and the Dementia Behaviour Disturbance Scale [269] are questionnaires used to measure the behavioural anomalies that the AD patients can undergo [270].

[7] Noise-to-Harmonics Ratio
[8] Harmonics-to-Noise Ratio

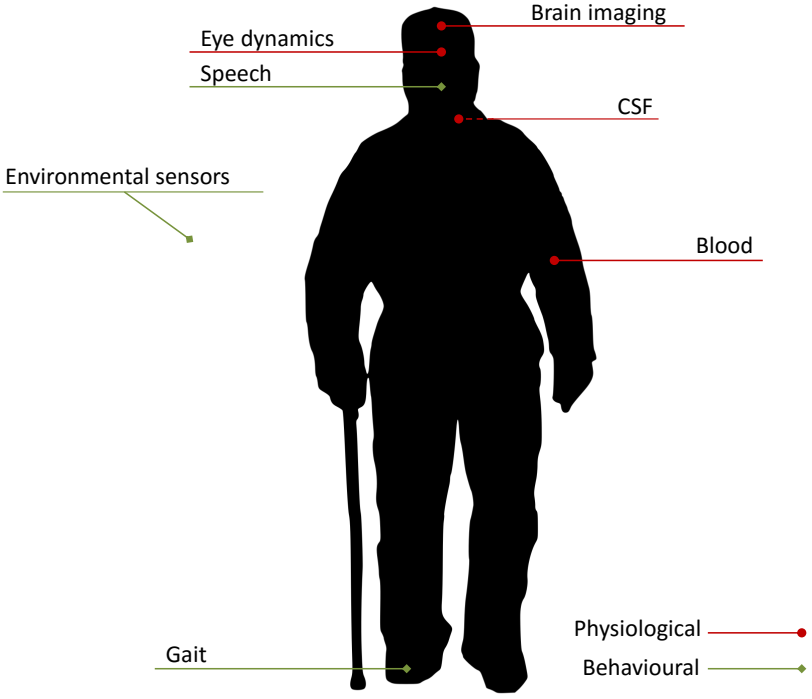


Figure 2.5: Physiological and behavioural measurements for AD diagnosis

Nevertheless, as these assessments are based on questionnaires, they are not realistic for a continuous monitoring and thus, their diagnosis might come too late.

Cognition analysis

There are some specific tests to measure the cognitive abilities of the people at risk of AD. Some examples are the MMSE [271] which is the most frequently used test for AD diagnosis, the Severe Cognitive Impairment Scale [272], the Alzheimer's Disease Assessment Scale - Cognitive (ADAS-Cog) [273] which focuses on attention, orientation, language, executive functioning and memory skills, the Neuropsychological Test Battery [274] which includes treatment effects' measurements, the Blessed Test which assesses memory, attention, concentration, and the ability to complete Activities of Daily Living (ADL)s and the Severe Impairment Battery [275] which alternatively focuses on measuring the unaffected cognitive functions [270].

These type of neuropsychological tests have shown to be effective in the assessment of AD. Nevertheless, they present some drawbacks. The most important one is that the assessment by means of these tests is lengthy and complicated [270]. Furthermore, they are not suitable for all the patients in all the stages of the dementia. Moreover, even if they can measure the state of progress of the dementia in a certain moment, it can be complicated to early detect AD because they may not show enough sensitivity or because as in many cases, it may be too late when the test is performed.

Activities of Daily Living (ADL) scales

Tests that aim at measuring the progress of dementia by analysing the abilities of the patients to perform typical ADLs with normality have been designed. These tests offer additional information to the one given by cognitive tests, because patients may have problems integrating visual, motor and cognitive skills, performing poorly in ADLs. Many times, the real state of the dementia can be better

assessed and the level of support needed can be much better understood seeing them in action and recording the level of cognitive support required to complete a certain task successfully [276]. ADL performance can be assessed both by means of questionnaires and by specifically designed tasks.

Some tests are based on the most basic activities (*i.e.* ADL), like feeding, walking or dressing, while others measure the abilities for more complex tasks, called instrumental activities (IADL) [270]. Examples of IADL are cooking, tasks which involve the use of money, and so on. Katz Index of ADL [277] and the ADCS-ADL19 [278] tests are examples of the former group while the ADCS-ADL23 [278] is suited for the IADL activities.

Among the specific tasks that allow to evaluate the abilities of the patient in vivo, the most well known is probably the Kitchen Task Assessment [276]. It is a functional measure that aims to evaluate the processing skills of initiation, organization, inclusion of all steps, sequencing, safety and judgement, and completion of a cooking task to measure the cognitive aspects of performance by means of behaviour.

Smart Homes

A smart home is a regular home that has been augmented with various types of sensors and actuators [279], being its main objective to overcome the cognitive disorders of people to enhance their autonomy [280].

Smart devices and environments allow to capture the actions of the residents while actuators can serve for automation or for providing comfort, making tasks easier or finishing the tasks that have not been accomplished by the patients, for example, for security reasons (turn off the oven after a certain time). Prompts or suggestions can also be made to recall to the patients how to continue an interrupted task and to provide them punctual assistance when needed [281]. All these actions should be carried out in a non-intrusive [282] and transparent way, respecting the privacy of the patients, to make it easier for adults to accept this technology in their daily life. Hence, monitoring systems such as video cameras are not desirable and the selected system should not interfere at all with the normal activities of the patients.

Even if specialized institutions where caregivers are available 24 hours a day exist, both seniors with dementia and their family members normally prefer the patients to be at home as long as possible [281]. Governments also prefer this option due to economical [283] and social reasons. Because of these reasons, Smart Homes and Ambient Intelligence (AmI) technology are being increasingly used in order to give assistance to elderly who suffer from dementia or cognitive impairment, to help them accomplish their ADL successfully and to reduce workload to the caregivers. For this purpose, it is necessary to predict these people's actions, and therefore, to learn their frequent behaviour patterns [3]. Learning these patterns can also be useful to detect abnormal behaviours [282] and to ease AD diagnosis.

Smart Home projects aimed at assisted living both for demented and non-demented elderly currently exist [279] in Europe (Grenoble Health Smart Home [284], Dem@care project [285]) and beyond (CASAS [286], DOMUS [287], MavHome [288]).

Smart Homes have been considered as a possibility for MCI detection by a few researchers. Some biomarkers have been found, indicating that this technology could be successfully used in early AD detection. Suzuki et al. [289] placed infra-red sensors in a smart home for monitoring ADL with an emphasis on people's sleep patterns and concluded that MCI patients went out of home with less frequency and had a shorter sleep time. Wadley et al. [290] measured the performance of healthy people and MCI patients carrying out ADLs like using the telephone, locating nutrition information on food labels, dealing with money, grocery shopping or medication managing and have observed that it took significantly longer for MCI patients to complete the tasks. Hayes et al. [291] also measured healthy people's and MCI patients' behaviour patterns in smart homes, including walking speed and measures of daily activities. Several markers were found in these case: walking times' variation of a

week showed to be twice as high in the MCI group compared to the healthy group, the time spent out of home was less for the MCI group than for the healthy group and the day-to-day pattern of activity of MCI subjects was more variable than for healthy subjects. Furthermore, MCI subjects had longer walking activities in the evening while this was not true for healthy subjects. Galambos *et al.* [4] discovered associations between overall in-home activity and outing patterns with both dementia and depression, which is also known to be a common AD symptom. GDS, MMSE and Short Form Health Survey-12 scales were used to determine subjects' state. Dawadi *et al.* found that the overall cognitive and mobility skills of older adults could be predicted by unobtrusively collected in-home behavioural data [292]. For that purpose, they introduced an algorithm called Clinical Assessment using Activity Behaviour (CAAB) and tested its validity for global cognition measured by the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) and mobility measured by the Timed Up and Go (TUG) scores' prediction using time-series statistics of several activities of daily living as predictors.

Moreover, previous research has demonstrated that longitudinal monitoring of smart home-based behavioural data can also be useful to assess older adults' overall health status. Petersen *et al.* [293] found emotional states in terms of mood and loneliness to be correlated to outing patterns, whereas they also verified the possibility of predicting other overall health predictors such as physical activity from these data. Loneliness of older adults has also been predicted by analysing behavioural data by Austin *et al.* [294].

Gait monitoring

As seen in the precedent section, smart homes can be used, among other things, to monitor walking activity of the demented elderly. Nevertheless, parameters such as walking speed may not be accurate enough to predict dementia. Gait monitoring takes into account the manner of walking of the person, where much more parameters apart from speed can be analysed. It has been recently found out that cognitive functioning and gait are closely related, so gait should not be longer considered a simple motor activity that is independent from cognition but as a complex cognitive task [295]. This hypothesis has been reinforced by dual-task tests [296]. This relationship is achieving more and more importance, and scientists are recently focusing on gait analysis for early AD diagnosis.

Gait can be monitored using an electronic walkway or force platforms placed on the floor, using cameras and image processing algorithms or by means of wearable sensors like force sensors, accelerometers, gyroscopes, extensometers, inclinometers, goniometers, active markers or electromyography [297].

Changes in gait behaviour have been reported in AD. Decreased velocity and step length, static and dynamic postural instability, and hesitation in starting and in turning and a widened base have already be found. Nonetheless, all these symptoms are part of a cautious walking, and they can also be found in normal ageing elderly. Scherder et al. [298] reported that AD patients differ from the healthy elderly in that they may show gait apraxia/ataxia, shuffling gait, lymbic discoordination, bradykinesia and rigidity. Increased support time has also been found [298, 299]. More recently, stride-to-stride variability has been reported to be an even more specific biomarker [300, 301, 302].

There are not many references in the literature affirming gait disturbances at the early stages of AD. Some report that these are inexistent and that gait is not useful for AD prediction [303], while others report interesting results that could be used for AD prediction. Camicioli et al. [304] affirmed that subjects developing cognitive decline walked more slowly than healthy people, and that they presented limbic coordination impairment. Scherder et al. [298] affirmed this information and added that rigidity is already present in the first stages of the dementia. Further research is needed to verify the predictability of gait disturbances for AD.

Table 2.4: AD biomarkers of the literature

Signal	Biomarkers
CSF	Aβ42 $\downarrow$, Aβ40, Aβ42/Aβ40 $\downarrow$, total tau $\uparrow$, p-tau231 $\uparrow$, p-tau181 $\uparrow$, p-tau199 $\uparrow$, [p-tau231, p-tau181, p-tau199]/Aβ42 $\uparrow$, Isoprostanes $\uparrow$, α1- antichymotrypsin, Interleukin-6, markers of inflammation, amount of CSF in the hippocampus $\uparrow$
Blood	Aβ42, Aβ40, Aβ42/Aβ40, α1- antichymotrypsin, various markers of inflammation, Isprostanes and Interleukin-6
sMRI	MTL's atrophy, vascular damage
fMRI	Activity in the MTL $\downarrow$, prefrontal cortex $\uparrow$, capacity of deactivation in posteromedial cortex $\downarrow$
MRSI	NAA $\downarrow$, NAA/Cr$\downarrow$, mI/Cr$\uparrow$, Cho/Cr$\uparrow$, Glu $\downarrow$, NAA/mI $\downarrow$
TCD	Carotid intima-media thickness $\uparrow$, total CBF $\downarrow$, cerebrovascular reserve capacity $\downarrow$, mean flow velocity$\downarrow$, pulsatility index $\uparrow$, cerebral microenbolization $\uparrow$
DTI	MD$\uparrow$, FA$\downarrow$
PET	Aβ plaques $\uparrow$, glucose consumption $\uparrow$, anomalies in neurotransmitter systems, inflammation in the central nervous system $\uparrow$
SPECT	CBF $\downarrow$,CSF $\uparrow$, perfusion $\downarrow$
MEG/EEG	Delta and theta activity $\uparrow$, alpha activity $\downarrow$, complexity $\downarrow$, synchrony$\downarrow$
Speech	Word recall/finding difficulties, repetitions, reading and writing skills $\downarrow$, problems following a conversation, non-verbal communication skills $\downarrow$
Eye dynamics	Prosaccades' and antisaccades' latency $\uparrow$, velocity $\downarrow$ and accuracy $\downarrow$, n°of incorrect saccades $\uparrow$, n°of corrections $\downarrow$, obliquity$\uparrow$, frequency$\uparrow$ and amplitude$\uparrow$ of microsaccades and saccadic intrusions, gaze-fixations' stability $\downarrow$, anomalies in smooth pursuit function, pupillary responses' latency$\uparrow$, amplitude$\downarrow$, velocity$\downarrow$ and acceleration$\downarrow$
Gait	Velocity $\downarrow$,step length $\downarrow$, support time$\uparrow$, postural stability $\downarrow$, hesitation $\uparrow$, apraxia/ataxia, shuffling gait, lymbic coordination $\downarrow$, bradykinesia, rigidity $\uparrow$, stride-to-stride variability$\uparrow$

Speech

The speaking and conversational skills of the AD patients deteriorate from the early stages of the disease [127]. They are likely to lose vocabulary, make big pauses while they are speaking or just stop abruptly because they are not able to continue the conversation. Thereby, speech recording aims at detecting these difficulties in speaking from the very beginning of the symptoms to facilitate an early diagnosis of AD. More specifically, their objective is to detect aphasia, which is the inability to communicate effectively [305]. Speech can be recorded continuously and in a non-invasive way, and can be analysed automatically with speech recognition and signal processing techniques.

Language and communication disturbances suffered by AD patients include [306] word recall and word-finding difficulties [307, 308, 309, 310], repetitions during speech [307, 308, 309, 311], loss of both reading and writing skills [310], problems to follow a conversation due to deterioration in concentration and comprehension skills [309, 310] and decline in non-verbal communication skills [312]. These problems are present from the very early stages of the disease, and progress and worsen at the same time that the cognitive decline [306, 313].

Due to the ease with which voice recordings can be obtained, in the recent years, speech features have been used in CAD systems for MCI and AD diagnosis [127, 262, 263, 314, 315, 316, 317].

One of the drawbacks of speech analysis for AD recognition is that aphasias are not unique to AD, but they can be caused by other factors. Nevertheless, it is true that "AD may be one of the primary causes responsible for a high proportion of aphasic patients in the human population" [318], so it can be of great interest to continuously keep an eye on people's speech features automatically, to later verify or discard the presence of dementia. Furthermore, speech can be easily and non-invasively measured, nowadays, almost continuously, which can be a big advantage for continuous monitoring and early diagnosis systems. It remains to be seen if the accuracy reported above would be reachable in a general population, but it is clear that speech features can provide important clues for a diagnosis.

Strengths and Weaknesses of Behavioural Responses to Detect AD

The same way as it happens in stress detection, behavioural measurements for AD detection are much less in number than the physiological ones. Two kinds of measurements can be distinguished in this group: the ones that are done by means of questionnaires, or specific tasks' performance and the ones that rely on emerging technologies. Whereas the former present the same drawbacks as psychological questionnaires (not realistic for a continuous assessment, and consequently, diagnosing too late), the latter provide a basis for creating an unobtrusive and ubiquitous monitoring systems, without the need of expensive equipment. Nonetheless, ethics and privacy issues can arise as with all the systems that are continuously gathering data, so the necessary measures must be taken into account. Despite this, the great advantages that this kind of technologies offer should definitely be seized, and further research is required so as to arrive to a correct AD diagnosis using such a system.

Summary

- AD's symptoms are evidenced psychologically, physiologically and on patients' behaviour.
- Psychological symptoms are assessed by several tests.
- Physiological symptoms' measurement rely mostly on CSF biomarkers and brain imaging biomarkers. Whereas the first method is very intrusive and might take a long time, the latter are very costly and obtrusive.
- Behavioural symptoms have been measured by assessment tests for a long time, but nowadays ways of measuring them in a completely unobtrusive and transparent way are being developed.

A summary of the reviewed AD markers can be found in the Table 2.4.

2.4 Models for the diagnosis and prediction

This section aims at responding the following RQs:

How are the disorders under study modeled? How are the data of the reviewed measurements exploited?

Introduction

Other than reviewing information about the kind of signals and features that can be used for detecting disorders such as stress and AD, it is important to analyse how these signals and features are exploited and used to create CAD systems that help make a correct decision about the diagnosis.

For this purpose, it is necessary to follow a data analysis process. Data analysis refers to extracting or "mining" knowledge from large amounts of data [319], producing insights and understanding about their structure. The term "structure" means statistical patterns, predictive models and hidden relationships. Data analysis is also known as knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology, and data dredging. In other words, the objective of data analysis is to identify valid novel, potentially useful, and understandable correlations and patterns in existing data [320].

In the knowledge discovery process, additional steps apart from the data analysis itself should always be followed [321]. These include data preparation, data selection, data cleaning, and proper interpretation of the results of the data analysis process, in order to ensure that useful knowledge enabling to make good final decisions is derived from the data. Figure 2.6 shows this process.

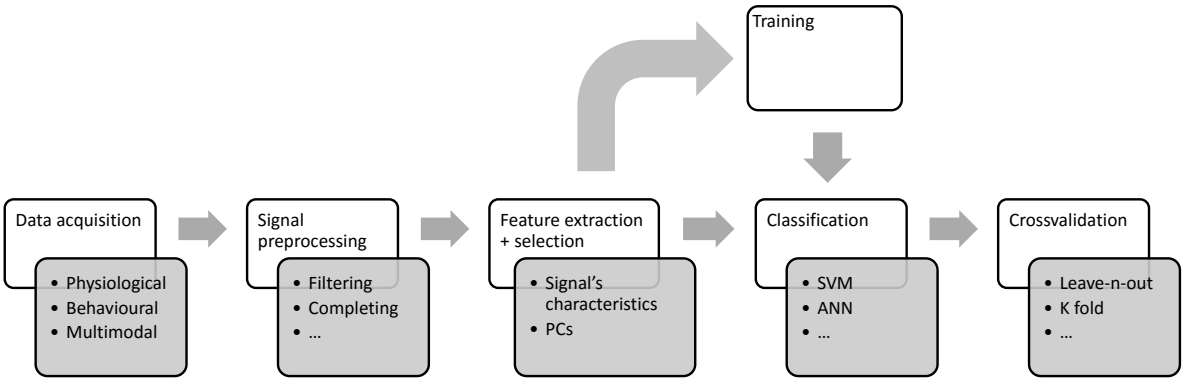


Figure 2.6: The common process of the stress detection and AD diagnosis researches

2.4.1 Data acquisition

The first step of such a system is the data acquisition, which has to be meticulously carried out in order to ensure its quality.

These data can come from both physiological measurements or behavioural measurements, and might be used alone or combined to make a multimodal analysis. These signals have been reviewed in the preceding sections, namely, in the Subsections 2.2.2 and 2.2.3 for stress detection from physiological and behavioural data, and in the Subsections 2.3.2 and 2.3.3 for AD detection. These data are frequently the result of cross-sectional experiments, where subjects with the selected disorders and normal controls take part so as to spot the differences between them.

For data to be of high quality, it must be "accurate, complete, relevant, timely, sufficiently detailed, appropriately represented, and must retain sufficient contextual information to support decision making" [322]. Nowadays' physiological monitoring devices, such as the BIOPAC System [323] or the FlexComp System [324], allow high quality data acquisition. Nonetheless, some varying factors can affect and thus, they have to be taken into account.

The incorrect placement of electrodes would derive in meaningless measurements, so in order to avoid ambiguities, standards for the correct measurement of physiological data have been defined and are internationally used. It is the case of the International 10-20 EEG System [325] or the standard 12-lead ECG. For sensors that do not have any standards defined, trials must be done to verify the best placement. It has been verified that different body placement of the sensors result in different signal patterns and classification accuracies [326]. Sensor placement is also crucial to the quality of the behavioural signals' recording in smart environments [327].

Sampling frequency of the data must also be adequate to the signal being collected, in order to establish a compromise between the amount of data to be treated and the quality obtained from them. Khusainov et al. [328] affirmed that for ADL monitoring a sampling frequency of 20 Hz is sufficient, while audio, speech and biomedical signals must be sampled with a higher frequency of up to 40KHz.

2.4.2 Signal preprocessing

Signal preprocessing step includes dealing with some issues, such as missing values in the data, noise or artifacts. In case of disease-related datasets, difficulties like inter-subject variability, imbalanced data distributions or temporal nature of the data are often present and must be taken into account in order to take full advantage of the information that they contain.

Signals are easily corrupted by instrumentation noise, random noise, electric and magnetic noise, etc., as well as by poor electrode-skin contact and body movements [328], resulting in noisy and artefact containing data. Signal processing techniques are needed to remove all these undesired effects from the signals. Noise can be filtered by means of several filters, like Kalman filters, Butterworth low-pass filters, Median filers, Wiener filters, Wavelet Decomposition, etc. The selection of the best filter in each case depends on the nature of the signal, the features to be extracted, and on the type of noise [328]. Power line interference can be removed by means of a notch filter.

For artefact removal, algorithms like least mean squares algorithm, regression analysis, independent component (ICA) and Principal Component Analysis (PCA) can be used, or pressure sensors or accelerometers can be used in order to detect movement artefacts [329] and reject the corresponding recordings.

Several normalization methods have been used to deal with inter-subject variability. These are necessary when the data of several subject is being used to create a general model. These methods include the use of range-corrected scores, the use of the proportion of maximal response and transforming the data into standard values using the z-scores [330].

In order to overcome the problems that missing values may pose, these must be replaced by some other values. Depending on the nature of the data, these missing values can be computed by means of interpolation, or can be replaced by the average values of the corresponding signal or dataset.

Other data transformations might also be required in some cases to correct the signs of non-normal distributions.

2.4.3 Feature extraction and selection

Feature selection consists on choosing the best set of features from the data. These subsets contain only those features that provide complementary information regarding the data classes, such that adding new features to the classification should not improve the overall result, and removing the

chosen ones should degrade it [331]. Feature selection helps in improving classification accuracy, in making the processing faster, in creating simpler classifiers which are more generalizable and in gaining insight into the models when they are analysed more closely.

In order to select the most relevant features from the initial data set, algorithms such as the Sequential Forward Selection (SFS) [6, 332] and the Sequential Backward Selection (SBS) [79] have been used. These techniques start from an empty or complete subset of features, adding or removing them one by one until any further addition or removal does not improve the result significantly. Other algorithms, such as the Correlation Based Feature Selection (CBFS), rank feature subsets according to the correlation with the class and with the other variables, so as to select the best uncorrelated subset of features. This algorithm has been used to select a subset of features in the context of mental workload monitoring in an office environment [37], as well as in the keystroke dynamics' feature selection with the purpose of detecting emotional states [85]. Alternative algorithms like the Random Subset Feature Selection exist, where "features are obtained by repetitively classifying the data with a k Nearest Neighbors (kNN) classifier while using randomly chosen subsets of all possible features and adjusting the relevance of each feature according to the classification performance of the subset that the feature participates in" [331]. This algorithm tests each feature in different contexts so that the result is less dependent on the order in where features are tested.

Genetic Algorithms choose a subset of features, starting from a random subset and adding and eliminating features iteratively using genetic operations like crossover and mutation. This technique works well in selecting the optimal set of features, but it normally takes a long time to find the minimal subset [65, 333].

An alternative approach to reduce the dimensionality of the data is to use a set of Principal Component (PC)s instead of directly using the features of the data. This way, PCA has been used to analyse PCs of ECG and sound signals [334], as well as of PET Scans [209]. Even if this type of analysis can work well in many cases, its performance depends on the nature of the data [81].

2.4.4 Classification

Data analysis uses both classical statistical procedures (such as logistic regression) and machine learning techniques which often overlap. Machine learning is the study of computational methods for improving performance by mechanizing the acquisition of knowledge from experience [335], and it has a major role in the process of data analysis. Statistical procedures are not less important, as they are needed to develop and asses models, construct rules and trees, and for validation and evaluation processes. Correlation analysis and regression analysis, fall under the umbrella of classical statistic procedures. The reader is referred to [336] for further information on this subject.

Both stress and AD detection are considered in most of the cases as a two-class classification problem, where stressed and not-stressed subjects and AD-patients and normal controls should be distinguished. For this purpose, several machine-learning algorithms have been used in the literature.

Decision Trees have been used to recognise stress from speech [24] and from physiological signals [52], as well as to recognise emotions from Keystroke dynamics [81]. These algorithms are very transparent and can be visualized graphically [337]. Random Forests (RF) were also used by Salmeron-Majadas et al. [338] to predict affective states from keyboard and mouse dynamics.

Naïve Bayes algorithms have been used with the same purposes [52, 62, 338]. Both kNN [37, 38] and Artificial Neural Networks (ANN) [65, 88, 339] are also some of the commonest algorithms for stress detection. Algorithms based on ANN have also been tested for AD and MCI diagnosis [229, 230]. Whereas the main characteristic of kNN and Naïve Bayes is their simplicity, ANNs are powerful tools able to model complex relationships between variables but which lack transparency in the sense that they do not provide information to the user about the contribution of each feature to the model.

K-means has been tested for three $2-$class stress classification problems from EDA and speech

data [24], but the achieved results suggest that the algorithm might be too simplistic for the problem aimed to be solved. Linear Discriminant Analysis (LDA), which is an easy method to use but with a limited ability to only model linear or quadratic relationships between variables, has also been used for AD recognition with MEG features [246].

Support Vector Machines (SVM) has been widely used both in stress and emotion recognition research [37, 50, 65, 78, 340] with satisfactory results in most cases. This algorithm follows a similar principle to LDA, but enables the use of kernels in order to model non-linear relationships. In the case of AD diagnosis, it is definitely the dominant method. It has been used with fMRI [158, 159] and DTI imaging alone [200], as well as with combination of both [201]. Nuclear imaging has also been analysed with SVMs, both PET [209, 210, 211, 212] and SPECT [210]. Prediction from EEGs [231] and eye dynamics [258] has also been tested with these algorithms, as well as from behavioural features derived from speech [262].

More sophisticated algorithms have started to be used in this area. For instance, De Santos Sierra et al. [72] used fuzzy logic to detect stress and no stress situations from EDA and HR signals, achieving very satisfying results (99.5% correct stress detection rate) and demonstrating the validity of the algorithm for stress recognition. ANFIS, a neuro-fuzzy algorithm, has also been proposed for AD recognition with MEG features [244]. These algorithms are notable because they do not require a training process [337].

Hidden Markov Models are another promising choice for these purposes [32], due to their ability to model and classify temporal sequences, such as stress and dementia. It has already been successfully used to detect stress-related facial patterns [89] and to calculate subjects' instantaneous stress levels within a scale of 1-7 [42] and of 1-5 [74] based this latter only on ECG related features.

Salmeron-Majadas et al. [338] used AdaBoost classification (*i.e.* a "strong" classifier built from a linear combination of "weak" classifiers) for emotion recognition from Keyboard and mouse interaction patterns, being the one that achieved the best results together with a Decision Tree.

Gaussian Mixture Models have been used by Kurniawan et al. [24] and Lu et al. [93] for stress detection purposes. Nevertheless, this algorithm may fail when the dimensionality of the data is big, and furthermore, it requires a prior knowledge about the existing number of mixture models in the data.

2.4.5 Cross-validation

Cross-validation is a model validation technique that aims at estimating the prediction accuracy of the input models. It consists of training the classifier's model with part of the data and their corresponding labels and evaluating the model in the remaining data, and comparing its estimated response to the real labels' values [341]. This process is repeated several times changing the training and testing datasets to be used, and the average performance of the classifier is computed. The commonest cross-validation methods in the State of the Art are on one hand, the "leave-one-out" [245, 342, 343], which consists on using each time a single data observation to test and training with n-1 observations, repeated n times. On the other hand, K-fold cross-validation can be found, normally with K=10 [7, 62, 77], which consists of separating the data into K subsets, and using one of the subsets for testing while the rest are used for training purposes. The process is repeated K times, once for each fold. The former is useful for small datasets and it provides unbiased error estimates but with high variance, whereas K-fold cross-validation with a small K offers much less variance in the error estimates, but the bias of the estimator will be large [344].

Summary

- The common process of the state of the art consists of: data acquisition in cross-sectional studies, preprocessing, feature extraction and selection (dimensionality reduction), classification and cross-validation.
- Care must be taken in order to ensure high quality data in the data collection process, and specific signal processing techniques should be applied to overcome common health-related datasets' problems such as noisy, incomplete, imbalanced, subject-dependent or temporal data.
- The most repeated algorithms for dimensionality reduction include both feature selection methods like the SFS, SBS or CBFS and alternative dimensionality reduction methods such as PCA.
- The final modeling and classification is done by a wide range of algorithms, starting form the very simple ones such as the Naïve Bayes, to the more complex ones like Fuzzy algorithms or Hidden Markov Models.
- Two cross-validation methods are the commonest: leave-one-out and 10-fold, which should be chosen by a compromise between bias and variance.

2.5 Early detection of disorders: The screening methodology

Current strategy for the early diagnosis of disorders, also called secondary prevention, is based on population screening [345]. Together with *opportunistic detection* or *case finding* strategy of primary care, in which a series of tests are performed according to age, sex and possible risk factors present in the person consulting for any reason, forms the passive methodology group for the detection of diseases. It is the only non-specific methodology for the early detection of disorders contemplated to date and it is still actively used [14]. Its purpose is the detection and treatment of the disease in very early stages by detecting potential disease indicators, in a large number of asymptomatic, but potentially at risk individuals [346].

Many diseases have a slow clinical evolution during which, although the disease already exists, it does not produce any manifestation (symptoms) that makes the person suffering from it suspect its presence. This phase of the disease is said to be asymptomatic. However, it is sometimes possible to apply certain diagnostic techniques to demonstrate the presence of the disease. Some diseases are congenital, that is, people are born with them, although they may take several years to manifest, even when the patient is already adult (*e.g.* some heart disease). Other diseases are acquired, they arise after birth, but are slow to debut (*e.g.* cancer) although they evolve over months or years. Some diagnostic procedures (*e.g.* mammography in the case of breast cancer) can identify these lesions when the carrier is not yet aware of them.

The activities linked to the screening method will depend on the nature of the disease. They can be activities of diagnostic anticipation or early detection when an effective treatment for the disease exists, while they will be of postposition when attempting to delay the course of the injury because there is no possible cure in the stage where the disease is detected [347].

2.5.1 Benefits and drawbacks

The use of screening methodology for the early detection of disorders brings certain benefits [345]. These include the following:

- It can reduce mortality from certain diseases that can be detected in the early stages.
- The chances of success in treatment increase.
- The complications and sequels of the disease decrease.
- It can let us know the incidence of some diseases that were previously unknown, such as in congenital diseases. When pre-testing all newborns, you can know the actual number of these diseases.
- The complications and sequels of the treatment decrease.
- In many cases complete restoration of health is achieved.
- Healthcare costs are reduced.

Nonetheless, the use of screening method also implies some inconveniences, such as the following:

- Technical limitations: Sometimes early detection techniques are not completely accurate. Sometimes the result is positive but then it is checked that there is no disease (false positive). In the case of a positive result, additional tests must be carried out. In contrast, sometimes a false negative occurs and a sick patient is not diagnosed. Both the sensitivity and the specificity of all the diagnostic tests accepted in the clinic are very high and this inconvenience is well outweighed by the benefits achieved.

- In the specific case of cancer, tumors that may never have evolved may be diagnosed, *e.g.* many cases of prostate cancer. In order to reduce over-treatment as much as possible, other more exhaustive studies are currently being considered so as to assess whether or not this diagnosed cancer may evolve, and especially if an intervention is justified when all tests indicate that the cancer has no tendency to worsen.
- Side effects and complications due to treatments (they are rare) may arise.
- It can increase psychological and emotional burden during the period of time that lasts the emission of a diagnosis or while a false positive result is not discarded by a diagnostic test.
- False negatives (they are minimized) might give a false sense of security and even delay a real diagnostic.
- It involves an "unnecessary" cost for people who will never be diagnosed.

2.5.2 Indications

In order to justify the use of this methodology, and to make sure that its use may actually benefit the disadvantages it may entail, the disease to be detected has to meet certain requirements.

Wilson and Jungner of the World Health Organization (WHO) published back in 1968 a guide about the principles and practice guidelines of the screening method for the early detection of diseases [14], also known as the Wilson's criteria. In these guidelines, they specified the requirements to fulfill for a screening process to be justifiable and launched. These principles are still applicable today, but usually, an updated principle list presented by Frame and Carslon in 1975 [348] is used. Both criteria are summarized in Table 2.5 [349]:

Recently, Andermann et al. [350] reviewed the modifications presented since the introduction of Wilson's criteria, and proposed an additional list of principles, more in line with current reality, to be considered. These additional terms are listed in Table 2.6.

2.5.3 Test methods

Screening test methods should not be confused with diagnostic methods [346]. The main purpose of screening tests is to detect early disease or risk factors for disease in large numbers of apparently healthy individuals, while the purpose of a diagnostic test is to establish the presence (or absence) of disease as a basis for treatment decisions in symptomatic or screen positive individuals (confirmatory test). Therefore, a positive result in a screening test indicates suspicion of disease, not a confirmed case. A diagnostic test is required for a confirmed diagnosis.

The test methods used for screening depend on each specific disease. Screening instruments can vary from technological procedures (*e.g.* radiography or laboratory tests), simple clinical examinations (*e.g.* blood pressure) to a set of standardised questions (*e.g.* depression screening in primary care).

Nonetheless, all screening tests must have certain characteristics. First, they must always be simple and acceptable to patients and staff. They must also be cheap since large numbers of people will need to be screened to identify a small number of potential cases. The tests must ensure sensitivity (true positives) in order not to miss any potential disease case, hence, the positive result is not a confirmed case. In contrast to screening test-methods, diagnostic tests have high specificity (true negative rate) and might be invasive or expensive, but yet justifiable as they are necessary for a confirmed diagnosis.

Screening has been used for the early detection of a wide variety of disorders, like, cervical [351], breast [352] or colorectal cancer [353], tuberculosis [354], depression [355], social anxiety disorder, dental caries [356], diabetic retinopathy [357], or abdominal aortic aneurysm [356]. Test-methods

Table 2.5: Wilson's criteria and Frame and Carlon's criteria for screening

Wilson and Jungner [14]	**Frame and Carslon [348]**
1.- The condition should be an important health problem.	1.- The disease represents a significant health problem with a marked effect on the quality and duration of life.
2.- There should be a treatment for the condition.	2.- There should be a treatment for the condition.
3.- Facilities for diagnosis and treatment should be available.	-
4.- There should be a latent stage of the disease.	3.- There must be a latent or early recognizable period in which detection and treatment reduce morbidity and / or morbidity.
5.- There should be a test or examination for the condition.	5.- A diagnostic test should be available at a reasonable cost.
6.- The test should be acceptable to the population.	-
7.- The natural history of the disease should be adequately understood.	3.- There must be a latent or early recognizable period in which detection and treatment reduce morbidity and / or morbidity.
8.- There should be an agreed policy on whom to treat.	-
9.- The total cost of finding a case should be economically balanced in relation to medical expenditure as a whole.	4.- The cost of early detection and treatment should be lower than the late equivalent.
10.- Case-finding should be a continuous process, not just a "once and for all" project.	-

Table 2.6: Additional list of principles summarizing the guidelines of the last 40 years

Andermann et al. [350]
1.- The screening program should respond to a recognized need.
2.- The objectives of screening should be defined at the outset.
3.- There should be a defined target population.
4.- There should be scientific evidence of screening program effectiveness.
5.- The program should integrate education, testing, clinical services and program management.
6.- There should be quality assurance, with mechanisms to minimize potential risks of screening.
7.- The program should ensure informed choice, confidentiality and respect for autonomy.
8.- The program should promote equity and access to screening for the entire target population.
9.- Program evaluation should be planned from the outset.
10.- The overall benefits of screening should outweigh the harm.

used for these screenings include liquid-based cytology, mammographies, colonoscopies, PPD (*i.e.* purified protein derivative) tests or inventories such as the Beck Depression or Social Phobia Inventory.

Nonetheless, all these screening test methods share some drawback. First, they are all obtrusive, even some are highly invasive. Second, not all of them are inexpensive as they should be. Moreover, they also require the patient to move to a medical center to get tested, limiting the performance of the tests to a certain periodicity.

Summary

- Current early detection of disorders is based on screening.
- So that the benefits gained outweigh the drawbacks of the screening method, the disease to detect must fulfill a certain criteria.
- Screening test-methods are not diagnostic methods: the former are simpler and cheaper than the latter, but only a diagnostic test can confirm the presence or absence of a disease.
- Most of the current screening tests are obtrusive, invasive, too costly and they require the patient to move to a medical center.

2.6 Open Research Areas

This section aims at highlighting the research gaps and open research areas on the early disorders' detection based on the literature reviewed in Sections 2.2, 2.3, 2.4 and 2.5.

2.6.1 Diagnosis of the disorders

Nowadays stress is diagnosed psychologically by means of self-report questionnaires or by being interviewed by a psychologist. The former is one of the most widely used ways to measure stress levels in humans and it is considered a reliable method. A more objective alternative to the use of questionnaires is the measurement of the salivary cortisol levels.

As in the case of stress detection, nowadays, AD diagnosis relies on cognitive assessment by means of tests such as the MMSE, on the use of CSF-based biomarkers and in the last years, on the use of some medical imaging modalities, namely, PET, computed tomography and sMRI for brain imaging. All these methods are considered as reliable biomarkers. However, they present some drawbacks that make impossible their use for early stress detection.

On one hand, they only offer information about the current health condition of the patient and not about the evolution of the disorder (nor about the stressors in the case of stress). Data can be sampled from time to time, but may not be suitable for detecting the subtle changes which could indicate an early stage of a major problem [111] neither realistic to carry out a continuous monitoring of the disease progress [29]. Actually, they are only measured when the affected themselves or the people around them realize or suspect about the severity of the situation, and this is too late in the vast majority of the cases. Consequently, the appreciation of both being over-stressed or suffering from cognitive impairment often comes too late, when health problems already manifest themselves [48].

On the other hand, psychological or cognitive assessment questionnaires can be too subjective and may lack sensitivity [358] whereas they require the full attention of the user. Regarding both cortisol and CSF measurements, they are intrusive, costly and slow methods of analysis [46]. Furthermore, all of these current tests, are "usually administered in a physician's office or a rehabilitation facility, causing inconvenience for the patient, using valuable healthcare resources, making frequent monitoring unrealistic" [111] and therefore, precluding an early diagnosis.

An early detection of these disorders would bring many benefits. First, if the disorder is detected at an early stage, symptoms might be in some cases reversible, and in other cases, at least, treatable. There are enough evidences that show that treatments are much more effective when they are applied in the early stages. In the case of stress, for example, the treatment might consist on putting aside work for a while or simply, reducing workloads, and avoiding this way, the long-term health consequences that may provoke. Moreover, enterprises would save much money. It should be noted that stress is the second most frequent work-related health problem in Europe [359] and that annual work-related stress costs of 20 billion euro have been reported by the enterprises of EU15[9] [360].

In the case of AD, the efficacy of medical treatments can highly increase, stopping or slowing down the cognitive decline. Furthermore, diagnosis can be more accurate in its early stages, when the patient is still able to answer to questions and to recall the order in where symptoms appeared. Patients' quality of life can also greatly improve, allowing them to make choices about their future (legal and financial decisions, how they want to be cared,...). It is estimated that AD will double its frequency in the next 20 years [233] and that 115.4 million people will suffer from it in 2050 [361], due to the increasing life expectancy. Moreover, while deaths attributed to other health problems such as heart disease have decreased in the last years, deaths attributed to AD between 2000 and

[9]EU-15 area countries are: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden and United Kingdom.

2010 have increased in 68 % [118]. This will bring serious consequences both in terms of medical services' cost and the increasing need for caregivers.

All these facts show the great importance of the early detection of these disorders.

As a conclusion, it is necessary to develop an ubiquitous monitoring system for these diseases so that even the possible decisive subtle changes can be detected. Such a system should work in a completely unobtrusive and transparent manner, *i.e.* embedded in every day's environment, in order to be practical the massive real use.

2.6.2 Multimodal analysis

Most of the recent research on both stress detection and AD diagnosis has been mainly focused on the search for biomarkers in physiological signals. A field much less present in the literature is that of behavioural markers. Furthermore, historically, behaviour assessment has been done by means of tests and scales whereas automatic behaviour assessment is a much more recent research subject.

This latest development has allowed to incorporate behavioural features in systems that were once created using only physiological features, leading to multimodal analyses. Examples include the works of Okada et al. [29] and Kocielnik et al. [48], who have made use of accelerometer data in combination of physiological signals for stress detection, or Kaklauskas et al. [25] who combined computer usage data with physiological signals. Nonetheless, these studies use both physiological and behavioural features as input for a supervised learning algorithm for the sole purpose of improving the detection accuracy, whereas the underlying relationships between the variables remain unanalysed.

Correlational studies of the literature between physiological and behavioural or psychological symptoms affirm that there exist relationships between symptoms of the different domains. Examples include the work of Tagai et al. [362], who used MRI and SPECT imaging modalities to relate anxiety of AD patients to the brain biomarkers or Poulin et al. [363] who also studied anxiety in relation to MRI markers. Delusions, apathy and agitation were also compared to markers on MRI images by others [364], as well as disinhibition and eating disorders with fluorodeoxyglucose-PET. In all of these cases, psychological and behavioural symptoms were assessed by means of tests such as the BEHAVE-AD or the Neuropsychiatric Inventory.

These type of studies have highly contributed in understanding the nature of these disorders. Nonetheless, as the emergence of ubiquitous computing and smart environments is very recent, there are not yet studies in where these both types of symptoms are related using automatic behaviour assessment methods. Therefore, such a study would be desired, not only to increase knowledge about the disorders and their effects, but also to progress towards an ubiquitous system for the early detection of these affections.

2.6.3 Temporal nature of the disorders

Current work related to both AD diagnosis and stress detection using physiological signals, are mostly cross-sectional studies. The problem is posed as a classic supervised classification problem, where samples of people belonging to different groups (stressed and no stressed for the first case, and control, MCI and AD groups for the second case) are taken at a given time, and after applying signal processing algorithms and feature extraction techniques, part of the data is used for training purposes for the selected classifier whereas the remaining data is used for the final classification and validation purposes. This way, the validity of the signals or image modalities, the signal processing techniques, the selected features and the chosen classifiers and other parameters used in the classification model are evaluated. This process has allowed for a long time to increase our understanding and knowledge levels about the physiological process behind these disorders, as well as to move towards an earlier and more accurate detection.

Both stress and AD are disorders that progress over time, so that their state in a certain point in time is not independent from the state in a previous point in time. Nonetheless, the vast majority of the research do not take their temporal/sequential nature into account and only a few exceptions that have used Hidden Markov Models have been found in the literature. Furthermore, latencies from triggers to the occurrence of symptoms are never taken into account: The correlation between multivariate signals is only analysed taking into account their values in paired moments, and not analysing how they evolve over time. Longitudinal studies allow to see these changes over the course of time, both to analyse how the situation under investigation affects an individual or to see the group differences that can be found over time, as well as to clarify the sequences in variables and deduce correlations and causalities.

Therefore, it is necessary to focus more on methods that exploit the behaviour of symptoms longitudinally, treating them as temporal or sequential signals and applying the correspondent analysis techniques, which could help discover heretofore unknown patterns.

2.6.4 A multidomain methodology

A final remarkable gap in the State of the Art is the lack of multidomain methodologies for the early detection of disorders based on AmI technologies. All the aforementioned studies focus on a specific disease and in the search of biomarkers and interesting patterns for their diagnosis. Nontheless, it has never been contemplated the possibility of taking advantage of the many analogies that can be found in several diseases so as to draw global conclusions. Stress and AD are only two examples of a collection of disorders that provoke behavioural changes, as are also Parkinson, depression, Attention Deficit Hyperactivity Disorder and many others. All of them could benefit from an ubiquitous monitoring system that learns behaviours of the patients and detects interesting shifts in their usual behaviour patterns. Nonetheless, traditional screening method ignores the fact that most of these disorders provoke behavioural shifts. We believe that a test-method based on the detection of unobtrusively detected physiological and behavioural shifts can provide many benefits to the current screening methodology. Hence, it would be desired to determine a multimodal methodology that defines the steps to be followed for the physiological and behavioural data collection and analysis, as well as for the correct interpretation of the discovered correlations and patterns. Among other things, the proposed approach can provide a cheap, ubiquitous and transparent method for screening, which can be continuously supervising, without requiring the patients to get out from their usual environment.

3

A method to predict Occupational Stress in Smart Office environments

Contents

This chapter exposes the analysis performed for the first case study: stress detection in smart office environments. First, in Section 3.1, it presents an introduction to the scenario. Then, in Section 3.2, a dataset composed by some of the unobtrusively collected measurements reviewed in the State of the Art (see Section 2.2) is presented and a proposal for the data processing and stress detection model building is done. Finally, in Section 3.3, the methodology of proposal is evaluated for the scenario under consideration and the resulting conclusions are exposed in Section 3.4.

3.1 Introduction

The pace of modern-day life, the competitiveness in the workplace, poor working conditions and the immense number of tasks with inaccessible deadlines that are assigned to workers are causing work-related stress to become increasingly frequent in our work environment.

The International Labour Organization defines stress as the harmful physical and emotional response caused by insufficient perceived resources and abilities of individuals to cope with the perceived demands, and is determined by work organization, work design and labour relations [365]. It is the second most frequent work-related health problem in Europe [359], presenting in 2005 a prevalence of 22% among working Europeans. In a recent opinion poll [366], 51% of the workers confessed that stress is common in their workplace and the 6th European Working Conditions Survey [367] exposed that 36% of European workers deal "(almost) all of the time" with high pressure to meet tight deadlines.

If timely action is not taken, occupational stress can provoke serious physical and mental problems on the worker [26], but also important economic losses in the companies. Musculoskeletal disorders, depression, anxiety, increased probability of infections [27], chronic fatigue syndrome, digestive problems, diabetes, osteoporosis, stomach ulcers and coronary heart disease [28, 42, 360] are only a few examples of occupational stress' long-term health consequences. Occupational stress can also result in increased absenteeism and presenteeism, reduced motivation, satisfaction and commitment, along with a greater rate of staff turnover and intention to quit, costing high amounts of money to the enterprises [368]. An estimate of €617 billion a year is what work-related depression costs to European enterprises, including costs of absenteeism and presenteeism (€272 billion), loss of productivity (€242 billion), healthcare costs (€63 billion) and social welfare costs in the form of disability benefit payments (€39 billion) [359]. An estimate of 50-60% of all lost working days in European enterprises are due to work-related stress and psychosocial risks [359].

In this context, methods to detect occupational stress in time so as to take the required measures and to avoid its negative health-related and economic consequences are necessary. Often, stress levels are evaluated by means of self-reported questionnaires, which are performed from time to time, and therefore, are not adequate to detect subtle changes that might end up in a more serious problem [369]. Usually, the diagnosis comes too late with these methods, when damage has been done. Moreover, self-reported questionnaires are subjective and rely on subjects' recall abilities and awareness of the situations, which is not guaranteed [47], leading sometimes to incorrect stress level measurements.

In recent years, technology to unobtrusively and ubiquitously monitor users' behaviour is being developed as Smart Environments [370]. Future work environments are supposed to be intelligent, adaptive, intuitive and interactive [371]. In this sense, a smart office has been defined as an environment that is able to adapt itself to the user's needs, release the users from routine tasks they should perform, change the environment to suit their preferences and access services available at each moment by customized interfaces [372]. In addition, we also see an opportunity based on its potential to avoid health-related problems for workers and improve their quality of life. As a great percentage of workers develop their tasks in an office environment, smart offices represent a useful infrastructure to continuously monitor workers' behaviour in a completely transparent way, gathering real work-life

data throughout the working day and therefore, to overcome the main disadvantages of the usual assessment methods. The collected data can provide a complete view of workers' behaviour in a real-world work environment, the efficiency and ecological validity of the resulting stress assessments and reducing stress detection delays.

Our goal in this chapter is to build and validate stress and mental workload prediction models based on unobtrusively collected physiological and behavioural data in a smart office environment. As all other disorders, stress progresses over time. Usually, in stress detection research, the temporal nature of the disorder is not taken into account, and only a snapshot of the symptoms is considered for prediction. In contrast, in this work we hypothesize that changes over time of these symptoms can predict the mental states of the subjects and the conditions they are undergoing.

To support this hypothesis, we propose the use of the Clinical Assessment using Activity Behaviour (CAAB) approach adapted to smart office environments to create stress prediction models [114]. This algorithm consists of the application of a sliding window to extract five different time-series statistics from physiological and behavioural data, describing the change and variability of these patterns. This allows the construction of models to predict self-assessed stress and workload levels from the change features instead of using the usual instantaneous feature values. Although it is out of the scope of this work, the computation of these behavioural and physiological change parameters not only provides a method to take the temporal nature of stress into account, but it is also a way to standardize data coming from different subjects, facilitating generalization of the models over a population group.

As a second goal of this work, we also determine the possibility of automatically detecting a workload condition change using these changes in physiological and behavioural data.

The CAAB algorithm has been validated in other scenarios and has been shown to be useful for cognitive state and everyday functioning assessment [114]. The validation of the approach for early stress detection would result in a system that could alert both workers and managers enabling to take timely action. Moreover, this would define the path to follow towards the final development and implementation of a global early detection system for disorders that provoke behavioural changes, among which stress is just an example.

Therefore, the Research Question (RQ)s we aim to address in this chapter are:

- Can we predict users' perceived stress and mental workload level from changes in their unobtrusively collected behavioural and physiological data?
- Which physiological or behavioural changes are the most informant about stress and mental workload levels?
- Can physiological and behavioural variability as monitored by ambient sensors be used to detect the conditions under which a participant is working, both from a predefined set of conditions and from reliably differently perceived conditions?
- Can these data be used to detect a change in workload settings? Can they also detect the direction of these changes? And a reliably perceived workload change?

The main contributions of this chapter are: 1) Use of the CAAB algorithm to evaluate the possibility of measuring self-assessed and standardized stress and mental workload from changes in unobtrusively collected real-life smart office data. 2) Analysis of the predictability of a wide variety of stress and mental workload assessment scores. 3) A feature selection-based analysis of the contribution of each type of behavioural and physiological change to the prediction of each of the self-assessment test scores. 4) Analysis of the predictability of an objective and reliable workload condition, change in these conditions and their directionality from unobtrusively collected data. 5) Testing of specific algorithms (*i.e.* SMOTEBoost [373] and RUSBoost [374]) to boost models' sensitivity for mental workload detection.

The remaining part of the chapter proceeds as follows. First, Section 3.2 explains the methods used for the data collection, preprocessing and model building process. Next, in Section 3.3, prediction models' results are presented. Finally, in Section 3.4, results are discussed and the conclusions are presented.

3.2 Proposed approach

3.2.1 Data collection in the Smart Office environment

The SWELL Knowledge Work Dataset for Stress and User Modeling Research (SWELL-KW) [375][1] collected in the 'Smart Reasoning for Well-being at Home and at Work' (SWELL) project was used for the current study. We decided to use this dataset for two reasons. First, it reflects real office workers' state performing their natural office work under real-life stressors, instead of being collected in an experiment where they are asked to perform artificial tasks or being submitted to non-common stressors. Second, the data gathered in the experiment can be easily collected with unobtrusive and easily accessible sensors that could be deployed in real office environments. Thus, this could facilitate the exploitation of the results obtained from this analysis.

SWELL-KW consists of multimodal data of 25 people who were submitted to a real work-setting experiment in a smart office environment. The participants were asked to perform common office work while they were being subjected to different workloads and different stress levels elicited by means of e-mail interruptions and time pressure. In addition to an initial relaxed state (R), three different conditions were simulated: a neutral condition where the subjects were asked to perform some 'normal work' without any stressors (N), a condition where they were forced to work under time pressure (T), and a third condition with e-mail interruptions as stressors (I). In the meanwhile, their physiological signals, computer use patterns, facial expressions and body posture were recorded by means of computer logging, video recordings, a Kinect 3D sensor and specific minimally-intrusive body sensors. Participants' perceived levels of stress and mental workload were assessed once per condition by a variety of self-reported questionnaires: Self Assessment Manikin (SAM) [377], Rating Scale Mental Effort (RSME) [378], NASA Task Load Index (NasaTLX) [379] and a stress level assessment by means of a visual analog scale. Table 3.1 summarizes the data collected in SWELL-KW.

3.2.2 Preprocessing

Minute-level feature extraction

Physiological and behavioural data of the 25 participants were collected continuously during the experiments, resulting in a raw data collection of 138 min (3 x 6 min R + 45 min N + 45 min I + 30 min T) for each one of the participants in the form of a computer log file, a FaceReader [380] log file, a Kinect SDK [381] joint coordinates file and a log registering the angles of the upper body and physiological data from Mobi [382]. Along with this raw dataset, SWELL-KW provided aggregated minute-length features as specified in Table 3.1 and whose extraction is explained in detail in the literature [375]. In this study, we made use of these minute-level features, but other time-window lengths for data aggregation could also be considered.

[1] Available online at [376].

Table 3.1: Raw-level and minute-level data available in the SWELL-KW dataset.

Modality	**Source**	**Minute-length aggregated features**
Physiology	Body sensors (3 features)	Heart Rate (HR), Heart Rate Variability (HRV), Skin Conductance Level (SCL)
Behaviour	Personal Computer (16 features)	Mouse use patterns (all mouse events, left clicks, right clicks, double clicks, wheel scrolling, drag events, distance), keyboard use patterns (all key events, n°of letter types, n°of special keys, n°direction keys, n°error keys, n°shortcut keys, n°of spaces typed) and applications (n°of app. changes, n°of tabfocus changes)
	Facial expressions (8 features)	The degree of detection of the following emotions: neutrality, happiness, sadness, anger, surprise, scare, disgust and valence
	Head and facial movements (32 features)	Head orientation (3), mouth opening, eye opening(2), eyebrow raising (4), gaze direction (3) and amount of activation of several facial points (20)
	Body posture and movements (94 features)	Proximity to the computer, forward inclination, shoulders' state (2), relative skeletal angles' average values describing the participants' posture (43) and standard deviations describing movements (47)
Subjective/ Psychological	Self-reported tests (12 features)	SAM scores (Valence, arousal, dominance), stress, RSME score (mental effort), NasaTLX scores (mental demand, physical demand, temporal demand, effort, performance, frustration and global NasaTLX)

Behaviour statistics' computation

As mentioned previously, we had available a set of minute-length physiological and behavioural features for each participant, as well as the subjective levels of perceived stress and mental workload under each condition for each participant. From this minute-level dataset, we computed two different summarizing datasets with two different goals using two different configurations for the CAAB algorithm. First, a dataset summarizing the physiological and behavioural responses of each participant under each condition was computed, aimed at creating prediction models for the self-reported mental states, perceived stress and conditions. Second, a dataset summarizing the physiological and behavioural response of the participants every 5 minutes was computed. The goal of this second component was to create prediction models for objective and reliable condition changes from physiological and behavioural data.

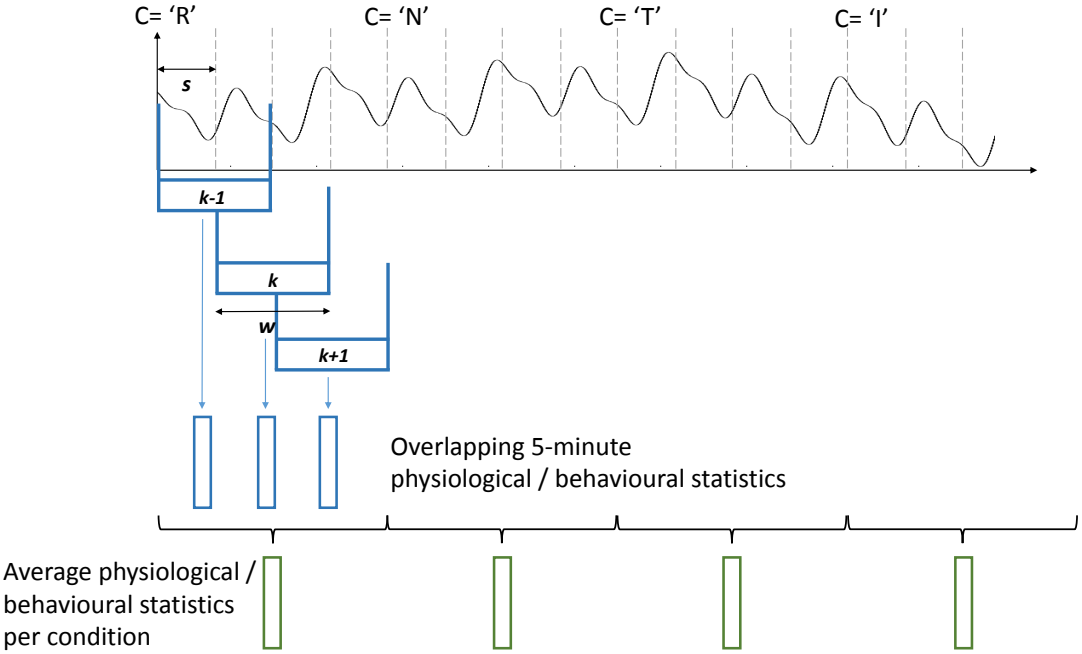
In order to extract the physiological and behavioural statistics for each participant under each condition, we implemented the CAAB [114] algorithm adapted to smart office data in Matlab. The minute-length physiological and behavioural data was processed using this algorithm as follows. First, each participant's minute-length physiological and behavioural features for each condition were extracted. Second, five summarizing time-series statistics were computed for each physiological and behavioural feature in this period using a sliding window of length (w) 5 minutes with a skip size (s) of 1 minute: variance, skewness, kurtosis, autocorrelation and change. While the first four are well-known time-series processing methods [383], the change statistic was first introduced by Dawadi et al. [114]. In brief, computation of the change feature is to apply a change detection algorithm between the two halves of the piece of time-series data that falls into the sliding window, so that we receive a '1' if a significant change is found between the two halves, and a '0' otherwise. For this purpose, we used an implementation of the Hotelling-T test [384] change algorithm available for Matlab[2]. In order to stabilize data variance and remove the effect of non-stationary (*e.g.* periodic) components, a log-transform followed by a linear detrending was applied to each physiological and behavioural variable falling inside the sliding window before the computation of summarizing statistics. Finally, the average of each time-series statistic for the length of the condition period was computed. The set of time-series statistics' averages was used for the final predictions. Note that the sliding window length (w=5) was selected empirically in a preliminary test, but other window sizes could also be considered. This process is highlighted in Figure 3.1a.

For the second dataset, only the last two steps differed from the previous process: after the application of the log-transform and linear detrending, the same five summarizing time-series statistics were computed, but this time, using a non-overlapping sliding window of five minutes' length (s=5, w=5). Condition-level averages were not computed this time, and the five minutes'-level dataset was considered as the final version for the condition change detection (see Figure 3.1b).

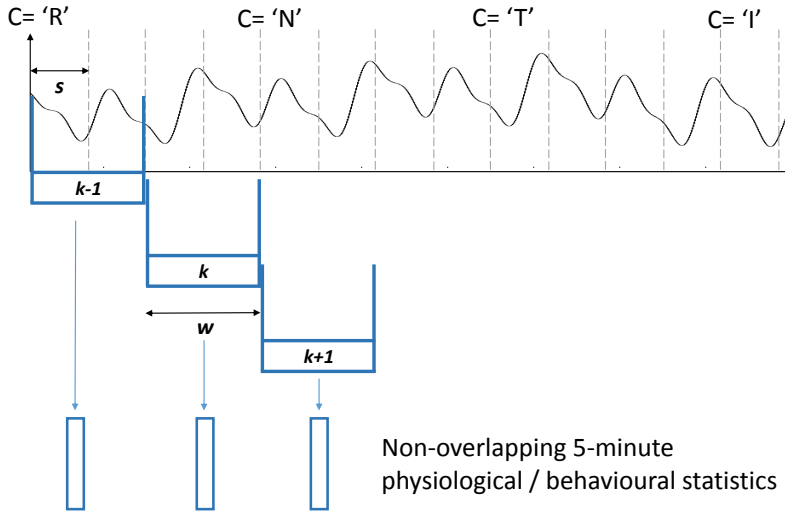
Thus, the resulting preprocessed datasets for further analysis were: 1) a collection 100 data instances of 780 (5 time-series statistics of 156 physiological and behavioural features) summary statistics modeling each one of the 25 participants who went through the four conditions of the SWELL experiment, and 2) a collection of 616 data instances of 780 summary statistics describing the physiological and behavioural output of the 25 participants for 5 minute non-overlapping intervals during the length of the whole experiment.

We made sure that none of the variables in any of the two datasets exceeded 30% of missing data, to remove it from the analysis if it was so. The remaining missing values were imputed by the mean value for each attribute using the 'ReplaceMissingValues' filter in Weka.

[2]Available online at https://github.com/brian-lau/multdist



(a) The first data preprocessing configuration.



(b) The second data preprocessing configuration.

Figure 3.1: The two configurations used for the computation of the two summarizing behavioural statistics datasets. The first configuration extracts summarizing statistics by applying a sliding-window of length w = 5 minutes with a skip-size s = 1 minute, and computes statistics' averages for each condition. The second configuration uses s = 5 minutes to compute non-overlapping statistics and use them as final predictors.

Stress and mental workload assessment scores' set up

The goal of this chapter is to create prediction models that map physiological and behavioural changes of data collected in a smart office to the subjective stress and mental workload ratings self-reported by the participants, as well as to objectively measure working conditions and condition changes. Our target variables are therefore defined as explained hereafter.

- **Self-reported stress and workload levels:**

  The self-reported valence, arousal and dominance levels measured by the SAM test, the stress level, the mental effort measured by the RSME questionnaire, and the mental demand, the physical demand, the temporal demand, the effort, the performance, the frustration and the global task load levels measured by the NasaTLX questionnaire were all collected once for each condition setting. As self-reported questionnaires might be very subject-dependent, we also computed the standardized version of the ratings by applying min-max normalization per subject to the questionnaire responses.

- **Simulated workload condition settings:**

  On the other hand, a label objectively indicating the condition under which the data were collected was used. This one takes the form of a four-class nominal variable, representing the four workload condition settings implemented during the experiments: R, N, T and I. Nonetheless, the effect provoked by each condition setting may depend on each subject, *i.e.* a participant might feel much more stressed under time pressure (T) than under a condition with frequent e-mail interruptions (I) while another one feels the opposite. To reduce this type of inter-subject variability, we computed the standardized versions of the condition settings. For this purpose, we ordered the conditions from the least to the most stressful for each participant (as measured by the 'stress' label) and assigned corresponding numbers: '0' for the least stressful and '3' for the most stressful one (see Figure 3.2).

- **Change in workload condition settings:**

  For the second configuration, a condition change variable was computed, indicating whether the subject was being submitted to a workload condition change in each one of the five-minute length data instances. Data were labeled with '1' if this was true and with a '0' otherwise. Finally, we also decided to make an attempt on detecting the directions of these condition changes, *i.e.*: for each 5-minute period, we computed whether the user was increasing (positive label), decreasing (negative label) or maintaining (neutral label) his/her self-reported perceived workload levels (as measured by the 'NasaTLX' label), and assigned '-1', '0' or '1' to each data instance.

- **Reliable change in perceived workload levels:**

  Despite the more ecologically valid experimental conditions that are used, the objectively measured condition might not necessarily be reflecting a significant workload change for all of the participants. To standardize the effect of each condition on the perceived task load for each participant, we computed the Reliable Change Index (RCI)es [385] for the NasaTLX scores. RCI informs whether a participant's perception (in this case, perceived workload levels) has experienced a significant change in an assessment score based on his/her own previous perception. RCI discards changes that might have appeared due to reasons other than an actual change in scores (such as measurement unreliability, repeated-testing or practice effects) by applying a threshold to the scores' differences. We looked for two different RCIs, one for each post-processed dataset. For the first case, we computed whether each participant was reporting a reliable change in the perceived task loads for each condition compared to the relaxed state (R). We assumed the NasaTLX score to be null for that initial condition. The RCI per condition and subject was computed as shown in Equation 3.1,

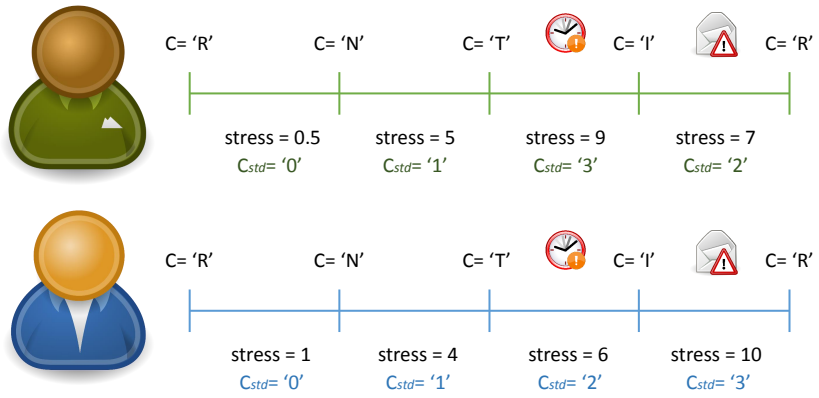


Figure 3.2: Two examples for the standardization of the 'condition' (C) label. The C variable shows the objective condition the participant is undergoing, whereas stress variable shows the perceived stress levels for that condition by the participant. C_{std} represents the standardized condition value calculated from the user's perceived stress level in each condition.

Table 3.2: Test-retest reliability (r) and standard deviation (SD) of the NasaTLX scores.

	r	SD
NasaTLX [386]	0.77	14.6

$$RCI_{baseline}(i) = \frac{Nasa_{TLX}(i) - Nasa_{TLX}(R)}{\sqrt{2SEm_{Nasa_{TLX}}}} \tag{3.1}$$

where $Nasa_{TLX}(i)$ and $Nasa_{TLX}(R)$ are the self-reported task-load level for the condition i and for the relaxed condition respectively, and $SEm_{Nasa_{TLX}}$ or Standard Error of Measurement represents the expected variation of the observed NasaTLX scores due to measurement error, being computed as shown in Equation 3.2,

$$SEm_{Nasa_{TLX}} = SD_{Nasa_{TLX}}\sqrt{1 - r_{Nasa_{TLX}}} \tag{3.2}$$

where $r_{NasaTLX}$ is the test-retest reliability measuring the consistency of the NasaTLX scores over time. Test-retest reliability parameters for the NasaTLX scores can be found in Table 3.2.

For the second case, we analysed whether the participants were undergoing a significant workload change in each 5-minute length period. For this purpose, we computed the RCI in self-reported NasaTLX scores at the beginning and at the end of each consecutive 5-minute time slot. This change was computed as shown in Equation 3.3,

$$RCI_{cons.}(j) = \frac{Nasa_{TLX}(j_{end}) - Nasa_{TLX}(j_{init})}{\sqrt{2SEm_{Nasa_{TLX}}}} \tag{3.3}$$

where $Nasa_{TLX}(j_{end})$ is the self-reported task-load index at the end of the 5-minute length period j and $Nasa_{TLX}(j_{init})$ is the self reported task-load index at the beginning of the 5-minute length period j.

3.2.3 Building stress and mental workload prediction models

The preprocessed datasets resulting from the previous steps were analysed using Weka [387].

Self-reported stress and workload levels: First, a regression analysis between the self-assessed stress and mental workload levels and smart office based physiological and behavioural data was performed. The models utilized linear and Radial Basis Function (RBF) kernel Support Vector Machines (SVM), Linear Regression (LR) and k Nearest Neighbors (kNN) algorithms. For this purpose, four models were built for each self-assessed score using all features extracted from the experiment data. The models were validated following a 10-fold CV approach and their correlation coefficients (r) and Mean Absolute Error (MAE) were compared. We searched for statistical predictability of the smart office data models comparing the results to a baseline model based on the ZeroR algorithm with a paired t-test. ZeroR is an algorithm aimed at creating prediction models based only on the distribution of the response variable and ignoring the data attributes [388]. It is commonly used as a basis of comparison for the other algorithms that have to overcome its performance to be considered useful. When it is being used for regression purposes, its error metric must be beaten. Adjusted p-values (*$p<0.01$, **$p<0.001$) were used to check for statistical significance in order to avoid Type 1 error rate due to the number of correlation analyses being run. Unless otherwise stated, the same validation approach based on 10-fold CV and t-test comparison to the corresponding ZeroR baseline classifier was used for all models in this work.

We then performed feature selection by analysing the predictive power of each type of feature for each self-assessment score. For that purpose, we built source-specific models based on only: (1) physiological features, (2) computer use patterns, (3) facial expressions, (4) head and facial movements, and (5) body posture and movements. A RBF SVM algorithm was used to build the models.

Next, all the previous steps were repeated to build prediction models of the standardized self-assessment scores. Models using all the collected data and source-specific models were created and validated.

The huge number of features coming from only five sources that are being used as attributes in this work, might result in highly collinear models which have the risk of not being optimal. To avoid this issue, we computed a Principal Component (PC) based reduced dataset explaining the 95% of the variability of the whole dataset. We built and evaluated the prediction models for this PC-reduced dataset.

Simulated workload condition settings: Regarding the detection of the objective mental-workload conditions from smart office data, we built and evaluated several classification algorithms. In this case, as subjects were submitted to four different workload conditions, we were facing a multi-class classification problem, where a random guess classifier would yield 25% accuracy. Naïve Bayes, linear SVM, AdaBoost and C4.5 tree algorithms were selected for this purpose. In addition, we tested a multi-class classifier trained following the one-against-all approach with a logistic classifier as weak classifier. As all the four conditions were considered of equal importance, the weighted versions of the area under the ROC curve ($wROC_{auc}$), the area under the precision-recall curve (wPR_{auc}), and Fscore were computed for comparison, as well as the overall accuracy (*Acc.*) of the models. We considered the classification models useful when they beat baseline models' accuracy and ROC_{auc} values. This process was then repeated for the standardized condition labels.

Change in workload condition settings: In the case of the second configuration, there were very few data instances representing a workload condition change available: only 17.8% of all data instances were of this type, resulting in highly imbalanced data. This is a very common problem in health-related machine learning tasks, where a disease is a rare event, and it is very difficult to collect

enough data instances representing the affected class. Usual machine learning algorithms tend to create biased models towards the majority class when being applied to imbalanced datasets, resulting in high prediction accuracies but, very low sensitivity. Notwithstanding, the main goal is often to detect the rare event, *i.e.* the presence of the disease or disorder.

To overcome this imbalanced data issue, alternative machine learning approaches must be used. In this work, in addition to some usual machine learning algorithms, two alternative algorithmic approaches called SMOTEBoost and RUSBoost were tested aiming at improving models' sensitivity. SMOTEBoost [373], is a method that combines boosting techniques with SMOTE [389] oversampling techniques. The objective of boosting is to create a "strong" classifier using a set of "weak" classifiers while SMOTE aims at reducing class imbalance by creating synthetic data instances to oversample the minority class. By combining these processes iteratively, SMOTEBoost often improves the sensitivity of the models without affecting the overall accuracy of the models.

In contrast, the second approach, uses the combination of boosting and RUS undersampling technique to reduce class imbalance [374]. RUS, randomly removes data instances from the majority class until a desired balance is achieved, resulting in training datasets of smaller size, and thus, greatly reducing complexity and training time of the models. Despite its simplicity, RUSBoost has demonstrated its effectiveness in previous works [390]. Therefore, we first built condition change prediction models using usual machine learning algorithms, namely, Naïve Bayes, linear SVM, AdaBoost and C4.5 tree. We evaluated the accuracy, ROC_{auc}, PR_{auc}, Fscore and sensitivity metrics of the models by means of a 10-fold CV approach. Next, we built models based on SMOTEBoost and RUSBoost algorithms using kNN, logistic algorithm, linear SVM and C4.5 tree as weak classifiers. This time, a 5-fold CV was used for validation purposes and the performance of the models was compared to a baseline algorithm by means of a McNemar's test.

Not all condition changes aimed to detect in the previous part imply the same risks: whereas a condition change from neutral or relaxed to stressful is an event of "high risk", the change in the opposite direction means an improvement in the workers' status. Both events are of interest, being the first one necessary to be detected in order to take preventive measures, and the second one, useful to track workers' status. Thus, we aimed at detecting the direction of the condition changes previously modeled. For that purpose, we built models based on Naïve Bayes, linear SVM, AdaBoost and C4.5 tree algorithms to solve the three-class classification problem (negative class: change to a more stressful condition, neutral class: no change, positive class: change to a less-stressful condition). We also added a logistic multiclass algorithm based-model which follows a one-vs-all approach.

Reliable change in perceived workload levels: Finally, we performed the detection analyses for the RCIs in perceived task-loads for each participant. First, we built and evaluated prediction models for the reliable NasaTLX score changes from baseline (*i.e.* classification of data instances representing relaxed states vs. significant workload states) using unobtrusively collected smart office data and Naïve Bayes, linear SVM, AdaBoost, C4.5 tree and Multilayer Perceptron (MLP) algorithms. We repeated the process for source-specific models. Second, we performed reliable perceived task-load change detection among consecutive 5 minute-length time periods using unobtrusively collected physiological and behavioural smart office data and the same algorithmic approaches as in the previous case.

3.3 Validation

This section presents the results obtained from the regression and classification models described in Subsection 3.2.3, which analyse the predictability of the self-reported and objective stress and workload condition levels from smart office data.

Table 3.3: Regression results for the self-reported test scores using all smart office features for 10-fold CV. (Statistically significant improvement (adjusted *p<0.01,**p<0.001) in comparison to a baseline algorithm.))

	Linear SVM		**RBF SVM**		**LR**		**kNN**	
	r	*MAE*	*r*	*MAE*	*r*	*MAE*	*r*	*MAE*
SAM								
Valence	**0.65****	**1.87***	**0.71****	**1.75****	**0.62****	**1.94****	0.13	3.22
Arousal	**0.53****	1.88	**0.56****	**1.91***	**0.55****	**1.86***	0.10	2.87
Dominance	**0.57****	**2.06***	**0.66****	**1.83****	**0.54****	2.15	0.24	3.09
Stress								
Stress	**0.31***	2.01	**0.35****	1.78	**0.33****	1.96	0.06	2.41
RSME								
MentalEffort	**0.64****	**2.13****	**0.68****	**2.06****	**0.62****	**2.16***	**0.30***	2.66
NasaTLX								
MentalDemand	**0.59****	**1.90****	**0.62****	**1.86****	**0.58****	**1.92****	**0.33***	3.02
PhysicalDemand	**0.47****	1.35	**0.49****	**1.23****	**0.47****	1.36	0.07	1.65
TemporalDemand	**0.53****	**2.47***	**0.59****	**2.37****	**0.54****	**2.45***	**0.26***	3.48
Effort	**0.74****	**1.72****	**0.75****	**1.70****	**0.75****	**1.70****	0.02	2.85
Performance	**0.37****	2.66	**0.50****	2.37	**0.32****	2.78	0.13	3.01
Frustration	**0.43****	1.89	**0.50****	**1.70***	**0.42****	1.87	**0.31***	2.61
NasaTLX	**0.71****	**16.27****	**0.71****	**16.41****	**0.72****	**16.12****	0.11	27.73

Self-reported stress and workload levels:

Table 3.3 shows the results of the regression analyses for the self-reported scores using all features available in the first-configuration dataset. Regarding algorithms, kNN is the one which is performing worst, and its MAE never beats the baseline classifier. On the contrary, RBF SVM is working the best in almost all cases, but the linear SVM and LR algorithms are also giving competitive results. Valence, mental effort, effort and global NasaTLX scores were found to be strongly correlated to the smart office data while dominance and mental demand were showing moderate to strong correlations. Arousal, physical demand, temporal demand and frustration were moderately correlated to the unobtrusively collected data, whereas correlation for performance label was weak to moderate and for stress only weak. In fact, for these last two scores, enough statistical significance was not found after adjusting the p-values, and therefore, they can not be considered to be predictable from the collected data.

Table 3.4 shows the prediction results for the self-assessed scores from the dataset of 82 PCs explaining the 95% of the variance of the whole dataset. Overall, correlation results are low, and none of the MAE values has shown enough statistical significance to be considered a useful model.

Table 3.5 shows the results for the regression analyses on the standardized self-reported scores. Generally speaking, the correlations obtained by these standardized scores are higher than the ones obtained using absolute values. Effort raised up to very strong correlation levels, while dominance raised to strong correlation levels. Valence, mental effort and NasaTLX were also found to be strongly correlated to the collected data. Moderate to strong correlation were found for arousal, mental demand, physical demand, temporal demand and performance, while frustration was only showing moderate correlations. Stress was the score showing the lowest correlations, but this time was found to correlate weak to moderately. Moreover, this time, all the scores showed statistically significant improvement in terms of prediction error compared to a baseline classifier, concluding that all scores can be predicted from smart office data after standardization.

Table 3.6 shows the results of the feature selection analysis for the self-reported scores. Valence and dominance were best predicted by body posture and movements followed by computer use patterns. Dominance was also predictable by facial and head movements, and arousal only showed

Table 3.4: Regression results for the self-reported test scores using the PC-reduced dataset for 10-fold CV. (Statistically significant improvement (adjusted *p<0.01,**p<0.001) in comparison to a baseline algorithm.))

	Linear SVM		**RBF SVM**		**LR**		**kNN**	
	r	*MAE*	*r*	*MAE*	*r*	*MAE*	*r*	*MAE*
SAM								
Valence	0.05	3.89	0.12	2.60	0.33	3.77	0.07	3.75
Arousal	0.01	4.06	0.00	2.40	0.13	4.30	0.00	3.17
Dominance	0.01	4.21	0.04	2.45	**0.36***	3.41	0.06	4.11
Stress								
Stress	0.11	3.60	0.11	1.93	0.07	3.63	0.11	2.56
RSME								
MentalEffort	0.04	4.75	0.06	2.75	0.15	4.86	0.11	3.52
NasaTLX								
MentalDemand	0.03	4.31	0.01	2.45	0.10	5.02	0.03	2.59
PhysicalDemand	0.16	3.00	0.16	1.40	0.09	4.39	0.09	1.96
TemporalDemand	0.22	3.89	0.03	3.03	0.16	5.17	0.10	3.47
Effort	0.24	3.57	**0.24***	2.51	**0.40***	3.53	0.06	3.05
Performance	0.11	5.67	0.08	2.80	0.16	4.25	0.04	4.31
Frustration	0.18	3.56	0.15	2.03	0.23	4.04	0.07	2.63
NasaTLX	0.21	31.03	0.07	22.35	0.19	57.12	0.05	25.49

Table 3.5: Regression results for the standardized self-reported test scores using all smart office features for 10-fold CV. (Statistically significant improvement (adjusted *p<0.01,**p<0.001) in comparison to a baseline algorithm.))

	Linear SVM		**RBF SVM**		**LR**		**kNN**	
	r	*MAE*	*r*	*MAE*	*r*	*MAE*	*r*	*MAE*
SAM								
Valence	**0.69****	**23.80****	**0.74****	**22.02****	**0.66****	**24.93****	0.19	37.34
Arousal	**0.56****	**26.66***	**0.64****	**25.52****	**0.56****	**26.66***	0.03	39.32
Dominance	**0.65****	**25.35****	**0.72****	**23.06****	**0.63****	**26.19****	0.16	37.04
Stress								
Stress	**0.38****	33.55	**0.51****	**30.80****	**0.39****	32.95	0.10	48.56
RSME								
MentalEffort	**0.66****	**25.08****	**0.73****	**23.27****	**0.65****	**25.21****	0.11	35.64
NasaTLX								
MentalDemand	**0.59****	**27.24****	**0.68****	**24.61****	**0.58****	**27.50****	0.11	38.81
PhysicalDemand	**0.49****	**30.50***	**0.61****	**27.17****	**0.49****	**30.92***	0.01	42.03
TemporalDemand	**0.55****	**28.48****	**0.63****	**26.83****	**0.56****	**28.47****	0.14	37.91
Effort	**0.78****	**20.68****	**0.81****	**20.24****	**0.78****	**20.88****	0.10	35.88
Performance	**0.53****	29.14	**0.62****	**26.92****	**0.50****	30.33	0.20	36.36
Frustration	**0.52****	**29.26**	**0.57****	**28.71****	**0.55****	**28.62****	0.05	45.71
NasaTLX	**0.67****	**23.84****	**0.70****	**23.23****	**0.67****	**23.87****	0.06	34.56

enough statistical significance for the computer use pattern-based models. Self reported stress was only found to be predictable by computer use patterns, beating the results obtained with the whole set of features. Mental effort measured by the RSME test was best predicted by the body posture and movement parameters, but was also statistically significant for the model based on computer use patters.

Next, regression results for the self-reported NasaTLX score and subscores are reviewed. The global score was best predicted by body posture and movements, followed by computer use patterns, as well as the mental and temporal demand. Performance was also predictable by computer use patterns and body posture and movements, in decreasing order. Physical demand was most correlated to facial and head movements, followed by body posture and movements and computer use patterns whereas effort was best predicted body posture and movements, computer use and facial and head movements. Finally, frustration was found to be only predictable by facial and head movements, followed by computer use patterns.

Table 3.7 shows the results of the feature selection analysis for the standardized scores. Overall, results improved, but follow the same trend. In this case, arousal became more predictable by means of body posture and movement-based models instead of computer use pattern-based models as in the previous case, which now occupies the second place. Standardized stress scores also showed statistically significant predictability using RBF SVM models based on only body posture and movements in addition to the one built using only computer use patterns. Standardized mental effort as measured by the RSME score was found to be predictable using only body posture and movement-, computer use pattern- and facial and head movement-based models, in decreasing order of performance. For the standardized NasaTLX questionnaire responses, models based on body posture and movements gained importance, as all subscores as well as the global task load index showed highest correlations with this feature type. The global score followed the same trend as the non-standardized scores, but correlations were slightly improved. Standardized mental demand was found to be only predictable by body posture and movements, and physical demand became most predictable with body posture and movements, followed by computer use patterns and facial and head movements. Standardized temporal demand and effort were most correlated to body posture and movements followed by computer use patterns as in the non-standardized case, but effort also showed enough statistical significance to be considered predictable by means of facial and head movements. Standardized performance scores became only predictable by the body posture and movement-based model whereas frustration gained enough statistical significance to be considered predictable by body posture and movements, computer use patterns and facial and head movements, in decreasing correlation order.

Simulated workload condition settings:

Table 3.8 shows the results for the objective and standardized workload condition detection models using all physiological and behavioural features and by feature type. Regarding the objective scores, overall, Naïve Bayes and AdaBoost based models were achieving the highest accuracies and the highest number of models with enough statistical significance. In fact, Naïve Bayes based models using all features, only computer use patters, only facial expressions and only body posture and movements were able to predict the workload condition. In the case of AdaBoost, physiological data-based models also showed statistical significance for prediction but facial expression-based model did not. Linear SVM based models were only useful using body posture and movement data, whereas C4.5 tree algorithm only resulted in statistically significant models using computer use patterns and the combination of all features. Against expectations, the logistic multi-class classifier based on one-vs-all approach was not overcoming the rest of the algorithms and only computer use pattern-based models showed enough statistical significance to accept workload condition predictability.

For the standardized scores, models show improved prediction accuracy compared to the non-

Table 3.6: Regression results for the absolute test scores by behavioural feature type for 10-fold CV and RBF SVM. (Statistically significant improvement (adjusted *p<0.01,**p<0.001) in comparison to a baseline algorithm.))

	Physiology		**Computer use**		**Facial expressions**		**Facial and head movements**		**Body posture and movements**	
	r	*MAE*	*r*	*MAE*	*r*	*MAE*	*r*	*MAE*	*r*	*MAE*
SAM										
Valence	0.21	2.55	**0.55****	**2.35***	0.10	2.64	**0.38****	2.48	**0.71****	**1.75****
Arousal	0.17	2.33	**0.55****	**2.03****	0.03	2.38	**0.42****	2.20	**0.47****	2.05
Dominance	0.29	**2.34***	**0.53****	**2.28***	0.00	2.42	**0.42****	**2.25***	**0.64****	**1.85****
Stress										
Stress	0.17	1.88	**0.50****	**1.63****	0.14	1.91	**0.33****	1.81	0.25	1.85
RSME										
MentalEffort	0.21	2.70	**0.58****	**2.46****	0.16	2.71	**0.45****	2.54	**0.62****	**2.22***
NasaTLX										
NasaTLX	**0.39****	21.45	**0.64****	**19.6****	0.03	22.58	**0.47****	21.15	**0.67****	**16.22****
MentalDemand	**0.35****	2.30	**0.52****	**2.16***	0.05	2.44	**0.35****	2.29	**0.59****	**1.84****
PhysicalDemand	0.19	1.40	**0.36****	**1.28***	0.03	1.44	**0.49****	**1.26***	**0.41****	**1.26***
TemporalDemand	**0.33****	2.93	**0.53****	**2.66****	0.04	3.03	**0.42****	2.80	**0.56****	**2.38****
Effort	**0.31***	2.46	**0.64****	**2.14****	0.08	2.52	**0.55****	**2.20****	**0.70****	**1.79****
Performance	0.17	2.75	**0.57****	**2.45****	0.08	2.83	**0.25***	2.72	**0.53****	**2.28***
Frustration	0.19	1.99	**0.46****	**1.75****	0.16	2.03	**0.48****	**1.76***	**0.40****	1.84

Table 3.7: Regression results for the standardized test scores by behavioural feature type for 10-fold CV and RBF SVM. (Statistically significant improvement (adjusted *p<0.01,**p<0.001) in comparison to a baseline algorithm.))

	Physiology		**Computer use**		**Facial expressions**		**Facial and head movements**		**Body posture and movements**	
	r	*MAE*	*r*	*MAE*	*r*	*MAE*	*r*	*MAE*	*r*	*MAE*
SAM										
Valence	0.17	33.48	**0.60****	**31.13***	0.06	34.06	**0.41****	32.08	**0.72****	**22.58****
Arousal	0.21	32.56	**0.57****	**31.08***	0.12	32.40	**0.40****	31.26	**0.62****	**25.39****
Dominance	0.22	32.77	**0.59****	**31.30***	0.04	33.96	**0.43****	**31.21***	**0.73****	**22.67****
Stress										
Stress	0.12	38.74	**0.56****	**34.07****	0.07	38.86	**0.25**	37.55	**0.46****	**31.86***
RSME										
MentalEffort	0.19	31.64	**0.63****	**30.11***	0.13	31.76	**0.46****	**29.71***	**0.70****	**24.06****
NasaTLX										
NasaTLX	**0.29***	30.80	**0.59****	**29.25***	0.13	32.06	**0.36***	31.40	**0.69****	**22.80****
MentalDemand	**0.27***	33.88	**0.54****	**32.30**	0.04	35.13	**0.42****	32.58	**0.67****	**24.11****
PhysicalDemand	0.02	38.06	**0.48****	**34.65***	0.05	37.58	**0.38****	**34.67***	**0.61****	**27.16****
TemporalDemand	0.26	35.11	**0.54****	**32.37***	0.06	35.89	**0.38****	34.00	**0.61****	**26.64****
Effort	**0.34***	31.99	**0.66****	**27.59****	0.04	33.66	**0.52****	**30.05****	**0.77****	**20.60****
Performance	0.14	35.30	**0.56****	32.88	0.08	36.07	**0.36****	33.39	**0.63****	**26.41****
Frustration	0.20	35.85	**0.51****	**32.14****	0.07	36.94	**0.45****	**33.00***	**0.58****	**28.09****

		Non-standardized				Standardized			
		Acc.	$wROC_{auc}$	wPR_{auc}	*wFscore*	*Acc.*	$wROC_{auc}$	wPR_{auc}	*wFscore*
Naïve Bayes	*All features*	**45.10****	**0.71****	**0.58****	**0.44****	**39.10****	**0.67****	**0.55****	**0.38****
	Physiology	27.10	0.54	**0.45****	**0.27****	30.70	0.59	**0.50****	**0.30****
	Computer use	**41.60****	**0.66****	**0.57****	**0.41****	**47.20****	**0.68****	**0.60****	**0.46****
	Facial expressions	**35.70***	**0.63***	**0.53****	**0.34****	**38.60****	**0.63***	**0.54****	**0.36****
	Face and head movements	26.00	0.55	**0.46****	**0.25****	23.40	0.53	**0.45****	**0.22****
	Body posture and movements	**51.90****	**0.77****	**0.63****	**0.48****	**39.50****	**0.68****	**0.54****	**0.35****
Linear SVM	*All features*	**34.90****	0.58	**0.45****	**0.34****	**40.10****	0.61	**0.47****	**0.39****
	Physiology	28.80	0.54	**0.38****	**0.25****	**34.50***	0.58	**0.40****	**0.31****
	Computer use	**41.00****	0.61	**0.47****	**0.39****	**44.30****	**0.62***	**0.47****	**0.42****
	Facial expressions	**33.20***	0.55	**0.38****	**0.30****	22.90	0.50	**0.33***	**0.20***
	Face and head movements	24.50	0.50	**0.38****	**0.23***	26.10	0.52	**0.39****	**0.25****
	Body posture and movements	**40.60****	**0.63****	**0.46****	**0.38****	**36.60****	**0.63****	**0.45****	**0.35****
AdaBoost	*All features*	**45.40****	**0.70****	**0.54****	**0.41****	**55.40****	**0.79****	**0.65****	**0.52****
	Physiology	**35.30****	**0.63****	**0.39****	**0.26****	**39.90****	**0.64****	**0.41****	**0.29****
	Computer use	**45.40****	**0.70****	**0.54****	**0.40****	**55.40****	**0.79****	**0.65****	**0.52****
	Facial expressions	22.50	0.47	0.27	0.12	**35.80****	**0.60****	**0.37****	**0.24****
	Face and head movements	**29.30***	0.57	**0.34****	**0.20****	**31.40***	0.58	**0.34***	**0.21****
	Body posture and movements	**36.30****	**0.62****	**0.39****	**0.26****	**35.40****	**0.63****	**0.39****	**0.25****
C4.5	*All features*	**40.90****	**0.67****	**0.55****	**0.39****	**46.60****	**0.70****	**0.59****	**0.45****
	Physiology	**37.60***	0.61	**0.48****	**0.35****	**35.30***	0.59	**0.48****	**0.32****
	Computer use	**38.70****	**0.61***	**0.52****	**0.38****	**46.90****	**0.68****	**0.57****	**0.45****
	Facial expressions	25.20	0.50	**0.37****	**0.23****	**34.80***	0.55	**0.43****	**0.32****
	Face and head movements	31.00	0.57	**0.45****	**0.30****	27.70	0.54	**0.43****	**0.26****
	Body posture and movements	30.70	0.53	**0.39****	**0.29****	27.60	0.53	**0.38****	**0.25****
Logistic Multi-class	*All features*	**34.50****	0.57	**0.48****	**0.28****	**37.80****	**0.61***	**0.52****	**0.31****
	Physiology	28.36	0.57	**0.47****	**0.27****	**37.50****	0.61	**0.53****	**0.35****
	Computer use	**40.60****	**0.65****	**0.59****	**0.33****	**40.30****	**0.68****	**0.61****	**0.33****
	Facial expressions	22.60	0.43	**0.38****	**0.21***	20.20	0.43	**0.38****	**0.19***
	Face and head movements	23.50	0.44	**0.38****	**0.21***	23.90	0.48	**0.41****	**0.22***
	Body posture and movements	28.20	0.57	**0.49****	**0.26****	**33.30***	**0.63***	**0.53****	**0.30****

Table 3.8: Classification results for the actual and standardized workload conditions by behavioural feature type for 10-fold CV. (Statistically significant improvement (adjusted *p<0.01,**p<0.001) in comparison to a baseline algorithm.))

standardized scores, and more statistical significances are found. In addition to those significances found for the non-standardized case, computer use pattern-based linear SVM, facial expression-based AdaBoost and logistic multiclass algorithm using all features and only body posture and movement features also showed prediction power. AdaBoost seems to be the best working algorithm for this case.

Change in workload condition settings:

Table 3.9 shows the results for the workload condition change detection using the usual machine-learning algorithms, whereas Table 3.10 shows the results for the SMOTEBoost and RUSBoost algorithms aimed at dealing with class imbalance. Usual algorithms gave better results than expected. Whereas some of the models showed too low sensitivities for the negative class, others where able to detect these events within an acceptable rate ($\geqslant$0.60). A computer use pattern-based Naïve Bayes model showed enough statistical significance to accept predictability of the objective workload changes, with a good sensitivity for the negative class. AdaBoost showed predictability of the target variable for all feature-, computer use pattern-, and body posture and movement-based models, in decreasing order of accuracy and sensitivity. C4.5 tree was the best in predicting the condition changes with a computer-use pattern-based model, followed by a model built using all the features. Linear SVM was not showing enough statistical significance in terms of accuracy to accept it was working better than a baseline model. Regarding SMOTEBoost and RUSBoost models, overall, we achieved higher sensitivity rates towards the negative class: some models even yielded 100% sensitivity. Nonetheless, only two of them showed enough statistical significance to accept predictability of the workload change, which were a SMOTEBoost based model using computer use patterns and a C4.5 tree as weak classifier, and a RUSBoost based model using the combination of all features and a C4.5 tree algorithm as weak classifier. Note that these significances were tested by means of a McNemar's test instead of the t-test as in the other models. However, these models were not highly improving the results obtained previously with the usual algorithms.

Table 3.9: Classification results for the workload condition change by behavioural feature type for 10-fold CV. (Statistically significant improvement (adjusted *p$<$0.01,**p$<$0.001) in comparison to a baseline algorithm.)

	Naïve Bayes					**Linear SVM**				
	Acc.	*ROC_{auc}*	*PR_{auc}*	*Fscore*	*Sens.*	*Acc.*	*ROC_{auc}*	*PR_{auc}*	*Fscore*	*Sens.*
All features	83.30	**0.80****	**0.47****	**0.57****	**0.62****	84.84	**0.74****	**0.42****	**0.57****	**0.57****
Physiology	81.62	0.56	**0.30****	**0.23****	**0.16****	82.20	0.50	0.18	0.00	0.00
Computer use	**88.02****	**0.89****	**0.73****	**0.70****	**0.78****	84.02	**0.63****	**0.33****	**0.40****	**0.31****
Facial expressions	78.24	**0.62****	**0.31****	**0.24****	**0.20****	82.17	0.50	0.18	0.00	0.00
Facial and head movements	73.05	**0.63****	**0.32****	**0.28****	**0.30****	80.91	**0.55***	0.23	**0.21****	**0.15****
Body posture and movements	80.58	**0.75****	**0.41****	**0.46****	**0.48****	79.50	**0.66****	**0.31****	**0.44****	**0.46****
	AdaBoost					**C4.5**				
	Acc.	*ROC_{auc}*	*PR_{auc}*	*Fscore*	*Sens.*	*Acc.*	*ROC_{auc}*	*PR_{auc}*	*Fscore*	*Sens.*
All features	**89.29****	**0.90****	**0.77****	**0.66****	**0.60****	**86.56***	**0.77****	**0.56****	**0.60****	**0.58****
Physiology	82.01	**0.64****	**0.32****	0.01	0.00	81.60	0.52	0.22	0.06	0.04
Computer use	**87.24****	**0.86****	**0.66****	**0.58****	**0.51****	**90.44****	**0.84****	**0.69****	**0.71****	**0.66****
Facial expressions	81.13	**0.63****	**0.31****	**0.15****	0.05	78.27	0.53	**0.24***	**0.15****	**0.12****
Facial and head movements	80.65	**0.65****	**0.32****	0.11	0.08	81.00	0.56	0.23	0.06	0.04
Body posture and movements	**86.53****	**0.85****	**0.64****	**0.55****	**0.48****	83.35	**0.69****	**0.45****	**0.52****	**0.50****

Table 3.11 shows the results for the task load change directionality detection. AdaBoost algorithm was performing worst, as all models based on this algorithm were biased towards the majority class. Some other models were showing statistically significant improvement in terms of accuracy compared to a baseline classifier, but were performing very poor in terms of F-score and/or sensitivity, making them useless for our purpose. Only a Naïve Bayes- and a linear SVM-based model built using the combination of all features showed statistical significance for all metrics, leading us to accept their

		SMOTEBoost					RUSBoost				
		Acc.	ROC_{auc}	PR_{auc}	*Fscore*	*Sens.*	*Acc.*	ROC_{auc}	PR_{auc}	*Fscore*	*Sens.*
kNN	*All features*	0.76	0.56	0.19	0.13	0.18	0.68	0.50	0.16	0.17	0.16
	Physiology	0.75	0.53	0.19	0.15	0.19	0.71	0.54	0.19	0.23	0.21
	Computer use	0.82	0.55	0.23	0.02	1.00	0.82	0.50	0.09	0.00	0.00
	Facial expressions	0.51	0.54	0.19	0.28	0.19	0.74	0.52	0.20	0.24	0.25
	Face and head movements	0.78	0.51	0.18	0.11	0.19	0.68	0.49	0.16	0.17	0.16
	Body movements	0.82	0.50	0.00	0.00	0.00	0.61	0.61	0.18	0.33	0.24
Logistic	*All features*	0.87	0.87	0.59	0.61	0.64	0.83	0.84	0.53	0.57	0.51
	Physiology	0.70	0.56	0.21	0.27	0.24	0.81	0.58	0.23	0.23	0.39
	Computer use	0.85	0.88	0.59	0.63	0.57	0.77	0.83	0.36	0.52	0.41
	Facial expressions	0.72	0.63	0.26	0.31	0.28	0.72	0.61	0.25	0.29	0.26
	Face and head movements	0.71	0.63	0.24	0.33	0.28	0.56	0.60	0.22	0.33	0.23
	Body movements	0.71	0.63	0.25	0.33	0.28	0.54	0.51	0.18	0.27	0.19
Linear SVM	*All features*	0.86	0.88	0.61	0.57	0.61	0.83	0.86	0.52	0.60	0.53
	Physiology	0.74	0.56	0.22	0.26	0.26	0.83	0.51	0.46	0.04	1.00
	Computer use	0.87	0.89	0.62	0.68	0.61	0.85	0.89	0.26	0.66	0.55
	Facial expressions	0.68	0.56	0.22	0.33	0.27	0.78	0.62	0.27	0.23	0.31
	Face and head movements	0.72	0.62	0.24	0.28	0.26	0.73	0.63	0.26	0.34	0.30
	Body movements	0.75	0.64	0.26	0.30	0.30	0.76	0.67	0.29	0.38	0.35
C4.5	*All features*	0.85	0.89	0.65	0.58	0.59	**0.86****	**0.88****	**0.60****	**0.64****	**0.60****
	Physiology	0.77	0.65	0.28	0.25	0.30	0.78	0.62	0.29	0.36	0.38
	Computer use	**0.87****	**0.91****	**0.74****	**0.66****	**0.60****	0.87	0.86	0.60	0.66	0.64
	Facial expressions	0.69	0.54	0.19	0.20	0.18	0.75	0.62	0.25	0.36	0.33
	Face and head movements	0.73	0.63	0.23	0.22	0.23	0.74	0.64	0.25	0.26	0.26
	Body movements	0.83	0.81	0.53	0.54	0.51	0.81	0.83	0.45	0.53	0.48

Table 3.10: Classification results for the workload condition change by behavioural feature type for 5-fold CV using SMOTEBoost and RUSBoost algorithms. (Statistically significant improvement (adjusted *p<0.01,**p<0.001) in comparison to a baseline algorithm.)

prediction power for the positive, negative and null task load changes. Nonetheless, these models were yet showing low sensitivity rates.

Reliable change in perceived workload levels:

Table 3.12 shows the results for the reliable perceived task load index change detection. The reliably different task-loads from baseline situation (relaxing vs. stressful) were found to be detectable using models based on all features, computer use patters and body posture and movements, whereas physiology-, facial and head movement- and facial expression-based models did not show enough statistical significance to accept the hypothesis. The highest accuracies were achieved by means of MLP, AdaBoost and Naïve Bayes-based models, and regarding feature selection, computer use patterns were found to be more useful than body posture and movement based models. A reliable change between consecutive 5-minute periods was harder to detect and fewer useful models were found. Naïve Bayes classifier was performing best by means of body posture and movement-based models, followed by computer use and facial and head movement based models. Models built using linear SVM, AdaBoost and C4.5 algorithms were not significantly improving the detection accuracy achieved by a baseline classifier. Regarding MLP based models, only a computer use pattern-based model was showing significantly increased performance compared to a baseline model. Nonetheless, it still showed a low sensitivity rate.

3.4 Discussions and conclusion

In this chapter, we analysed the possibility of predicting workers' stress and workload levels, as well as changes in these conditions, by means of time-series statistics computed from unobtrusively collected physiological and behavioural data in a smart office environment. The RQs in hands are of great interest to today's society where stress is becoming increasingly present and harmful, but are also pertinent to the current state of the art in Ambient Intelligence (AmI) and smart environments. Unobtrusive monitoring of peoples' behaviour and physiology is already possible, but we yet need to associate these patterns to the disorder of interest. Moreover, it is still necessary to clarify and limit the use of the proposed system to avoid ethical and privacy issues before is implementation [369]. Results show that the prediction of perceived stress and workload levels is possible using change and variability patterns of data collected unobtrusively from smart offices.

A regression analysis of the target scores from smart office data showed many statistically significant results, enforcing the hypothesis that this kind of collected data can actually predict the perceived stress and workload levels. The correlations found by this analysis vary from moderate to strong, depending on the nature of the objective label. NasaTLX scores, together with effort, mental effort and valence were the best-predicted scores, whereas self-reported stress and performance did not show enough statistical significance to be considered predictable. In case of stress prediction, this is not surprising, as this label was acquired by means of a single-question visual analog scale, which unlike NasaTLX, RSME or VAS questionnaires, is not a questionnaire whose reliability has been verified and might be too subjective to be well capturing the real perceived stress levels of the users. Nonetheless, the analyses on the standardized scores improved the previous results, even demonstrating predictability for the self-reported stress and performance levels. This reasserts the fact that there is some inter-subject variability present on every score used for the study, but also suggests that controlling for this variability by means of standardization methods, can make their prediction possible.

A reduced dataset using PC approach showed a highly decreased performance on the predictability of the models. This might be due to several reasons. On one hand, it suggests that actual feature values are much more correlated to the self-reported scores than the PCs representing this data.

On the other hand, it might also suggest that there is no much collinearity among the initial set of features. Nonetheless, the reason can also be an excessive standardization of the input data which might have provoked the loss of machine-learning algorithms' mapping ability to subject-specific response data. This can be verified by validating the previous models based on actual feature values following a Leave-One-Subject-Out Cross-Validation (LOSOCV), which is a well-known procedure in the field. LOSOCV consists of excluding one participant at each time from the model-training step, while using their data in the model-testing part. This process is repeated until all users' data is used both for training and for testing, and models' average performance is computed. This would allow to verify the usability of the current approach to detect stress and workload levels of new workers without the need of collecting their data. The literature shows that LOSOCV based validation usually gives much more moderate results [375]. Hence, the importance of building user-specific models, models based on data from a small group of people which is as similar as possible to the final user or to build general models that can benefit from users' feedback to adapt gradually to each of them.

Regarding feature selection analyses performed in the regression models, computer use patterns and body posture and movements are the most correlated type of behaviour, followed by head and facial movements. These results agree with previous research that report a relationship between perceived stress levels and computer-use patterns [5, 63, 86, 111], body posture [88] and head and facial movements [63]. In fact, models based on only physiological measurements and facial expressions were never significant by themselves, while literature affirms the predictability of stress levels both from facial expressions [89, 91, 391] and physiological signals [23, 27, 52, 65, 392]. This is an important finding, as physiological measurements based on Skin Conductance Level (SCL) and Electrocardiogram (ECG)s are the most widely used signals in stress detection [369]. These results suggest that behaviour might be much better in predicting stress under the circumstances of this case study. However, we must first understand the nature of the experiment used to collect the data of the current study and the steps taken to process it, to interpret the results consequently. The reason why physiological signals might not be showing high correlations as usual, can be that the time-series statistics extracted from them are not reflecting an increase or decrease in the signals but the amount of absolute change.

The directionality of the change in physiological signals might be very important as far as stress detection is concerned. For example, it is well known that stress provokes an increase in SCL signals, or a reduction on the Heart Rate Variability. Due to the data processing approach used herein, we might be missing this valuable information. Furthermore, the amount of imputed data in physiological signals was higher than in the behavioural statistics, which might have also blurred the correlations in this domain causing a significance loss. Moreover, results based on the computer use patterns must be interpreted carefully: *i.e.*, due to the nature of the experiment, where the participants were asked to perform a set of specific computer tasks under each condition and then evaluate the perceived stress and workload levels per condition too, results based on computer use patterns are much more likely to be correlated to the self-reported scores.

Unlike body posture and movements, facial expressions, head and facial movements and physiological signals, computer use patterns were not varying completely freely but were being conditioned by the tasks that had been assigned to the participants. It would be interesting to analyse whether the same patterns of behaviour are repeated in an experiment where other methods are used to induce stress in the users, or in a longitudinally collected dataset where no stress is being induced in the participant nor is being subjected to any special condition, but all their behaviour only depends on their daily work and hypothetically, their stress levels. Another solution would be to use alternative statistical analysis methods to control the variability on the behavioural data caused by the condition to which the participants are subjected and to quantify the part of behavioural variability that corresponds to the level of stress suffered. Moreover, the insufficient predictability of facial expressions for the self-reported stress and workload levels might not be due to the lack of correlation among the two but to the lack of reliability of the method used to estimate the facial expressions from video

	Naïve Bayes					Linear SVM				
	Acc.	ROC_{auc}	PR_{auc}	*Fscore*	*Sens.*	*Acc.*	ROC_{auc}	PR_{auc}	*Fscore*	*Sens.*
All features	**90.49****	**0.89****	**0.44****	**0.37****	**0.35****	**91.54****	**0.83****	**0.26****	**0.35****	**0.36****
Physiology	**93.53****	**0.93****	**0.47****	0.14	0.09	**93.19****	**0.94****	**0.38****	0.00	0.00
Computer use	79.77	**0.78****	**0.34****	**0.31****	**0.36****	82.43	0.49	0.08	0.01	0.01
Facial expressions	74.01	0.62	**0.17***	0.11	0.13	82.17	0.52	0.08	0.00	0.00
Face and head movements	69.11	0.56	**0.20***	**0.14***	**0.16***	81.18	0.54	0.13	0.13	0.10
Body posture and movements	77.08	0.57	**0.20***	**0.21***	**0.22***	76.97	0.54	0.10	0.10	0.11

	AdaBoost					C4.5				
	Acc.	ROC_{auc}	PR_{auc}	*Fscore*	*Sens.*	*Acc.*	ROC_{auc}	PR_{auc}	*Fscore*	*Sens.*
All features	**93.19****	**0.94****	**0.38****	0.00	0.00	**90.84****	**0.91****	**0.42****	**0.20***	0.19
Physiology	**93.19****	**0.94****	**0.38****	0.00	0.00	**92.92****	**0.94****	**0.38****	0.01	0.01
Computer use	82.38	**0.84****	**0.28****	0.00	0.00	**85.27***	**0.87****	**0.40****	0.13	0.10
Facial expressions	82.17	0.60	0.11	0.00	0.00	74.55	0.48	0.16	**0.15***	0.15
Face and head movements	82.17	0.58	0.09	0.00	0.00	80.84	0.50	0.09	0.00	0.00
Body posture and movements	81.56	**0.78****	**0.26****	0.00	0.00	80.78	0.52	**0.20***	**0.24****	**0.23***

	Multiclass logistic				
	Acc.	ROC_{auc}	PR_{auc}	*Fscore*	*Sens.*
All features	80.18	**0.74****	**0.36****	**0.32****	**0.46****
Physiology	**92.88****	**0.94****	**0.49****	0.11	0.08
Computer use	79.15	**0.77****	**0.28****	**0.21***	**0.21***
Facial expressions	80.78	**0.63***	**0.19****	0.05	0.03
Face and head movements	67.44	0.54	**0.20***	**0.14***	**0.22***
Body posture and movements	53.74	0.55	**0.12***	**0.13****	**0.28****

Table 3.11: Classification results for the positive, negative and null workload condition change by behavioural feature type for 10-fold CV. (Statistically significant improvement (adjusted *p<0.01,**p<0.001) in comparison to a baseline algorithm.)

		$RCI_{baseline}$					$RCI_{consecutive}$				
		Acc.	ROC_{auc}	PR_{auc}	*Fscore*	*Sens.*	*Acc.*	ROC_{auc}	PR_{auc}	*Fscore*	*Sens.*
Naïve Bayes	*All features*	**96.00****	**1.00****	**1.00****	**0.88****	**0.84****	85.77	**0.79****	**0.35****	**0.45****	**0.55****
	Physiology	80.30	**0.74****	**0.65****	**0.42****	**0.35****	87.49	0.55	0.20*	**0.14***	0.11
	Computer use	**95.00****	**1.00****	**1.00****	**0.85****	**0.79****	**83.55****	**0.81****	**0.47****	**0.47****	**0.68****
	Facial expressions	84.00	**0.85****	**0.79****	**0.60****	**0.59****	**85.42****	0.51	0.17	0.09	0.07
	Head and facial movements	82.90	**0.80****	**0.75****	**0.57****	**0.56****	**78.87****	**0.61***	**0.23****	**0.21****	**0.26****
	Body posture and movements	**87.70***	**0.92****	**0.76****	**0.77****	**0.85****	**84.46***	**0.73****	**0.33****	**0.36****	**0.41****
Linear SVM	*All features*	**93.00****	**0.86****	**0.79****	**0.79****	**0.72****	87.04	**0.63****	**0.22***	**0.34****	**0.33****
	Physiology	77.20	0.54	0.31	0.11	0.08	89.29	0.50	0.11	0.00	0.00
	Computer use	**90.80****	**0.84****	**0.72****	**0.76****	**0.70****	89.31	0.50	0.11	0.00	0.00
	Facial expressions	73.60	0.51	0.28	0.07	0.05	89.31	0.50	0.11	0.00	0.00
	Head and facial movements	79.90	**0.66***	**0.45***	**0.41****	**0.38***	88.51	0.51	0.13	0.05	0.03
	Body posture and movements	**89.80****	**0.82****	**0.71****	**0.72****	**0.65****	86.99	**0.57***	**0.17***	**0.23****	**0.19****
AdaBoost	*All features*	**97.00****	**0.98****	**0.92****	**0.95****	**1.00****	90.39	**0.88****	**0.58****	**0.46****	**0.41****
	Physiology	79.90	**0.75****	**0.65****	**0.51****	**0.50****	89.19	**0.71****	**0.28****	0.01	0.01
	Computer use	**97.00****	**0.98****	**0.92****	**0.95****	**1.00****	89.04	**0.81****	**0.42****	**0.19***	0.15
	Facial expressions	73.30	0.52	**0.42***	0.17	0.14	89.03	**0.62****	**0.20****	0.01	0.01
	Head and facial movements	81.20	**0.85****	**0.75****	**0.48****	**0.43****	89.03	**0.65****	**0.20****	0.01	0.01
	Body posture and movements	84.00	**0.89****	**0.81****	**0.58****	**0.54****	88.36	**0.73****	**0.33****	**0.17***	**0.13***
C4.5	*All features*	**92.50****	**0.99****	**0.97****	**0.80****	**0.72****	87.70	**0.67***	**0.31****	**0.26****	**0.22****
	Physiology	81.70	**0.75****	**0.61****	**0.54****	**0.52****	89.13	0.51	0.11	0.01	0.01
	Computer use	**93.00****	**1.00****	**0.99****	**0.80****	**0.71****	89.03	0.55	0.15	0.01	0.01
	Facial expressions	69.40	0.58	**0.43***	**0.32***	**0.33***	**85.56***	0.47	0.13	0.04	0.04
	Head and facial movements	73.20	**0.68***	**0.52***	**0.35****	**0.34****	89.18	0.51	0.11	0.00	0.00
	Body posture and movements	73.80	0.63	**0.45***	**0.45****	**0.49****	**84.60****	0.54	**0.21***	**0.21****	**0.20****
MLP	*All features*	**97.30****	**1.00****	**1.00****	**0.91****	**0.88****	87.73	**0.79****	**0.39****	**0.28****	**0.23****
	Physiology	70.60	0.57	**0.48****	0.25	0.23	**85.24****	0.53	0.18	0.08	0.07
	Computer use	**99.90****	**1.00****	**1.00****	**1.00****	**1.00****	**83.92****	**0.74****	**0.32****	**0.24****	**0.26****
	Facial expressions	68.20	0.47	**0.43***	**0.25***	**0.26***	**83.2****	0.53	0.18	**0.12***	**0.11***
	Head and facial movements	81.50	**0.80****	**0.73****	**0.50****	**0.48****	**83.78****	0.59	**0.21***	**0.14***	**0.13***
	Body posture and movements	**87.60****	**0.86****	**0.83****	**0.66****	**0.60****	86.76	**0.67****	**0.29****	**0.25****	**0.22****

Table 3.12: Classification results for the Reliable Task Load Index Change from the relaxed state and from the previous state by feature type for 10-fold CV (Statistically significant improvement (adjusted *p<0.01, **p<0.001) in comparison to a baseline algorithm).

recordings. Other methods to map each segment of the recordings to a facial expression should first be tested before discarding an existing useful correlation between these data.

Following with feature selection, self-reported stress levels were found to be best predicted by computer use patterns, even better than using the whole set of features, to the point of becoming a statistically significant prediction model. As mentioned, this is something to be interpreted cautiously. The rest of the labels were best predicted by models based on the whole set of physiological and behavioural features.

Examining the results of the feature analysis performed on the standardized self-reported scores, in addition to finding higher correlations than in the non-standardized case, the use of body posture and movements to build prediction models showed improved results. In fact, self-reported stress showed enough statistical significance to be considered predictable by means of these measures. Interestingly, all NasaTLX questionnaire-based responses showed to be best predicted by body posture and movements, above models based on computer use and facial and head movements. Again, physiological measurements and facial expressions by themselves were not found to be useful to create prediction models for the target labels.

In terms of objective condition detection from smart office data, results show a highly significant prediction ability of the models. The Naïve Bayes and AdaBoost algorithms appear to be the best algorithms for this problem, whereas against expectations, a specific multiclass classifier following a one-vs-all approach was not showing any advantage over the rest of the algorithms. In this case, both facial expressions and physiological features also showed prediction ability, whereas head and facial movements did not. Therefore, we notice a difference in features' ability to predict self-reported stress and workload levels to objective condition settings' prediction. However, as in previous cases, the most repeated feature sets in terms of statistical significance are computer use patterns, body posture and movements and the whole set of features.

Results for the standardized condition detection case were improved in comparison to the non-standardized versions. In terms of algorithms, AdaBoost was found to be the most effective for this purpose, and regarding feature types, all except head and facial movement-based models were found to be statistically significantly predicting the target labels. Note that the standardization technique used for this purpose is similar to performing a discretization of the self-reported stress values. Therefore, results are transferable to the prediction of these scores.

Regarding objective condition change detection, overall, usual algorithms were performing better or similar than the SMOTEBoost and RUSBoost class-imbalance specialized algorithms. Notwithstanding, a significant improvement in the sensitivity of the models was noticed with these latter algorithms, as promised. Useful prediction models were achieved for computer use pattern-based models, as well as for models based on the combination of all features and on only using body posture and movement features. As for the detection of the direction of these changes, models' performance is worsened. Only models built using all features extracted from the experiments were showing enough prediction power, along with a fairly reduced sensitivity. This is not surprising, because, on one hand, the three-class classification problem that poses the detection of changes' directionality is more complex than the two-class classification problem of the absolute changes' detection, both due to an added class to classify and to the reduced number of instances available for each class. On the other hand, the time-series statistics extracted from the data are not necessarily reflecting the directionality of the physiological and behavioural features, but an absolute change. As the directionality of some of the features used in the study can be directly related to the outputs' directionality (*e.g.*, increased SCL levels to increased stress levels), the use of only absolute change statistics might difficult the resolution of this problem.

Reliable change detection was found to be predictable both from a relaxed state and between consecutive 5-minute time intervals. For the first case, we saw that the best predictors were computer use patterns followed by body posture and movement features. The rest of the source-specific models did not show enough statistical significance to accept their predictability of this target, but

the combination of all features also showed to be useful for this purpose. Regarding reliable change detection between consecutive 5-minute time intervals, we found less significances, thus a harder problem to solve. Also, we noticed a decrease in the sensitivity of the models for this detection problem compared to the previous approach. Nevertheless, computer use patterns, body posture and movements and head and facial movements were predicting this change. Surprisingly, the combination of all features was yielding lower and not statistically significant results.

Note that we only performed all our analyses with a single time-window length combination (1 minute for data aggregation, 5 minutes for time-series statistics' computation). Results might vary depending on the length of these temporal windows, and therefore, an analysis of the effects of these window-size choices and the estimation of the best values to use would be highly required.

Summing up, this work has demonstrated the possibility of predicting the perceived stress and workload levels of office workers, as well as the objectively measured conditions they might be undergoing or the significant workload condition changes that they might be suffering from changes in unobtrusively collected smart office-based physiological and behavioural data. Three main conclusions can be drawn from all these analyses: first, the importance of the use of standardization methods to reduce the intrinsic inter-subject variability of stress and workload assessment methods. Overall, all analyses of this work found improved results for these type of labels. Second, the repeated statistical significance of the computer use patterns and body posture and movements suggest the relevance of these data for stress and workload prediction, while surprisingly, physiological measurements did not highly contribute to the task. Nonetheless, as previously mentioned, computer use patterns might be biased due to the experiment's nature and must, therefore, be verified with alternative datasets or data analysis methods. Also, physiological signals might better reflect users' stress levels when time-series statistics that take into account the directionality of their change are used. Finally, the importance of the use of highly-reliable and well-established stress and/or mental workload assessment methods must be ensured to build the final models. Results presented herein suggest that NasaTLX questionnaire captures in a relatively objective way the perceived mental workload levels of the workers and thus, is a good candidate for this purpose.

Summary

- Stress and mental workload levels can be predicted from changes in smart office behavior data, as well as a reliable change in the analysed scores.
- Behavior showed much more predictability than physiological measurements.
- Computer-use patterns and body posture and movements are the most predictive behaviors for this purpose.
- NasaTLX test is a good basis for ground-truth, but all the analysed standardized tests' results can be predicted from behavior.

4

Adapting and validating the system to predict Alzheimer's Disease (AD) in Smart Home environments

Contents

This chapter explains the work carried out for the second case study: AD diagnosis in smart home environments, and contributes to the State of the Art presented in Section 2.3. After a brief introduction of the context in Section 4.1, the use and adaptation of the approach presented in Chapter 3 for this specific case is exposed in Section 4.2. In Section 4.3, models for the detection of AD symptoms' are built and validated in an unobtrusively collected real-life dataset. Finally, in Section 4.4, conclusions for this scenario are drawn.

4.1 Introduction

Increasing life expectancy in developed countries has resulted in a growing number of cases of people affected by age-related neurodegenerative diseases, such as AD. An estimate of 115.4 million people will suffer from AD in 2050 [361], which can result in devastating consequences in terms of health-care costs and quality of life of patients and relatives given that there is no known cure [227]. As a matter of general interest, the search for methods of early detection and a cure for AD are currently high priority issues.

AD manifests symptoms in multiple domains, including mood, behaviour, and cognition [393]. These symptoms and the associated pathology are usually measured by means of self- and informant-report questionnaires, clinical assessments conducted by health care professionals and medical examinations that may involve brain imaging. Often evaluations are initiated after symptoms have been prominent for some time, resulting in a delayed diagnosis [394]. Currently, only treatments to delay and reduce cognitive and behavioural symptoms of AD are available [395]. Given that AD pathology in the brain accumulates slowly over time, a key for these treatments to be effective is early detection of the disease and implementation of available treatments.

Smart homes are an emerging technological solution, enabling the monitoring of people's behaviour unobtrusively and ubiquitously [396]. Real-life data can be gathered non-stop throughout the day in a completely naturalistic way for the user, offering a complete view of older adults' behaviour and allowing the detection of changes that might indicate the onset of a disorder. If smart home-based behaviour shifts were mapped to AD, the main disadvantages of the usual assessment methods could be overcome, making an early diagnosis of the disorder possible.

Our goal in this chapter is to assess the possibility of detecting changes in psychological, cognitive and behavioural symptoms making use of unobtrusively collected smart home behaviour data. The affirmation of this hypothesis would result in development and implementation of an early detection system for disorders that provoke behavioural changes, such as AD. Such a system could alert patients and relatives of likely changes, making it possible to take timely action.

The main contributions of this chapter can be summarized as follows. We analyse the predictability of several multimodal symptoms often found to be impaired in AD, we analyse the contribution of behavioural features to the prediction of these health assessment scores, and we introduce and assess new smart home-based behaviour features to quantify global daily routine. In addition, we offer an approach to detect a reliable change in the health assessment scores based on unobtrusively collected behavioural data and to address the imbalanced class distribution problem that is common in health-related data.

4.2 Proposed approach

4.2.1 Data collection in the Smart Home environment

First, we unobtrusively collected in-home behavioural data of 40 older adults living in 38 smart homes from two senior-living communities and we gathered biannual neuropsychological assessment

Table 4.1: Modality, test-retest reliability, and standard deviations of the scores used in the study

Domain	Score	r_{score}	SD_{score}	Ref.
Mobility	Arm Curl	0.96	4.98	[403]
	TUG	0.96	3.18	[402]
Cognition / Memory	RBANS - total	0.84	15.58	[404]
	RBANS - attention	0.16	19.0	
	RBANS - delayed memory	0.77	13.29	
	RBANS - immediate memory	0.75	16.58	
	RBANS - visuospatial	0.76	15.31	
	RBANS - language	0.33	15.31	
	PRMQ - total	0.89	9.15	[405]
	PRMQ - prospective memory	0.85	4.91	
	PRMQ - retrospective memory	0.89	4.98	
	Digit Cancellation	0.85	37.20	[406]
Mood	GDS	0.68	2.20	[407]

data. This data was collected by the Center for Studies in Adaptive Systems (CASAS) and the Neuropsychology and Aging Laboratory at Washington State University (WA, USA). Part of this data (N= 18 older adults) was analysed in previous work [114]. For this work, a larger sample is available thanks to a longer monitoring time and to the inclusion of more subjects in the study.

The current study focuses on cognition, mobility, and mood (depression) scores (see Table 4.1), which were collected as part of the biannual assessment and have been found to be affected by AD [393]. Cognitive abilities of the older adults were measured by means of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) [397], the Prospective and Retrospective Memory Questionnaire (PRMQ) [398] and a Digit Cancellation test, while mobility was assessed by Timed Up and Go (TUG) [399] and Arm Curl [400] tests. The Geriatric Depression Scale - Short Form (GDS-15) [401] was used to assess the depression level of the elderly under study.

The smart home sensor data collection used for this study was from 2011 through 2016, a period in which the data were collected continuously for lengths ranging from <1 month to 60 months (M= 19.95 months, SD=17.98 months) depending on each apartment. As data coming from homes with multiple habitants poses some additional challenges for correctly estimating each individual's activity level, these participants were not included in our analyses. Subjects with missing health assessment data or with behavioural data collected for less than 6 months were also removed. Hence, the final dataset contained the behavioural and health assessment data (cognition, mobility, and mood) of 29 older adults (21 females, 8 males) who were living independently and alone in their own smart home residences (M=26 months, SD=17.5 months, range=6-60 months). All participants were 73 years of age or older (M = 84.34, SD = 5.70, range 73-97) and have a mean education level of 17.46 years (SD = 2.06, range 12-20). Participants were classified as either cognitively healthy (N = 13), at risk for cognitive difficulties (N = 10) or experiencing cognitive difficulties (N = 6). One participant in the cognitively compromised group was diagnosed with a brain tumor with marked reductions in cognition proceeding diagnosis. The remaining 5 individuals in the cognitively compromised group met criteria for Mild Cognitive Impairment (MCI) as outlined by the National Institute on Aging-Alzheimer's Association workgroup [402]. Participants in the risk group had data suggestive of lowered performance on one or more cognitive tests (relative to an estimate of premorbid abilities), but did not meet criteria for MCI or dementia.

```
2011-06-18 12:45:10.959930 LivingRoom LivingRoom MA003 ON Work
2011-06-18 12:45:12.653415 LivingRoom LivingRoom MA003 OFF Work
2011-06-18 12:45:13.406246 LivingRoom LivingRoom MA003 ON Work
2011-06-18 12:45:14.544843 LivingRoom LivingRoom MA003 OFF Work
2011-06-18 13:23:16.338151 WorkArea WorkArea M005 ON Work
2011-06-18 13:23:18.041338 WorkArea WorkArea M005 OFF Relax
2011-06-18 13:23:18.587922 WorkArea WorkArea M005 ON Relax
2011-06-18 13:23:24.956352 WorkArea WorkArea M005 OFF Relax
2011-06-18 13:23:31.533091 Kitchen Kitchen MA006 ON Cook
2011-06-18 13:23:34.531439 Kitchen Kitchen MA006 OFF Cook
2011-06-18 13:23:35.468449 Kitchen Kitchen MA006 ON Cook
2011-06-18 13:23:37.725703 Kitchen Kitchen MA006 OFF Cook
2011-06-18 13:23:55.332816 Kitchen Kitchen MA006 ON Cook
2011-06-18 13:23:56.459425 Kitchen Kitchen MA006 OFF Cook
2011-06-18 13:24:03.583423 Kitchen Kitchen MA006 ON Cook
2011-06-18 13:24:05.268921 Kitchen Kitchen MA006 OFF Cook
2011-06-18 13:24:11.082793 WorkArea WorkArea M005 ON Eat
2011-06-18 13:24:18.597894 WorkArea WorkArea M005 OFF Eat
…
2011-07-28 08:39:39.277723 Bathroom Bathroom MA008 ON Personal_Hygiene
2011-07-28 08:39:40.408324 Bathroom Bathroom MA008 OFF Personal_Hygiene
2011-07-28 08:40:39.410092 Bedroom Bedroom MA007 ON Sleep
2011-07-28 08:40:41.827160 Bedroom Bedroom MA007 OFF Sleep
```

Figure 4.1: Extract of an AR activity-labeled raw sensor data stream

4.2.2 Preprocessing

Day-level behaviour feature extraction

Smart homes were set up to collect all sensor events that took place in each residence during the study period. Each raw-sensor data stream event was an entry specifying the event's timestamp, ID of the sensor detecting the event and type of event (activation/deactivation). In order to make the raw-sensor data streams interpretable, it was first necessary to assign a specific activity to each sensor entry. For that purpose, the AR activity recognition algorithm specified in [408] was used. This algorithm maps each one of the sensor events to a value from a predefined set of activity labels in real-time, by applying an adaptive length sliding window to the raw sensor data stream. The predefined set of activities include both ambulatory activities (such as mobility inside the home) and specific Activities of Daily Living (ADL)s (*e.g.* cook, eat, sleep, or relax). This approach allows not only to take into account the actual sensor events to identify the activity being performed but also contextual information such as the activity performed in the previous time-window. The reliability of this algorithm has been demonstrated in previous work, where accuracy greater than 98% was achieved on 30 testbed smart homes using three-fold cross validation [408]. Figure 4.1 shows an extract of an AR activity labeled sensor data stream.

Once the activity-level information was available, we computed 17 daily behaviour features for each subject, explaining their daily sleep and mobility patterns, time spent in several specific ADLs (*e.g.*, cook, eat) and overall characteristics of their routines. A detailed list of the computed features can be seen in Table 4.2.

The daily distance that the subjects were traveling inside their homes was estimated by creating sensor mapping files based on the floor plan and sensor layout for each residence (see example in Figure 4.2), where the x-y coordinates of the motion sensor's positions were specified. Three of the apartments lacked specific information about the positioning of the sensors and/or the distribution within the houses. In those cases, it was first necessary to estimate the positions of the sensors, which was done by considering the apartments to be of a similar shape to the rest and checking the activation order of the sensors in the raw sensor data files. Once all sensor positioning information was available, we computed the daily sum of the Euclidean distances between the consecutively

Table 4.2: Day-level activity features included in the study

Type	Day-level features
Duration of specific activities (6 features)	Time spent per day in cooking, eating, relaxing, carrying out personal hygiene activities, being out of home and nighttime toileting activities
Sleep-related (2 features)	The daily sleep duration and frequency
Mobility-related (2 features)	The total number of activated sensors and the total distance covered walking inside the apartment per day
Routine-related (7 features)	Complexity of the daily routine, number of total and of non-repeated activities performed per day, maximum and minimum inactivity times, day length and similarity with the previous day

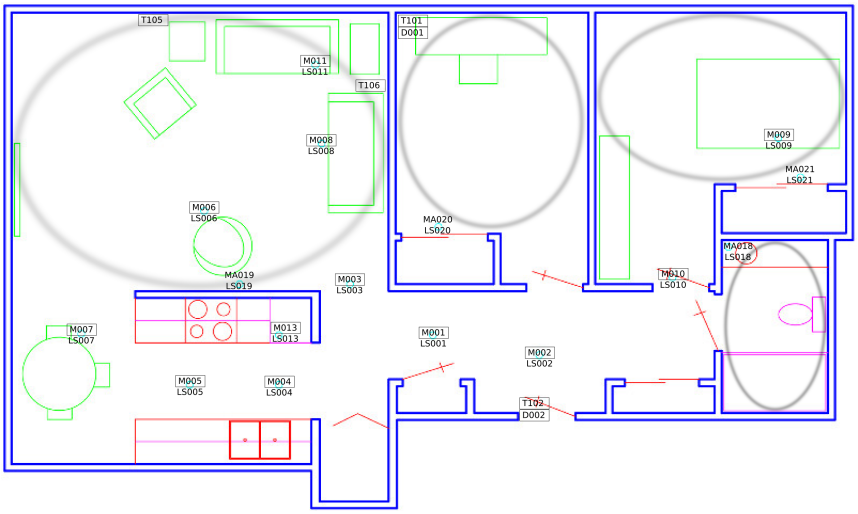


Figure 4.2: Floor plan and sensor layout of one of the residences of the study

activated motion sensors in order to estimate the total walking distance traveled by the inhabitants. Note that this approach only provides an approximation of the real covered distance, as it doesn't take into account the existence of walls or other obstacles between the sensors that must be avoided or surrounded.

To compute daily-routine features, we first extracted the daily activity sequence from the AR-labeled sensor data stream. We then encoded the daily activity sequence by replacing each activity with a number from 1 to 12 (*i.e.*, Sleep=1, Cook=2, Relax=3, ..., Other = 12). Shannon entropy was used as the measure of complexity of the daily routine. To get this entropy value, we computed the daily probability distribution (histogram) of the activity sequence (P) and we then applied the Shannon entropy as shown in Equation 4.1,

$$\text{Complexity}_{\text{routine}} = \sum_{\text{activity}=1}^{12} \text{P}_{\text{activity}} - \log_2 \text{P}_{\text{activity}} \tag{4.1}$$

where $\text{P}_{\text{activity}}$ was the probability of a certain activity to happen during the day based on the actual day's histogram.

The same encoded activity-sequence was used to compare the daily routines of consecutive days. For this purpose, we used an implementation of the "gestalt pattern matching" algorithm [409]

available in Python as a "SequenceMatcher" function, which expresses the similarity of any two sequences as a ratio between 0 and 1. This allowed us to measure the degree of similarity between consecutive days. Finally, we checked the timestamps of the daily activity events and computed the day-length as the time elapsed from the first to the last detected activity of the day. We believe that the remaining features in Table 4.2 are self-explanatory.

Between-assessments behaviour statistics' computation

At the end of the previous step, we had available a set of daily activity features for each subject. We then applied the Clinical Assessment using Activity Behaviour (CAAB) algorithm, which was introduced in [114], to the daily activity data in order to extract the behavioural statistics of each between-assessment period. RStudio for R was the selected environment for this purpose.

In short, the CAAB algorithm was used to apply the following processing steps to the daily behaviour data: 1) Take each subject's between-assessment daily behaviour data (which was 6 months in length as assessments were performed twice a year), 2) Apply a log transform and a Gaussian detrending to each time-series (behavioural variable), 3) Compute five summarizing time-series statistics (variance, skewness, kurtosis, autocorrelation, and change) for each behavioural feature in this period using a sliding window of length 7 days and 4) Compute the average of each time-series statistic for the 6-month period and use the set of averages for the final predictions.

The resulting preprocessed dataset for further analysis was a collection of 85 (5 time-series statistics of 17 behavioural features) biannual summary behaviour statistics for each of 29 older adults who were living alone in their sensorized apartments for a period of $24.0 \pm 13.68(\mathrm{SD})$ months.

Health assessment scores

Our goal is to create prediction models that map smart home-based behaviour features to health assessment values that might capture AD symptoms. In this study, our target variables are the Arm Curl and TUG mobility test scores, cognition assessment based on Digit-Cancellation test, RBANS and PRMQ scores and subscores, as well as depression symptoms represented as GDS test-scores. All these values were collected from the participants at the end of each corresponding 6-month period.

In order to take into account the inter-subject variability between time points, we standardized the data by computing a Reliable Change Index (RCI) [385] that informs whether a participant's performance has suffered a significant change in an assessment score based on his/her own previous performance that cannot be accounted by repeat testing or practice effects. The RCI discards changes that might have appeared due to reasons other than an actual change in scores (such as measurement unreliability) by applying a threshold to the scores' differences. We looked for both reliable absolute changes compared to baseline values ($\mathrm{RCI_{baseline}}$) and compared to the previous assessment point ($\mathrm{RCI_{consecutive}}$) of each subject for all tests' outputs.

In order to calculate the RCIs for the scores used herein, we gathered test-retest reliability ($\mathrm{r_{score}}$) and standard deviation ($\mathrm{SD_{score}}$) that the tests have shown in their development cohorts and/or in previous works, as shown in Table 4.1. Therefore, the RCIs for each subject were computed as:

$$\mathrm{RCI_{baseline}(i)} = \frac{\mathrm{Score_i} - \mathrm{Score_{baseline}}}{\sqrt{2\mathrm{SEm}}} \tag{4.2}$$

$$\mathrm{RCI_{consecutive}(i)} = \frac{\mathrm{Score_i} - \mathrm{Score_{i-1}}}{\sqrt{2\mathrm{SEm}}} \tag{4.3}$$

where SEm or Standard Error of Measurement represents the expected variation of the observed test scores due to measurement error and is computed as $\mathrm{SEm} = \mathrm{SD_{score}}\sqrt{1 - \mathrm{r_{score}}}$, $\mathrm{r_{score}}$ is the test-retest reliability measuring the consistency of the test-scores over time, $\mathrm{Score_i}$ is the test score

at assessment point *i*, $Score_{baseline}$ is the test score at the first/baseline assessment and $Score_{i-1}$ is the test score at the previous assessment point.

There were few positive instances (data instances where a reliable change was observed) for some of the assessment scores, resulting in highly imbalanced data. For the following analyses, we removed from the study those tests which were extremely imbalanced ($<$5% of positive instances). We distinguished the remaining tests as imbalanced (5%-30% of positive instances) and not-imbalanced data (30%-50% of positive instances).

Additionally, we also considered the possibility of detecting improvement and decline in tests' scores among consecutive assessment-points as a method to reduce inter-subject variability. In this case, the difference between each consecutive assessment point was computed for each self-reported test score of each subject. Every data instance with an improvement in the scores ($\geqslant$0) was considered as a positive point whereas a decline in the performance of the skill being evaluated by tests ($<$0) was labeled as a negative point.

4.2.3 Building cognition and mobility change prediction models

The preprocessed dataset resulting from the previous steps was analysed using Weka [387]. First, we performed a correlation analysis between the mobility, cognition, and mood assessment scores and the smart home behaviour data. For this purpose, we used four different regression models using all behavioural features computed in the previous step for each one of the scores. The four models we evaluated were Support Vector Regression (SVR) with a linear kernel, Linear Regression (LR), SVR with a Radial Basis Function (RBF) kernel and k Nearest Neighbors (kNN) algorithms. We compared the correlation coefficients (r) and Mean Absolute Error (MAE) of the models using 10-fold Cross Validation (CV) approach. Corresponding pairwise random algorithms were built and evaluated in our dataset following the same process. These random algorithms provided a basis of comparison to ensure that performance results are not due to chance. The random algorithms were built using a uniformly distributed random data-matrix of the same size as the real behavioural data while respecting each variable's data range as in the original dataset. A corrected paired t-test was used to detect a significant improvement of smart home-based algorithms in comparison to the random data algorithms. Adjusted p-values (*p$<$0.01, **p$<$0.001) were used to avoid Type 1 error when checking for significance.

In order to analyse the types of behavioural features that are most correlated with each one of the tests, we built and evaluated activity-specific models for the main test scores with 10-fold CV. The behavioural features that were included in each one of the models are shown in Table 4.3. Again, we searched for statistically significant improvement in comparison to pairwise random algorithms using a corrected paired t-test and adjusted p-values (*p$<$0.01, **p$<$0.001).

Regarding RCI detection, we used different approaches for the imbalanced and not-imbalanced datasets. First, not imbalanced datasets containing all behavioural features were reduced by means of a Principal Component Analysis (PCA). Principal Component (PC)s explaining 95% of the variability in the behaviour data were kept to create the reduced datasets. Support Vector Machines (SVM), AdaBoost, Multilayer Perceptron (MLP) and Random Forests (RF) algorithms were trained and validated following a 10-fold CV approach. Area under the ROC curve (ROC_{auc}), area under the Precision-Recall curve (PR_{auc}), Fscore and sensitivity were selected as the metrics for model evaluation. The combination of these metrics offers an excellent overview of both the models' overall performance and the capability to detect the event of interest (the reliable change event), and are especially suitable when the data distribution is skewed. A corrected paired t-test was used to detect a significant improvement of smart home-based algorithms in comparison to the pairwise random data algorithms, and an adjusted p-value (*p$<$0.0125) was used to avoid Type 1 error.

For the imbalanced datasets, a different approach was required. Common machine-learning algorithms tend to create biased models towards the majority class when being applied to imbalanced

Table 4.3: Task-specific grouping of the daily features

Group	Day-level features
Daily-routine	Complexity of the daily routine, number of total activities and number of non-repeated activities performed per day, maximum and minimum inactivity times, day length and similarity with the previous day
Mobility	The total number of activated sensors and the total distance covered walking inside the apartment per day
Outings	Time spent per day in being out of home
Mobility & outings	Mobility + Outings
Sleep	The daily sleep duration and frequency
Overnight toileting	Time spent per day in nighttime toileting activities
Overnight patterns	Sleep + Overnight toileting
Cook & eat	Time spent per day in cooking and eating

datasets, resulting in high-accuracies but, very low sensitivity. In most of the health-related machine learning applications, the event in which we are more interested is the rare event or the minority class, highlighting the need to use alternative methods to improve the detection of these minority events. Two algorithmic approaches are tested in the current work to overcome this issue. The first one, SMOTEBoost [373], is a method combining boosting techniques with SMOTE [389] oversampling techniques. Whereas boosting aims at creating a "strong" classifier using a set of "weak" classifiers, SMOTE is a technique to oversample the minority class by creating synthetic data instances and thus, reduce class imbalance. SMOTEBoost combines these processes iteratively in order to improve the sensitivity of the models without the overall accuracy being affected.

The second approach, the wrapper-based Rapidly Converging Gibbs sampler (wRACOG) [410], is a minority-class oversampling algorithm based on Gibbs sampling. Unlike SmoteBOOST and most of the minority-class oversampling techniques, wRACOG takes into account the underlying probability distribution of the minority class and the interdependencies of the data attributes when synthetically generating rare-event samples. This results in a better representation of the minority class. Moreover, wRACOG learns the models iteratively, selecting from the Markov chain generated by the Gibbs sampler the samples that have the highest probability of being misclassified by a learning model (wrapper) at each step, often leading to better classification rates. wRACOG stops iterating when there is no further improvement with respect to a chosen performance metric.

First, we built prediction models for imbalanced datasets using SMOTEBoost and kNN with k=5 as the "weak" classifier. A 3-fold CV was performed for validation purposes. Pairwise random algorithms were also built using the previously mentioned random data and were validated for prediction of our data following the same 3-fold CV process. Again, ROC_{auc}, PR_{auc}, Fscore and sensitivity of the models were computed for models' performance screening. McNemar's test was applied to check whether a significant improvement (for an adjusted p-value ($*p<0.005$)) was observed using smart home data in the prediction of reliable change in the scores in comparison to random data algorithms.

Next, we built the prediction models for the same imbalanced datasets following the second approach, *i.e.* using the wRACOG algorithm. For this purpose, it was first necessary to discover the interdependencies of the data attributes. In order to reduce the dimensionality of the data and to make it easier to map the interdependencies between the attributes, we used the PCA-based reduced datasets explaining the 95% of the data variance. Moreover, wRACOG assumes that the data attribute values are categorical, so we first discretized all of the PCs into five uniform bins. We then constructed the Bayesian tree of dependencies following the Chow-Liu algorithm in Weka. The Chow-Liu algorithm [411] aims at constructing a maximal weighted spanning tree in a graph, allowing each attribute to have exactly one parent on which its value depends. Thus, the interdependencies

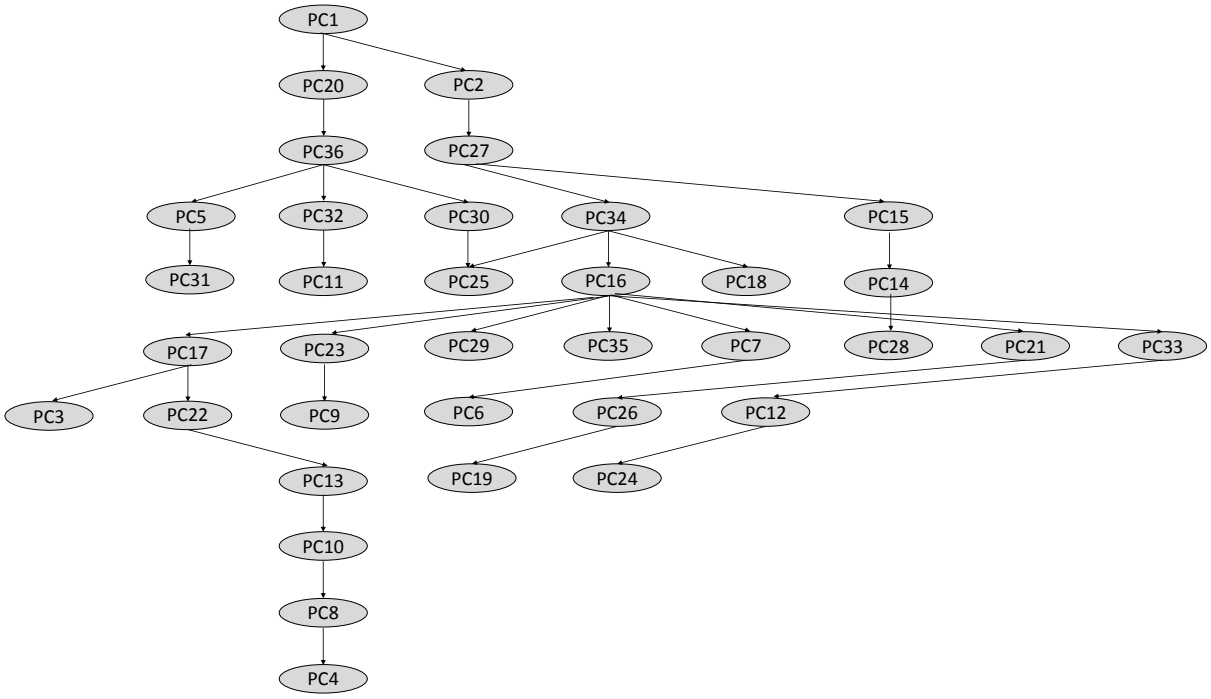


Figure 4.3: Chow-Liu tree for the PCA-reduced dataset

between the PCs were discovered. Figure 4.3 shows the Chow-Liu interdependency tree for the PCA-reduced and discretized baseline dataset.

A kNN algorithm was used as the wrapper classifier and two different stop criteria for the iterative process were tested: 1) First, as in many applications the detection of the reliable change might be critical, we searched the maximum sensitivity of the models. 2) Second, for cases where the overall prediction ability of the models might be more interesting, we used the maximized ROC_{auc} metric as the stop criteria for the algorithm. A 5-fold CV was performed this time for validation purposes and ROC_{auc}, PR_{auc}, Fscore, and sensitivity of the models were evaluated. As in previous cases, in order to check for statistically significant prediction of reliable change in the scores using smart home data, we compared models' outputs to those of their pairwise random algorithms by means of a McNemar's test. Adjusted p-value (*$p<0.005$) was used to avoid family-wise (Type 1) error rate. The PCA-reduced random dataset was discretized following the same process as the actual smart home dataset.

Finally, for the analysis on elderly's skills' improvement/decline detection from smart home data, we used the PCA-based reduced dataset as in the previous case. SVM, AdaBoost, MLP and RF algorithms were trained and validated following a 10-fold CV approach for the labels indicating an improvement (positive instance) or a decline (negative instance) in older adults' scores. ROC_{auc}, PR_{auc}, and Fscore were computed for each one of the algorithms and compared to the ones of their pairwise-random algorithms. As the detection of a decline in the performance of the skills aimed at measuring with the self-reported test might be more important than the detection of an improvement, we also computed the sensitivity of the algorithms towards these negative events. All statistical significances were checked for adjusted p-values (*$p<0.01$, **$p<0.001$).

Table 4.4: Regression results for the absolute test scores using all behavioural features for 10-fold CV (Statistically significant improvement (adjusted *p<0.01,**p<0.001) in comparison to the corresponding pairwise random algorithm.)).

	SVR		**LR**		**SVR RBF**		**kNN**	
	r	*MAE*	*r*	*MAE*	*r*	*MAE*	*r*	*MAE*
Mobility								
Arm Curl	0.17	5.62	0.06	4.99	**0.29***	4.18	0.14	5.91
TUG	**0.51****	3.89	**0.36****	3.92	**0.57****	**3.03****	**0.42****	3.72
Cognition								
PRMQ	**0.26***	9.28	0.20	8.42	**0.31****	7.01	**0.31***	8.40
Prospective Memory	**0.30***	4.72	0.12	4.55	**0.26***	3.72	**0.28***	4.50
Retrospective Memory	0.15	5.15	0.14	5.17	**0.39****	**3.57***	0.27	4.53
RBANS	0.27	15.97	**0.34****	15.50	**0.40****	13.07	**0.34****	15.74
Attention	**0.40****	16.14	**0.33****	17.12	**0.31****	15.52	0.22	16.80
Delayed Memory	0.13	15.01	0.04	17.65	**0.31***	11.19	**0.31***	16.39
Immediate Memory	0.00	16.23	0.03	15.43	0.08	12.04	**0.29****	14.52
Language	**0.47****	10.77	**0.35****	13.25	**0.47****	**10.10***	**0.26***	12.41
Visuospatial	0.01	18.25	0.04	17.51	**0.21***	12.28	0.18	15.12
Digit Cancel - Speed	0.22	35.05	0.17	46.10	0.18	26.88	0.23	31.26
Mood								
GDS	0.02	2.79	0.02	2.84	0.21	**1.67***	0.12	1.97

4.3 Validation

Absolute test scores' prediction

Table 4.4 shows the results of the regressions for all the absolute test scores using all smart home behavioral features. For mobility tests, whereas Arm Curl had low correlation with behavioral data, TUG demonstrated a moderate to strong correlation with behavioral data. For the cognition overall scores and subscores, the measures showed mostly moderate correlations with behavioral data. Exceptions included the visuospatial and immediate memory subscores of the RBANS test and the digit cancellation test scores, which were found to correlate weakly. In fact, the digit cancellation test did not show any statistically significant improvement compared to random models. Finally, depression showed a weak correlation with the global set of smart home behavioral data with only enough statistical significance for a reduced prediction error on the SVR RBF regression model.

Regressions based on specific activities, which can be seen in Table 4.5, showed some interesting results. The Arm Curl mobility test showed weak but, statistically significant correlations only with outings and cooking and eating features. In contrast, the TUG test showed significant moderate correlations with daily routine, overnight toileting and the combination of overnight toileting and sleep, as well as a significant weak correlation with cooking and eating features.

Regarding the self-reported cognition questionnaires, the global PRMQ score was moderately associated with daily routine and to the overnight patterns, as well as weakly correlated to sleep and overnight toileting. RBANS was moderately correlated with overnight patterns, whereas it was also showing weak yet statistically significant correlations with mobility, daily routine, and overnight toileting behaviors. Digit Cancellation processing speed was found to be moderately correlated to sleep and overnight patterns, and weakly yet significantly to overnight toileting features.

Finally, for the geriatric depression assessment, we did not find any significant correlations but we perceived a significant reduction of the MAE of the models for mobility alone, mobility and outings and sleep feature-sets.

Table 4.5: Regression results for the absolute test scores by behavioural feature type for 10-fold CV (Statistically significant improvement (adjusted *p<0.01,**p<0.001) in comparison to the corresponding pairwise random algorithm.)).

(a) Mobility - Arm Curl

	SVR		LR		SVR RBF		kNN	
	r	*MAE*	*r*	*MAE*	*r*	*MAE*	*r*	*MAE*
Daily routine	0.28	4.35	0.17	4.83	0.16	4.43	0.05	6.52
Mobility	0.09	4.93	0.12	4.71	0.02	4.61	0.18	5.80
Mobility & outings	0.14	4.58	0.11	5.01	0.11	4.42	0.15	5.79
Outings	**0.28***	4.30	0.20	4.46	0.23	4.40	0.00	6.45
Sleep	0.17	4.99	0.11	4.74	0.00	4.57	0.17	5.77
Overnight patterns	0.03	4.81	0.16	4.63	0.10	4.65	0.04	6.51
Overnight toileting	0.10	4.67	0.14	4.66	0.06	4.57	0.12	7.21
Cook & eat	0.06	4.85	0.20	4.55	**0.27***	4.32	0.08	6.19

(b) Mobility - TUG

	SVR		LR		SVR RBF		kNN	
	r	*MAE*	*r*	*MAE*	*r*	*MAE*	*r*	*MAE*
Daily routine	**0.35****	3.77	**0.33***	3.97	**0.37****	**3.52***	0.32	3.77
Mobility	0.12	3.74	0.04	4.21	0.14	3.68	0.08	4.71
Mobility & outings	0.18	3.74	0.11	4.19	0.16	3.70	0.10	4.61
Outings	0.00	3.96	0.07	4.08	0.08	3.76	0.08	4.73
Sleep	0.30	3.66	0.30	3.75	0.25	3.71	0.13	4.53
Overnight patterns	**0.33***	3.59	0.30	3.78	**0.32****	3.53	0.19	4.21
Overnight toileting	**0.26***	**3.54***	**0.30***	3.81	**0.29***	**3.50***	0.14	4.62
Cook & eat	0.16	3.78	**0.24***	3.91	0.10	3.73	0.19	5.92

(c) Cognition - PRMQ

	SVR		LR		SVR RBF		kNN	
	r	*MAE*	*r*	*MAE*	*r*	*MAE*	*r*	*MAE*
Daily routine	0.18	7.96	0.20	7.93	0.21	7.42	**0.32***	8.15
Mobility	0.12	7.84	0.21	7.45	0.16	7.35	0.01	10.63
Mobility & outings	0.09	8.03	0.14	7.91	0.14	7.44	0.17	10.28
Outings	0.05	7.74	0.10	7.88	0.04	7.59	0.19	8.73
Sleep	**0.28***	7.13	0.19	7.56	0.24	7.57	0.11	9.86
Overnight patterns	**0.30***	7.37	0.27	7.42	**0.29***	7.47	0.25	8.51
Overnight toileting	**0.29***	7.18	0.23	7.43	0.25	7.54	**0.28****	8.13
Cook & eat	0.11	7.65	0.07	7.92	0.09	7.54	0.01	12.29

(d) Cognition (Self-Report) - RBANS

	SVR		LR		SVR RBF		kNN	
	r	*MAE*	*r*	*MAE*	*r*	*MAE*	*r*	*MAE*
Daily routine	0.13	15.64	0.13	16.43	**0.25***	14.27	0.18	17.86
Mobility	0.18	14.51	0.23	14.59	**0.26***	14.21	0.18	19.00
Mobility & outings	0.14	15.24	0.22	14.73	0.25	14.11	0.15	17.69
Outings	0.01	14.62	0.09	14.6	0.01	14.45	0.16	16.37
Sleep	0.20	14.36	0.20	14.72	0.24	14.12	0.19	17.12
Overnight patterns	**0.30***	14.18	**0.31***	13.46	**0.31***	13.61	0.17	17.17
Overnight toileting	**0.26***	13.66	0.22	13.91	**0.25***	**13.72***	0.03	18.61
Cook & eat	0.02	14.78	0.10	14.72	0.04	14.23	0.02	18.83

Table 4.5: Regression results for the absolute test scores by behavioural feature type for 10-fold CV (Statistically significant improvement (adjusted *p<0.01,**p<0.001) in comparison to the corresponding pairwise random algorithm.)).

(e) Cognition - Digit Cancellation

	SVR		**LR**		**SVR RBF**		**kNN**	
	r	*MAE*	*r*	*MAE*	*r*	*MAE*	*r*	*MAE*
Daily routine	0.23	26.86	0.20	32.44	0.16	27.08	0.05	37.28
Mobility	0.07	27.77	0.07	29.04	0.08	27.33	0.03	36.70
Mobility & outings	0.00	29.35	0.13	31.26	0.05	27.62	0.13	33.43
Outings	0.12	28.94	0.11	30.22	0.06	27.55	0.02	36.84
Sleep	0.22	26.98	**0.30***	26.44	0.20	26.92	0.04	37.91
Overnight patterns	0.22	28.52	**0.31***	28.03	0.20	27.10	0.09	33.83
Overnight toileting	0.22	27.49	**0.29***	27.68	0.14	27.12	0.17	31.81
Cook & eat	0.21	30.06	0.17	31.59	0.18	27.56	0.02	46.91

(f) Mood - GDS

	SVR		**LR**		**SVR RBF**		**kNN**	
	r	*MAE*	*r*	*MAE*	*r*	*MAE*	*r*	*MAE*
Daily routine	0.16	1.85	0.15	2.38	0.16	1.67	0.07	1.92
Mobility	0.12	1.72	0.25	1.9	0.19	**1.66***	0.13	1.97
Mobility & outings	0.21	**1.68***	0.24	1.95	0.25	**1.62***	0.07	2.07
Outings	0.14	1.73	0.17	1.98	0.19	1.67	0.03	2.55
Sleep	0.26	**1.67***	0.21	1.97	0.25	**1.66***	0.08	2.07
Overnight patterns	0.19	1.77	0.13	2.07	0.22	1.67	0.06	2.05
Overnight toileting	0.02	1.76	0.08	1.91	0.09	1.74	0.14	2.01
Cook & eat	0.08	1.83	0.04	2.07	0.09	1.70	0.01	2.24

Table 4.6: Reliable change detection of Arm Curl scores from baseline (*: Statistically significant improvement (adjusted p<0.0125) in comparison to the corresponding pairwise random algorithm.)

	ROC_{auc}	PR_{auc}	Fscore	Sens.
RF	0.58	**0.73***	**0.77***	**0.92***
SVM	0.59	**0.69***	**0.77***	**0.89***
AdaBoost	0.64	**0.76***	**0.76***	**0.84***
MLP	0.58	**0.75***	**0.69***	**0.71***

RCI detection

The detection of reliable change on attention and language skills were excluded from our objectives due to the uncertainty that their low test-retest reliability would introduce in the results obtained for these labels. Global PRMQ and subscores, consecutive global RBANS scores, RBANS subscores related to immediate memory, Digit cancellation, and the GDS test score were excluded from the RCI detection analyses as they were capturing less than 5% of the reliable change instances. Among the remaining labels, only the reliable change in Arm Curl score from baseline had enough positive instances to be considered a balanced dataset. The remaining scores (RBANS, RBANS delayed memory, RBANS visuospatial and TUG change from baseline, and RBANS delayed memory, RBANS visuospatial and TUG change between consecutive assessments) were considered imbalanced and processed as such.

Table 4.6 shows the results for Arm Curl reliable change detection from baseline using 37 PCs explaining the 95% variability of the data. All four classifiers showed a statistically significant improvement in terms of PR_{auc}, Fscore and sensitivity for the adjusted p-value, whereas ROC_{auc} showed reasonable results surpassing the 0.6 barrier.

Table 4.7 summarizes the results for the prediction models for the imbalanced datasets that are sampled based on the SMOTEBoost algorithm. McNemar's tests found significant improvement of the smart home based prediction models compared to random classifiers for an adjusted p-value of 0.005 for the reliable change detection between consecutive assessments in mobility measured by the TUG test. However, and even having used a method to overcome class-imbalance, models remain yet biased and lacking sensitivity.

Table 4.8 shows the results of the RCI detection models based on the wRACOG algorithms for the imbalanced datasets, using the sensitivity maximization as the criteria for the algorithm to stop. Compared to previous SMOTEBoost based algorithms, the sensitivity of the models is highly improved, which might be very interesting for some applications. However, some models' ROC_{auc} lie below or near 0.5, while their PR_{auc} show low values, which might be again an indicator of a biased model, in this case, towards the minority class. McNemar's tests for an adjusted p-value of 0.005 only found enough statistical significance to accept predictability of delayed memory skills between consecutive assessment points.

Table 4.9 shows the results of the RCI detection models based on the wRACOG algorithms for the imbalanced datasets, using the ROC_{auc} metric as the stop criteria for the iterative algorithm. The sensitivity of the models using this second approach is, overall, higher than the SMOTEBoost based models and lower than the models presented in Table 4.8. Interestingly, in some cases, the ROC_{auc} and PR_{auc}, as well as the Fscores, are greater than the ones obtained with the previous approaches. This suggests a better suitability of the wRACOG based models maximizing ROC_{auc} for some of the RCI detection problems. After controlling for the p-value to reduce the family-type error rate, only the model for the detection of reliable changes on consecutive Arm Curl mobility scores' was showing a statistically significant prediction ability.

Table 4.7: Reliable change detection of the imbalanced scores using SMOTEBoost (*: Statistically significant improvement (adjusted p<0.005) in comparison to the corresponding pairwise random algorithm.)

	ROC_{auc}	**PR_{auc}**	**Fscore**	**Sens.**
$RBANS_{baseline}$ - *total*	0.52	0.05	0.00	0.00
$RBANS_{baseline}$ - *delayed memory*	0.69	0.18	0.31	0.5
$RBANS_{baseline}$ - *visuospatial*	0.45	0.09	0.08	0.08
$TUG_{baseline}$	0.48	0.17	0.06	0.11
Arm $Curl_{consecutive}$	0.40	0.18	0.13	0.12
$RBANS_{consecutive}$ - *delayed memory*	0.40	0.03	0.00	0.00
$RBANS_{consecutive}$ - *visuospatial*	0.68	0.20	0.35	0.50
$TUG_{consecutive}$	**0.56***	**0.22***	**0.15***	**0.50***

Table 4.8: Reliable change detection of the imbalanced scores using wRACOG and sensitivity maximization as stop criteria for the algorithm (*: Statistically significant improvement (adjusted p<0.005) in comparison to the corresponding pairwise random algorithm.)

	ROC_{auc}	**PR_{auc}**	**Fscore**	**Sens.**
$RBANS_{baseline}$ - *total*	0.72	0.07	0.09	1.00
$RBANS_{baseline}$ - *delayed memory*	0.63	0.10	0.13	0.60
$RBANS_{baseline}$ - *visuospatial*	0.72	0.20	0.21	1.00
$TUG_{baseline}$	0.52	0.21	0.32	0.84
Arm $Curl_{consecutive}$	0.54	0.22	0.40	0.83
$RBANS_{consecutive}$ - *delayed memory*	**0.69***	**0.06***	**0.11***	**0.80***
$RBANS_{consecutive}$ - *visuospatial*	0.52	0.09	0.17	1.00
$TUG_{consecutive}$	0.48	0.18	0.35	0.96

Table 4.9: Reliable change detection of the imbalanced scores using wRACOG and ROC_{auc} maximization as stop criteria for the algorithm (*: Statistically significant improvement (adjusted p<0.005) in comparison to the corresponding pairwise random algorithm.)

	ROC_{auc}	**PR_{auc}**	**Fscore**	**Sens.**
$RBANS_{baseline}$ - *total*	0.77	0.07	0.17	1.00
$RBANS_{baseline}$ - *delayed memory*	0.66	0.10	0.19	1.00
$RBANS_{baseline}$ - *visuospatial*	0.64	0.14	0.20	0.23
$TUG_{baseline}$	0.51	0.17	0.39	0.6
Arm $Curl_{consecutive}$	**0.62***	**0.22***	**0.49***	**0.63***
$RBANS_{consecutive}$ - *delayed memory*	0.67	0.03	0.08	1.00
$RBANS_{consecutive}$ - *visuospatial*	0.53	0.09	0.19	0.80
$TUG_{consecutive}$	0.59	0.18	0.29	0.48

Detection of improvement/decline in cognition and mobility skills

Table 4.10 shows the results of the mobility and cognition scores' improvement and decline detection. After adjusting the p-value for reduced family-wise error rate (*p<0.01,**p<0.001), only the detection of improvement and decline in mobility as measured by the Arm Curl test seemed to be possible. A significant improvement both in ROC_{auc} and PR_{auc} was detected using RF and AdaBoost classifiers in comparison to their pairwise random data classifiers, as well as a significant improvement in Fscore and sensitivity of the RF based model.

4.4 Discussions and conclusion

The problem addressed in this work is not an easy task to solve. Our goal was to predict the multimodal symptoms commonly seen in AD from unobtrusively collected behavioral data inside older adults' apartments. Despite the complexity of the task, our results show that measures of cognition, mobility, and depression are predictable using activity-labeled smart home data.

A regression analysis of the smart home-based behavior data with all the tests under analysis has shown several moderate yet significant correlations. As expected, behavioral data were the most correlated to mobility assessment scores, followed by cognitive skills, whereas the most difficult task seems to be mood prediction. Nonetheless, almost all models, with the exception of cognition level prediction based on Digit Cancellation scores, showed a significant improvement compared to models based on random data.

The feature selection analysis has brought to light such valuable information as the predictability of mobility scores from outing patterns, daily routine and cooking and eating patterns. In the specific case of TUG score, there was a significant correlation with global overnight activities including bed-to-toilet transitions. This finding suggests that individuals who take longer to complete the TUG (indicative of slowed movement) tend to be more active at night. This is supported by the AD literature that finds both impaired mobility and sleep disturbances to be related to dementia [412,413]. In [292], TUG showed significant correlations with mobility, outings, sleep and ADL (cook, eat, relax and personal hygiene activities) features. In our case, we did not achieve enough statistical significance for outings, mobility and sleep after adjusting the p-value for reduced family-wise error rate, but we did for global daily routine patterns, which were not analysed previously, and for cooking and eating activities, which likely reflect part of the ADLs of the previous work. Cognition was mainly correlated to sleep and overnight patterns, but also to daily routine, mobility, and outings. These results also agree with previous work [292], where correlations between total RBANS score and smart home activity data were analysed and statistical significance for sleep, mobility, outings, and ADLs was found. Also in agreement with these results, sleep and sleep-related disturbances have been found to be related to cognitive impairment in other research [414, 415], as well as time spent out of home to cognitive state as measured by the Clinical Dementia Rating (CDR) scale [293]. Finally, yet lacking statistical significance for the correlation scores, depression assessed with the GDS scale was found to be predictable with mobility, outings and sleep features in terms of prediction models' error. This agrees with previous work [4] where correlation of GDS score with overall in-home mobility and outing patterns was discovered. Thus, our results validate those reported in the literature, in addition to analysing more in detail each aspect of mobility and cognition skills thanks to the use of more tests and their subscores. Part of the data used for these correlation analyses overlaps with the data used in a previous work (N=18) [292], so similar conclusions would be expected. Nonetheless, we have reaffirmed and given more strength to most of those conclusions by including data collected over a longer period and from more subjects (*i.e.* using a bigger sample size), as well as discovering new correlations with daily routine patterns. In fact, the novel overall daily-routine features presented in this paper have shown predictability both for mobility and cognition skills of the elderly.

Table 4.10: Positive and negative fluctuations' detection between consecutive assessment points (Statistically significant improvement (adjusted *p<0.01,**p<0.001) in comparison to the corresponding pairwise random algorithm.)

	RF				**SVM**			
	ROC_{auc}	*PR_{auc}*	*Fscore*	*Sens.*	*ROC_{auc}*	*PR_{auc}*	*Fscore*	*Sens.*
Mobility								
Arm Curl	**0.65****	**0.54****	**0.33***	**0.28***	0.60	0.38	0.38	0.34
TUG	0.41	0.49	0.38	0.39	0.46	0.45	0.45	0.48
Cognition								
PRMQ	0.54	0.47	0.29	0.25	0.56	0.38	0.39	0.35
Prospective Memory	0.58	0.44	0.26	0.21	0.50	0.31	0.19	0.16
Retrospective Memory	0.55	0.44	0.22	0.18	0.60	0.40	0.41	0.35
RBANS	0.38	0.46	0.31	0.29	0.39	0.44	0.21	0.19
Attention	0.54	0.55	0.39	0.35	0.56	0.49	0.44	0.39
Delayed Memory	0.58	0.53	0.34	0.27	0.48	0.40	0.18	0.15
Immediate Memory	0.50	0.51	0.37	0.34	0.43	0.42	0.20	0.18
Language	0.48	0.50	0.32	0.30	0.47	0.44	0.26	0.23
Visuospatial	0.48	0.51	0.32	0.30	0.57	0.52	0.43	0.39
Digit Cancellation - Speed	0.44	0.45	0.28	0.25	0.51	0.43	0.36	0.32
	AdaBoost				**MLP**			
	ROC_{auc}	*PR_{auc}*	*Fscore*	*Sens.*	*ROC_{auc}*	*PR_{auc}*	*Fscore*	*Sens.*
Mobility								
Arm Curl	**0.59****	**0.47****	0.36	0.36	0.57	0.46	0.42	0.47
TUG	0.45	0.49	0.46	0.50	0.51	0.57	0.46	0.49
Cognition								
PRMQ	0.51	0.45	0.41	0.43	0.55	0.46	0.44	0.49
Prospective Memory	0.58	0.42	0.35	0.37	0.48	0.39	0.43	0.58
Retrospective Memory	0.56	0.46	0.38	0.37	0.55	0.45	0.43	0.54
RBANS	0.36	0.42	0.32	0.35	0.39	0.46	0.44	0.49
Attention	0.53	0.56	0.44	0.46	0.61	0.62	0.55	0.61
Delayed Memory	0.55	0.48	0.35	0.35	0.53	0.50	0.50	0.61
Immediate Memory	0.51	0.48	0.38	0.45	0.40	0.45	0.38	0.45
Language	0.51	0.49	0.31	0.33	0.45	0.50	0.40	0.45
Visuospatial	0.43	0.48	0.35	0.36	0.59	0.61	0.51	0.56
Digit Cancellation - Speed	0.43	0.43	0.32	0.34	0.44	0.48	0.41	0.47

Regarding reliable change detection, we see that activity-labeled smart home data can actually be used to build quite accurate models when a complete and balanced dataset is available. This is the case for the Arm Curl test change from baseline, which has been seen to be predictable in a quite accurate manner and with a high sensitivity. We have verified in all four models built for this reliable change prediction that the use of smart-home activity data significantly contributes to the detection of such events. Unfortunately, a balanced dataset was not available for all cases. Despite that problem, by applying the SMOTEBoost technique to overcome class imbalance, we were able to demonstrate that consecutive reliable change on mobility measured by TUG test is predictable using smart-home activity labeled data. A McNemar's test with an adjusted p-value has supported this hypothesis, yet we are aware that the model lacks sensitivity to be considered a final model. The use of the wRACOG algorithm has resulted in some models with better prediction characteristics: improved sensitivity and ROC_{auc} and PR_{auc}-s and Fscores have been found in some cases. Changes in consecutive Arm Curl and delayed memory scores have also shown enough statistical significance compared to random classifiers in a McNemar's test to be considered reliably predictable from smart home data. Now that we know that behavioral data can be used to at least automatically assess changes in mobility and memory skills, we can keep collecting more longitudinal data to create better models in the future. This might also result in the discovery of other significant associations. Note that these results were also achieved by using all the behavioral features, whereas a feature-selection process can also help in improving them. Also, we used a kNN algorithm as the wrapper model for the wRACOG approach, but other algorithms can also be considered and might improve the results. Maximization of PR_{auc} and of Fscores could also be tested as stop criteria for the iterative process, possibly leading to different conclusions.

The analysis on the detection of positive and negative changes in the various cognitive and mobility skills has demonstrated the possibility of predicting a decline or an improvement in the mobility of older adults' skills' as measured by the Arm Curl test. This not only confirms the results of the previous RCI analysis, where we have seen that reliable changes in the Arm Curl tests were detectable by smart home activity-labeled data but also adds value to the results suggesting that the direction of the change is also predictable. Literature also supports the idea of the relationship between Arm Curl test scores and ADLs [400]. This can be extremely helpful not only to monitor the progress of a disorder like dementia but also to, for example, closely follow-up subjects who are undergoing rehabilitation. None of the other tests have shown enough evidence of predictability after adjusting the significance level. This highlights the difficulty of the task which might be due to several reasons. On one hand, in this case, we were considering all fluctuations as labels (either positive or negative) without considering their magnitude or without taking into account their reliability (*i.e.* not only reliable changes were considered but all changes). This might have included "noise" in the dataset by considering changes that might have appeared due to reasons other than an actual change in the skills (such as low reliability on tests), making the classification task more difficult. On the other hand, the time-series statistics that we're extracting from the smart home behavior data are not necessarily reflecting a positive or negative change in behavior, but an absolute change.

Summing up, this work has demonstrated the possibility of predicting mobility, cognitive, and mood-related symptoms from unobtrusively collected in-home behavioral data. We believe that the results shown herein are of high relevance, as they suggest the possibility of implementing a system that could bring huge benefits to our aging society which is suffering increasing AD incidence. The models shown in this paper are early models aimed at demonstrating the feasibility of such a system and providing insight into the behavioral features that might be used for this purpose.

Summary

- Common AD symptoms of mobility, cognition and mood decline can be predicted from changes in smart home behavior data.
- Mobility symptoms are the most predictable, followed by cognition and mood.
- Mobility was found to be correlated to outing patterns, daily routine, cooking and eating patterns and global overnight activities. Cognition showed correlations with sleep and overnight patterns, daily routine, mobility and outings. Depression levels were found to be related to mobility, outings and sleep.
- A reliable change in mobility and memory skills can be detected from smart home behavior data, as well as the improvement and decline of mobility skills as measured by the Arm Curl test.

5

A multidomain method for the early detection of disorders in Smart Environments: a cross-case study

Contents

This chapter presents a multidomain methodology for the early detection of disorders based on a cross-case analysis of the two scenarios analysed in this PhD thesis: detection of stress in smart offices (Chapter 3) and diagnosis of Alzheimer's Disease (AD) in smart home environments (Chapter 4). Therefore, it contributes to the State of the Art on the current methodologies for the early detection of disorders presented in Section 2.5. After a brief introduction of the context in Section 5.1, the procedure for the cross-case analysis is presented (see Section 5.2). Based on this analysis, a global methodological approach for the early detection of disorders is abstracted and presented in Section 5.3, and the strengths, weaknesses and limitations of the proposed methodology are exposed. Finally, in Section 5.4, conclusions about the work are drawn.

5.1 Introduction

There are many disorders that besides affecting patients' health status, directly alter users' behaviour, for instance, AD, stress, Attention Deficit Hyperactivity Disorder or Parkinson's Disease. People suffering such a disorder are not aware of their situation, and usually, the problem is identified by relatives or co-workers who notice behavioural shifts. However, when these changes become noticeable, it is often too late, when the disorder has progressed too much and irreversible harm has been caused. Early detection of these disorders allows one to take measures to prevent them completely or to slow down their progress, reducing the consequent health-related damages.

In the last decade, thanks to the emergence of ubiquitous computing paradigm and the interaction of users with many electronic devices, peoples' activities can be unobtrusively monitored. Based on Ambient Intelligence (AmI) technology and this monitoring, a system that detects users' frequent behaviours and shifts has been developed [3] and validated with AD patients. Nonetheless, the possibility to use these systems for the early detection of disorders is currently under research. In this sense, there are different EU projects (*e.g.* i-Prognosis [11] for the early detection of Parkinson's disease, ICT4Life [12] aimed at the early detection of anomalous behaviour related to Alzheimer's, Parkinson's and other dementias or Brainview [13], to early detect signs of autism and attention deficit hyperactivity disorder in babies) aligned with this objective. In turn, the miniaturization of technology has allowed some of the physiological signals to be monitored in a totally transparent way thanks to the development of wearables and E-textiles [416].

In order to early detect this kind of disorders from unobtrusively collected physiological and behavioural data it is first necessary to analyse and understand the underlying relationships between the progress of the disorders and their behavioural and physiological symptoms. Once enough knowledge gained, Artificial Intelligence (AI) and machine learning techniques can be applied to the gathered data to construct robust disease detection and monitoring systems. This would lead to developing ubiquitous and unobtrusive monitoring systems for specific symptoms of disorders, resulting in their early detection.

Many efforts have been done in the recent years to apply AI and machine learning techniques to the detection of several disorders. Nevertheless, most of the disease-related datasets show some common particularities (*e.g.* missing data, repeated-measures data, inter-subject variability, imbalanced data distributions, ...) that make the effort of pattern discovery and prediction model building more complicated than in other research areas. However, there are some other characteristics that make each disease-related dataset a particular problem. The awareness of all these shared characteristics, as well as of the variability among specific cases, would allow to define a robust global methodology able to deal with a big number of disorders.

Current literature in early disorders' detection is mostly based on the predictability analyses of specific diseases, without taking into account the similarities and analogies that can be found among a wide variety of disorders, as, notably, the fact that they provoke behavioural shifts. In this sense, we have spotted a gap in the current literature, where a global methodology for the early detection of

disorders based on unobtrusively collected physiological and/or behavioural data is required. People suffering from many different disorders can gain from such a methodology, with the consequent benefits for the whole society, by means of a reduced number of ill people, reduced health-care costs, reduced dependability and an increase in the global welfare.

In this chapter, we aim at presenting a novel screening methodology for the early detection of disorders based on physiological and behavioural shifts. The proposed methodology, bridges the gap between the traditional (yet currently used) screening methodology for the early detection of disorders and the use of AmI techniques. For this purpose, an AmI-based test-method development guideline which relies on the physiological and/or behavioural changes provoked by the disorders and which is not specific to any disorder is specified.

The suggested screening method is the result of a cross-case analysis of two scenarios: stress in smart offices and AD in smart home environments. Individual validation of the methodology on each case study has given promising results. The thorough analysis of the similarities and differences of the characteristics of each individual case-study has set the path to ask global questions and to abstract the underlying methodology for its use in multiple applications.

The remaining of the chapter is structured as follows. In Section 5.2, the method followed for the cross-case analysis is exposed. The chapter follows in Section 5.3 by explaining the answers obtained for the Research Question (RQ)s. For that purpose, it first provides a summary (Subsection 5.3.1) and a comparative analysis of the two case studies (see Subsections 5.3.2 and 5.3.3), followed by the definition of the steps for a global methodology (Subsection 5.3.4) and the strengths and weaknesses (Subsections 5.3.5 and 5.3.6) of the new approach. Methodological and technological issues and challenges are also reviewed in Subsection 5.3.7. The chapter concludes in Section 5.4 by summarizing the inferred ideas and suggesting new lines for future work.

5.2 Definition of the cross-case analysis procedure

After reviewing the wide variety of disorders that provoke unobtrusively measurable physiological and/or behavioural shifts, we selected two case studies for our research: occupational stress in knowledge workers and AD in older adults living alone. The main reasons to select these scenarios were, first, that both cases are relatively easy to bound spatially: whereas knowledge workers usually work in office environments, older adults often spend most of the time at home. This makes monitoring easier. Second, the monitoring of workplace and home is of particular interest as these sites are the ones where people usually spend most of their time. This facilitates the use of the deployed system for other future applications. Finally, the implication of such disparate scenarios allows to validate a global methodology which is not specific to any disorder or environment an thus, will be useful for a wide variety of disorders.

First, the most important ongoing research for both cases under study was reviewed. This allowed us to spot the main characteristics of health-related datasets, namely, the temporal nature of the data, high presence of missing data, imbalanced data distributions or subject-dependence.

Based on these particularities of health-related data, a first proposal of the algorithmic approach for the use of unobtrusively collected physiological and behavioural data for the stress detection case study was done. The proposed algorithm was aware of the common characteristics of health-related data: temporal nature of the data was taken into account and a sliding window was used to extract time-series statistics describing change over time. Methods to deal with missing data and imbalanced datasets were also considered, as well as inter-subject standardization methods. The methodology was first validated in that particular scenario. The validation was then performed for the second scenario on AD detection where the required changes and refinements were done. Case-specific reports can be found in Chapters 3 and 4.

Based on these two case-studies, a cross-case analysis was performed. The goal of this analysis

was to answer the following RQs:

- RQ 1: Was the proposed approach successful (*i.e.* were statistically significant correlations found following the proposed approach) in both scenarios?
- RQ 2: What are the similarities that can be found between the two cases?
- RQ 3: What differences are there between the two cases?
- RQ 4: Based on RQ2 and RQ3, how would be a global set of steps to follow for the early detection of 'any' disorder defined?
- RQ 5: Which are the strengths of the proposed methodology?
- RQ 6: Which are the weaknesses of the methodology?
- RQ 7: Are there any methodological and/or technological challenges that prevent the immediate widespread deployment and application of the proposed methodology?

After reviewing the results of the two case-studies, the required comparisons were made and similarities and differences between the two cases were discovered, answering the proposed RQs. Consequently, general conclusions about the steps to follow for the detection of any disorder from unobtrusively collected physiological and/or behavioural data were drawn, as well as the strengths, weaknesses and main issues that the methodology implies.

To bound the cross-case study, which could otherwise be extended indefinitely, the criteria to use to perform the analysis has been defined (see Table 5.1). These criteria include disease-related characteristics, as well as algorithmic and result-related characteristics.

5.3 Results of the cross-case analysis

This section exposes the answers given to the RQs defined in the previous section (Section 5.2).

5.3.1 Summary of the case-studies

This section aims at responding to RQ 1. For that purpose, we offer an overview and summary of the two case studies considered for this work by reviewing the comparison criteria listed in Table 5.1.

Stress detection in smart office environments

The first case study considered for this work was the early detection of occupational stress. In this case, office workers were identified as the risk group, and therefore, the office was selected as monitoring environment. Short-term stress (*i.e.* acute stress) was analysed in the study, which is reversible if timely-action is taken. Co-workers were identified as people who usually detect behavioural shifts in the worker suffering from stress, while managers were spotted as responsible for taking-action following specialists' recommendations along with the affected themselves.

Physiological and behavioural data were collected through minimally-obtrusive body-sensors, computer logging, a Kinect 3D sensor and video recordings from a camera. A facial expression recognition algorithm was applied to the video recordings before the extraction of all physiological and behavioural minute-level features. The physiological signals were also preprocessed following their specific requirements. Five time-series statistics were computed by means of a 5-minute length sliding window, and condition-level statistics' averages were computed. Missing data were imputed

with features' mean values where necessary. A manual feature selection analysis was performed, as well as an investigation on the effect of self-reported scores' standardization. A variety of machine learning algorithms were tested for general purposes and SMOTEBoost [373] and RUSBoost [374] algorithms were used for imbalanced classification problems.

Results were validated by means of cross-validation, and significant correlations were found for the target labels. Nonetheless, results differed depending on the label, algorithm or on the use or not of standardization methods. For more details in the work, the reader is referred to Chapter 3.

AD detection in smart homes

The second case study under analysis herein, was dementia due to AD. Older adults were spotted as the risk group for this disease, and home environments were selected as the best scenario for this group to be monitored. Although the speed of progress depends on each individual case, AD evolves relatively slowly over time and is not reversible. Family and relatives are usually the ones noticing behavioural shifts on people suffering it, and following the recommendations of a specialist, they will also be the responsible along with a caretaker to ensure the well-being of the affected as far as possible.

For this case-study, behaviour was monitored by means of environmental sensors embedded in a smart-home environment and interpreted using an Activity Recognition [408] algorithm. Daily behaviour patterns were extracted, and five time-series statistics were computed by means of a week-length sliding window. Data instances with missing self-assessments were removed from the resulting dataset, and manual feature selection and standardization analysis were performed. Several well-known machine-learning algorithms were used to build the prediction models, and SMOTEBoost and wRACOG [410] algorithms were tested to overcome the class-imbalance issue.

Results were validated through cross-validation, and even though significant correlations and predictability were found for some models, overall, the lack of sensitivity was highlighted.

The reader can find the detailed explanation of this work in Chapter 4.

Table 5.1 summarizes the two case studies under analysis in this work from several points of view. First, a comparison of the disorders themselves is given, followed by the methodological comparison of the two case studies. Finally, a comparison of the results obtained in the two cases is shown.

5.3.2 Similarities between the case-studies

After comparing the two case-studies under analysis, we have listed the most notorious common characteristics of both scenarios, and therefore, considered to be generalizable for the whole set of disorders that provoke unobtrusively measurable physiological and/or behavioural changes. Following, these shared characteristics for the two cases are shown classified in terms of disorder, approach and results:

- Disorder

  Naturally, no similarities have been spotted in terms of disorder's characteristics between the two cases. Therefore, we conclude that disorder-related traits depend on each specific case.

- Algorithmic approach

  - The presence of **missing** data. This is a very common problem when data is being collected from multiple sources and non-stop, as any sensor can fail at some point. Fortunately, this was not a big deal in our two case studies. In the stress-detection work, we had low rates of missing data spread across almost all variables, which were replaced by the mean value. Nonetheless, we realized that the variables with the biggest

Table 5.1: Comparison of the two case-studies under analysis in terms of disorders' characteristics, the analysed approach and the results obtained.

Comparison criteria		**Stress**	**AD**
Disorder	*Risk group*	Office workers	Older adults
	Environment	Office	Home
	Time granularity	Short-term	Long-term
	Reversibility	Yes	No
	Stakeholders	Oneself, co-workers, managers, specialist	Oneself, family & relatives, caretaker, specialist
Algorithmic approach	*Ground-truth*	Valence, Arousal and Dominance measured by the SAM test, Stress measured by a visual analog scale, Mental Effort measured by the RSME test and task-load measured by the NasaTLX test.	Mobility measured by the Arm Curl and TUG tests, cognition level measured by the PRMQ, RBANS and Digit Cancel tests and Mood measured by the GDS scale.
	Data-source	Body-sensors, computer, camera, Kinect 3D	Environmental smart-home sensors
	Unobtrusivity	✓	✓
	Raw-data interpretation	Physiological-data event extraction, facial expression recognition	Activity recognition
	Temporal windows	-Data aggregation: 1 minute, -Sliding window: 5 minutes, -Summarizing: length of simulated condition	-Data aggregation: 1 day, -Sliding window: 1 week, -Summarizing: length of between assessment period (6 months)
	Temporal statistics	Variance, skew, kurtosis, autocorrelation, change	Variance, skew, kurtosis, autocorrelation, change
	Missing data	Mean imputation	Remove data instances with missing labels
	Feature selection	Manual, per-type	Manual, per-type
	Standardization	Min-max, Reliable Change Index (RCI), others	RCI
	Machine-learning algorithms	kNN, SVM, LR, NB, AdaBoost, C4.5, MLP, Multiclass logistic	kNN, SVM, LR, RF, AdaBoost, MLP
	Algorithms for imbalanced datasets	SMOTEBoost, RUSBoost	SMOTEBoost, wRACOG
Results	*Validation*	Cross-validation.	Cross-validation.
	Existing correlations	✓	✓
	Feature selection	Computer-use patterns and body posture & movements show the highest predictability	-Mobility prediction: Outing patterns, daily routine, cooking & eating and overnight patterns -Cognition prediction: sleep & overnight patterns, daily routine, mobility & outings -Depression prediction: mobility, outings and sleep for depression prediction
	Models' performance	Label-dependent, but even strong correlations were found	Significant correlations were found, but final RCI detection models were still lacking sensitivity.

missing data amount were the ones extracted from body sensors (yet remaining below the 30%), possibly indicating some failure in the physiological data acquisition during some experiment. A sensor-failure might be critical for some applications and must, therefore, be detected on time. In the second case study, we might have had some missing behavioural data, but as our starting point was a sensor log file where a sensor failure would be translated into a missing entry line, we could not identify them. The result of such a failure will be a misinterpretation of the performed activities by the activity recognition algorithm, and thus, a misinterpretation of the cognitive and mobility skills of the user. Therefore, it would be interesting to implement a sensor-failure detection system before the data acquisition process. Nevertheless, we did have some data instances with missing self-assessment scores for this case study, which were removed from the model building step.

- Importance of **feature selection**: Naturally, not all measurements and extracted features contribute equally to the prediction of the objective labels. The feature analysis performed in the two case studies confirmed this statement, as some groups of features were more correlated to stress and mental workload assessments in the first case, as were for the mobility, cognition and mood assessments in the second case. This highlights the importance of the feature selection analysis. In our case-studies, we performed this analysis by manually separating features of interest depending on their nature. This has contributed to gain insight and knowledge of each specific disorder. Nonetheless, the use of automatic feature selection algorithms might help in selecting the best subset of features for the prediction of the disorder under study. A reduced set of selected features can benefit the final system in many aspects: on one hand, this can reduce the number of sensors or monitoring devices to be deployed in the selected environment, gaining in aspects of unobtrusiveness and cost. On the other hand, the final prediction models can see their accuracy enhanced, as redundant measurements only contribute to overfit the developed models reducing their effectiveness for new cases. Finally, the resulting models will also be computationally less expensive, as they will pose fewer problems of storage and on-line processing, and will require less time to be built and used. Related to this, the importance of the sensor placement must be remarked. No feature selection will be efficient enough if the sensors used to collect the data are not placed in strategic locations and therefore are not collecting useful data.
- **Imbalanced** datasets: In the two cases under study, we were facing a class-imbalance problem. This is something usual in health-related machine learning problems, as considerable health changes are rare events. Nevertheless, they are most of the time the events of interest. Traditional machine learning algorithms tend to give biased results towards the majority class when they are trained with imbalanced data, resulting in high accuracies but very low or null sensitivity towards the events of interest. This results in the need to use specific algorithms for this type of data. In our case studies, we made use of three different algorithms to overcome class imbalance: SMOTEBoost [373] based on Synthetic Minority Over-sampling Technique (SMOTE), RUSBoost [374] based on Random Undersampling and wRACOG [410], a Gibbs sampling-based oversampling technique. Whereas SMOTEBoost and RUSBoost were used for the first case study, the reduced sample size in the second case study was not the most suitable for the application of an undersampling method such as RUSBoost, so the wRACOG algorithm was tested instead. Overall, the sensitivity of the models was improved by using these algorithms.
- **Inter-subject variability**: The physiological and behavioural data required by the system under proposal might be highly subject dependent. And not only the sensor-collected data but also the self-reported questionnaires that might be used as ground truth are. The data processing approach proposed herein considers temporal shifts on the gathered

sensor-data instead of actual data values, which might help in reducing the inter-subject variability that might bias the models. Furthermore, we have been standardizing the different objective labels in both cases in order to reduce this type of variability. In both cases but, especially in the stress-detection case-study, we have noticed the benefits of performing label standardization. Although in this case it is has not been prioritized, it would also have been interesting to see an analysis on the effect of the inter-subject variability by performing a leave one subject out cross validation (LOSOCV). It is something to consider for future applications.

- Results
  - The **feasibility** of the proposed approach: After comparing the two case studies, the hypothesis on the feasibility of such a system has been affirmed. In both scenarios, physiological and/or behavioural measurements that could be performed in a transparent way for the user were spotted, as well as significant correlations between the unobtrusively collected and processed data with the corresponding objective or self-reported labels found.
  - **Low-cost** systems: The hardware used in both case studies have been sensors and devices of relatively low cost and wide availability. For the first scenario, a personal computer, a Kinect 3D sensor, a webcam and Electrocardiogram (ECG) and Electrodermal Activity (EDA) body sensors were used whereas for the second, only proximity/presence sensors and monitoring of some specific home appliances and furniture were used. Despite the simplicity of the selected monitoring hardware, satisfactory results have been obtained in both cases, as this has been compensated by means of the algorithmic approach.
  - **Validation from a specialist** is always needed: After validating our models by means of a cross-validation approach, we have seen that a prediction accuracy of 100% is too far for the method to be considered a definitive diagnostic test. Therefore, the proposed approach must be considered as an early alarming system so that the disease specialist can be aware and can evaluate the situation to take the necessary measures. Such an automatic disease detection system should never be considered as absolute truth or definite diagnostic test.

5.3.3 Differences between the case-studies

After comparing the two individual case studies under analysis, we spotted some differences that may be generalizable to the global set of disorders that provoke behavioural shifts. The following list summarizes the terms in which the proposed methodology might vary for its use in a general manner, and answers, therefore, to the RQ3:

- Disorder
  - **The risk group**: Each disorder under analysis will have one or more specific risk groups that will differ between disorders. In our first case study, office workers were in the spotlight, while older adults living alone were in the second case study. This risk group will have to be specified by a disease specialist and will have to be considered to decide the details of the system deployment, such as the best environment for the purpose, or the most suitable sensor elements to use.
  - **The environment**: The environment in which the system should be deployed may vary depending on the disorder that is aimed at detecting. The selected environment for the two scenarios under analysis in the current work differed: whereas a smart office was

used for early stress detection, a smart home was the preferred choice for people at risk of dementia. These choices meet the requirements just mentioned: knowledge workers spend most of their time in their office, as older adults with a slight cognition decline do at their home. Similarly, occupational stress is caused and therefore most manifested in office environments, while AD symptoms are manifested at home. Some interesting AD symptoms would also be acquired from the monitoring of the external trips of the older adults (*e.g.* wandering) but as it is not a limited space, the monitoring of the behaviour patterns should be performed otherwise (*e.g.* by means of smartphones).

- **Time granularity**: Each disorder has its characteristic rate of progress. Although this may vary from subject to subject, it will remain in the order of magnitude corresponding to the disease. Consequently, whereas progress of some disorders need to be monitored in terms of seconds, for others, it will suffice to acquire data daily or even monthly. Not only that but also, this time granularity will affect on the sliding window-length to select for the algorithmic approach.
- The **reversibility** of the disease: The early detection of a disorder will not bring the same consequences in all cases. While some disorders can only be prevented from worsening by the provision of the corresponding treatment, others are treatable and reversible, so an early disorder can completely cure the disorder. In our case studies, the latter was the case for stress. Stress can be cured and prevented from making irreversible damages when it is detected and treated early. On the contrary, AD can not yet be cured, but its symptoms can be alleviated and the progress slowed down. Whether the disease is reversible or not, the benefits of the early detection of the disorder must be clear before starting with the process presented herein.
- **Stakeholders**: People who take part in the development of such a methodology will slightly differ between cases. Whereas there will always be a patient or a target person to be monitored and a disease-specialist who will supervise the whole process, people who interact with the target person in the selected scenario and consequently with the monitoring process vary among cases. In our first case study, the patients interact mostly with their co-workers and these are the ones that best see the alarming behavioural shifts suffered by the patient. In addition to the workers themselves, their managers must also be aware of the system and situation, to take the required measures. In the second scenario, family and relatives are the ones who can best evaluate the daily functioning of the patients, and both family or an external caretaker may appear in the scene as people in charge of the wellbeing of the patient.

▶ Algorithmic approach

- The **ground truth**: In order to train supervised prediction models, data to be used as ground truth must be collected. This ground truth must reliably measure the symptoms of the specific disease under study. In our two case studies we were using well known self-reported tests for this purpose, which were designed specifically to detect symptoms related to each disease: whereas stress and mental workload levels were measured by the Self Assessment Manikin (SAM) [377], Rating Scale Mental Effort (RSME) [378], NASA Task Load Index (NasaTLX) [379] and a visual analog scale, AD symptoms were measured by means of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) [397], the Prospective and Retrospective Memory Questionnaire (PRMQ) [398], a Digit Cancel test, Timed Up and Go (TUG) [399], Arm Curl [400] and the Geriatric Depression Scale - Short Form (GDS-15) [401].
- The **source** of the collected data: Disease-related data can be collected from a wide variety of sensors, starting from wearable devices that continuously collect physiological

data to sensors embedded in a multitude of different points in the environment, which can be related to a great collection of diverse activities. Each one of the signals must be processed accordingly to create useful data for further use, *i.e.* applying signal-specific feature extraction and activity-recognition algorithms. In the case-studies under analysis, we have seen that unobtrusive measurement of behaviour was feasible in both cases, but this was not true for the monitoring of physiological signals. The physiological signals of interest in AD detection are currently measured by means of expensive tests (*e.g.* brain imaging) or invasive methods (*e.g.* Cerebrospinal Fluid (CSF) extraction) in health centers, or by obtrusive (*e.g.* Electroencephalogram (EEG) cap) and therefore not useful technologies for the daily use. On the contrary, some signals such as ECGs or EDAs can already be unobtrusively measured thanks to wearable technology. Therefore, the types of signals to be used may vary among different cases, and an analysis on which ones to select must be performed for every new case.

- **Interpretation** of raw data: Data coming from different sources must be preprocessed and treated differently. It is very unlikely any collected signal to be usable in its raw form. All physiological and behavioural data used in the two cases under study were first interpreted and represented in a feature form. The way to make this interpretation or translation to a feature form differ between data types. Physiological signals must usually be preprocessed to extract events of interest (*e.g.* RR intervals from ECGs or EDRs from EDAs) so that they can be derived in new variables and aggregated with the desired frequency. Video recordings might require the use of specific processing algorithms depending on our objective, such as FaceReader [380], for facial expression extraction. Sensors embedded in the environment usually require the use of activity recognition algorithms adapted for each specific case, in order to translate sensor activation/deactivation events on human interpretable behaviour or activity patterns. These can then be used to derive new information or features of interest or can be aggregated per time periods. Other computer or smartphone based behaviour measurements might luckily be given in a more interpretable way. Otherwise, the specific algorithms to convert raw data to useful data for each case must be developed.
- **Temporal-windows'** length: As previously mentioned, the progress rate of the disorders will affect the length of the temporal windows used for the computation of change statistics. Regarding our two case studies, we could say that stress (or acute stress) is a disorder that varies in short periods of time (even during the day) whereas AD is a disorder that progresses over the years. Usually, the time-varying nature of the data and of the disorders is ignored, and approaches for the prediction of disorders are built using only instantaneous snapshots of the data. Instead, we considered and applied in the two scenarios a methodology that is aware of the temporal nature of this data and which exploits statistics describing behavioural or physiological shifts over time instead of single data snapshots. The time windows to compute these statistics were chosen based on the disorders' progression pace and a coherent empirical analysis: the window-lengths for the first case were of the order of minutes, whereas for the second case were of the order of days, weeks and months. The AD methodology has been found to be effective in the two scenarios, and as all disorders share this characteristic, it can be considered a good approach for generalization.
- Performance of **machine-learning algorithms**: A wide variety of supervised machine-learning algorithms have been developed in recent years [388]. Literature shows that it is not possible to select a single best algorithm for every case: their performance varies depending on the application. The same conclusion can be drawn from our cross-case analysis. Whereas in the first case-study even a Naïve Bayes was doing a good job, the second case-study showed much better results by means of linear and radial Support

Vector Regression (SVR), Random Forests (RF) or AdaBoost. This means that for each new specific case, a preliminary analysis on the performance of the different algorithms before fitting the final models must be performed. A good starting point would be the application of Support Vector Machines (SVM) or AdaBoost algorithms as they are robust and often one of the best-performing methods across different applications areas. However, care must be taken to avoid overfitting especially when AdaBoost is being used.

- Results
  - **Accuracies of the final models**: The detection and monitoring of each specific disease is by itself an individual problem and an individual research area. Each disorder has its own characteristics and poses different levels of difficulties on their diagnosis even for the traditional medicine and for the best specialists. Therefore, we can not expect the methodology proposed herein to perform equally for all cases. From the comparison of the two case-studies we might be tempted to affirm that stress detection is more feasible with the proposed method: but this might not be true. We have to acknowledge that in order to train our models, we have been considering some variables as ground truth. In the first case of stress detection, we had available a set of self-reported questionnaire ratings, as well as some objective variables indicating the conditions that the subjects were undergoing every moment. In the second case, we only had available a set of self-reported questionnaire scores, which were being reported by a group of users at risk of or even suffering some cognitive decline. The considerable difference in the prediction abilities might not be due to the lack of efficiency of the proposed model or the behavioural data of the users, but due to the unreliability or subjectivity of the ground truth. For the application of this methodology in a new case study, it would be interesting to use, as far as possible, a reliable and objectively measured variable (*e.g.* resulting from a specialist or some traditional diagnostic method) as ground truth to build efficient and well-trained models.

5.3.4 Methodology for the early detection of disorders

Based on the answers given to RQ2 and RQ3, we have abstracted the underlying traits of the methodology and we have defined a generalizable path to follow towards the detection of a new disorder from physiological and/or behavioural shifts. This section answers the RQ4, by defining the steps to follow and the things to take into account for the application of our methodology to a new case-study.

The proposed methodology for the early detection of the disorders is a four-step process, which has to be applied individually for every disorder under research: First, an important step where a thorough analysis of the disorder under research will be performed is proposed, which will lead to the definition of the iterative process. The second step is about the setting-up of the environment, where knowledge from the first analysis will be used to define the monitoring environment. A learning step where data about the disorder will be gathered and where prediction models for the next step will be built follows. Finally, in the screening step, the developed models will be used for suspicious changes' detection purposes. The process can be seen in Figure 5.1.

1. **Perform preliminary analysis**

   Before starting with anything, it will be necessary to get to know the disease to be detected more closely. For this purpose, a first analysis of the disease, the affected people and the environment in which it develops will be convenient. As stated in the World Health Organization's (WHO) screening criteria [350], we will have to make sure that an early detection of the disorder can bring benefits to the affected person. This is an essential requirement to begin the process,

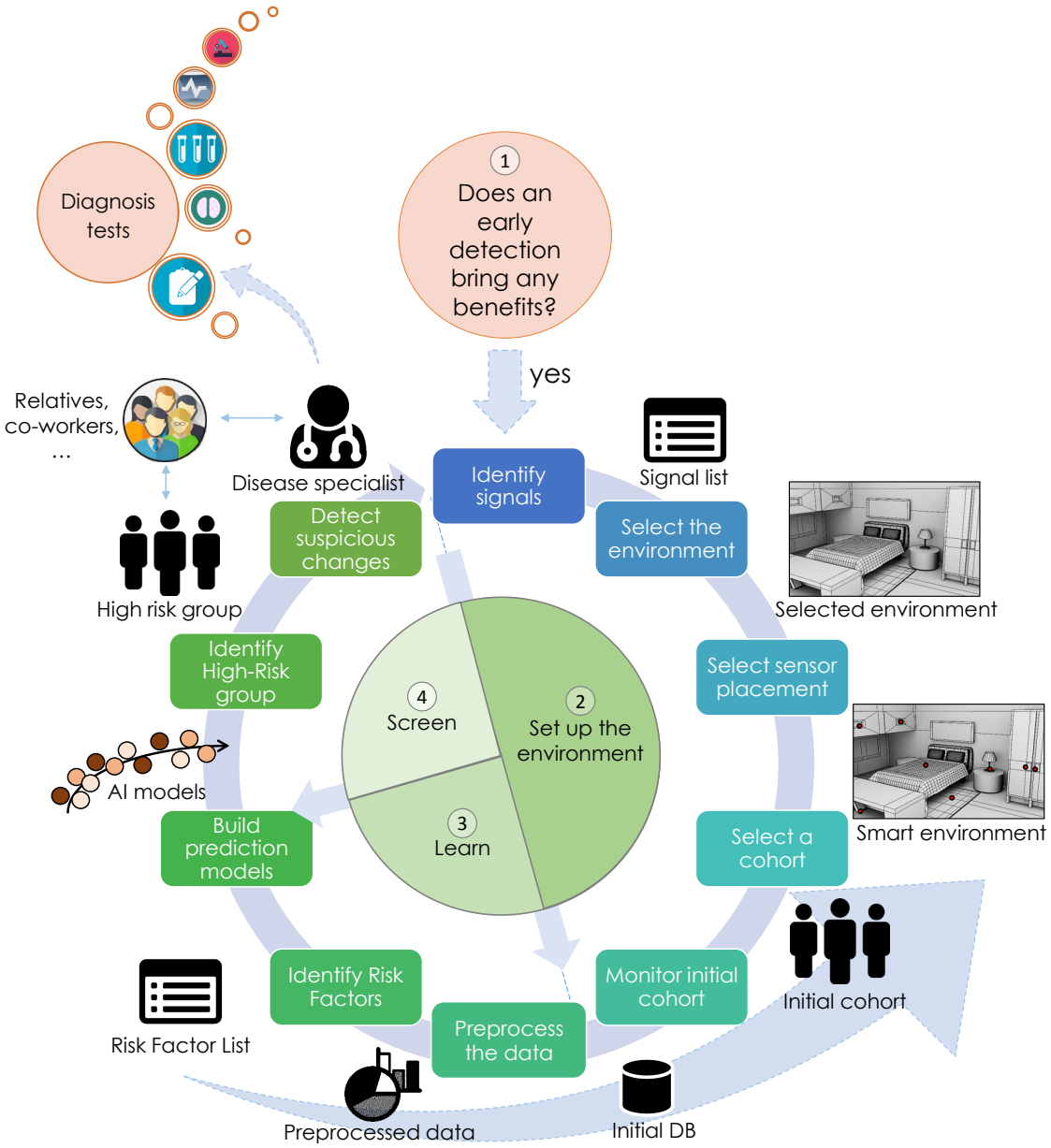


Figure 5.1: Workflow for the early detection of disorders based on physiological and/or behavioural shifts

and all candidate disorders with no existing effective symptomatic treatment or definitive cure will be discarded. Once assured that an early detection would improve patients' quality of life, we will think about the usual pace of progress of the disease, the people who might be at risk of suffering it and how and where these people spend their daily life. Finally, we will identify the stakeholders of the process and define their role on it. Usually, the participants of the methodology will be the following:

- The doctor or **disease-specialist**: Depending on the disease willing to detect, a general-purpose doctor or a disease specialist will be the one in charge of selecting the cohort to monitor and to evaluate the results given by the implemented approach and feedback from people around the patient.
- The **people at risk**: Will be the protagonists of the story. They will have to collaborate with the disease-specialist and engineers as much as possible and will have to accept the ethical issues and risks that being continuously in a monitored environment poses. In return, their progress will be closely monitored and the possible onset of the disease will be detected in time.
- The **engineers** and data scientists: Will be in charge of designing and implementing the smart environment and of building the prediction models.
- Family, **relatives, co-workers**, friends, others: People with whom the patient shares his/her daily life. They will be in charge of collaborating with the disease specialist to provide complementary information to the one offered by the monitoring system and of taking care of the affected person when necessary.

2. **Set-up the Environment**

   Second, it will be necessary to set-up the monitoring environment. This step is crucial for the success of the methodology, as the gathered data and conclusions drawn from the data will totally depend on the decisions that are taken in this step. Therefore, the execution of this step is worth time to ensure that is being performed correctly. To gain reliability, to save time and to improve the final results, the engineers and data scientists will perform the steps regarding this section with the assistance or support of a specialist in the disease under research or a physician.

   - **Identify signals to be measured:** Each disorder will require different signals to be measured, depending on its symptoms. For this purpose, it is first necessary to review the literature about the disorder and discuss with the disease-specialist to identify those symptoms that are convenient to be measured in each case, both those coming from the physiological sources as well as those coming from the behavioural ones. Moreover, it will be necessary to decide and identify if any psychological assessment test should be periodically performed.
   - **Select the environment:** Depending on the disease, the symptoms to be measured, and the cohort to be monitored, an environment or another will be more convenient. The selection of the adequate environment is crucial for the system to be useful and effective. No measurement of interest can be acquired from a place where the subject is spending almost no time. The selected environment for each specific case must be a spot where the subject at risk is spending a large part of his time, (almost) daily, where the physiological or behavioural symptoms are manifested and which can be easily bounded.
   - **Select sensor placement:** Research has shown that it is of great importance to cautiously select the placement of the sensors in smart environments in order to gather data of interest [417]. The bad selection of sensor placements could lead us to miss important physiological and/or behavioural shifts that might be highly correlated with the progression of the disease.

- **Select a cohort:** The disease specialist or doctor can help in selecting a cohort of patients that can be at risk of developing the disease under research regarding their age, gender, work position, etc. Cohort selection criteria for many disorders is already defined for classic screening processes [346]. The doctor may take advantage of this information at the same time that he can make use of case finding. Moreover, the inclusion of control subjects (people with a minimum probability of suffering the disorder under research) in the screening process may be required for the statistical validation of the following steps, *i.e.* discovery of physiological and/or behavioral biomarkers and validation of the detection models. A screening consent and authorization form will have to be signed by the selected cohort before starting the monitoring process.
- **Monitor cohort of interest:** The designed smart environment must be implemented for each one of the selected subjects to collect data automatically. In addition to physiological and/or behavioural data, information to use as ground truth will also be collected. This ground truth must come from a reliable and preferably objective assessment method, which will be selected by the disease-specialist. Enough ground truth data will have to be collected until the reliability of the prediction models is assured. Assessment methods other than self-reported test can also be used, *e.g.* reliable diagnostic tests based on medical checkups or the diagnosis given by the disease specialist.

3. **Learn**

   The execution of the first step in this methodology will give as result a preliminary database of a specific disorder consisting of behavioural data collected from subjects at risk during an established period of time. This database will be exploited in this second step to identify the most useful measurements for the prediction of the disorder and to create disease-progression monitoring models by AI and machine learning algorithms.

   - **Data preprocessing:** The suggested data preprocessing approach for the computation of physiological and/or behavioural and physiological time series' statistics can be seen in detail in Figure 5.2. As can be seen from the figure, the first step of the data preprocessing is the raw data interpretation. Depending on the nature of the collected data, one algorithm or another will be necessary, *e.g.* QRS detection algorithms for ECGs, skin conductance response events' detection algorithm for EDA, Activity Recognition algorithms for environmental sensors or facial expression recognition algorithms for facial camera recordings. Coming up next, is the feature extraction step where data will be aggregated to make it significant. A time-window will be selected for data aggregation, and features like the number of repetitions of an action or event (*e.g.* number of outings per day, number of keyboard events per hour, number of eye-blinks per minute, etc.), or time spent performing an action (*e.g.* time spent sleeping per day, time spent smiling per minute, etc.) will be computed. Another type of aggregated data features can also be considered, and ideally, an analysis on the effect of the data aggregation window-length will be performed to search for optimal results. Then, a missing data imputation will be performed if necessary. Depending on the approach followed for this purpose, this step can be performed earlier or later. In this work, we have been using mean data imputation to replace the few missing behavioural and physiological data instances whereas we removed those instances where self-reported data was missing. Other approaches like the predictive mean matching [418] might give better results when more data is missing, and interpolation can be used to estimate missing assessment data. Once all feature-level data is available, the Clinical Assessment using Activity Behaviour (CAAB) algorithm [114] will be used to extract the five time-series statistics from physiological and/or behavioural data. In short, the CAAB algorithm consists of applying the following

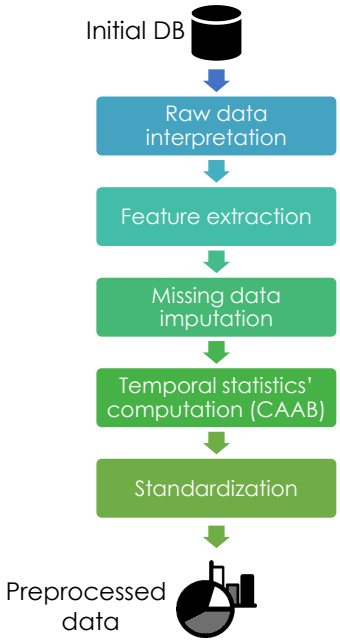


Figure 5.2: Detailed data preprocessing approach for the methodology of proposal

processing steps to the aggregated behavioural or physiological data: 1) Take each subject's aggregated behavioural and/or physiological data for the desired analysis period, 2) By means of a sliding window, apply a log transform and a Gaussian detrending to each time-series (behavioural/physiological variable), and 3) Compute five summarizing time-series statistics (variance, skewness, kurtosis, autocorrelation, and change) for each physiological and/or behavioural feature in this period 4) Either compute the average of each time-series statistic for the desired period and use the set of averages for the final predictions or use the CAAB statistics directly without computing their average. In the latter case, the use of a non-overlapping sliding window will be required so as to avoid the use of redundant data. Moreover, the length of the sliding-window will have to be selected, and ideally, an analysis or at least an empirical search for the best option will be performed. Finally, the self-reported labels will be standardized to reduce the inter-subject variability that might bias models' results. For this purpose, Reliable Change Index [385] can be computed, or in the absence of information to compute these values, a min-max normalization based on each subjects' baseline data can also work.

- **Identify risk factors:** Correlation analysis performed with the models built using activity-specific or type-specific features allows to find out the physiological and/or behavioural features that might have the best predictive power for the health-assessment of interest. This might help giving an insight of the signals to be measured depending on the objective of the system under development, as well as of the features to be extracted from the collected data. Automatic feature selection algorithms can also help in this process. Some examples include Sequential Forward Selection (SFS) [6, 332], Sequential Backward Selection (SBS) [79] and Ridge Regression-based regularization [419]. Depending on the results achieved in this step, we might also want to refine the implemented smart environment to improve sensor selection and placement or to reduce the number of sensors.
- **Model disease with machine learning:** Once the best subset of signals and features to be used for the prediction models have been selected, machine-learning based prediction models can be built. It is interesting to test different algorithms based on different

approaches, as there is no best algorithm for every case and their accuracy may really vary from case to case. Usually, Naïve Bayes, SVM or AdaBoost based models are a good starting point. In case we are facing an imbalanced data problem, algorithms that aim at boosting models' sensitivity (*e.g.* SMOTEBoost, RUSBoost or wRACOG) must be used.

4. **Screen**

   The knowledge gained and the prediction models developed in the previous step will allow to monitor people over time. This can be used to detect anomalies that might indicate the onset of the disease or to monitor behaviours and/or physiological signals related to a health improvement or decline. The progress being monitored by the system should be closely followed by the disease-specialist or a physician to take the necessary decisions and perform additional diagnostic tests when an alarm is triggered or at any given time when the specialists considers appropriate.

   - **Identify people at high risk:** Based on physicians or specialist's recommendations, the people at risk to develop the selected disorder will be identified and the system designed in the previous steps will be deployed in the specified environment, where the selected cohort can be monitored in their natural settings. All the participants must give consent and authorization to be monitored in order to get involved in the process.
   - **Monitor high-risk group:** Once the system deployed, the subjects under risk can be longitudinally monitored in their natural settings as specified in the first step. The disease-progression models that have been created in the second step will be applied to the data being gathered from the smart environments, to see the progress that the disorder is making on the subjects of interest and to early detect the onset of the disease in case it is given. This progress must be closely followed by a specialist who will also require feedback from the people around the patient. When an alarm is triggered by the behaviour monitoring system, the specialist will perform the required diagnostic tests to confirm or discard the presence of the disease.

Note that the proposed process can iterate in three ways. On one hand, once the first iteration finished, the disease specialist can ask to start the process over (back to step 2), by applying the requirements that might have appeared during the process. On the other hand, the data being collected in the last step can be used to feed the initial database resulting from step 2, so that new analyses can be performed iteratively, and more reliable knowledge and detection models can be derived from it (back to step 3). Finally, new high-risk subjects can be invited continuously and the monitoring process can be launched for them, repeatedly performing the screening process (keep iterating in step 4).

5.3.5 Strengths of the new methodology

One of the main strengths of the methodology being presented herein is its independence from any specific disease. The steps that have been defined should be repeated for any disease that provokes physiological and/or behavioural shifts. Moreover, the consideration of AmI as a basis for monitoring makes the systems generated have two very important advantages. On one hand, the resulting system will be transparent to the user, improving peoples' acceptance towards the system, causing them to forget about it and consequently gathering more ecologically valid data. On the other hand, the system can be deployed with a relatively low cost. Usually, smart environments can be implemented by means of inexpensive sensors or even by means of already existing devices such as smartphones or personal computers. This is a breakthrough compared to the use of more traditional diagnostic methods, which are often costly and require patients to move to a specialized center. Linked to

this, the proposed methodology will result in systems that are able to monitor patients non-stop, in much of their daily lives. This will result in earlier diagnoses compared to traditional periodical checkings. Moreover, if the deployed system would be verified to be reliable enough, all routine checks for specific disorders could be avoided and be only applied when anomalies are detected, with the resulting health-care savings.

Finally, the methodology proposed in this work is completely aligned with the path being followed by the current technologies: internet of things, smart environments and big data are already the order of the day. In the current information age, it is completely natural, useful and realistic the use of the information being collected non-stop to prevent health-related issues and their consequences.

5.3.6 Weaknesses of the new methodology

As a specific case of a screening methodology, the proposed approach inherits some of the weaknesses of screening methods. First, the implementation of the designed smart environment, as well as the use of medical resources from people who will not develop the disease under research, involves an 'unnecessary' cost. Nevertheless, our approach minimizes the cost of the screening procedure by using inexpensive sensors or existing equipment as test method instead of the usual expensive medical procedures. Second, false positives can provoke the patient to suffer from stress and anxiety, as well as unnecessary investigations and treatments to be initiated. Finally, the false negatives might provide a false sense of security which might consequently lead to a delayed diagnosis. To avoid this, we have focused on the maximization of models' sensitivity using specific algorithms for the purpose, therefore reducing as much as possible the number of false negatives.

Furthermore, the effectiveness of a screening methodology always relies on the selected high-risk group, and therefore, the selection of the wrong cohort will lead to a useless effort. The lesser the people who convert in the selected cohort, the less information about the diseases' physiological and behavioural responses will be available and the worse prediction models will be built. Moreover, more affected people will be out of the monitoring process and therefore too late diagnosed. To minimize this problem, our methodology specifies that the groups at high-risk of disease will be selected by a disease-specialist.

Depending on the objective disease, we might also need several iterations to build reliable models, which might take a long time, especially, if the smart environment has to be completely redefined.

Finally, regarding the algorithmic approach, the selection of the best window-lengths for the computation of the temporal statistics might be a bit tricky: ideally, a combination of the best data aggregation window-length, sliding window-length and summarizing window length should be selected, which might require to perform many analyses before the system can be implemented.

5.3.7 Issues and challenges

The development of a system based on the methodology being proposed herein involves a number of technical and methodological difficulties. Although the technological advancement of the last years has greatly facilitated the development of such a system, there are still some issues that must be solved to enable the widespread use of smart environments and AI technologies for monitoring purposes. Issues related to legislation, ethics or privacy policy must also be solved for the applicability of the proposed system. Below, we list the main technical difficulties that must be taken into account when building a system based on the methodology proposed herein, as well as non-technical issues that must be solved.

- **Data collection and quality**

  One of the first steps of such a system is the *data collection*, which has to be meticulously carried out in order to ensure its quality. For data to be of high-quality, it must be "accurate,

complete, relevant, timely, sufficiently detailed, appropriately represented, and must retain sufficient contextual information to support decision making" [322]. Some of the current physiological monitoring devices [323, 324] allow high-quality data acquisition. However, there are a few things to keep in mind to assure this.

Regarding physiological measurements, the correct placement of electrodes or monitoring devices must be guaranteed. In order to avoid ambiguities, standards for the correct measurement of physiological data have been defined and are internationally used. It is the case of the International 10-20 EEG System [325] or the standard 12-lead ECG. For sensors that do not have any standards defined, trials must be done to check the best placement. It has been affirmed that different body placement of the sensors result in different signal patterns and classification accuracies [326]. Sensor placement is also crucial to the quality of the behavioural signals' recording in smart environments [327].

Moreover, the sampling frequency of the data must also be adequate to the signal being collected, in order to establish a compromise between the amount of data to be treated and the quality obtained from them. This sampling frequency will vary for each type of signal. Khusainov et al. [328] affirmed that to monitor Activities of Daily Living (ADL)s a sampling frequency of 20 Hz is sufficient, while audio, speech and other biomedical signals must be sampled with a higher frequency of up to 40KHz.

Furthermore, signals are easily corrupted by instrumentation noise, random noise, electric and magnetic noise, etc., as well as by poor electrode-skin contact and body movements [328], resulting in noisy and artifact containing data. Signal processing techniques are needed to remove all these undesired effects from the signals. Noise can be filtered by means of several filters, like Kalman filters [420], Butterworth low-pass filters [421], Median filters [422], Wiener filters [328], Wavelet Decomposition [328,420], etc. The selection of the best filter in each case depends on the nature of the signal, the features to be extracted, and on the type of noise [328]. Power line interference can be removed by means of a notch filter [423]. For artifact removal, algorithms such as the independent component analysis (ICA) can be used [424], as well as additional sensors to detect unwanted movement artifacts (*e.g.* pressure sensors in chairs [329]) and reject the corresponding recordings.

- **Integration of multimodal data**

The data collected in a health monitoring system based on multimodal measurements will come from a wide variety of sensors and devices, and the *integration* of all these data still poses some challenges.

Different acquisition systems, rely on different physical phenomena and, thereby, the resulting data is represented in different physical units. Furthermore, they do not offer the same time and space resolutions, and what is more, datasets do not have the same dimensions. Even so, these difficulties can be overcome working with features extracted from the data instead of working with the raw data by itself, as proposed herein.

Another important issue when working with data coming from several sources is the presence of missing data. This makes impossible to compare multimodal data at the same time points and to ensure synchrony. However, possibilistic data fusion frameworks could allow to overcome this issue [425, 426, 427] as well as a data imputation method as used in our case studies.

Moreover, inconsistency in the final decisions may arise from data of different sources. This might be solved with a voting system [428], or by building single unified models from multimodal data as we did in our approach. In fact, in which step to integrate the multimodal data in the detection system is another issue. Data can be processed separately, and merged in the final decision step; they can be sequentially processed and merged, adding new data to constrain the prior solution or they might be fused from the very beginning using a few variables from

each modality, multivariate features or minimally reduced raw data. This decision depends on the nature of data to be fused.

For more details about the current challenges in data fusion, the reader is referred to [429] and [430].

- **"Big data" issues**

  One of the biggest problems of such a system is probably the huge amount of data being collected non-stop. This problem might be avoided with on-board signal processing algorithms, and thus, avoiding the transmission and storage of big amounts of data. Nevertheless, this affects power consumption and battery lifespan of the individual devices, as well as their storage requirements, and complexity of algorithms, so this would be a choice to make according to the equipment availability. Moreover, in our approach storage of data is desired to iteratively build more accurate prediction models. *Dimensionality reduction* of the data by means of a manual feature selection as we did herein might help, to only use the pertinent features and ease transmission, processing and storage. Automatic feature selection algorithms (*e.g.* SFS [6,332], SBS [79], Correlation Based Feature Selection (CBFS) [85] or Genetic Algorithms [54]. For an empirical study on feature selection methods the reader is referred to [431]) can also work for dimensionality reduction, yet not that much to reduce the number of required sensors. Algorithms such as Principal Component Analysis (PCA) [334] might also help in dimensionality reduction but our case-study analyses showed that this might not be suitable for our approach, as prediction accuracy was highly reduced, especially in the stress detection scenario. For the successful use of these algorithms in behavioural data, data segmentation may also be a critical issue: an incorrect selection of the segmentation window can lead to incorrectly infer ADLs. The use of sliding window techniques as the one in our approach has been recommended to avoid this issue [328].

- **Unobtrusiveness and noninvasiveness for users' acceptance**

  The unobtrusiveness and noninvasiveness of biomedical measuring devices are key factors on acceptance and satisfaction from the subjects [432].

  Nowadays, technology for making this monitoring system ubiquitous and completely transparent to the user exists and that is what we have prioritized in our approach. Smart environments allow to record behavioural data without disturbing the user by means of sensors integrated both in the ambient and in objects (smart objects), for example, in ceilings [101], chairs [88,100,329], or doors [100], and with simple monitoring software installed in computers allowing to sense the interaction activity with the computer, *i.e.* computer exposure [5], keyboard and mouse dynamics [46], etc. Physiological monitoring has been usually much more obtrusive and, thereby, a bigger issue for this kind of applications. Nonetheless, the unobtrusive measurement of physiological signals is also becoming increasingly easy thanks to the development of wearable devices and physiological sensors integrated into devices and textiles (E-textiles) of everyday use. Among the examples, it is possible to find a computer mouse with photo-plethysmographic surfaces that allows to measure RR intervals enough accurately for computing Heart Rate Variability (HRV) parameters [433], a belt for sensing breath [434], a shirt for Electromyogram (EMG) sensing [435] or a wearable ECG recorder with acceleration sensors [29] .

  Smart wearable systems are becoming more and more present. In fact, they have already been considered for the monitoring of several diseases' progress like cardiovascular [436], renal [437] or respiratory diseases [438], diabetes [439], and even cancer [440]. They can be used to monitor patients 24h a day, recording physiological or behavioural data, with sensors integrated into jewelry, wrist watches [441], armbands [442], shoes [443], embedded in clothes [444] and implanted in vivo [445, 446]. Unfortunately, not all the physiological data can yet be unobtrusively acquired with wearable devices [447], *e.g.* EEG requires electrodes or an electrode

cap to be worn, making its use in daily-life unrealistic. We refer the reader to the recent review of Chan et al. [416] for a more detailed information about current wearable technology's state.

Nowadays, smartphones also allow to acquire big amounts of data without the user being aware of it. Furthermore, apart from monitoring the characteristic behavioural features of a smartphone (*e.g.* the number of sent SMS), these devices can be used to unobtrusively monitor other physiological or behavioural features as, for example, speech features [448] or traveled distance [6].

- **Interoperability**

  The system being proposed herein, apart from detecting specific disorders, could be useful for a global health status checking, and data being collected and treated can be practical for many other purposes. These secondary objectives might imply exchanging data between other medical devices, or with experts of the health area. Consequently, it may be interesting to consider the existing standards for physiological and medical data coding and storage during the system's design process.

  Many standards have been defined in order to overcome interoperability problems and improve the communication and data exchange between different devices all around the world. For example, the European Data Format (EDF), which has already an extended version (EDF+), was originally created for EEG and PSG recordings, but the new version also allows to store information of ECG, EMG, and Evoked Potential data, as well as annotations [449, 450]. The General Data Format (GDF) for Biomedical Signals [451] derived from EDF, aiming at satisfying the needs of all the biomedical research community. An ISO standard has also been defined to assign medical waveforms' description rules in order to ensure interoperability between devices. This standard is known as Medical waveform Format Encoding Rules (MFER). As it is a general specification, it is compatible with other standards. The Standard Communication Protocol for Computer Assisted ECG (SCP-ECG) is also defined by the ISO [452] and specifies the conventions to interchange ECG signal data, measurements and interpretation results. Nowadays, the most known standards are the Digital Imaging and Communications in Medicine (DICOM) standard [453], which was created for the communication and management of medical imaging information and related data, and the standard of annotated ECGs (aECG) of the international organization Health Level 7 (HL7) [454], which is an XML-based format for the exchange of data in hospitals.

  Data to be appropriately represented and to avoid them to be lost or messed up, it is advisable the use of common terminology, and for this purpose, standards like the Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT) [455] of the International Health Terminology Standards Development Organisation (IHTSDO) and the LOINC [456] which stands for "Laboratory Observations, Identifiers, Names and Codes" have been created.

  Nonetheless, standards are not yet available for all the necessary aspects of telemedicine [457, 458]. The European Commission affirms that interoperability problems are one of the most important issues that avoid investments in these devices to be well-used and therefore, limit the scalability of this kind of solutions. Interoperability is not guaranteed without globally accepted standards, and hence, the existing standards must be adopted by systems all over the world. This might be complicated due to the wide heterogeneity of health information systems, and because millions of terminologies and vocabulary are required to describe and codify health data [459]. Nevertheless, as it is a priority for the successful development of emerging health services, the first steps towards interoperability of electronic health systems in the EU have already been taken [460].

- **Privacy, security and ethical issues**

  As stated before, such a system implies people to be continuously monitored. Therefore, huge amounts of information about the individuals and their lives can be inferred, and unfortunately, this information may be the target of many people interested in things that have little to do with the health of individuals [461]. In fact, a Financial Times investigation revealed that 9 of the top 20 health-related mobile apps have been used to transmit data to a company interested on people's mobile phone usage [462]. Currently, approaches are being developed in order to avoid this information to be used for evil or non-ethical purposes [463, 464, 465]. This type of solutions must contain safety precautions such as the encryption of data and patient authentication mechanisms. The awareness of the subjects being monitored must also be assured [461], and their autonomy must be respected [466]. The current Personal Data Protection Directive of the European Union is being revised in order to give a better response to these issues posed by the development and globalization of the new technologies [459].

- **Efficiency and reliability**

  Efficiency and reliability are key characteristics for the widespread use of the developed technology [416]. It has recently been affirmed that many algorithms, for example, those used in smart environments for activity recognition, need improvements in order to become more reliable and more accurate for real life [279]. Other reports have also remarked that some solutions do not work as expected or they have not been properly tested, which in some cases may pose a risk to people's safety [467]. The European Commission adverts that errors may arise from many different sources due to the large list of stakeholders involved in the development and use of these medical devices, such as the doctor who may make an incorrect diagnosis due to inaccurate data, the IT engineer who might have introduced a bug in the code or the patient who might have misused the device [459]. This becomes a real problem if due to any of these reasons the patient is harmed, and in order to limit these risks the legal responsibilities of each stakeholder should be clearly stated. Furthermore, safety must be demonstrated by safety standards such as the IEC 82304-1 [468] or specific quality labels, and certifications might be used for ensuring the credibility of the health solutions [459].

- **Cost**

  Costs associated with such a system include the first investments, as well as maintenance and operational costs. Even if the research stage of the disorder detection system might be funded, the lack of financing structure for the continuation of the project can make all the work come to nothing as it has already happened with some telemedicine applications [457, 469, 470, 471]. It has been affirmed that the high cost of current wearable system services limits their expansion and that these economic issues have to be addressed to ensure the opening of the market to these new technological systems [416]. The European Commission also accepts that the lack of innovative and adequate refund models for electronic health solutions is a major obstacle in their development and in their spread. Even if some insurance companies are adopting measures, most of them do not yet have standard tariffs for these applications [459, 472, 473].

- **Legislation**

  Legislation and policy for certain aspects of telemedicine are not yet available [279, 457], albeit they are a prerequisite for the development of the system described herein. Licensure, certification and protection must be standardized in terms of laws of the European Communities, especially, if services are to be given over the internet [416].

5.4 Conclusions

This work has presented a global screening methodology for the early detection of disorders based on physiological and behavioural change. For this purpose, a cross-case analysis of two case-studies analysed individually in previous work has been performed. These works were about early detection of occupational stress from physiological and behavioural data monitored in a smart office environment, and early detection of AD from the monitoring of older adults' behaviour in a smart home environment. The analysis of such disparate scenarios has allowed to abstract the methodology from its use in specific case-studies, and to draw conclusions on the steps to take from a global point of view.

Research work aimed at predicting individual disorders from physiological data are widely present in the literature, whereas unobtrusively collected behavioural data is gradually opening up a space thanks to the advancement of smart environments. Nonetheless, each disorder is individually treated and similarities among many disorders that provoke behavioural shifts are not taken into consideration. To overcome this issue, this work has proposed the use of unobtrusive physiology and behaviour data monitoring system as a test-method for a global screening process.

Therefore, the work has updated the current screening methodology [14] which relies on specific test-methods for each specific disorder, to a new one which relies on a single test-method based on unobtrusively collected physiological and behavioural shifts. This test-method of proposal meets the requirements specified by the WHO [14, 350], as it is simple, cheap and can be acceptable for patients and staff with a little awareness thanks to its unobtrusive nature. Sensitivity requirements can also be fulfilled by selecting the correct algorithms for prediction model building, as explained in this work. An overview of the guidelines to follow in order to design a smart monitoring system and make use of changes detected in unobtrusively collected physiological and behavioural data has been given.

The methodology consists of an iterative four-step process that will have to be applied for each individual disorder. After a first reflexion step, the monitoring environment will have to be set-up. Once the technological part is ready, a first cohort will be monitored, and their preprocessed data will be used to build initial prediction models. The data preprocessing step has been specifically defined to extract change statistics from the physiological and behavioural data. Once the initial prediction models built, they will be used to detect suspicious changes on people at risk of the disease, while the process will be monitored by a specialist who will also receive information and feedback from the people around the patient. In case a positive alarm is triggered, the disease specialist will have to perform the corresponding diagnostic test to verify or discard the presence of the disease. The process can then iterate in three ways. If the disease specialist deems it necessary, the process will start over, re-defining everything from the very beginning. If the technological approach is approved, the data used in the last monitoring process can be used to feed the initial database and rebuild more accurate prediction models. Once prediction models are stable and accurate enough, the iteration can be limited to continuously identifying new people at risk of the disease and monitoring their health status.

This new approach offers several advantages in comparison to the traditional screening method. On one hand, it provides a way to monitor users continuously, detecting every subtle but suspicious change and minimizing detection delay as much as possible. On the other hand, users can be monitored in their usual environments, avoiding to make regular visits to medical or specialized centers. Moreover, both physiological and behavioural symptoms can be monitored completely unobtrusively or by means of minimally invasive body sensors, which greatly improves the methods' acceptance among the users. For the same reason, the gathered data is ecologically more valid than tests performed in artificial settings. Furthermore, the solution brings many economic advantages, due to the inexpensiveness of the required hardware, and to the reduced number of periodical checkups in medical centers that will have to be performed. Finally, the independence of the methodology

from any disease makes it widely useful.

In contrast, the methodology also shows some disadvantages shared by all screening methodologies. These include the implementation cost and work for the people who will never be diagnosed, the anxiety that a false positive might provoke in the patient and the false security sense provided by a false negative. Moreover, the efficiency of the method will highly depend on the selected high-risk group and temporal window lengths for signal processing, as well as the number of iterations performed. In our approach, measures to reduce these disadvantages as far as possible have been taken.

Technological and methodological issues and challenges that must be overcome for the widespread use of this methodology have also been listed. Technological challenges include the necessity to provide high-quality data, to integrate data from several modalities, to process and store big amounts of data and to assure models' reliability. Moreover, problems related to users' acceptance towards the developed systems, the interoperability with other medical systems and devices, privacy, security and ethical issues, initial investments and legislation must be addressed.

For the best of our knowledge, this is the first time that a methodological approach for the early detection of disorders based on unobtrusively collected physiological and behavioural data from a global point of view, *i.e.* non-specific to any disorder, is presented. The proposed approach updates the current method for the early detection of diseases to the technological level that we already have today. The widespread use of such a methodology can help the healthcare system to get closer to the technological solutions that are available to them, resulting in benefits for health specialists, patients and people around them, and therefore, for society in general.

This chapter intends to be a first approach of a non-specific screening methodology based on AmI and AI technologies, and will certainly need to be refined in the upcoming years. As it is not possible to cover each and every one of the points that a methodology of such dimensions and such applicability requires to be specified, we followed some predefined criteria for the analysis and therefore, we admit that it has some limitations and there are several interesting points have been left unanalysed in this work. For example, the inclusion of a sensor failure detection system would be highly desired, as all the results will rely on the data collected from the sensors. Moreover, we did not test our case studies following a Leave-One-Subject-Out cross-validation (LOSOCV) which might be very interesting to validate the performance of the detection system for completely unseen people by the system. Furthermore, some smart-environments might pose difficulties when there is more than one person living or spending time in it. The differentiation of the activities performed by one person or another by means of sensors is a very complicated task. Thus, in our second case study, we were only considering older adults living alone. When designing the methodology, this factor must be taken into account and either methods that allow individualized monitoring must be implemented or its use must be limited to environments where there is no possibility of confusion in the person being monitored. Finally, the signal preprocessing approach suggested herein only computes change statistics reflecting the absolute value of change but not its directionality. For some of the features, especially for the ones extracted from physiological data, the direction of the change might be highly informant. It would be a valuable work to include new time-series statistics reflecting the direction of the change (*e.g.* differential) to the five used in our approach, and analyse their effectiveness both in physiological and behavioural data. This might help improving final models' performance. Research aimed at overcoming these limitations of the approach presented herein can be a good starting point for future work.

Summary

- A multidomain screening methodology for the early detection of disorders from unobtrusivelly collected physiological and/or behavioral data has been presented.
- The presented approach provides a basis for the inexpensive early detection of many disorders.
- There are still some methodological and technological issues to overcome for the widespread use of this methodology.
- The extension of such a methodology is huge by nature and has therefore been limited following some predefined criteria. Future work must focus on expanding the work presented herein.

6

Conclusions, Contributions and Future Work

Contents

A general abstract of the main results obtained from the attainment of the thesis is given in this chapter. Section 6.1 presents a global description of the research work. The scientific contributions that resulted from this work are then listed in Section 6.2 whereas the relevant publications are presented in Section 6.3. Section 6.4 identifies some future research lines for further development of the area and finally, Section 6.5 states the final remarks.

6.1 Conclusions

This dissertation has been centered on the definition of a global methodology for the early detection of disorders based on unobtrusively collected physiological and behavioural data by means of a cross-case analysis.

Since many disorders are detected by relatives and co-workers due to the behavioural changes that the patient suffers, a system for the early detection of diseases that benefits from these behavioural symptoms is desired. Towards this goal, we first selected two very disparate scenarios where this fact is reflected, in order to analyse them individually: stress in office environments and AD at home.

We then performed a comprehensive literature review where we analysed the symptoms and their assessment methods for the two case studies from a multimodal point of view. We analysed the strengths and weaknesses of each assessment method, focusing, in addition to their prediction abilities for the specific cases, on their usability for the implementation of a ubiquitous and unobtrusive user-centered monitoring system. Moreover, we identified the common practices in current disorders' early detection research, which allowed us to spot some gaps and room for future improvements in the current literature.

Once the main characteristics of health-related datasets and current disorders' detection models identified, we proposed a first algorithmic approach for early stress detection from unobtrusively collected physiological and behavioral data in realistic work settings in a smart office environment, which was aware of the identified common characteristics of health-related data. We performed specific prediction analyses for this case-study, focusing on the contribution of each type of physiological and behavioral measurement to the prediction of each of the self-assessed stress and workload levels. We draw conclusions about the usability of the presented approach for the specific goal of early stress detection, contributing to this precise research area.

Later, we applied the previously presented algorithmic approach in the second scenario on the early AD detection in Smart Homes. This phase required the refinement of the approach by adapting each data processing step to the characteristics of the new dataset. We then performed and evaluated the predictions based on the presented approach for this specific case study and draw conclusions about its usefulness for this individual objective. This allowed us to contribute a little to the understanding of AD and to the development of early detection systems for this disorder.

Finally, we performed a cross-case analysis of the two case-studies. For this purpose, we analysed their similarities and differences starting from the definition of the problem, through the application and adaptation of the algorithmic approach, to the validation of the predictive models. This process allowed us to abstract a global methodology for the early detection of disorders based on unobtrusively collected physiological and/or behavioral data and to define the complete path to follow for the analysis of a new case-study and the implementation of a ubiquitous and unobtrusive monitoring system for its early detection.

The work exposed herein has filled the research gaps presented in Section 2.6 by presenting a multidomain method for the early detection of disorders based on temporal statistics of multimodal symptoms.

6.2 Contributions

Based on the aforementioned conclusions, the main contributions of this work can be summarized in the following items:

- An extensive review of stress' and AD's symptoms' multimodal measurement techniques has been performed, focusing on their usefulness for unobtrusive and ubiquitous monitoring systems (see Chapter 2).
- Workload and stress prediction models have been built and evaluated based on unobtrusively collected physiological and behavioral data in a smart office, in addition to performing an analysis of the best predictor for this purpose (see Chapter 3).
- AD's common symptoms' prediction models have been built and evaluated based on unobtrusively collected behavioral data in a smart home, in addition to analysing the best predictors for each symptom (see Chapter 4).
- A cross-case analysis of the previous stress and AD detection works has been done, resulting in the definition of a multidomain methodology for the early detection of disorders based on unobtrusively detected physiological and behavioral shifts (see Chapter 5).

6.3 Relevant publications

Parts of the works covered in this dissertation have already been published or have been sent for review in different international peer-reviewed scientific journals. We now list the scientific publications that are directly related to the work in this thesis:

- Ane Alberdi, Asier Aztiria and, Adrian Basarab. Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. *Journal of Biomedical Informatics*, 2015;59;49-75.
- Ane Alberdi, Asier Aztiria and, Adrian Basarab. On the early diagnosis of Alzheimer's Disease from multimodal signals: A survey. *Artificial Intelligence in Medicine*, 2016;71;1-29.
- Ane Alberdi, Alyssa Weakley, Asier Aztiria, Maureen Schmitter-Edgecombe and, Diane J. Cook. Automatic assessment of functional health decline in older adults: a smart home approach. *Submitted for publication to the Journal of Biomedical Informatics*, 2017.
- Ane Alberdi, Asier Aztiria, Adrian Basarab and, Diane J. Cook. Using Smart Offices to Predict Occupational Stress. *Submitted for publication to the International Journal of Industrial Ergonomics*, 2017.
- Ane Alberdi, Alyssa Weakley, Maureen Schmitter-Edgecombe, Diane J. Cook, Asier Aztiria, Adrian Basarab and, Maitane Barrenechea. Smart Homes predicting the Multi-Domain Symptoms of Alzheimer's Disease. *Submitted for publication to the IEEE Journal of Biomedical and Health Informatics*, 2017.

6.4 Future work

In this section we outline future research directions that can lead to additional contributions in the field of early detection of disorders, both from a multi-domain point of view and regarding the

specific case-studies analysed in this PhD thesis. This dissertation, though limited in scope, offers development opportunities that deserve the attention of the scientific community for further advances in the field. We now list the aforementioned opportunities arranged by topic. These opportunities can be seen as a direct continuation of the work presented in this dissertation, and can therefore be complemented with the research gaps identified in Section 2.6.

6.4.1 Improvements on the early detection of stress

The state of the art on stress detection has been mainly focused on the use of physiological measurements, whereas our approach has shown much higher predictability from changes in behavioral patterns. As stated in Chapter 3, this can be due to a higher prediction ability of behavior compared to physiological signals, or can also be because our approach was not capturing well the main information given by the physiolgical signals. We believe that the inclusion of time-series statistics describing the directionality of change can highly improve the predictability of physiological signals, thus, being interesting their analysis in future applications. Moreover, the validation of our approach in a longitudinal dataset, as well as following a "leave-one-subject-out" Cross Validation (CV) approach, would be desired. Furthermore, some of the data were collected by means of techniques that are, even if minimally invasive, on the borderline of obtrusivity (*i.e.* body sensors) or respect for privacy (*i.e.* webcam). Methods to capture these data without any possibility of having problems in this regard (*e.g.* with smart devices measuring physiological signals) must be analysed and tested for the presented approach. In addition, we did not perform an analysis of the effect of the temporal windows for time-series statistics' computation and processing, but would be highly valuable to maximize the prediction ability of the models from the proposed approach. Finally, a more in-depth feature analysis by means of specific algorithms for this purposes, and analysis on the best time-series statistics to be used has not been performed, but would be highly required to reduce the amount of data being processed by the final system.

6.4.2 Improvements on the early diagnosis of AD

Regarding the second case study, some improvements that could further contribute to this specific research field have been identified. First, as mentioned in Chapter 4, the proposed system should be enriched with a sensor failure detection system. With the current approach, a sensor failure would be translated into incorrect activity recognitions, that could lead, at the same time, to incorrect health status predictions. A sensor failure detection system could avoid this chain effect. Moreover, the validation of our system in a bigger sample size would be required, mainly, because there were no AD-diagnosed individuals in our dataset, but only elderly with commonly impaired symptoms in AD. The availability of a bigger dataset would allow to develop more accurate detection models. Furthermore, this could also contribute to overcome the imbalanced dataset issue that we have been facing for this case study, although the testing of specific algorithms for this type of difficulties is highly recommended. Finally, as in the previous case study, an analysis of the best behavioral predictors by means of automatic feature selection algorithms or of the best temporal statistics for AD symptoms' detection has not been performed, nor has been analysed the effect of the temporal widow's size to compute the time series statistics for prediction, which would be highly interesting as future work. A "leave-one-subject-out" CV approach for validation would also help in completing the current work.

6.4.3 Improvements on the definition of a multidomain methodology

Regarding the methodological proposal for the early detection of disorders given in Chapter 5, there are many threads that can be pulled to improve and complete, until it can be implemented and

used in everyday life. Such a methodology, is huge in its nature, and has been strictly limited to a few items (which resulted from the comparison of the two case studies from the point of view of the selected disorders, applied algorithmic approach and obtained results) in this PhD thesis due to restrictions on time, space and areas of knowledge. Subsection 5.3.7 has exposed the main challenges that must be overcome for such a system to be widely used. Among other things, future work to polish the proposed methodology and to make it usable in real life must give a solution to the big data issues, to the uncertainty in their efficiency and reliability and to the privacy and ethical issues. In addition, the extension of the presented methodology to multiple-user environments would be required to implement the monitoring environment in some specific scenarios.

6.5 Final remarks

With this research work we have tried to do our part towards the widespread use of Ambient Intelligence (AmI) technologies for the early detection of diseases, an area in which increasing research is being done due to its importance. Interesting conclusions have been drawn both for the individual case-studies and for the global applicability of these technologies for a wide variety of disorders. Different future research lines have also been defined to go a step further and get closer to the definitive implementation and widespread use of smart environments to reliably control our health status.

List of Symbols

EEG Symbols	
α_{eeg}	Alpha band energy
β_{eeg}	Beta band energy
δ_{eeg}	Delta band energy
θ_{eeg}	Theta band energy
All	The whole spectral band

ECG Symbols	
All	The whole spectral band
HF	High frequency (0.15 - 0.5 Hz band) power.
HR	Heart Rate, Number of beats (or R peaks) per minute.
I	Interruption-based stressor state
LF	Low frequency (0.00 - 0.08 Hz band) power.
MF	Medium frequency (0.08 - 0.15 Hz) power.
NN50	Number of NN intervals that are greater than 50 ms.
ULF	Ultra low frequency ($<$ 0.003 Hz) power.
VLF	Very low frequency (0.003 - 0.04 Hz band) power.
pnn50 (%)	Percent of NN50.
SDANN	SD of the averages of NN intervals in all 5-minute segments of the entire recording.
SDNN	The SD of all NN intervals (normal R-R intervals).
SDNN index	The mean of all the 5 minute SDs of NN intervals.
SDSD	SD of differences between adjacent NN intervals.

Other Symbols	
Aβ	Amyloid-beta
All	The whole spectral band
ApEn	Approximate Entropy
Cho	Choline
Cr	Creatine
Glu	Glutamine
κ_n	n-th order cumulant
maxRatio	n°of maxima / n°of total signal values
mI	Myo-Inositol
minRatio	n°of minima / n°of total signal values
μ	Mean
N	Neutral state
NAA	N-acetyl Aspartate
p-tau	Tau protein

r	Correlation coefficient
R	Relaxed state
RASTA-PLP	Relative Spectral Transform - Perceptual Linear Perception
RMSSD	The square root of the mean of the sum of the squares of differences between adjacent NN intervals.
s	Skip-size
SampEn	Sample-Entropy
SD	Standard Deviation
T	Time-pressure stressor state
TEO-CB-AutoEnv	Teager Energy Operator based non-linear transformation
w	Length of sliding window

Bibliography

[1] D. J. Cook and S. K. Das, "How smart are our environments? An updated look at the state of the art," *Pervasive and Mobile Computing*, vol. 3, no. 2, pp. 53–73, 2007.

[2] C.-W. Chen, A. Aztiria, and H. Aghajan, "Learning human behaviour patterns in work environments," in *CVPR 2011 Workshops*. IEEE, jun 2011, pp. 47–52.

[3] A. Aztiria, J. C. Augusto, R. Basagoiti, A. Izaguirre, and D. J. Cook, "Learning Frequent Behaviors of the Users in Intelligent Environments," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 43, no. 6, pp. 1265–1278, nov 2013.

[4] C. Galambos, M. Skubic, S. Wang, and M. Rantz, "Management of dementia and depression utilizing in-home passive sensor data," *Gerontechnology*, vol. 11, no. 3, pp. 457–468, 2013.

[5] B. H. W. Eijckelhof, M. a. Huysmans, B. M. Blatter, P. C. Leider, P. W. Johnson, J. H. van Dieën, J. T. Dennerlein, and A. J. van der Beek, "Office workers' computer use patterns are associated with workplace stressors." *Applied ergonomics*, vol. 45, no. 6, pp. 1660–7, nov 2014.

[6] A. Sano and R. W. Picard, "Stress Recognition Using Wearable Sensors and Mobile Phones," *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pp. 671–676, sep 2013.

[7] A. Muaremi, B. Arnrich, and G. Tröster, "Towards Measuring Stress with Smartphones and Wearable Devices During Workday and Sleep," *BioNanoScience*, vol. 3, no. 2, pp. 172–183, may 2013.

[8] K.-C. Lan and W.-Y. Shih, "Early Diagnosis of Parkinson's Disease Using a Smartphone," *Procedia Computer Science*, vol. 34, pp. 305–312, 2014.

[9] P. Lorenzi, R. Rao, G. Romano, A. Kita, M. Serpa, F. Filesi, R. Parisi, A. Suppa, M. Bologna, A. Berardelli, and F. Irrera, "Smart Sensing Systems for the Detection of Human Motion Disorders," *Procedia Engineering*, vol. 120, pp. 324–327, 2015.

[10] A. Anzulewicz, K. Sobota, and J. T. Delafield-Butt, "Toward the Autism Motor Signature: Gesture patterns during smart tablet gameplay identify children with autism," *Scientific Reports*, vol. 6, no. 1, p. 31107, nov 2016.

[11] "i-PROGNOSIS Project," 2016, [Online] http://www.i-prognosis.eu/, (Date retrieved: 2017-08-21).

[12] "ICT4Life," 2016, [Online] http://www.ict4life.eu, (Date retrieved: 2017-08-21).

[13] "Brainview Project," 2015, [Online] http://www.brainview.org/, (Date retrieved: 2017-08-21).

[14] J. M. G. Wilson and G. Jungner, "Principles and Practice of Screening for Disease," p. 168, 1968.

[15] R. K. Yin, "Case Study Research: Design and Methods," *Applied Social Research Methods Series*, vol. 5, p. 219, 1994.

[16] B. J. Oates, *Researching Information Systems and Computing*. SAGE Publications, 2006.

[17] Collins English Dictionary, "Psychological," p. 1888, 2007, [Online] http://www.collinsdictionary.com/dictionary/english/psychological?showCookiePolicy=true, (Date retrieved: 2015-10-05).

[18] Oxford Dictionaries, "Physiology," p. 2112, 2015, [Online] http://www.oxforddictionaries.com/definition/english/physiology, (Date retrieved: 2015-10-05).

[19] Merriam-Webster, "Behavior," p. 939, 2004, [Online] http://www.merriam-webster.com/dictionary/behavior, (Date retrieved: 2015-10-05).

[20] H. Selye, *The stress of life*, mcgraw-hil ed., New York, 1956.

[21] B. Mishra, S. Mehta, N. D. Sinha, S. K. Shukla, N. Ahmed, and A. Kawatra, "Evaluation of work place stress in health university workers: a study from rural India," *Indian journal of community medicine : official publication of Indian Association of Preventive & Social Medicine*, vol. 36, no. 1, pp. 39–44, jan 2011.

[22] J. Bakker, M. Pechenizkiy, and N. Sidorova, "What's your current stress level? Detection of stress patterns from GSR sensor data," in *Proceedings - IEEE International Conference on Data Mining, ICDM*, no. 1. Ieee, dec 2011, pp. 573–580.

[23] N. Hjortskov, D. Rissén, A. K. Blangsted, N. Fallentin, U. Lundberg, and K. Søgaard, "The effect of mental stress on heart rate variability and blood pressure during computer work." *European journal of applied physiology*, vol. 92, no. 1-2, pp. 84–9, jun 2004.

[24] H. Kurniawan, A. V. Maslov, and M. Pechenizkiy, "Stress detection from speech and Galvanic Skin Response signals," *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, pp. 209–214, jun 2013.

[25] A. Kaklauskas, E. K. Zavadskas, M. Seniut, G. Dzemyda, V. Stankevic, C. Simkevičius, T. Stankevic, R. Paliskiene, A. Matuliauskaite, S. Kildiene, L. Bartkiene, S. Ivanikovas, and V. Gribniak, "Web-based Biometric Computer Mouse Advisory System to Analyze a User's Emotions and Work Productivity," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 6, pp. 928–945, sep 2011.

[26] M. Milczarek, E. Schneider, and E. R. González, "OSH in figures. Stress at work. Facts and figures," European Agency for Safety and Health at Work, Luxembourg, Tech. Rep., 2009.

[27] J. Wijsman, B. Grundlehner, H. Liu, J. Penders, and H. Hermens, "Wearable Physiological Sensors Reflect Mental Stress State in Office-Like Situations," in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, no. iii. IEEE, sep 2013, pp. 600–605.

[28] M. Bickford, "Stress in the workplace: A general overview of the cases, the effects and the solutions," Canadian Mental Health Association, Tech. Rep. August, 2005.

[29] Y. Okada, T. Y. Yoto, T.-A. Suzuki, S. Sakuragawa, and T. Sugiura, "Wearable ECG recorder with acceleration sensors for monitoring daily stress: Office work simulation study," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, jul 2013, pp. 4718–4721.

[30] D. Carneiro, J. C. Castillo, P. Novais, A. Fernández-Caballero, and J. Neves, "Multimodal behavioral analysis for non-invasive stress detection," *Expert Systems with Applications*, vol. 39, no. 18, pp. 13 376–13 389, dec 2012.

[31] T. Hayashi, Y. Mizuno-Matsumoto, E. Okamoto, M. Kato, and T. Murata, "An fMRI study of brain processing related to stress states," in *World Automation Congress (WAC), 2012*. Puerto Vallarta, Mexico: IEEE, 2012, pp. 1–6.

[32] N. Sharma and T. Gedeon, "Objective measures, sensors and computational techniques for stress recognition and classification: a survey." *Computer methods and programs in biomedicine*, vol. 108, no. 3, pp. 1287–301, dec 2012.

[33] S.-H. Seo and J.-T. Lee, "Stress and EEG," in *Convergence and Hybrid Information Technologies*, M. Crisan, Ed., 2010, no. March, p. 426.

[34] S. J. Lupien and F. Seguin, "How to Measure Stress in Humans," Centre for Studies on Human Stress, Quebec, Tech. Rep., 2013, [Online] http://www.humanstress.ca/documents/pdf/SalivaLab/HOWTOMEASURESTRESS_CSHS.pdf, (Date retrieved: 2015-11-09).

[35] L. Hadjileontiadis, "Biosignals and Compression Standards," *M-Health. Emerging Mobile Health Systems*, pp. 277–292, 2006.

[36] K. Asai, "The Role of Head-Up Display in Computer- Assisted Instruction," *Human Computer Interaction: New Developments*, oct 2010.

[37] B. Cinaz, B. Arnrich, R. La Marca, and G. Tröster, "Monitoring of mental workload levels during an everyday life office-work scenario," *Personal and Ubiquitous Computing*, vol. 17, no. 2, pp. 229–239, oct 2013.

[38] K. Palanisamy, M. Murugappan, and S. Yaacob, "Multiple Physiological Signal-Based Human Stress Identification Using Non-Linear Classifiers," *Electronics and Electrical Engineering*, vol. 19, no. 7, pp. 80–85, sep 2013.

[39] W. J. Ray and H. W. Cole, "EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes." *Science (New York, N.Y.)*, vol. 228, no. 4700, pp. 750–752, may 1985.

[40] K. S. Rahnuma, A. Wahab, N. Kamaruddin, and H. Majid, "EEG analysis for understanding stress based on affective model basis function," in *2011 IEEE 15th International Symposium on Consumer Electronics (ISCE)*. IEEE, jun 2011, pp. 592–597.

[41] J. Peuscher, "Galvanic skin response (GSR)," TMSI, Tech. Rep. November, 2012, [Online] http://www.tmsi.com/products/accessories?task=callelement&format=raw&item_id=43&element=fe0c95f3-af08-4719-bc51-36917715660d&method=download, (Date retrieved: 2015-11-12).

[42] K. Peternel, M. Pogačnik, R. Tavčar, and A. Kos, "A presence-based context-aware chronic stress recognition system." *Sensors (Basel, Switzerland)*, vol. 12, no. 11, pp. 15 888–906, jan 2012.

[43] D. Alexander, C. Trengove, P. Johnston, T. Cooper, J. August, and E. Gordon, "Separating individual skin conductance responses in a short interstimulus-interval paradigm," *Journal of Neuroscience Methods*, vol. 146, no. 1, pp. 116–123, jul 2005.

[44] D. Fowles, "Mechanisms of electrodermal activity: Methods in Physiological Psychology," in *Bioelectric Recording Techniques*, M. P. RF Thompson, Ed. New York: Academic Press, 1974, ch. Vol 1., pp. 231–71.

[45] J. A. Healey and R. W. Picard, "Detecting Stress During Real-World Driving Tasks Using Physiological Sensors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 156–166, jun 2005.

[46] J. Hernandez, P. Paredes, A. Roseway, and M. Czerwinski, "Under pressure," in *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*. New York, New York, USA: ACM Press, 2014, pp. 51–60.

[47] D. McDuff, A. Karlson, A. Kapoor, A. Roseway, and M. Czerwinski, "AffectAura: An intelligent System for Emotional Memory," in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*. New York, New York, USA: ACM Press, 2012, p. 849.

[48] R. Kocielnik, N. Sidorova, F. M. Maggi, M. Ouwerkerk, and J. H. D. M. Westerink, "Smart technologies for long-term stress monitoring at work," in *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*. IEEE, jun 2013, pp. 53–58.

[49] T. Pickering, "Principles and techniques of blood pressure measurement," *Cardiology clinics*, vol. 28, no. 4, pp. 571–586, 2002.

[50] C. Maaoui, A. Pruski, and F. Abdat, "Emotion recognition for human-machine communication," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, sep 2008, pp. 1210–1215.

[51] F. Seoane, I. Mohino-Herranz, J. Ferreira, L. Alvarez, R. Buendia, D. Ayllón, C. Llerena, and R. Gil-Pita, "Wearable biomedical measurement systems for assessment of mental stress of combatants in real time." *Sensors (Basel, Switzerland)*, vol. 14, no. 4, pp. 7120–41, jan 2014.

[52] J. Zhai and A. Barreto, "Stress detection in computer users based on digital signal processing of noninvasive physiological variables." in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, vol. 1, jan 2006, pp. 1355–8.

[53] C. H. Vinkers, R. Penning, J. Hellhammer, J. C. Verster, J. H. G. M. Klaessens, B. Olivier, and C. J. Kalkman, "The effect of stress on core and peripheral body temperature in humans," *Stress*, vol. 16, no. 5, pp. 520–530, sep 2013. [Online]. Available: http://www.tandfonline.com/doi/full/10.3109/10253890.2013.807243

[54] N. Sharma, A. Dhall, T. Gedeon, and R. Goecke, "Thermal spatio-temporal data for stress recognition," *EURASIP Journal on Image and Video Processing*, vol. 2014, no. 1, p. 28, 2014.

[55] J. A. Levine, I. T. Pavlidis, L. MacBride, Z. Zhu, and P. Tsiamyrtzis, "Description and clinical studies of a device for the instantaneous detection of office-place stress." *Work (Reading, Mass.)*, vol. 34, no. 3, pp. 359–64, jan 2009.

[56] D. Shastri, M. Papadakis, P. Tsiamyrtzis, B. Bass, and I. Pavlidis, "Perinasal Imaging of Physiological Stress and Its Affective Potential," *IEEE Transactions on Affective Computing*, vol. 3, no. 3, pp. 366–378, jul 2012.

[57] M. Norzali, H. Mohd, M. Kashima, K. Sato, and M. Watanabe, "Facial Visual-Infrared Stereo Vision Fusion Measurement as an Alternative for Physiological Measurement," *Journal of Biomedical Image Processing*, vol. 1, no. May, pp. 34–44, 2014.

[58] T. Chen, P. Yuen, M. Richardson, G. Liu, and Z. She, "Detection of Psychological Stress Using a Hyperspectral Imaging Technique," *IEEE Transactions on Affective Computing*, vol. 5, no. 4, pp. 391–405, oct 2014.

[59] C. Z. Wei, "Stress Emotion Recognition Based on RSP and EMG Signals," *Advanced Materials Research*, vol. 709, pp. 827–831, jun 2013.

[60] Y. Shi, M. H. Nguyen, P. Blitz, B. French, S. Fisk, F. D. Torre, A. Smailagic, D. P. Siewiorek, M. Absi, E. Ertin, T. Kamarck, and S. Kumar, "Personalized Stress Detection from Physiological Measurements," *International Symposium on Quality of Life Technology*, 2010.

[61] E. Peper, R. Harvey, I.-m. Lin, H. Tylova, and D. Moss, "Is There More to Blood Volume Pulse Than Heart Rate Variability , Respiratory Sinus Arrhythmia , and Cardiorespiratory Synchrony ?" *Biofeedback*, vol. 35, no. 2, pp. 54–61, 2007.

[62] J. Ramos, J.-H. Hong, and A. K. Dey, "Stress Recognition - A Step Outside the Lab," in *Proceedings of the International Conference on Physiological Computing Systems*. SCITEPRESS - Science and and Technology Publications, 2014, pp. 107–118.

[63] W. Liao, W. Zhang, Z. Zhu, and Q. Ji, "A Real-Time Human Stress Monitoring System Using Dynamic Bayesian Network," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, vol. 3, pp. 70–70, 2005.

[64] A. Barreto, J. Zhai, N. Rishe, and Y. Gao, "Measurement of Pupil Diameter Variations as a Physiological Indicator of the Affective State in a Computer User," in *Proceedings of the 44th Annual Rocky Mountain Bioengineering Symposium*, P. John S. Sollers, III, PhD and Julian F. Thayer, Ed., vol. 43. Denver, Colorado: ISA, 2007, pp. 146–151.

[65] N. Sharma and T. Gedeon, "Hybrid Genetic Algorithms for Stress Recognition in Reading," in *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics*, ser. Lecture Notes in Computer Science, L. Vanneschi, W. S. Bush, and M. Giacobini, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, vol. 7833, pp. 117–128.

[66] M. E. Jabon, J. N. Bailenson, E. Pontikakis, L. Takayama, and C. Nass, "Facial expression analysis for predicting unsafe driving behavior," *IEEE Pervasive Computing*, vol. 10, no. 4, pp. 84–95, apr 2011.

[67] M. Haak, S.Bos, S.Panic, and L.J.M.Rothkrantz, "Detecting Stress Using Eye Blinks and Brain Activity from EEG Signals," in *GAMEON*, L. Breitlauch, Ed. EUROSIS, 2009, pp. 75–82.

[68] A. M. Burhan, R. Bartha, C. Bocti, M. Borrie, R. Laforce, P. Rosa-Neto, and J.-P. Soucy, "Role of emerging neuroimaging modalities in patients with cognitive impairment: a review from the Canadian Consensus Conference on the Diagnosis and Treatment of Dementia 2012," *Alzheimer's Research & Therapy*, vol. 5, no. Suppl 1, p. S4, 2013.

[69] K. A. Johnson, N. C. Fox, R. A. Sperling, and W. E. Klunk, "Brain Imaging in Alzheimer Disease," *Cold Spring Harbor Perspectives in Medicine*, vol. 2, no. 4, pp. a006 213–a006 213, apr 2012.

[70] T. Hayashi, E. Okamoto, H. Nishimura, Y. Mizuno-Matsumoto, R. Ishii, and S. Ukai, "Beta Activities in EEG Associated with Emotional Stress," *International Journal of Intelligent Computing in Medical Sciences & Image Processing*, vol. 3, no. 1, pp. 57–68, jan 2009.

[71] S. Koldijk, M. A. Neerincx, and W. Kraaij, "Detecting work stress in offices by combining unobtrusive sensors," *IEEE Transactions on Affective Computing*, no. c, pp. 1–1, 2016.

[72] A. de Santos Sierra, C. Sanchez Avila, G. Bailador del Pozo, and J. Guerra Casanova, "Stress detection by means of stress physiological template," *2011 Third World Congress on Nature and Biologically Inspired Computing*, pp. 131–136, oct 2011.

[73] J. Choi and R. Gutierrez-Osuna, "Using Heart Rate Monitors to Detect Mental Stress," in *2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks*. Berkeley, CA: IEEE, jun 2009, pp. 219–223.

[74] X. Li, Z. Chen, Q. Liang, and Y. Yang, "Analysis of Mental Stress Recognition and Rating Based on Hidden Markov Model," *Journal of Computational Information Systems*, vol. 10, no. 18, pp. 7911–7919, 2014.

[75] V. Malhotra and M. K. Patil, "Mental Stress Assessment of ECG Signal using Statistical Analysis of Bio-Orthogonal Wavelet Coefficients," *International Journal of Science and Research (IJSR)*, vol. 2, no. 12, pp. 430–434, 2013.

[76] J. Sztajzel, "Heart rate variability: a noninvasive electrocardiographic method to measure the autonomic nervous system." *Swiss medical weekly*, vol. 134, no. 35-36, pp. 514–22, sep 2004.

[77] N. Sharma and T. Gedeon, "Modeling observer stress for typical real environments," *Expert Systems with Applications*, vol. 41, no. 5, pp. 2231–2238, apr 2014.

[78] O. Sourina and Y. Liu, "EEG-enabled Affective Human-Computer Interfaces," in *Universal Access in Human-Computer Interaction. Design and Development Methods for Universal Access*, ser. Lecture Notes in Computer Science, C. Stephanidis and M. Antona, Eds. Cham: Springer International Publishing, 2014, vol. 8513, pp. 536–547.

[79] D. Giakoumis, D. Tzovaras, and G. Hassapis, "Subject-dependent biosignal features for increased accuracy in psychological stress detection," *International Journal of Human-Computer Studies*, vol. 71, no. 4, pp. 425–439, apr 2013.

[80] J. Wijsman, B. Grundlehner, J. Penders, and H. Hermens, "Trapezius muscle EMG as predictor of mental stress," in *Wireless Health 2010 on - WH '10*, vol. V. New York, New York, USA: ACM Press, 2010, p. 155.

[81] R. L. M. C. Epp, M. Lippold, "Identifying emotional states using keystroke dynamics," *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, p. 715, 2011.

[82] N. Ahmad, A. Szymkowiak, and P. a. Campbell, "Keystroke dynamics in the pre-touchscreen era." *Frontiers in human neuroscience*, vol. 7, no. December, p. 835, jan 2013.

[83] M. Curtin, C. Tappert, M. Villani, G. Ngo, J. Simone, H. S. Fort, and S.-H. Cha, "Keystroke Biometric Recognition on Long-Text Input: A Feasibility Study," in *Proceedings of Student/Faculty Research Day*, 2006, pp. 1–5.

[84] P. Zimmermann, S. Guttormsen, B. Danuser, and P. Gomez, "Affective computing–a rationale for measuring mood with mouse and keyboard." *International journal of occupational safety and ergonomics : JOSE*, vol. 9, pp. 539–551, 2003.

[85] R. L. M. C. Epp, M. Lippold, "Identifying emotional states using keystroke dynamics," *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, p. 715, 2011.

[86] A. Kolakowska, "A review of emotion recognition methods based on keystroke dynamics and mouse movements," in *2013 6th International Conference on Human System Interactions (HSI)*. IEEE, jun 2013, pp. 548–555.

[87] A. Kapoor and R. W. Picard, "Multimodal affect recognition in learning environments," *Proceedings of the 13th annual ACM international conference on Multimedia - MULTIMEDIA '05*, p. 677, 2005.

[88] B. Arnrich, C. Setz, R. La Marca, G. Tröster, and U. Ehlert, "What does your chair know about your stress level?" *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 207–214, 2010.

[89] D. F. Dinges, R. L. Rider, J. Dorrian, E. L. McGlinchey, N. L. Rogers, Z. Cizman, S. K. Goldenstein, C. Vogler, S. Venkataraman, and D. N. Metaxas, "Optical computer recognition of facial expressions associated with stress induced by performance demands," in *Aviation Space and Environmental Medicine*, vol. 76, no. 6, 2005.

[90] H. Madokoro and K. Sato, "Facial Expression Spacial Charts for Describing Dynamic Diversity of Facial Expressions," *Journal of Multimedia*, vol. 7, no. 4, pp. 314–324, aug 2012.

[91] S. Das and K. Yamada, "Evaluating Instantaneous Psychological Stress from Emotional Composition of a Facial Expression," *Jounal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 17, no. 4, 2013.

[92] M. Hagmueller, E. Rank, and G. Kubin, "Evaluation of the Human Voice for Indications of Workload Induced Stress in the Aviation Environment," European Organisation for the Safety of Air Navigation, Brétigny-sur-Orge, France, Tech. Rep. 18, 2006.

[93] H. Lu, D. Frauendorfer, M. Rabbi, M. S. Mast, G. T. Chittaranjan, A. T. Campbell, D. Gatica-Perez, and T. Choudhury, "StressSense: Detecting Stress inUnconstrained Acoustic Environments using Smartphones," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*. New York, New York, USA: ACM Press, 2012, p. 351.

[94] V. P. Patil, K. K. Nayak, and M. Saxena, "Voice Stress Detection," *International Journal of Electrical, Electronics and Computer Engineering*, vol. 2, no. 2, pp. 148–154, 2013.

[95] P. Adams, M. Rabbi, T. Rahman, M. Matthews, A. Voida, G. Gay, T. Choudhury, and S. Voida, "Towards Personal Stress Informatics: Comparing Minimally Invasive Techniques for Measuring Daily Stress in the Wild," in *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*. ICST, 2014.

[96] G. Demenko and M. Jastrzebska, "Analysis of Voice Stress in Call Centers Conversations," in *Proceedings of the 6th International Conference on Speech Prosody*, no. 1, 2012, pp. 3–6.

[97] S. Rai and X. Hu, "Behavior Pattern Detection for Data Assimilation in Agent-Based Simulation of Smart Environments," in *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*. IEEE, nov 2013, pp. 171–178.

[98] N. K. Suryadevara, M. Quazi, and S. C. Mukhopadhyay, "Smart Sensing System for Human Emotion and Behaviour Recognition," in *Perception and Machine Intelligence*, ser. Lecture Notes in Computer Science, M. K. Kundu, S. Mitra, D. Mazumdar, and S. K. Pal, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, vol. 7143, pp. 11–22.

[99] S. J. O. Malley, R. T. Smith, and B. H. Thomas, "Data Mining Office Behavioural Information from Simple Sensors," in *Proceedings of the Thirteenth Australasian User Interface Conference - Volume 126*. Melbourne, Australia: Australian Computer Society, Inc., 2012, pp. 97–98.

[100] S. Puteh, C. Langensiepen, and A. Lotfi, "Fuzzy ambient intelligence for intelligent office environments," in *2012 IEEE International Conference on Fuzzy Systems*. IEEE, jun 2012, pp. 1–6.

[101] S. Tao, M. Kudo, H. Nonaka, and J. Toyama, "Person Authentication and Activities Analysis in an Office Environment Using a Sensor Network," in *Constructing Ambient Intelligence, Communications in Computer and Information Science*, ser. Communications in Computer and Information Science, R. Wichert, K. Van Laerhoven, and J. Gelissen, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, vol. 277, pp. 119–127.

[102] J. W. Pennebaker, R. J. Booth, and M. E. Francis, "Linguistic Inquiry and Word Count (LIWC)," [Online] http://www.liwc.net/, (Date retrieved: 2015-06-17).

[103] "Harvard General Inquirer," [Online] http://www.wjh.harvard.edu/~inquirer/, (Date retrieved: 2015-09-10).

[104] Lexalytics, "Semantria," [Online] https://semantria.com/, (Date retrieved: 2015-06-17).

[105] "SentiStrength," [Online] http://sentistrength.wlv.ac.uk/, (Date retrieved: 2015-10-05).

[106] "Synesketch: Free Open-Source Software for Textual Emotion Recognition and Visualization," [Online] http://synesketch.krcadinac.com/, (Date retrieved: 2015-09-10).

[107] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-Based Methods for Sentiment Analysis," *Computational Linguistics*, vol. 37, no. 2, pp. 267–307, 2011.

[108] M. D. Choudhury and M. Gamon, "Predicting depression via social media," *International AAAI Conference on Web and Social Media (ICWSM)*, vol. 2, 2013.

[109] M. Park, C. Cha, and M. Cha, "Depressive moods of users portrayed in twitter," *Proc. of the ACM SIGKDD Workshop on Health Informatics*, 2012.

[110] S. Saleem, R. Prasad, S. Vitaladevuni, M. Pacula, M. Crystal, B. Marx, D. Sloan, J. Vasterling, and T. Speroff, "Automatic Detection of Psychological Distress Indicators and Severity Assessment from Online Forum Posts," *Proceedings of COLING 2012*, vol. 5, no. December 2012, pp. 2375–2388, 2012.

[111] L. M. Vizer, L. Zhou, and A. Sears, "Automated stress detection using keystroke and linguistic features: An exploratory study," *International Journal of Human-Computer Studies*, vol. 67, no. 10, pp. 870–886, oct 2009.

[112] S. Salmeron-Majadas, O. C. Santos, and J. G. Boticario, "An Evaluation of Mouse and Keyboard Interaction Indicators towards Non-intrusive and Low Cost Affective Modeling in an Educational Context," *Procedia Computer Science*, vol. 35, pp. 691–700, 2014.

[113] R. V. Yampolskiy and V. Govindaraju, "Behavioural biometrics: a survey and classification," *International Journal of Biometrics*, vol. 1, no. 1, p. 81, 2008.

[114] P. Dawadi, D. J. Cook, and M. Schmitter-Edgecombe, "Automated clinical assessment from smart home-based behavior data," *IEEE Journal of Biomedical and Health Informatics*, vol. 99164, pp. 1–12, 2015.

[115] Alzheimer's Foundation of America, "About Alzheimer's. Alzheimer's Foundation of America," 2014, [Online] http://www.alzfdn.org/AboutAlzheimers/definition.html, (Date retrieved: 2015-02-04).

[116] P. M. M. Morgado, "Automated Diagnosis of Alzheimer's Disease using PET Images: A study of alternative procedures for feature extraction and selection," Ph.D. dissertation, Higher technical institute, Technical University of Lisbon, 2012.

[117] Society of Nuclear Medicine and Molecular Imaging, "Alzheimer's Disease and Molecular Imaging : Get the Facts," Society of Nuclear Medicine and Molecular Imaging, Reston, Virginia, Tech. Rep. June, 2009.

[118] J. Gaugler, B. James, T. Johnson, K. Scholz, and J. Weuve, "Alzheimer's Disease Facts and Figures," Alzheimer's Association, Chicago, Tech. Rep. 2, 2014.

[119] N.-J. Gong, C.-S. Wong, C.-C. Chan, L.-M. Leung, and Y.-C. Chu, "Correlations between microstructural alterations and severity of cognitive deficiency in Alzheimer's disease and mild cognitive impairment: a diffusional kurtosis imaging study," *Magnetic Resonance Imaging*, vol. 31, no. 5, pp. 688–694, jun 2013.

[120] L. O'Dwyer, F. Lamberton, A. L. W. Bokde, M. Ewers, Y. O. Faluyi, C. Tanner, B. Mazoyer, D. O'Neill, M. Bartley, D. R. Collins, T. Coughlan, D. Prvulovic, and H. Hampel, "Using Support Vector Machines with Multiple Indices of Diffusion for Automated Classification of Mild Cognitive Impairment," *PLoS ONE*, vol. 7, no. 2, p. e32441, feb 2012.

[121] T. Varghese, R. Sheelakumari, J. S. James, and P. Mathuranath, "A review of neuroimaging biomarkers of Alzheimer's disease." *Neurology Asia*, vol. 18, no. 3, pp. 239–248, 2013.

[122] L. K. McEvoy, C. Fennema-Notestine, J. C. Roddey, D. J. Hagler, D. Holland, D. S. Karow, C. J. Pung, J. B. Brewer, and A. M. Dale, "Alzheimer Disease: Quantitative Structural Neuroimaging for Detection and Prediction of Clinical and Structural Changes in Mild Cognitive Impairment 1," *Radiology*, vol. 251, no. 1, pp. 195–205, apr 2009.

[123] F. Raudino, "Non-cognitive symptoms and related conditions in the Alzheimer's disease: a literature review," *Neurological Sciences*, vol. 34, no. 8, pp. 1275–1282, aug 2013.

[124] Alzheimer's Association, "Basics of Alzheimer's Disease: What it is and what you can do," Alzheimer's Association, Chicago, Tech. Rep., 2013, [Online] http://www.alz.org/documents/national/brochure_basicsofalz_low.pdf, (Date retrieved: 2015-02-20).

[125] O. Ben Ahmed, J. Benois-Pineau, M. Allard, C. Ben Amar, and G. Catheline, "Classification of Alzheimer's disease subjects from MRI using hippocampal visual features," *Multimedia Tools and Applications*, vol. 74, no. 4, pp. 1249–1266, feb 2015.

[126] F. Shi, B. Liu, Y. Zhou, C. Yu, and T. Jiang, "Hippocampal volume and asymmetry in mild cognitive impairment and Alzheimer's disease: Meta-analyses of MRI studies," *Hippocampus*, vol. 19, no. 11, pp. 1055–1064, nov 2009.

[127] K. López-de Ipiña, J.-B. Alonso, C. Travieso, J. Solé-Casals, H. Egiraun, M. Faundez-Zanuy, A. Ezeiza, N. Barroso, M. Ecay-Torres, P. Martinez-Lage, and U. Lizardui, "On the Selection of Non-Invasive Methods Based on Speech Analysis Oriented to Automatic Alzheimer Disease Diagnosis," *Sensors*, vol. 13, no. 5, pp. 6730–6745, may 2013.

[128] K. Bethune, "Diagnosis and Treatment of Alzheimer's Disease: Current Challenges," Outstanding Honors Theses. Paper 42., University of South Florida (USF), 2010.

[129] T. Mullerthomsen, S. Arlt, U. Mann, R. Mas, and S. Ganzer, "Detecting depression in Alzheimer's disease: evaluation of four different scales," *Archives of Clinical Neuropsychology*, vol. 20, no. 2, pp. 271–276, mar 2005.

[130] R. Craig-Schapiro, A. M. Fagan, and D. M. Holtzman, "Biomarkers of Alzheimer's disease," *Neurobiology of Disease*, vol. 35, no. 2, pp. 128–140, aug 2009.

[131] H. M. Snyder, M. C. Carrillo, F. Grodstein, K. Henriksen, A. Jeromin, S. Lovestone, M. M. Mielke, S. O'Bryant, M. Sarasa, M. Sjøgren, H. Soares, J. Teeling, E. Trushina, M. Ward, T. West, L. J. Bain, D. W. Shineman, M. Weiner, and H. M. Fillit, "Developing novel blood-based biomarkers for Alzheimer's disease," *Alzheimer's & Dementia*, vol. 10, no. 1, pp. 109–114, jan 2014.

[132] C. Ibáñez, C. Simó, D. K. Barupal, O. Fiehn, M. Kivipelto, A. Cedazo-Mínguez, and A. Cifuentes, "A new metabolomic workflow for early detection of Alzheimer's disease," *Journal of Chromatography A*, vol. 1302, pp. 65–71, aug 2013.

[133] D. Ferreira, L. Perestelo-Pérez, E. Westman, L.-O. Wahlund, A. SarríÂa, and P. Serrano-Aguilar, "Meta-Review of CSF Core Biomarkers in Alzheimer's Disease: The State-of-the-Art after the New Revised Diagnostic Criteria," *Frontiers in Aging Neuroscience*, vol. 6, no. March, pp. 1–24, mar 2014.

[134] R. J. L. Kandimalla, S. Prabhakar, W. Y. Wani, A. Kaushal, N. Gupta, D. R. Sharma, V. K. Grover, N. Bhardwaj, K. Jain, and K. D. Gill, "CSF p-Tau levels in the prediction of Alzheimer's disease," *Biology Open*, vol. 2, no. 11, pp. 1119–1124, nov 2013.

[135] T. J. Montine, E. R. Peskind, J. F. Quinn, A. M. Wilson, K. S. Montine, and D. Galasko, "Increased Cerebrospinal Fluid F2-Isoprostanes are Associated with Aging and Latent Alzheimer's Disease as Identified by Biomarkers," *NeuroMolecular Medicine*, vol. 13, no. 1, pp. 37–43, mar 2011.

[136] L. M. Shaw, H. Vanderstichele, M. Knapik-Czajka, C. M. Clark, P. S. Aisen, R. C. Petersen, K. Blennow, H. Soares, A. Simon, P. Lewczuk, R. Dean, E. Siemers, W. Potter, V. M.-Y. Lee, and J. Q. Trojanowski, "Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects," *Annals of Neurology*, vol. 65, no. 4, pp. 403–413, apr 2009.

[137] L. Khedher, J. Ramírez, J. Górriz, A. Brahim, and F. Segovia, "Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images," *Neurocomputing*, vol. 151, pp. 139–150, mar 2015.

[138] S. Kitamura, "Characteristic Diagnostic Imaging Findings in Alzheimer's Disease," *Japan Medical Association Journal (JMAJ)*, vol. 46, no. 6, pp. 269–276, 2003.

[139] P. Vemuri and C. R. Jack, "Role of structural MRI in Alzheimer's disease," *Alzheimer's Research & Therapy*, vol. 2, no. 4, p. 23, 2010.

[140] F. Yepes-Calderon, F. Pedregosa, B. Thirion, Y. Wang, and N. Lepore, "Automatic pathology classification using a single feature machine learning support - vector machines," in *Proc. SPIE 9035, Medical Imaging 2014: Computer-Aided Diagnosis, 903524 (March 24, 2014)*, S. Aylward and L. M. Hadjiiski, Eds. Bellingham, Washington USA: Society of Photo-Optical Instrumentation Engineers (SPIE), mar 2014.

[141] W. Yang, X. Chen, H. Xie, and X. Huang, "ICA-Based Automatic Classification of Magnetic Resonance Images from ADNI Data," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, K. Li, L. Jia,

X. Sun, M. Fei, and G. W. Irwin, Eds. Berlin: Springer Berlin Heidelberg, 2010, vol. 6330 LNBI, pp. 340–347.

[142] N. Villain, B. Desgranges, F. Viader, V. de la Sayette, F. Mezenge, B. Landeau, J.-C. Baron, F. Eustache, and G. Chetelat, "Relationships between Hippocampal Atrophy, White Matter Disruption, and Gray Matter Hypometabolism in Alzheimer's Disease," *Journal of Neuroscience*, vol. 28, no. 24, pp. 6174–6181, jun 2008.

[143] S. Duchesne, A. Caroli, C. Geroldi, C. Barillot, G. B. Frisoni, and D. L. Collins, "MRI-Based Automated Computer Classification of Probable AD Versus Normal Controls," *IEEE Transactions on Medical Imaging*, vol. 27, no. 4, pp. 509–520, apr 2008.

[144] R. Mahmood and B. Ghimire, "Automatic detection and classification of Alzheimer's Disease from MRI scans using principal component analysis and artificial neural networks," in *2013 20th International Conference on Systems, Signals and Image Processing (IWSSIP)*. IEEE, jul 2013, pp. 133–137.

[145] R. Wolz, V. Julkunen, J. Koikkalainen, E. Niskanen, D. P. Zhang, D. Rueckert, H. Soininen, and J. Lötjönen, "Multi-Method Analysis of MRI Images in Early Diagnostics of Alzheimer's Disease," *PLoS ONE*, vol. 6, no. 10, p. e25446, oct 2011.

[146] A. Savio, M. García-Sebastián, C. Hernández, M. Graña, J. Villanúa, E. Corchado, and H. Yin, *Intelligent Data Engineering and Automated Learning - IDEAL 2009*, ser. Lecture Notes in Computer Science, E. Corchado and H. Yin, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, vol. 5788.

[147] A. Rueda, J. Arevalo, A. Cruz, E. Romero, and F. A. González, "Bag of Features for Automatic Classification of Alzheimer's Disease in Magnetic Resonance Images," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, L. Alvarez, M. Mejail, L. Gomez, and J. Jacobo, Eds. Berlin: Springer Berlin Heidelberg, 2012, pp. 559–566.

[148] B. C. Dickerson, D. H. Salat, D. N. Greve, E. F. Chua, E. Rand-Giovannetti, D. M. Rentz, L. Bertram, K. Mullin, R. E. Tanzi, D. Blacker, M. S. Albert, and R. a. Sperling, "Increased hippocampal activation in mild cognitive impairment compared to normal aging and AD," *Neurology*, vol. 65, no. 3, pp. 404–411, aug 2005.

[149] A. Hämäläinen, M. Pihlajamäki, H. Tanila, T. Hänninen, E. Niskanen, S. Tervo, P. A. Karjalainen, R. L. Vanninen, and H. Soininen, "Increased fMRI responses during encoding in mild cognitive impairment," *Neurobiology of Aging*, vol. 28, no. 12, pp. 1889–1903, dec 2007.

[150] K. a. Celone, V. D. Calhoun, B. C. Dickerson, A. Atri, E. F. Chua, S. L. Miller, K. DePeau, D. M. Rentz, D. J. Selkoe, D. Blacker, M. S. Albert, and R. a. Sperling, "Alterations in Memory Networks in Mild Cognitive Impairment and Alzheimer's Disease: An Independent Component Analysis," *Journal of Neuroscience*, vol. 26, no. 40, pp. 10 222–10 231, oct 2006.

[151] A. Golby, "Memory encoding in Alzheimer's disease: an fMRI study of explicit and implicit memory," *Brain*, vol. 128, no. 4, pp. 773–787, feb 2005.

[152] R. a. Sperling, "fMRI studies of associative encoding in young and elderly controls and mild Alzheimer's disease," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 74, no. 1, pp. 44–50, jan 2003.

[153] C. L. Grady, A. R. McIntosh, S. Beig, M. L. Keightley, H. Burian, and S. E. Black, "Evidence from functional neuroimaging of a compensatory prefrontal network in Alzheimer's disease." *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 23, no. 3, pp. 986–93, feb 2003.

[154] C. Solé-Padullés, D. Bartrés-Faz, C. Junqué, P. Vendrell, L. Rami, I. C. Clemente, B. Bosch, A. Villar, N. Bargalló, M. A. Jurado, M. Barrios, and J. L. Molinuevo, "Brain structure and function related to cognitive reserve variables in normal aging, mild cognitive impairment and Alzheimer's disease," *Neurobiology of Aging*, vol. 30, no. 7, pp. 1114–1124, jul 2009.

[155] P. S. Miettinen, M. Pihlajamäki, A. M. Jauhiainen, E. Niskanen, T. Hänninen, R. Vanninen, and H. Soininen, "Structure and function of medial temporal and posteromedial cortices in early Alzheimer's disease." *The European journal of neuroscience*, vol. 34, no. 2, pp. 320–30, jul 2011.

[156] C. Lustig, A. Z. Snyder, M. Bhakta, K. C. O'Brien, M. McAvoy, M. E. Raichle, J. C. Morris, and R. L. Buckner, "Functional deactivations: Change with age and dementia of the Alzheimer type," *Proceedings of the National Academy of Sciences*, vol. 100, no. 24, pp. 14 504–14 509, nov 2003.

[157] A. Khazaee, A. Ebrahimzadeh, and A. Babajani-Feremi, "Automatic classification of Alzheimer's disease with resting-state fMRI and graph theory," in *2014 21th Iranian Conference on Biomedical Engineering (ICBME)*. IEEE, nov 2014, pp. 252–257.

[158] E. E. Tripoliti, D. I. Fotiadis, and M. Argyropoulou, "An automated supervised method for the diagnosis of Alzheimer's disease based on fMRI data using weighted voting schemes," in *2008 IEEE International Workshop on Imaging Systems and Techniques*. IEEE, sep 2008, pp. 340–345.

[159] ——, "A supervised method to assist the diagnosis and monitor progression of Alzheimer's disease using data from an fMRI experiment," *Artificial Intelligence in Medicine*, vol. 53, no. 1, pp. 35–45, sep 2011.

[160] S. Posse, R. Otazo, S. R. Dager, and J. Alger, "MR spectroscopic imaging: Principles and recent advances," *Journal of Magnetic Resonance Imaging*, vol. 37, no. 6, pp. 1301–1325, jun 2013.

[161] D. Graveron-Demilly, "Magnetic Resonance Spectroscopic Imaging," Laboratory CREATIS-LRMN, Université Lyon 1, Lyon, Tech. Rep., 2012.

[162] C. Loos, E. Achten, and P. Santens, "Proton magnetic resonance spectroscopy in Alzheimer's disease, a review." *Acta neurologica Belgica*, vol. 110, no. 4, pp. 291–8, dec 2010.

[163] K. Kantarci and Graff Radford, "Magnetic resonance spectroscopy in Alzheimer's disease," *Neuropsychiatric Disease and Treatment*, vol. 9, p. 687, may 2013.

[164] W. E. Klunk, K. Panchalingam, J. Moossy, R. J. McClure, and J. W. Pettegrew, "N-acetyl-L-aspartate and other amino acid metabolites in Alzheimer's disease brain: A preliminary proton nuclear magnetic resonance study," *Neurology*, vol. 42, no. 8, pp. 1578–1578, aug 1992.

[165] K. Kantarci, C. R. Jack, Y. C. Xu, N. G. Campeau, P. C. O'Brien, G. E. Smith, R. J. Ivnik, B. F. Boeve, E. Kokmen, E. G. Tangalos, and R. C. Petersen, "Regional metabolic patterns in Mild Cognitive Impairment and Alzheimer's Disease: A (1)H MRS study," *Neurology*, vol. 55, no. 2, pp. 210–217, jul 2000.

[166] F. Jessen, W. Block, F. Traber, E. Keller, S. Flacke, A. Papassotiropoulos, R. Lamerichs, R. Heun, and H. H. Schild, "Proton MR spectroscopy detects a relative decrease of N-acetylaspartate in the medial temporal lobe of patients with AD," *Neurology*, vol. 55, no. 5, pp. 684–688, sep 2000.

[167] W. Huang, G. E. Alexander, L. Chang, H. U. Shetty, J. S. Krasuski, S. I. Rapoport, and M. B. Schapiro, "Brain metabolite concentration and dementia severity in Alzheimer's disease: A 1H MRS study," *Neurology*, vol. 57, no. 4, pp. 626–632, aug 2001.

[168] P. G. Antuono, J. L. Jones, Y. Wang, and S.-J. Li, "Decreased glutamate + glutamine in Alzheimer's disease detected in vivo with 1H-MRS at 0.5 T," *Neurology*, vol. 56, no. 6, pp. 737–742, mar 2001.

[169] N. Hattori, K. Abe, S. Sakoda, and T. Sawada, "Proton MR spectroscopic study at 3 Tesla on glutamate/glutamine in Alzheimer's disease," *Neuroreport*, vol. 13, no. 1, pp. 183–186, jan 2002.

[170] R. Rupsingh, M. Borrie, M. Smith, J. Wells, and R. Bartha, "Reduced hippocampal glutamate in Alzheimer disease," *Neurobiology of Aging*, vol. 32, no. 5, pp. 802–810, may 2011.

[171] K. Kantarci, S. D. Weigand, R. C. Petersen, B. F. Boeve, D. S. Knopman, J. Gunter, D. Reyes, M. Shiung, P. C. O'Brien, G. E. Smith, R. J. Ivnik, E. G. Tangalos, and C. R. Jack, "Longitudinal 1H MRS changes in mild cognitive impairment and Alzheimer's disease," *Neurobiology of Aging*, vol. 28, pp. 1330–1339, 2007.

[172] E. J. Mason, M. J. Donahue, and B. A. Ally, *Using Magnetic Resonance Imaging in the Early Detection of Alzheimer's Disease*. InTech, feb 2013.

[173] S. Rose, G. de Zubicaray, D. Wang, G. Galloway, J. Chalk, S. Eagle, J. Semple, and D. Doddrell, "A 1H MRS study of probable Alzheimer's disease and normal aging: implications for longitudinal monitoring of dementia progression," *Magnetic Resonance Imaging*, vol. 17, no. 2, pp. 291–299, feb 1999.

[174] L. Parnetti, R. Tarducci, O. Presciutti, D. T. Lowenthal, M. Pippi, B. Palumbo, G. Gobbi, G. P. Pelliccioli, and U. Senin, "Proton magnetic resonance spectroscopy can differentiate Alzheimer's disease from normal aging." *Mechanisms of ageing and development*, vol. 97, no. 1, pp. 9–14, jul 1997.

[175] P. J. Modrego, N. Fayed, and M. A. Pina, "Conversion from mild cognitive impairment to probable Alzheimer's disease predicted by brain magnetic resonance spectroscopy," *American Journal of Psychiatry*, vol. 162, no. 4, pp. 667–675, 2005.

[176] A. Metastasio, P. Rinaldi, R. Tarducci, E. Mariani, F. T. Feliziani, A. Cherubini, G. P. Pelliccioli, G. Gobbi, U. Senin, and P. Mecocci, "Conversion of MCI to dementia: Role of proton magnetic resonance spectroscopy," *Neurobiology of Aging*, vol. 27, no. 7, pp. 926–932, jul 2006.

[177] L. Rami, B. Gómez-Anson, R. Sanchez-Valle, B. Bosch, G. C. Monte, A. Lladó, and J. L. Molinuevo, "Longitudinal Study of Amnesic Patients at High Risk for Alzheimer's Disease: Clinical, Neuropsychological and Magnetic Resonance Spectroscopy Features," *Dementia and Geriatric Cognitive Disorders*, vol. 24, no. 5, pp. 402–410, 2007.

[178] N. Fayed, J. Dávila, A. Oliveros, J. Castillo, and J. J. Medrano, "Utility of Different MR Modalities in Mild Cognitive Impairment and Its Use as a Predictor of Conversion to Probable Dementia," *Academic Radiology*, vol. 15, no. 9, pp. 1089–1098, sep 2008.

[179] P. J. Modrego, N. Fayed, and M. Sarasa, "Magnetic resonance spectroscopy in the prediction of early conversion from amnestic mild cognitive impairment to dementia: a prospective cohort study," *BMJ Open*, vol. 1, no. 1, pp. e000 007–e000 007, sep 2011.

[180] T. den Heijer, P. E. Sijens, N. D. Prins, A. Hofman, P. J. Koudstaal, M. Oudkerk, and M. Breteler, "MR spectroscopy of brain white matter in the prediction of dementia," *Neurology*, vol. 66, no. 4, pp. 540–544, feb 2006.

[181] K. Kantarci, S. D. Weigand, R. C. Petersen, B. F. Boeve, D. S. Knopman, J. Gunter, D. Reyes, M. Shiung, P. C. O'Brien, G. E. Smith, R. J. Ivnik, E. G. Tangalos, and C. R. Jack, "Longitudinal 1H MRS changes in mild cognitive impairment and Alzheimer's disease," *Neurobiology of Aging*, vol. 28, no. 9, pp. 1330–1339, sep 2007.

[182] U. Pilatus, C. Lais, A. d. M. de Rochmont, T. Kratzsch, L. Frölich, K. Maurer, F. E. Zanella, H. Lanfermann, and J. Pantel, "Conversion to dementia in mild cognitive impairment is associated with decline of N-actylaspartate and creatine as revealed by magnetic resonance spectroscopy," *Psychiatry Research: Neuroimaging*, vol. 173, no. 1, pp. 1–7, jul 2009.

[183] T. Ernst, L. Chang, R. Melchor, and C. M. Mehringer, "Frontotemporal dementia and early Alzheimer disease: differentiation with frontal lobe H-1 MR spectroscopy." *Radiology*, vol. 203, no. 3, pp. 829–36, jun 1997.

[184] F. Gao and P. B. Barker, "Various MRS Application Tools for Alzheimer Disease and Mild Cognitive Impairment," *American Journal of Neuroradiology*, vol. 35, no. Supplement 6, pp. S4–S11, jun 2014.

[185] A. E. Roher, Z. Garami, S. L. Tyas, C. L. Maarouf, T. a. Kokjohn, M. Belohlavek, L. J. Vedders, D. Connor, M. N. Sabbagh, T. G. Beach, and M. R. Emmerling, "Transcranial Doppler ultrasound blood flow velocity and pulsatility index as systemic indicators for Alzheimer's disease," *Alzheimer's & Dementia*, vol. 7, no. 4, pp. 445–455, jul 2011.

[186] S. E. Bullard, M. Griss, S. Greene, and A. Gekker, "Encyclopedia of Clinical Neuropsychology," *Archives of Clinical Neuropsychology*, vol. 28, no. 1, pp. 92–92, feb 2013.

[187] M. Silvestrini, B. Gobbi, P. Pasqualetti, M. Bartolini, R. Baruffaldi, C. Lanciotti, R. Cerqua, C. Altamura, L. Provinciali, and F. Vernieri, "Carotid atherosclerosis and cognitive decline in patients with Alzheimer's disease," *Neurobiology of Aging*, vol. 30, no. 8, pp. 1177–1183, aug 2009.

[188] M. van Oijen, F. Jan de Jong, J. C. M. Witteman, A. Hofman, P. J. Koudstaal, and M. M. B. Breteler, "Atherosclerosis and risk for dementia," *Annals of Neurology*, vol. 61, no. 5, pp. 403–410, may 2007.

[189] F. Doepp, J. M. Valdueza, and S. J. Schreiber, "Transcranial and extracranial ultrasound assessment of cerebral hemodynamics in vascular and Alzheimer's dementia," *Neurological Research*, vol. 28, no. 6, pp. 645–649, sep 2006.

[190] S. J. Schreiber, F. Doepp, E. Spruth, U. a. Kopp, and J. M. Valdueza, "Ultrasonographic measurement of cerebral blood flow, cerebral circulation time and cerebral blood volume in vascular and Alzheimer's dementia," *Journal of Neurology*, vol. 252, no. 10, pp. 1171–1177, oct 2005.

[191] A. Tomek, B. Urbanová, and J. Hort, "Utility of transcranial ultrasound in predicting Alzheimer's disease risk." *Journal of Alzheimer's disease : JAD*, vol. 42 Suppl 4, pp. S365–74, 2014.

[192] A. Stefani, G. Sancesario, M. Pierantozzi, G. Leone, S. Galati, A. H. Hainsworth, and M. Diomedi, "CSF biomarkers, impairment of cerebral hemodynamics and degree of cognitive decline in Alzheimer's and mixed dementia," *Journal of the Neurological Sciences*, vol. 283, no. 1-2, pp. 109–115, aug 2009.

[193] J. A. H. R. Claassen, R. Diaz-Arrastia, K. Martin-Cook, B. D. Levine, and R. Zhang, "Altered cerebral hemodynamics in early Alzheimer disease: a pilot study using transcranial Doppler." *Journal of Alzheimer's disease : JAD*, vol. 17, no. 3, pp. 621–9, 2009.

[194] A. Vilanova, S. Zhang, G. Kindlmann, and D. Laidlaw, "An Introduction to Visualization of Diffusion Tensor Imaging and Its Applications," in *In Visualization and Processing of Tensor Fields*, P. J. Weickert and P. H. Hagen, Eds. Berlin: Springer Berlin Heidelberg, 2006, pp. 121–153.

[195] L. d. G. Vasconcelos, S. M. D. Brucki, A. P. Jackowiski, and O. F. A. Bueno, "Diffusion Tensor Imaging in Alzheimer's disease," *Dement Neuropsychol*, vol. 3, no. 4, pp. 268–274, 2009.

[196] M. Bozzali and A. Cherubini, "Diffusion tensor MRI to investigate dementias: a brief review," *Magnetic Resonance Imaging*, vol. 25, no. 6, pp. 969–977, jul 2007.

[197] L. Minati, M. Grisoli, and M. G. Bruzzone, "MR Spectroscopy, Functional MRI, and Diffusion-Tensor Imaging in the Aging Brain: A Conceptual Review," *Journal of Geriatric Psychiatry and Neurology*, vol. 20, no. 1, pp. 3–21, mar 2007.

[198] H. Hanyu, T. Asano, H. Sakurai, Y. Imon, T. Iwamoto, M. Takasaki, H. Shindo, and K. Abe, "Diffusion-weighted and magnetization transfer imaging of the corpus callosum in Alzheimer's disease." *Journal of the neurological sciences*, vol. 167, no. 1, pp. 37–44, aug 1999.

[199] L. Clerx, P. J. Visser, F. Verhey, and P. Aalten, "New MRI markers for Alzheimer's disease: a meta-analysis of diffusion tensor imaging and a comparison with medial temporal lobe measurements." *Journal of Alzheimer's disease : JAD*, vol. 29, no. 2, pp. 405–29, 2012.

[200] M. Dyrba, M. Ewers, M. Wegrzyn, I. Kilimann, C. Plant, A. Oswald, T. Meindl, M. Pievani, A. L. W. Bokde, A. Fellgiebel, M. Filippi, H. Hampel, S. Klöppel, K. Hauenstein, T. Kirste, and S. J. Teipel, "Robust Automated Detection of Microstructural White Matter Degeneration in Alzheimer's Disease Using Machine Learning Classification of Multicenter DTI Data," *PLoS ONE*, vol. 8, no. 5, p. e64925, may 2013.

[201] C.-Y. Wee, P.-T. Yap, D. Zhang, K. Denny, J. N. Browndyke, G. G. Potter, K. a. Welsh-Bohmer, L. Wang, and D. Shen, "Identification of MCI individuals using structural and functional connectivity networks," *NeuroImage*, vol. 59, no. 3, pp. 2045–2056, feb 2012.

[202] H. Yasmin, Y. Nakata, S. Aoki, O. Abe, N. Sato, K. Nemoto, K. Arima, N. Furuta, M. Uno, S. Hirai, Y. Masutani, and K. Ohtomo, "Diffusion abnormalities of the uncinate fasciculus in Alzheimer's disease: diffusion tensor tract-specific analysis using a new method to measure the core of the tract," *Neuroradiology*, vol. 50, no. 4, pp. 293–299, apr 2008.

[203] M. E. Phelps, "PET: the merging of biology and imaging into molecular imaging." *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, vol. 41, no. 4, pp. 661–81, apr 2000.

[204] T. G. Turkington, *Clinical PET-CT in Radiology*, P. Shreve and D. W. Townsend, Eds. New York, NY: Springer New York, 2011.

[205] A. Nordberg, J. O. Rinne, A. Kadir, and B. Långström, "The use of PET in Alzheimer disease," *Nature Reviews Neurology*, vol. 6, no. 2, pp. 78–87, feb 2010.

[206] F. Nobili and S. Morbelli, "[18F]FDG-PET as a Biomarker for Early Alzheimer's Disease," *The Open Nuclear Medicine Journal*, vol. 2, no. 1, pp. 46–52, nov 2010.

[207] L. Sokoloff, "Relation Between Physiological Function and Energy Metabolism in the Central Nervous System," *Journal of Neurochemistry*, vol. 29, no. 1, pp. 13–26, jul 1977.

[208] S.-s. Su, K.-w. Chen, and Q. Huang, "Discriminant analysis in the study of Alzheimer's disease using feature extractions and support vector machines in positron emission tomography with 18F-FDG," *Journal of Shanghai Jiaotong University (Science)*, vol. 19, no. 5, pp. 555–560, oct 2014.

[209] H. Dehghan, "Detection of Alzheimer's Disease Using Multitracer Positron Emission Tomography Imaging," *International Journal of Engineering*, vol. 27, no. 1 (A), pp. 51–56, jan 2014.

[210] M. López, J. Ramírez, J. Górriz, I. Álvarez, D. Salas-Gonzalez, F. Segovia, R. Chaves, P. Padilla, and M. Gómez-Río, "Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer's disease," *Neurocomputing*, vol. 74, no. 8, pp. 1260–1271, mar 2011.

[211] I. Illán, J. Górriz, J. Ramírez, D. Salas-Gonzalez, M. López, F. Segovia, R. Chaves, M. Gómez-Rio, and C. Puntonet, "18F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis," *Information Sciences*, vol. 181, no. 4, pp. 903–916, feb 2011.

[212] M. López, J. Ramírez, J. Górriz, D. Salas-Gonzalez, I. Álvarez, F. Segovia, and C. Puntonet, "Automatic tool for Alzheimer's disease diagnosis using PCA and Bayesian classification rules," *Electronics Letters*, vol. 45, no. 8, p. 389, 2009.

[213] C. Turkington and D. Mitchell, *The Encyclopedia of Alzheimer's Disease*, 2nd ed. New York, USA: Facts On File Inc, 2003, 2010.

[214] M. M. Svedberg, O. Rahman, and H. Hall, "Preclinical studies of potential amyloid binding PET/SPECT ligands in Alzheimer's disease," *Nuclear Medicine and Biology*, vol. 39, no. 4, pp. 484–501, may 2012.

[215] J. M. Yeo, X. Lim, Z. Khan, and S. Pal, "Systematic review of the diagnostic utility of SPECT imaging in dementia," *European Archives of Psychiatry and Clinical Neuroscience*, vol. 263, no. 7, pp. 539–552, oct 2013.

[216] I. Goethals, C. Van De Wiele, D. Slosman, and R. Dierckx, "Brain SPET perfusion in early Alzheimer's disease: where to look?" *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 29, no. 8, pp. 975–978, aug 2002.

[217] J. W. Ashford, W. J. Shih, J. Coupal, R. Shetty, A. Schneider, C. Cool, A. Aleem, V. H. Kiefer, M. S. Mendiondo, and F. a. Schmitt, "Single SPECT measures of cerebral cortical perfusion reflect time-index estimation of dementia severity in Alzheimer's disease." *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, vol. 41, no. 1, pp. 57–64, jan 2000.

[218] M.-O. Habert, L. C. de Souza, F. Lamari, N. Daragon, S. Desarnaud, C. Jardel, B. Dubois, and M. Sarazin, "Brain perfusion SPECT correlates with CSF biomarkers in Alzheimer's disease," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 37, no. 3, pp. 589–593, mar 2010.

[219] R. Chaves, J. Ramírez, J. Górriz, M. López, D. Salas-Gonzalez, I. Álvarez, and F. Segovia, "SVM-based computer-aided diagnosis of the Alzheimer's disease using t-test NMSE feature selection with feature correlation weighting," *Neuroscience Letters*, vol. 461, no. 3, pp. 293–297, sep 2009.

[220] D. Kogure, H. Matsuda, T. Ohnishi, T. Asada, M. Uno, T. Kunihiro, S. Nakano, and M. Takasaki, "Longitudinal evaluation of early Alzheimer's disease using brain perfusion SPECT." *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, vol. 41, no. 7, pp. 1155–62, jul 2000.

[221] H. Braak and E. Braak, "Neuropathological stageing of Alzheimer-related changes," *Acta Neuropathologica*, vol. 82, no. 4, pp. 239–259, sep 1991.

[222] J. Ramírez, J. Górriz, D. Salas-Gonzalez, A. Romero, M. López, I. Álvarez, and M. Gómez-Río, "Computer-aided diagnosis of Alzheimer's type dementia combining support vector machines and discriminant set of features," *Information Sciences*, vol. 237, pp. 59–72, jul 2013.

[223] I. Illán, J. Górriz, M. López, J. Ramírez, D. Salas-Gonzalez, F. Segovia, R. Chaves, and C. Puntonet, "Computer aided diagnosis of Alzheimer's disease using component based SVM," *Applied Soft Computing*, vol. 11, no. 2, pp. 2376–2382, mar 2011.

[224] J. Górriz, F. Segovia, J. Ramírez, A. Lassl, and D. Salas-Gonzalez, "GMM based SPECT image classification for the diagnosis of Alzheimer's disease," *Applied Soft Computing*, vol. 11, no. 2, pp. 2313–2325, mar 2011.

[225] H. Matsuda, "Role of Neuroimaging in Alzheimer's Disease, with Emphasis on Brain Perfusion SPECT," *Journal of Nuclear Medicine*, vol. 48, no. 8, pp. 1289–1300, aug 2007.

[226] M. Weih, Ü. Degirmenci, S. Kreil, P. Lewczuk, D. Schmidt, J. Kornhuber, and T. Kuwert, "Perfusion Imaging with SPECT in the Era of Pathophysiology-Based Biomarkers for Alzheimer's Disease," *International Journal of Alzheimer's Disease*, vol. 2010, pp. 1–5, 2010.

[227] J. Dauwels and S. Kannan, "Diagnosis of Alzheimer's Disease using Electric Signals of the Brain. A Grand Challenge," *Asia-Pacific Biotech News*, vol. 16, no. 10n11, pp. 22–38, oct 2012.

[228] J. Dauwels, F. Vialatte, and a. Cichocki, "On the Early Diagnosis of Alzheimer's Disease from EEG Signals: A Mini-Review," *Advances in Cognitive Neurodynamics (II)*, no. Section 2, 2011.

[229] P. M. Rossini, M. Buscema, M. Capriotti, E. Grossi, G. Rodriguez, C. Del Percio, and C. Babiloni, "Is it possible to automatically distinguish resting EEG data of normal elderly vs. mild cognitive impairment subjects with high degree of accuracy?" *Clinical Neurophysiology*, vol. 119, no. 7, pp. 1534–1545, jul 2008.

[230] M. Buscema, P. Rossini, C. Babiloni, and E. Grossi, "The IFAST model, a novel parallel nonlinear EEG analysis technique, distinguishes mild cognitive impairment and Alzheimer's disease patients with high degree of accuracy," *Artificial Intelligence in Medicine*, vol. 40, no. 2, pp. 127–141, jun 2007.

[231] L. R. Trambaiolli, A. C. Lorena, F. J. Fraga, P. A. M. Kanda, R. Anghinah, and R. Nitrini, "Improving Alzheimer's Disease Diagnosis with Machine Learning Techniques," *Clinical EEG and Neuroscience*, vol. 42, no. 3, pp. 160–165, jul 2011.

[232] D. Al-Jumeily, S. Iram, F.-b. Vialatte, P. Fergus, and A. Hussain, "A novel method of early diagnosis of Alzheimer's disease based on EEG signals." *The Scientific World Journal*, vol. 2015, p. 931387, 2015.

[233] F. C. Morabito, "The compressibility of an electroencephalography signal may indicate Alzheimer's disease," *SPIE Newsroom*, pp. 3–5, jun 2013.

[234] J. Dauwels, F. Vialatte, and A. Cichocki, "Diagnosis of Alzheimer's Disease from EEG Signals: Where Are We Standing?" *Current Alzheimer Research*, vol. 999, no. 999, pp. 1–19, jun 2010.

[235] H. W. Berendse, J. P. Verbunt, P. Scheltens, B. W. van Dijk, and E. J. Jonkman, "Magnetoencephalographic analysis of cortical activity in Alzheimer's disease: a pilot study." *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, vol. 111, no. 4, pp. 604–12, apr 2000.

[236] A. Fernández, R. Hornero, A. Mayo, J. Poza, P. Gil-Gregorio, and T. Ortiz, "MEG spectral profile in Alzheimer's disease and mild cognitive impairment," *Clinical Neurophysiology*, vol. 117, no. 2, pp. 306–314, feb 2006.

[237] B. Wan, D. Ming, X. Fu, C. Yang, H. Qi, and B. Chen, "Study on a quantitative electroencephalography power spectrum typical of Chinese Han Alzheimer's disease patients by using wavelet transforms," *Journal of Neural Engineering*, vol. 3, no. 1, pp. 71–77, mar 2006.

[238] J. Poza, R. Hornero, D. Abásolo, A. Fernández, and M. García, "Extraction of spectral based measures from MEG background oscillations in Alzheimer's disease," *Medical Engineering & Physics*, vol. 29, no. 10, pp. 1073–1083, dec 2007.

[239] T. Montez, S.-S. Poil, B. F. Jones, I. Manshanden, J. P. a. Verbunt, B. W. van Dijk, A. B. Brussaard, A. van Ooyen, C. J. Stam, P. Scheltens, and K. Linkenkaer-Hansen, "Altered temporal correlations in parietal alpha and prefrontal theta oscillations in early-stage Alzheimer disease," *Proceedings of the National Academy of Sciences*, vol. 106, no. 5, pp. 1614–1619, feb 2009.

[240] C. Stam, "Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field," *Clinical Neurophysiology*, vol. 116, no. 10, pp. 2266–2301, oct 2005.

[241] J. F. Alonso, J. Poza, M. Á. Mañanas, S. Romero, A. Fernández, and R. Hornero, "MEG Connectivity Analysis in Patients with Alzheimer's Disease Using Cross Mutual Information and Spectral Coherence," *Annals of Biomedical Engineering*, vol. 39, no. 1, pp. 524–536, jan 2011.

[242] D. Osipova, K. Rantanen, J. Ahveninen, R. Ylikoski, O. Häppölä, T. Strandberg, and E. Pekkonen, "Source estimation of spontaneous MEG oscillations in mild cognitive impairment," *Neuroscience Letters*, vol. 405, no. 1-2, pp. 57–61, sep 2006.

[243] A. Fernández, R. Hornero, C. Gómez, A. Turrero, P. Gil-Gregorio, J. Matías-Santos, and T. Ortiz, "Complexity Analysis of Spontaneous Brain Activity in Alzheimer Disease and Mild Cognitive Impairment," *Alzheimer Disease & Associated Disorders*, vol. 24, no. 2, pp. 182–189, apr 2010.

[244] C. Gomez, M. Martinez-Zarzuela, J. Poza, F. J. Diaz-Pernas, A. Fernandez, and R. Hornero, "Synchrony analysis of spontaneous MEG activity in Alzheimer's disease patients," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, aug 2012, pp. 6188–6191.

[245] C. Gómez, R. Hornero, D. Abásolo, A. Fernández, and J. Escudero, "Analysis of MEG Background Activity in Alzheimer's Disease Using Nonlinear Methods and ANFIS," *Annals of Biomedical Engineering*, vol. 37, no. 3, pp. 586–594, mar 2009.

[246] R. Hornero, J. Escudero, A. Fernández, J. Poza, and C. Gómez, "Spectral and nonlinear analyses of MEG background activity in patients with Alzheimer's disease," *IEEE Transactions on Biomedical Engineering*, vol. 55, pp. 1658–1665, 2008.

[247] C. Gómez, D. Abásolo, J. Poza, A. Fernández, and R. Hornero, "MEG analysis in Alzheimer's disease computing approximate entropy for different frequency bands," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'10*, 2010, pp. 2379–2382.

[248] J. Escudero, R. Hornero, J. Poza, D. Abásolo, and A. Fernández, "Assessment of classification improvement in patients with Alzheimer's disease based on magnetoencephalogram blind source separation," *Artificial Intelligence in Medicine*, vol. 43, pp. 75–85, 2008.

[249] M. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa, "Magnetoencephalography-theory, instrumentation, and applications to noninvasive studies of the working human brain," *Reviews of Modern Physics*, vol. 65, no. 2, pp. 413–497, apr 1993.

[250] J. Currie, B. Ramsden, C. McArthur, and P. Maruff, "Validation of a Clinical Antisaccadic Eye Movement Test in the Assessment of Dementia," *Archives of Neurology*, vol. 48, no. 6, pp. 644–648, jun 1991.

[251] R. J. Molitor, P. C. Ko, and B. A. Ally, "Eye movements in Alzheimer's disease." *Journal of Alzheimer's disease : JAD*, vol. 44, no. 1, pp. 1–12, jan 2015.

[252] I. E. P. Stolerman, *Encyclopedia of Psychopharmacology*, 1st ed., I. P. Stolerman, Ed. Berlin: Springer-Verlag Berlin Heidelberg, 2010.

[253] T. J. Anderson and M. R. MacAskill, "Eye movements in patients with neurodegenerative disorders," *Nature Reviews Neurology*, vol. 9, no. 2, pp. 74–85, jan 2013.

[254] L. D. Kaufman, J. Pratt, B. Levine, and S. E. Black, "Executive deficits detected in mild Alzheimer's disease using the antisaccade task," *Brain and Behavior*, vol. 2, no. 1, pp. 15–21, jan 2012.

[255] D. Lagun, C. Manzanares, S. M. Zola, E. A. Buffalo, and E. Agichtein, "Detecting cognitive impairment by eye movement analysis using automatic classification algorithms," *Journal of Neuroscience Methods*, vol. 201, no. 1, pp. 196–203, sep 2011.

[256] Q. Yang, T. Wang, N. Su, Y. Liu, S. Xiao, and Z. Kapoula, "Long Latency and High Variability in Accuracy-Speed of Prosaccades in Alzheimer's Disease at Mild to Moderate Stage," *Dementia and Geriatric Cognitive Disorders Extra*, vol. 1, no. 1, pp. 318–329, 2011.

[257] Q. Yang, T. Wang, N. Su, S. Xiao, and Z. Kapoula, "Specific saccade deficits in patients with Alzheimer's disease at mild to moderate stage and in patients with amnestic mild cognitive impairment," *AGE*, vol. 35, no. 4, pp. 1287–1298, aug 2013.

[258] M. D. Crutcher, R. Calhoun-Haney, C. M. Manzanares, J. J. Lah, A. I. Levey, and S. M. Zola, "Eye Tracking During a Visual Paired Comparison Task as a Predictor of Early Dementia," *American Journal of Alzheimer's Disease and Other Dementias*, vol. 24, no. 3, pp. 258–266, jun 2009.

[259] M. D'Esposito, E. Zarahn, G. K. Aguirre, and B. Rypma, "The Effect of Normal Aging on the Coupling of Neural Activity to the Bold Hemodynamic Response," *NeuroImage*, vol. 10, no. 1, pp. 6–14, jul 1999.

[260] C. Gomez and R. Hornero, "Entropy and Complexity Analyses in Alzheimer's Disease: An MEG Study," *The Open Biomedical Engineering Journal*, vol. 4, no. 1, pp. 223–235, mar 2010.

[261] C. Gómez, A. Mediavilla, R. Hornero, D. Abásolo, and A. Fernández, "Use of the Higuchi's fractal dimension for the analysis of MEG recordings from Alzheimer's disease patients." *Medical engineering & physics*, vol. 31, no. 3, pp. 306–13, apr 2009.

[262] A. Khodabakhsh, S. Kusxuoglu, and C. Demiroglu, "Natural language features for detection of Alzheimer's disease in conversational speech," in *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE, jun 2014, pp. 581–584.

[263] K. Lopez-de ipina, J. Sol, J. B. Alonso, and C. M. Travieso, *Transactions on Computational Collective Intelligence XVII*, ser. Lecture Notes in Computer Science, N. T. Nguyen, R. Kowalczyk, A. Fred, and F. Joaquim, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, vol. 8790.

[264] J. C. McBride, X. Zhao, N. B. Munro, C. D. Smith, G. a. Jicha, L. Hively, L. S. Broster, F. a. Schmitt, R. J. Kryscio, and Y. Jiang, "Spectral and complexity analysis of scalp EEG characteristics for mild cognitive impairment and early Alzheimer's disease." *Computer methods and programs in biomedicine*, vol. 114, no. 2, pp. 153–63, apr 2014.

[265] B. Reisberg, J. Borenstein, S. P. Salob, S. H. Ferris, E. Franssen, and A. Georgotas, "Behavioral symptoms in Alzheimer's disease: phenomenology and treatment." *The Journal of clinical psychiatry*, vol. 48 Suppl, pp. 9–15, may 1987.

[266] J. E. Overall and D. R. Gorham, "The Brief Psychiatric Rating Scale," *Psychological Reports*, vol. 10, no. 3, pp. 799–812, jun 1962.

[267] P. N. Tariot, J. L. Mack, M. B. Patterson, S. D. Edland, M. F. Weiner, G. Fillenbaum, L. Blazina, L. Teri, E. Rubin, and J. A. Mortimer, "The Behavior Rating Scale for Dementia of the Consortium to Establish a Registry for Alzheimer's Disease. The Behavioral Pathology Committee of the Consortium to Establish a Registry for Alzheimer's Disease." *The American journal of psychiatry*, vol. 152, no. 9, pp. 1349–57, sep 1995.

[268] C. Steckl, "Diagnosis Of Alzheimer's Disease - Neuropsychological Testing," 2008, [Online] https://www.mentalhelp.net/articles/diagnosis-of-alzheimer-s-disease-neuropsychological-testing/, (Date retrieved: 2015-04-14).

[269] M. Baumgarten, R. Becker, and S. Gauthier, "Validity and reliability of the dementia behavior disturbance scale." *Journal of the American Geriatrics Society*, vol. 38, no. 3, pp. 221–6, mar 1990.

[270] P. Robert, S. Ferris, S. Gauthier, R. Ihl, B. Winblad, and F. Tennigkeit, "Review of Alzheimer's disease scales: is there a need for a new multi-domain scale for therapy evaluation in medical practice?" *Alzheimer's Research & Therapy*, vol. 2, no. 4, p. 24, jan 2010.

[271] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "Mini-mental state," *Journal of Psychiatric Research*, vol. 12, no. 3, pp. 189–198, nov 1975.

[272] G. M. Peavy, D. P. Salmon, V. A. Rice, D. Galasko, W. Samuel, K. I. Taylor, C. Ernesto, N. Butters, and L. Thal, "Neuropsychological assessment of severely demeted elderly: the severe cognitive impairment profile." *Archives of neurology*, vol. 53, no. 4, pp. 367–72, apr 1996.

[273] W. G. Rosen, R. C. Mohs, and K. L. Davis, "A new rating scale for Alzheimer's disease." *The American journal of psychiatry*, vol. 141, no. 11, pp. 1356–64, nov 1984.

[274] J. Harrison, S. L. Minassian, L. Jenkins, R. S. Black, M. Koller, and M. Grundman, "A neuropsychological test battery for use in Alzheimer disease clinical trials." *Archives of neurology*, vol. 64, no. 9, pp. 1323–9, sep 2007.

[275] J. Saxton and A. A. Swihart, "Neuropsychological assessment of the severely impaired elderly patient." *Clinics in geriatric medicine*, vol. 5, no. 3, pp. 531–43, aug 1989.

[276] C. Baum and D. F. Edwards, "Cognitive Performance in Senile Dementia of the Alzheimer's Type: The Kitchen Task Assessment," *American Journal of Occupational Therapy*, vol. 47, no. 5, pp. 431–436, may 1993.

[277] S. Katz, "Studies of Illness in the Aged," *JAMA*, vol. 185, no. 12, p. 914, sep 1963.

[278] D. Galasko, D. Bennett, M. Sano, C. Ernesto, R. Thomas, M. Grundman, and S. Ferris, "An inventory to assess activities of daily living for clinical trials in Alzheimer's disease. The Alzheimer's Disease Cooperative Study." *Alzheimer disease and associated disorders*, vol. 11 Suppl 2, pp. S33–9, 1997.

[279] P. Rashidi and A. Mihailidis, "A Survey on Ambient-Assisted Living Tools for Older Adults," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 3, pp. 579–590, may 2013.

[280] A. Serna, H. Pigot, and V. Rialle, "Modeling the progression of Alzheimer's disease for cognitive assistance in smart homes," *User Modeling and User-Adapted Interaction*, vol. 17, no. 4, pp. 415–438, aug 2007.

[281] J. Lapointe, B. Bouchard, J. Bouchard, A. Potvin, and A. Bouzouane, "Smart homes for people with Alzheimer's disease," in *Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments - PETRA '12*. New York, New York, USA: ACM Press, 2012, p. 1.

[282] A. Lotfi, C. Langensiepen, S. M. Mahmoud, and M. J. Akhlaghinia, "Smart homes for the elderly dementia sufferers: identification and prediction of abnormal behaviour," *Journal of Ambient Intelligence and Humanized Computing*, vol. 3, no. 3, pp. 205–218, sep 2012.

[283] H. Pigot, "The role of intelligent habitats in upholding elders in residence," in *Simulations in biomedicine V*, B. Lefebvre, J.-G. Meunier, B. Kerhervé, A. Mayers, and S. Giroux, Eds., mar 2003, pp. 497–506.

[284] N. Noury, G. Virone, and T. Creuzet, "The health integrated smart home information system (HIS2): rules based system for the localization of a human," in *2nd Annual International IEEE-EMBS Special Topic Conference on Microtechnologies in Medicine and Biology. Proceedings (Cat. No.02EX578)*. IEEE, 2002, pp. 318–321.

[285] "Dem@care - Dementia Ambient Care: Multi-Sensing Monitoring for Intelligent Remote Management and Decision Support," [Online] http://www.demcare.eu/, (Date retrieved: 2015-10-13).

[286] P. Rashidi and D. Cook, "Keeping the Resident in the Loop: Adapting the Smart Home to the User," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 39, no. 5, pp. 949–959, sep 2009.

[287] B. Bouchard, S. Giroux, and A. Bouzouane, "A Keyhole Plan Recognition Model for Alzheimer's patients: First Results," *Applied Artificial Intelligence*, vol. 21, no. 7, pp. 623–658, aug 2007.

[288] D. Cook, M. Youngblood, E. Heierman, K. Gopalratnam, S. Rao, A. Litvin, and F. Khawaja, "MavHome: an agent-based smart home," in *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003. (PerCom 2003).* IEEE Comput. Soc, 2003, pp. 521–524.

[289] T. Suzuki, S. Murase, T. Tanaka, and T. Okazawa, "New Approach for The Early Detection of Dementia by Recording In-House Activities," *Telemedicine and e-Health*, vol. 13, no. 1, pp. 41–44, feb 2007.

[290] V. G. Wadley, O. Okonkwo, M. Crowe, and L. A. Ross-Meadows, "Mild Cognitive Impairment and Everyday Function: Evidence of Reduced Speed in Performing Instrumental Activities of Daily Living," *American Journal of Geriatric Psychiatry*, vol. 16, no. 5, pp. 416–424, may 2008.

[291] T. L. Hayes, F. Abendroth, A. Adami, M. Pavel, T. a. Zitzelberger, and J. a. Kaye, "Unobtrusive assessment of activity patterns associated with mild cognitive impairment," *Alzheimer's & Dementia*, vol. 4, no. 6, pp. 395–405, nov 2008.

[292] P. N. Dawadi, D. J. Cook, and M. Schmitter-Edgecombe, "Automated Cognitive Health Assessment From Smart Home-Based Behavior Data," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 4, pp. 1188–1194, jul 2016.

[293] J. Petersen, D. Austin, N. Mattek, and J. Kaye, "Time out-of-home and cognitive, physical, and emotional wellbeing of older adults: A longitudinal mixed effects model," *PLoS ONE*, vol. 10, no. 10, pp. 1–16, 2015.

[294] J. Austin, H. H. Dodge, T. Riley, P. G. Jacobs, S. Thielke, and J. Kaye, "A Smart-Home System to Unobtrusively and Continuously Assess Loneliness in Older Adults," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 4, pp. 1–11, 2016.

[295] J. M. Hausdorff, G. Yogev, S. Springer, E. S. Simon, and N. Giladi, "Walking is more like catching than tapping: gait in the elderly as a complex cognitive task," *Experimental Brain Research*, vol. 164, no. 4, pp. 541–548, aug 2005.

[296] D. Maquet, F. Lekeu, E. Warzee, S. Gillain, V. Wojtasik, E. Salmon, J. Petermans, and J. L. Croisier, "Gait analysis in elderly adult patients with mild cognitive impairment and patients with mild Alzheimerâ€™s disease: simple versus dual task: a preliminary report," *Clinical Physiology and Functional Imaging*, vol. 30, no. 1, pp. 51–56, jan 2010.

[297] A. Muro-de-la Herran, B. García-Zapirain, and A. Méndez-Zorrilla, "Gait Analysis Methods: An Overview of Wearable and Non-Wearable Systems, Highlighting Clinical Applications," *Sensors*, vol. 14, no. 2, pp. 3362–3394, feb 2014.

[298] E. Scherder, L. Eggermont, D. Swaab, M. van Heuvelen, Y. Kamsma, M. de Greef, R. van Wijck, and T. Mulder, "Gait in ageing and associated dementias; its relationship with cognition," *Neuroscience & Biobehavioral Reviews*, vol. 31, no. 4, pp. 485–497, jan 2007.

[299] O. Beauchet, "Gait analysis in demented subjects: Interests and perspectives," *Neuropsychiatric Disease and Treatment*, vol. 4, no. 1 A, p. 155, mar 2008.

[300] G. Allali, R. W. Kressig, F. Assal, F. R. Herrmann, V. Dubost, and O. Beauchet, "Changes in gait while backward counting in demented older adults with frontal lobe dysfunction," *Gait & Posture*, vol. 26, no. 4, pp. 572–576, oct 2007.

[301] J. Verghese, C. Wang, R. B. Lipton, R. Holtzer, and X. Xue, "Quantitative gait dysfunction and risk of cognitive decline and dementia," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 78, no. 9, pp. 929–935, sep 2007.

[302] O. Beauchet, G. Allali, S. Thiery, J. Gautier, B. Fantino, and C. Annweiler, "Association Between High Variability of Gait Speed and Mild Cognitive Impairment: A Cross-Sectional Pilot Study," *Journal of the American Geriatrics Society*, vol. 59, no. 10, pp. 1973–1974, oct 2011.

[303] J. Verghese, R. B. Lipton, C. B. Hall, G. Kuslansky, M. J. Katz, and H. Buschke, "Abnormality of Gait as a Predictor of Non-Alzheimer's Dementia," *New England Journal of Medicine*, vol. 347, no. 22, pp. 1761–1768, nov 2002.

[304] R. Camicioli, D. Howieson, B. Oken, G. Sexton, and J. Kaye, "Motor slowing precedes cognitive impairment in the oldest old," *Neurology*, vol. 50, no. 5, pp. 1496–1498, may 1998.

[305] "About Alzheimer's," 2014, [Online] http://www.alzfdn.org/AboutAlzheimers/symptoms.html, (Date retrieved: 2015-06-17).

[306] M. Woodward, "Aspects of communication in Alzheimer's disease: clinical features and treatment options." *International psychogeriatrics / IPA*, vol. 25, no. 6, pp. 877–85, 2013.

[307] J. O. de Lira, K. Z. Ortiz, A. C. Campanha, P. H. F. Bertolucci, and T. S. C. Minett, "Microlinguistic aspects of the oral narrative in patients with Alzheimer's disease," *International Psychogeriatrics*, vol. 23, no. 03, pp. 404–412, apr 2011.

[308] M. F. Weiner, K. E. Neubecker, M. E. Bret, and L. S. Hynan, "Language in Alzheimer's Disease," *The Journal of Clinical Psychiatry*, vol. 69, no. 8, pp. 1223–1227, aug 2008.

[309] J. A. Powell, M. A. Hale, and A. J. Bayer, "Symptoms of communication breakdown in dementia: carers' perceptions." *European journal of disorders of communication : the journal of the College of Speech and Language Therapists, London*, vol. 30, no. 1, pp. 65–75, 1995.

[310] J. B. Orange and E. B. Ryan, "Alzheimer's disease and other dementias. Implications for physician communication." *Clinics in geriatric medicine*, vol. 16, no. 1, pp. 153–73, xi, feb 2000.

[311] K. A. Bayles, C. K. Tomoeda, and M. W. Trosset, "Relation of linguistic communication abilities of Alzheimer's patients to stage of disease," *Brain and Language*, vol. 42, no. 4, pp. 454–472, may 1992.

[312] G. Glosser, M. J. Wiley, and E. J. Barnoskir, "Gestural Communication in Alzheimer's Disease," *Journal of Clinical and Experimental Neuropsychology (Neuropsychology, Development and Cognition: Section A)*, vol. 20, no. 1, pp. 1–13, feb 1998.

[313] M. Blair, C. A. Marczinski, N. Davis-Faroque, and A. Kertesz, "A longitudinal study of language decline in Alzheimer's disease and frontotemporal dementia," *Journal of the International Neuropsychological Society*, vol. 13, no. 02, pp. 237–245, mar 2007.

[314] V. Baldas, C. Lampiris, C. Capsalis, and D. Koutsouris, "Early Diagnosis of Alzheimer's Type Dementia Using Continuous Speech Recognition," in *Wireless Mobile Communication and Healthcare*, J. C. Lin and K. S. Nikita, Eds. Berlin: Springer Berlin Heidelberg, 2011, pp. 105–110.

[315] R. Coulston, E. Klabbers, J. Villiers, and J.-p. Hosom, "Application of Speech Technology in a Home Based Assessment Kiosk for Early Detection of Alzheimer's Disease," in *INTERSPEECH 2007, 8th Annual Conference of the International Speech Communication Association*. Antwerp, Belgium: International Speech Communication Association (ISCA), 2007, pp. 2420–2423.

[316] B. Roark, J.-P. Hosom, M. Mitchell, and J. a. Kaye, "Automatically Derived Spoken Language Markers for Detecting Mild Cognitive Impairment," in *Proceedings of the 2nd International Conference on Technology and Aging (ICTA)*, Toronto, Canada, 2007, pp. 1–4.

[317] B. Roark, M. Mitchell, J.-P. Hosom, K. Hollingshead, and J. Kaye, "Spoken Language Derived Measures for Detecting Mild Cognitive Impairment," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2081–2090, sep 2011.

[318] M. Teichmann and S. Ferrieux, "Aphasia(s) in Alzheimer." *Revue neurologique*, vol. 169, no. 10, pp. 680–6, oct 2013.

[319] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, second edi ed., A. Stephan, Ed. San Francisco: Diane Cerra, 2006.

[320] H. M. Chung and P. Gray, "Special Section: Data Mining," in *Journal of Management Information Systems*. M.E. Sharpe, Inc., 1999, vol. 16, pp. 11–16.

[321] J. Jackson, "Data mining: A conceptual overview," *Communications of the Association for Information Systems*, vol. 8, pp. 267–296, 2002.

[322] J. C. Wyatt and J. L. Y. Liu, "Basic concepts in medical informatics." *Journal of epidemiology and community health*, vol. 56, no. 11, pp. 808–812, 2002.

[323] BIOPAC Systems, "BIOPAC Systems Inc." [Online] www.biopac.com, (Date retrieved: 2015-06-09).

[324] Thought Technology Ltd., "FlexComp System with/ BioGraph Infiniti Software - T7555M," [Online] http://thoughttechnology.com/index.php/flexcomp-system-with-biograph-infiniti-software-t7555m.html, (Date retrieved: 2015-06-09).

[325] Trans Cranial Technologies Ltd., "10 / 20 System Positioning Manual," Tech. Rep., 2012, [Online] www.trans-cranial.com, (Date retrieved: 2015-06-10).

[326] U. Lindemann, A. Hock, M. Stuber, W. Keck, and C. Becker, "Evaluation of a fall detector based on accelerometers: A pilot study," *Medical and Biological Engineering and Computing*, vol. 43, no. 5, pp. 548–551, 2005.

[327] P. Zappi, C. Lombriser, L. Benini, and G. Tröster, "Collecting Datasets from Ambient Intelligence Environments," in *Innovative Applications of Ambient Intelligence: Advances in Smart Systems*, K. Curran, Ed. IGI Global, 2012, 2012, p. 332 pages.

[328] R. Khusainov, D. Azzi, I. E. Achumba, and S. D. Bersch, "Real-time human ambulation, activity, and physiological monitoring: taxonomy of issues, techniques, applications, challenges and limitations." *Sensors (Basel, Switzerland)*, vol. 13, no. 10, pp. 12 852–12 902, 2013.

[329] B. Arnrich, C. Kappeler-setz, J. Schumm, and G. Tröster, "Design , Implementation and Evaluation of a Multimodal Sensor System Integrated Into an Airplane Seat," *Sensor Fusion - Foundation and Applications*, p. 159, 2010.

[330] J. J. Braithwaite, D. G. Watson, R. Jones, and M. Rowe, "A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments," Selective Attention & Awareness Laboratory (SAAL) Behavioural Brain Sciences Centre, University of Birmingham, UK, Birmingham, Tech. Rep., 2013.

[331] O. Räsänen and J. Pohjalainen, "Random subset feature selection in automatic recognition of developmental disorders , affective states , and level of conflict from speech," *Interspeech2013*, no. August, pp. 210–214, 2013.

[332] J. Pärkkä, "Analysis of Personal Health Monitoring Data for Physical Activity Recognition and Assessment of Energy Expenditure, Mental Load and Stress," Ph.D. dissertation, Tampere University of Technology, Finland, 2011.

[333] M. Karagiannopoulos, D. Anyfantis, S. B. Kotsiantis, and P. E. Pintelas, "Feature Selection for Regression Problems," in *The 8th Hellenic European Research on Computer Mathematics & its Applications, HERCMA 2007*, no. i. Athens, Greece: University of Patras, 2007.

[334] M. D. Bugdol and A. W. Mitas, "Multimodal biometric system combining ECG and sound signals," *Pattern Recognition Letters*, vol. 38, pp. 107–112, mar 2014.

[335] P. Langley and H. A. Simon, "Applications of machine learning and rule induction," *Communications of the ACM*, vol. 38, no. 11, pp. 54–64, nov 1995.

[336] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed. New Jersey: Pearson Prentice Hall, 2007.

[337] D. Novak, M. Mihelj, and M. Munih, "A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing," *Interacting with Computers*, vol. 24, no. 3, pp. 154–172, 2012.

[338] S. Salmeron-Majadas, O. C. Santos, and J. G. Boticario, "Exploring indicators from keyboard and mouse interactions to predict the user affective state," in *Proceedings of the 7th International Conference on Education Data Mining*, S. D'Mello, R. Calvo, and A. Olney, Eds., 2014, pp. 365–366.

[339] A. Gaggioli, P. Cipresso, S. Serino, G. Pioggia, G. Tartarisco, G. Baldus, D. Corda, M. Ferro, D. Tzovaras, A. Riera, and G. Riva, "A Decision Support System for Real-Time Stress Detection During Virtual Reality Exposure," *Studies in Health Technology and Informatics*, vol. 196, pp. 114–120, 2014.

[340] F. Shah A, R. Sukumar A, and B. Anto P, "Automatic Stress Detection from Speech by Using Support Vector Machines and Discrete Wavelet Transforms," in *International Conference on VLSI Design and Communication Systems (ICVLSICOM)*, Chennai, India, 2010.

[341] S. Haller, K. O. Lovblad, P. Giannakopoulos, and D. Van De Ville, "Multivariate Pattern Recognition for Diagnosis and Prognosis in Clinical Neuroimaging: State of the Art, Current Challenges and Future Trends," *Brain Topography*, pp. 329–337, 2014.

[342] J. Young, G. Ridgway, K. Leung, and S. Ourselin, "Classification of Alzheimer's disease patients with hippocampal shape, wrapper based feature selection and support vector machine," in *Proc. SPIE 8314, Medical Imaging 2012: Image Processing, 83140Q*, D. R. Haynor and S. Ourselin, Eds., vol. 8314. Bellingham, Washington USA: SPIE, feb 2012, p. 83140Q.

[343] J.-F. Horn, M.-O. Habert, A. Kas, Z. Malek, P. Maksud, L. Lacomblez, A. Giron, and B. Fertil, "Differential automatic diagnosis between Alzheimer's disease and frontotemporal dementia based on perfusion SPECT images." *Artificial intelligence in medicine*, vol. 47, pp. 147–158, 2009.

[344] A. Smola and S. Vishwanathan, *Introduction to Machine Learning*. Cambridge: Cambridge university press, 2008.

[345] Sanitas, "Diagnóstico precoz de las enfermedades," [Online] http://www.sanitas.es/sanitas/seguros/es/particulares/biblioteca-de-salud/prevencion-salud/san041825wr.html, (Date retrieved: 2017-08-24).

[346] M. Ruf and O. Morgan, "Differences between screening and - diagnostic tests, case finding," Public Health Action Support Team (PHAST), Bucks, England, Tech. Rep., 2008.

[347] P. Gil, *Medicina Preventiva y Salud Pública* , 11th ed. New Jersey: MASSON, 2008.

[348] P. S. Frame and S. J. Carlson, "A Critical Review of Periodic Health Screening Using Specific Screening Criteria," *The Journal of Family Practice*, vol. 2, no. 1, pp. 29–36, 1975.

[349] J.-L. Pérez-Arellano and C. Carranza-Rodriguez, "Population screening strategies in recent immigrant to Spain," *Enfermedades Infecciosas y Microbiología Clínica*, vol. 34, no. 1, pp. 45–52, 2016.

[350] A. Andermann, I. Blancquaert, S. Beauchamp, and V. Déry, "Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years," *Bulletin of the World Health Organization*, vol. 86, no. 4, pp. 241–320, 2008.

[351] M. Quinn, P. Babb, J. Jones, and E. Allen, "Effect of screening on incidence of and mortality from cancer of cervix in England: evaluation based on routinely collected statistics," *BMJ*, vol. 318, no. 7188, pp. 904–904, apr 1999.

[352] P. C. Gøtzsche and K. J. Jørgensen, "Screening for breast cancer with mammography," in *Cochrane Database of Systematic Reviews*, P. C. Gøtzsche, Ed. Chichester, UK: John Wiley & Sons, Ltd, jun 2013.

[353] K. Bibbins-Domingo, D. C. Grossman, S. J. Curry, K. W. Davidson, J. W. Epling, F. A. R. García, M. W. Gillman, D. M. Harper, A. R. Kemper, A. H. Krist, A. E. Kurth, C. S. Landefeld, C. M. Mangione, D. K. Owens, W. R. Phillips, M. G. Phipps, M. P. Pignone, and A. L. Siu, "Screening for Colorectal Cancer," *JAMA*, vol. 315, no. 23, p. 2564, jun 2016.

[354] K. Bibbins-Domingo, D. C. Grossman, S. J. Curry, L. Bauman, K. W. Davidson, J. W. Epling, F. A. García, J. Herzstein, A. R. Kemper, A. H. Krist, A. E. Kurth, C. S. Landefeld, C. M. Mangione, W. R. Phillips, M. G. Phipps, and M. P. Pignone, "Screening for Latent Tuberculosis Infection in Adults," *JAMA*, vol. 316, no. 9, p. 962, sep 2016.

[355] A. L. Siu, K. Bibbins-Domingo, D. C. Grossman, L. C. Baumann, K. W. Davidson, M. Ebell, F. A. R. García, M. Gillman, J. Herzstein, A. R. Kemper, A. H. Krist, A. E. Kurth, D. K. Owens, W. R. Phillips, M. G. Phipps, and M. P. Pignone, "Screening for Depression in Adults," *JAMA*, vol. 315, no. 4, p. 380, jan 2016.

[356] M. L. LeFevre, "Screening for Abdominal Aortic Aneurysm: U.S. Preventive Services Task Force Recommendation Statement," *Annals of Internal Medicine*, vol. 161, no. 4, p. 281, aug 2014.

[357] "Diabetic eye screening - NHS Choices," NHS Choices, Tech. Rep., [Online] http://www.nhs.uk/Conditions/diabetic-eye-screening/Pages/Introduction.aspx, (Date retrieved: 2017-08-13).

[358] M. Simard, "The Mini-Mental State Examination: strengths and weaknesses of a clinical instrument," *The Canadian Alzheimer Disease Review*, no. December, pp. 10–12, 1998.

[359] European Agency for Safety and Health at Work, "Campaign Guide: Managing stress and psychosocial risks at work," 2013.

[360] B. C. Marlen Cosmar, Robert Gründler, Danny Flemming and K. Van, "Calculating the cost of work-related stress and psychosocial risks," European Agency for Safety and Health at Work, Luxembourg, Tech. Rep., 2014.

[361] M. Prince, E. Albanese, M. Guerchet, and M. Prina, "World Alzheimer Report 2014. Dementia and Risk Reduction. An analysis of protective and modifiable factors," Alzheimer's Disease International (ADI), London, Tech. Rep., 2014.

[362] K. Tagai, T. Nagata, S. Shinagawa, K. Nemoto, K. Inamura, N. Tsuno, and K. Nakayama, "Correlation between both Morphologic and Functional Changes and Anxiety in Alzheimer's Disease," *Dementia and Geriatric Cognitive Disorders*, vol. 38, no. 3-4, pp. 153–160, 2014.

[363] S. P. Poulin, R. Dautoff, J. C. Morris, L. F. Barrett, and B. C. Dickerson, "Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity," *Psychiatry Research: Neuroimaging*, vol. 194, no. 1, pp. 7–13, oct 2011.

[364] P. D. Bruen, W. J. McGeown, M. F. Shanks, and A. Venneri, "Neuroanatomical correlates of neuropsychiatric symptoms in Alzheimer's disease," *Brain*, vol. 131, no. 9, pp. 2455–2463, aug 2008.

[365] "Workplace Stress. A Collective Challenge." International Labour Organization, Tech. Rep. April 2016, 2016, [Online] www.ilo.org/safeday, (Date retrieved: 2016-08-25).

[366] European Agency for Safety and Health at Work, "European Opinion Poll on Occupational Safety and Health," European Agency for Safety and Health at Work, Bilbao, Spain, Tech. Rep. May, 2013, [Online] https://osha.europa.eu/en/safety-health-in-figures/eu-poll-press-kit-2013.pdf, (Date retrieved: 2016-01-10).

[367] European Foundation for the Improvement of Living and Working Conditions, *Sixth European Working Conditions Survey - Overview report*, A. Parent-Thirion, I. Biletta, J. Cabrita, O. Vargas, G. Vermeylen, A. Wilczynska, and M. Wilkens, Eds. Luxembourg:: Publications Office of the European Union, 2016, no. November.

[368] "Drivers and Barriers for Psychosocial Risk Management: An analysis of findings of the European survey of enterprises on new and emerging risks." European Agency for Safety and Health at Work, Luxembourg, Tech. Rep., 2012, [Online] https://osha.europa.eu/en/tools-and-publications/publications/reports/drivers-barriers-psychosocial-risk-management-esener, (Date retrieved: 2016-04-06).

[369] A. Alberdi, A. Aztiria, and A. Basarab, "Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review," *Journal of Biomedical Informatics*, vol. 59, pp. 49–75, 2015.

[370] C. Ramos, G. Marreiros, R. Santos, and C. F. Freitas, "Smart Offices and Intelligent Decision Rooms," in *Handbook of Ambient Intelligence and Smart Environments*, H. Nakashima, H. Aghajan, and J. C. Augusto, Eds. Boston, MA: Springer US, 2010, pp. 1–31.

[371] H. Strömberg, V. Pirttillä, and V. Ikonen, "Interactive scenarios-building ubiquitous computing concepts in the spirit of participatory design," *Personal and Ubiquitous Computing*, vol. 8, no. 3-4, jul 2004.

[372] I. Marsá Maestre, E. de la Hoz, B. Alarcos, and J. R. Velasco Pérez, "A Hierarchical, Agent-based Approach to Security in Smart O ces," in *International Conference on Ubiuitous Computing*, Orange County, CA, USA, 2006, pp. 1–9.

[373] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, "SMOTEBoost: Improving Prediction of the Minority Class in Boosting," 2003, no. December, pp. 107–119.

[374] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "RUSBoost: A hybrid approach to alleviating class imbalance," *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, vol. 40, no. 1, pp. 185–197, 2010.

[375] S. Koldijk, M. Sappelli, S. Verberne, M. A. Neerincx, and W. Kraaij, "The SWELL Knowledge Work Dataset for Stress and User Modeling Research," in *Proceedings of the 16th International Conference on Multimodal Interaction*. New York, New York, USA: ACM Press, 2014, pp. 291–298.

[376] W. Kraaij, S. Koldijk, and M. Sappelli, "The SWELL Knowledge Work Dataset for Stress and User Modeling Research. DANS." 2014, [Online] https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:58624, (Date retrieved: 2017-01-10).

[377] P. Lang, *Behavioral treatment and bio-behavioral assessment: computer applications.*, J. B. Sidowski, J. H. Johnson, and T. A. Williams, Eds., Norwood, NJ: Ablex., 1980.

[378] F. R H Zijlstra and L. Van Doorn, "The construction of a scale to measure subjective effort." Delft University of Technology, Department of Philosophy and Social Sciences., Tech. Rep., 1985.

[379] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," 1988, pp. 139–183.

[380] "Facereader™version 6.1." Wageningen, The Netherlands, pp. 1–208, 2015, [Online] http://www.noldus.com/human-behavior-research/products/facereader, (Date retrieved: 2016-09-13).

[381] "Kinect for Windows SDK 2.0," [Online] https://www.microsoft.com/en-us/download/details.aspx?id=44561, (Date retrieved: 2017-04-12).

[382] "Mobi." pp. 1–22, 2016, [Online] https://www.tmsi.com/images/products/mobi/UserManual_Mobi_EN-Rev2.pdf, (Date retrieved: 2017-04-12).

[383] V. Dakos, S. R. Carpenter, W. A. Brock, A. M. Ellison, V. Guttal, A. R. Ives, S. Kéfi, V. Livina, D. A. Seekell, E. H. van Nes, and M. Scheffer, "Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data," *PLoS ONE*, vol. 7, no. 7, 2012.

[384] H. Hotelling, "The Generalization of Student's Ratio," *The Annals of Mathematical Statistics*, vol. 2, no. 3, pp. 360–378, aug 1931.

[385] L. Christensen and J. L. Mendoza, "A method of assessing change in a single subject: An alteration of the RC index," *Behavior Therapy*, vol. 17, no. 3, pp. 305–308, jun 1986.

[386] V. Battiste and M. Bortolussi, "Transport Pilot Workload: A Comparison of Two Subjective Techniques," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 32, no. 2, pp. 150–154, 1988.

[387] E. Frank, M. A. Hall, and I. H. Witten, *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"*, fourth edi ed., M. Kaufmann, Ed., 2016.

[388] I. H. Witten, E. Frank, and M. a. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 2011.

[389] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif. Int. Res.*, vol. 16, no. 1, pp. 321–357, 2002.

[390] J. Van Hulse, T. M. Khoshgoftaar, and A. Napolitano, "Experimental perspectives on learning from imbalanced data," in *Proceedings of the 24th international conference on Machine learning - ICML '07*. New York, New York, USA: ACM Press, 2007, pp. 935–942.

[391] H. Otsu, K. Sato, H. Madokoro, and S. Kadowaki, "Analysis of psychological stress factors and facial parts effect on intentional facial expressions," in *2012 Proceedings of SICE Annual Conference (SICE)*, Akita, 2012, pp. 1144–1149.

[392] H. Al Osman, M. Eid, and A. El Saddik, "U-biofeedback: a multimedia-based reference model for ubiquitous biofeedback systems," *Multimedia Tools and Applications*, vol. 72, no. 3, pp. 3143–3168, jul 2013.

[393] A. Alberdi, A. Aztiria, and A. Basarab, "On the early diagnosis of Alzheimer's Disease from multimodal signals: A survey," *Artificial Intelligence in Medicine*, vol. 71, pp. 1–29, jul 2016.

[394] R. C. Petersen, "Early diagnosis of Alzheimer's disease: is MCI too late?" *Current Alzheimer research*, vol. 6, no. 4, pp. 324–30, aug 2009.

[395] K. G. Yiannopoulou and S. G. Papageorgiou, "Current and future treatments for Alzheimer's disease," *Therapeutic Advances in Neurological Disorders*, vol. 6, no. 1, pp. 19–33, jan 2013.

[396] M. Chan, D. Esteve, C. Escriba, and E. Campo, "A review of smart homes - Present state and future challenges," *Computer Methods and Programs in Biomedicine*, vol. 91, no. 1, pp. 55–81, 2008.

[397] C. Randolph, M. C. Tierney, E. Mohr, and T. N. Chase, "The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): Preliminary Clinical Validity," *Journal of Clinical and Experimental Neuropsychology (Neuropsychology, Development and Cognition: Section A)*, vol. 20, no. 3, pp. 310–319, jun 1998.

[398] J. Crawford, G. Smith, E. Maylor, S. Della Sala, and R. Logie, "The Prospective and Retrospective Memory Questionnaire (PRMQ): Normative data and latent structure in a large non-clinical sample," *Memory*, vol. 11, no. 3, pp. 261–275, jan 2003.

[399] D. Podsiadlo and S. Richardson, "The timed "Up & Go": a test of basic functional mobility for frail elderly persons." *Journal of the American Geriatrics Society*, vol. 39, no. 2, pp. 142–8, feb 1991.

[400] R. E. Rikli and C. J. Jones, "Development and Validation of a Functional Fitness Test for Community-Residing Older Adults," *Journal of Aging and Physical Activity*, vol. 7, no. 2, pp. 129–161, apr 1999.

[401] J. I. Sheikh and J. A. Yesavage, "Geriatric Depression Scale (GDS) Recent Evidence and Development of a Shorter Version." *Clinical Gerontologist*, vol. 5, no. 1-2, pp. 165–173, nov 1986.

[402] E. Smith, L. Walsh, J. Doyle, B. Greene, and C. Blake, "The reliability of the quantitative timed up and go test (QTUG) measured over five consecutive days under single and dual-task conditions in community dwelling older adults," *Gait & Posture*, vol. 43, pp. 239–244, jan 2016.

[403] J. M. Miotto, W. J. Chodzko-Zajko, J. L. Reich, and M. M. Supler, "Reliability and Validity of the Fullerton Functional Fitness Test: An Independent Replication Study," *Journal of Aging and Physical Activity*, vol. 7, no. 4, pp. 339–353, oct 1999.

[404] C. McKay, J. Casey, J. Wertheimer, and N. Fichtenberg, "Reliability and validity of the RBANS in a traumatic brain injured sample," *Archives of Clinical Neuropsychology*, vol. 22, no. 1, pp. 91–98, jan 2007.

[405] J. Crawford, G. Smith, E. Maylor, S. Della Sala, and R. Logie, "The Prospective and Retrospective Memory Questionnaire (PRMQ): Normative data and latent structure in a large non-clinical sample," *Memory*, vol. 11, no. 3, pp. 261–275, jan 2003.

[406] M. C. Salinsky, D. Storzbach, C. B. Dodrill, and L. M. Binder, "Test-retest bias, reliability, and regression equations for neuropsychological measures repeated over a 12-16-week period," *Journal of the International Neuropsychological Society*, vol. 7, no. 5, jul 2001.

[407] O. Pedraza, V. M. Dotson, F. B. Willis, N. R. Graff-Radford, and J. A. Lucas, "Internal Consistency and Test-Retest Stability of the Geriatric Depression Scale-Short Form in African American Older Adults," *Journal of Psychopathology and Behavioral Assessment*, vol. 31, no. 4, pp. 412–416, dec 2009.

[408] N. C. Krishnan and D. J. Cook, "Activity recognition on streaming sensor data," *Pervasive and Mobile Computing*, vol. 10, no. PART B, pp. 138–154, 2014.

[409] J. W. Ratcliff and D. E. Metzener, *Pattern-matching-the gestalt approach*. Miller Freeman, inc 411 Borel Ave, San Mateo, CA 94402-3522, 1988, vol. 13, no. 7.

[410] B. Das, N. C. Krishnan, and D. J. Cook, "WRACOG: A gibbs sampling-based oversampling technique," *Proceedings - IEEE International Conference on Data Mining, ICDM*, pp. 111–120, 2013.

[411] C. Chow and C. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 462–467, may 1968.

[412] M. I. Tolea, J. C. Morris, and J. E. Galvin, "Trajectory of Mobility Decline by Type of Dementia," *Alzheimer Disease & Associated Disorders*, vol. 30, no. 1, pp. 60–66, 2016.

[413] S. Ooms and Y.-E. Ju, "Treatment of Sleep Disorders in Dementia," *Current Treatment Options in Neurology*, vol. 18, no. 9, p. 40, sep 2016.

[414] T. L. Hayes, T. Riley, N. Mattek, M. Pavel, and J. A. Kaye, "Sleep habits in mild cognitive impairment." *Alzheimer disease and associated disorders*, vol. 28, no. 2, pp. 145–50, 2014.

[415] R. A. P. C. da Silva, "Sleep disturbances and mild cognitive impairment: A review," *Sleep Science*, vol. 8, no. 1, pp. 36–41, jan 2015.

[416] M. Chan, D. Estève, J.-Y. Fourniols, C. Escriba, and E. Campo, "Smart wearable systems: Current status and future challenges," *Artificial Intelligence in Medicine*, vol. 56, no. 3, pp. 137–156, 2012.

[417] B. L. Thomas, A. S. Crandall, and D. J. Cook, "A Genetic Algorithm approach to motion sensor placement in smart environments," *Journal of Reliable Intelligent Environments*, vol. 2, no. 1, pp. 3–16, apr 2016.

[418] G. Vink, L. E. Frank, J. Pannekoek, and S. van Buuren, "Predictive mean matching imputation of semicontinuous variables," *Statistica Neerlandica*, vol. 68, no. 1, pp. 61–90, feb 2014.

[419] E. Alpaydin, *Introduction to Machine Learning*. London, England: Massachusetts Institute of Technology, 2010.

[420] P. Ren, A. Barreto, J. Huang, Y. Gao, F. R. Ortega, and M. Adjouadi, "Off-line and on-line stress detection through processing of the pupil diameter signal," *Annals of Biomedical Engineering*, vol. 42, no. 1, pp. 162–176, 2014.

[421] I. Á. Illán, J. Górriz, J. Ramírez, D. Salas-Gonzalez, M. López, F. Segovia, P. Padilla, and C. Puntonet, "Projecting independent components of SPECT images for computer aided diagnosis of Alzheimer's disease," *Pattern Recognition Letters*, vol. 31, no. 11, pp. 1342–1347, 2010.

[422] A. Bulling, J. A. Ward, H. Gellersen, and G. Tröster, "Robust Recognition of Reading Activity in Transit Using Wearable Electrooculography," 2008, pp. 19–37.

[423] B. H. Tan, "Using a Low-cost EEG Sensor to Detect Mental States," Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, PA, 2012.

[424] T. P. Jung, S. Makeig, C. Humphries, T. W. Lee, M. J. McKeown, V. Iragui, and T. J. Sejnowski, "Removing electroencephalographic artifacts by blind source separation." *Psychophysiology*, vol. 37, pp. 163–178, 2000.

[425] H. Borotschnig, L. Paletta, M. Prantl, and A. Pinz, "Comparison of probabilistic, possibilistic and evidence theoretic fusion schemes for active object recognition," *Computing (Vienna/New York)*, vol. 62, no. 4, pp. 293–319, 1999.

[426] D. Dubois and H. Prade, "Possibility theory and data fusion in poorly informed environments," *Control Engineering Practice*, vol. 2, no. 5, pp. 811–823, 1994.

[427] M. C. Florea, A.-L. Jousselme, and É. Bossé, "Fusion of imperfect information in the unified framework of random sets theory," Defence R& D Canada, Valcartier, DRDC Valcartier TR 2003-319, Tech. Rep., 2007.

[428] I. Van Mechelen and A. K. Smilde, "A generic linked-mode decomposition model for data fusion," *Chemometrics and Intelligent Laboratory Systems*, vol. 104, no. 1, pp. 83–94, 2010.

[429] D. Lahat and C. Jutten, "Challenges in multimodal data fusion," in *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*. Lisbon: IEEE, 2014, pp. 101 – 105.

[430] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Information Fusion*, vol. 14, no. 1, pp. 28–44, 2013.

[431] A. Arauzo-Azofra, J. L. Aznarte, and J. M. Benítez, "Empirical study of feature selection methods based on individual feature evaluation for classification problems," *Expert Systems with Applications*, vol. 38, no. 7, pp. 8170–8177, 2011.

[432] O. Postolache, P. S. Girão, E. Pinheiro, and G. Postolache, "Unobtrusive and non-invasive sensing solutions for on-line physiological parameters monitoring," *Lecture Notes in Electrical Engineering*, vol. 75 LNEE, pp. 277–314, 2010.

[433] H. Chigira, M. Kobayashi, and A. Maeda, "Mouse with Photo-Plethysmographic surfaces for unobtrusive stress monitoring," in *2012 IEEE Second International Conference on Consumer Electronics - Berlin (ICCE-Berlin)*. IEEE, sep 2012, pp. 304–305.

[434] C. M. Yang, C. C. Wu, C. M. Chou, and C. W. Yang, "Textile-based breath-sensing belt," in *ICCE 2010 - 2010 Digest of Technical Papers International Conference on Consumer Electronics*, 2010, pp. 11–12.

[435] J. Taelman, T. Adriaensen, C. van der Horst, T. Linz, and A. Spaepen, "Textile integrated contactless EMG sensing for stress analysis." in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, vol. 2007, jan 2007, pp. 3966–9.

[436] B. Maric, A. Kaan, A. Ignaszewski, and S. A. Lear, "A systematic review of telemonitoring technologies in heart failure," *European Journal of Heart Failure*, vol. 11, no. 5, pp. 506–517, 2009.

[437] W. H. Fissell, A. J. Fleischman, H. D. Humes, and S. Roy, "Development of continuous implantable renal replacement: past and future," *Translational Research*, vol. 150, no. 6, pp. 327–336, 2007.

[438] M. Okubo, Y. Imai, T. Ishikawa, T. Hayasaka, S. Ueno, and T. Yamaguchi, "Development of automatic respiration monitoring for home-care patients of respiratory diseases with therapeutic aids," in *IFMBE Proceedings*, vol. 22, 2008, pp. 1117–1120.

[439] B. Guerci, P. Böhme, C. Halter, and C. Bourgeois, "Capteurs de glucose et mesure continue du glucose," *Medecine des Maladies Metaboliques*, vol. 4, no. 2, pp. 157–168, 2010.

[440] R. Islam, S. I. Ahamed, N. Talukder, and I. Obermiller, "Usability of Mobile Computing Technologies," *Third Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2007 Graz, Austria, November, 22, 2007. Proceedings*, pp. 227–240, 2007.

[441] U. Anliker, J. A. Ward, P. Lukowicz, G. Tröster, F. Dolveck, M. Baer, F. Keita, E. B. Schenker, F. Catarsi, L. Coluccini, A. Belardinelli, D. Shklarski, M. Alon, E. Hirt, R. Schmid, and M. Vuskovic, "AMON: A wearable multiparameter medical monitoring and alert system," *IEEE Transactions on Information Technology in Biomedicine*, vol. 8, no. 4, pp. 415–427, 2004.

[442] H. Miwa, S.-i. Sasahara, and T. Matsui, "Roll-over detection and sleep quality measurement using a wearable sensor." *Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, vol. 2007, pp. 1507–1510, 2007.

[443] S. J. M. Bamberg, A. Y. Benbasat, D. M. Scarborough, D. E. Krebs, and J. A. Paradiso, "Gait analysis using a shoe-integrated wireless sensor system," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 4, pp. 413–423, 2008.

[444] T. Giorgino, P. Tormene, F. Lorussi, D. De Rossi, and S. Quaglini, "Sensor evaluation for wearable strain gauges in neurological rehabilitation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 17, no. 4, pp. 409–415, 2009.

[445] R. D. Beach and F. Moussy, "Totally implantable real-time in vivo video telemetry monitoring system for implant biocompatibility studies," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 3, pp. 716–723, 2001.

[446] A. Chaudhary, M. J. McShane, and R. Srivastava, "Glucose response of dissolved-core alginate microspheres: towards a continuous glucose biosensor." *The Analyst*, vol. 135, no. 10, pp. 2620–2628, 2010.

[447] B. Kellihan, T. J. Doty, W. D. Hairston, J. Canady, K. W. Whitaker, C.-t. Lin, T.-p. Jung, and K. Mcdowell, "A Real-World Neuroimaging System to Evaluate Stress," in *Foundations of Augmented Cognition*, ser. Lecture Notes in Computer Science, D. D. Schmorrow and C. M. Fidopiastis, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, vol. 8027, pp. 316–325.

[448] K.-h. Chang, D. Fisher, J. Canny, and B. Hartmann, "How's my Mood and Stress ? An Efficient Speech Analysis Library for Unobtrusive Monitoring on Mobile Phones," in *BODYNETS 2011*, no. November, 2011, pp. 71–77.

[449] "European Data Format," [Online] http://www.edfplus.info/, (Date retrieved: 2015-10-13).

[450] B. Kemp and J. Olivan, "European data format 'plus' (EDF+), an EDF alike standard format for the exchange of physiological data," *Clinical Neurophysiology*, vol. 114, no. 9, pp. 1755–1761, 2003.

[451] A. Schlögl, O. Filz, H. Ramoser, and G. Pfurtscheller, "GDF - A general dataformat for biosignals Version 1.25," Tech. Rep., 2005.

[452] "ISO 11073-91064:2009," [Online] http://www.iso.org/iso/catalogue_detail.htm?csnumber=46493, (Date retrieved: 2015-05-06).

[453] "DICOM Digital Imaging and Communications in Medicine," [Online] http://medical.nema.org/standard.html, (Date retrieved: 2015-10-13).

[454] "HL7 annotated ECG," [Online] http://www.hl7.org/participate/tool-sandresources.cfm, (Date retrieved: 2015-10-13).

[455] "SNOMED CT," [Online] http://www.ihtsdo.org/snomed-ct, (Date retrieved: 2015-05-06).

[456] "LOINC from Regenstrief," [Online] https://loinc.org/, (Date retrieved: 2015-09-10).

[457] T. H. F. Broens, R. M. H. a. Huis in't Veld, M. M. R. Vollenbroek-Hutten, H. J. Hermens, A. T. van Halteren, and L. J. M. Nieuwenhuis, "Determinants of successful telemedicine implementations: a literature study." *Journal of telemedicine and telecare*, vol. 13, no. 6, pp. 303–309, 2007.

[458] J. Barlow, S. Bayer, B. Castleton, and R. Curry, "Meeting government objectives for telecare in moving from local implementation to mainstream services." *Journal of telemedicine and telecare*, vol. 11 Suppl 1, pp. 49–51, 2005.

[459] European Comission, "Green Paper on mobile Health ("mHealth")," European Commission, Brussels, Tech. Rep., 2014.

[460] A. S. Hansen, "Guidelines on Minimum / Non- Exhaustive Patient Summary Dataset for Electronic Exchange in Accordance With the Cross-Border Directive 2011 / 24 / Eu," European Commission, Tech. Rep. November, 2013.

[461] J. V. Hoof, H. D. Kort, P. Markopoulos, and M. Soede, "Ambient intelligence, ethics and privacy," *Gerontechnology*, vol. 6, no. 3, 2007.

[462] E. Steel and A. Dembosky, "Health apps run into privacy snags," *Financial Times*, 2013.

[463] G. Acampora, D. J. Cook, P. Rashidi, and A. V. Vasilakos, "A Survey on Ambient Intelligence in Health Care," in *Proc IEEE Inst Electr Electron Eng.*, vol. 101, no. 12, 2013, pp. 2470–2494.

[464] M. Laszlo and S. Mukherjee, "Minimum spanning tree partitioning algorithm for microaggregation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, 2005.

[465] J. W. J. Wang, Y. L. Y. Luo, Y. Z. Y. Zhao, and J. L. J. Le, "A Survey on Privacy Preserving Data Mining," *2009 First International Workshop on Database Technology and Applications*, 2009.

[466] B. Tannous and J. Teng, "Secreted blood reporters: insights and applications." *Biotechnology advances*, vol. 29, no. 6, pp. 997–1003, 2011.

[467] R. Sharp, "Lacking regulation, many medical apps questionable at best," New England Center for Investigative Reporting, Tech. Rep., 2012.

[468] "IEC/DIS 82304-1 Health software – Part 1: General requirements for product safety," International Organization for Standardization ; International Electrotechnical Commission, Tech. Rep., 2015, [Online] http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=59543, (Date retrieved: 2017-01-05).

[469] J. A. Brebner, E. M. Brebner, and H. Ruddick-Bracken, "Experience-based guidelines for the implementation of telemedicine services." *Journal of telemedicine and telecare*, vol. 11 Suppl 1, pp. 3–5, 2005.

[470] A. Tinker and P. Lansley, "Introducing assistive technology into the existing homes of older people: feasibility, acceptability, costs and outcomes." *Journal of telemedicine and telecare*, vol. 11 Suppl 1, pp. 1–3, 2005.

[471] T. H. Malasanos, J. B. Burlingame, L. Youngblade, B. D. Patel, and A. B. Muir, "Improved access to subspecialist diabetes care by telemedicine: cost savings and care measures in the first two years of the FITE diabetes project." *Journal of telemedicine and telecare*, vol. 11 Suppl 1, pp. 74–76, 2005.

[472] C. A. Kavamoto, C. L. Wen, L. R. Battistella, and G. M. Böhm, "A Brazilian model of distance education in physical medicine and rehabilitation based on videoconferencing and Internet learning." *Journal of telemedicine and telecare*, vol. 11 Suppl 1, pp. 80–82, 2005.

[473] I. H. M. Aas and J. T. Geitung, "Teleradiology and picture archiving and communications systems: changed pattern of communication between clinicians and radiologists." *Journal of telemedicine and telecare*, vol. 11 Suppl 1, pp. 20–22, 2005.