UNIVERSITÀ
POLITECNICA
DELLE MARCHE

PhD School in Engineering
PhD Course in Information Engineering

Measurement and processing of multimodal physiological signals in response to external stimuli by wearable devices and evaluation of parameters influencing data acquisition

Ph.D. Dissertation of:
Angelica Poli

Supervisor:
Prof. Susanna Spinsante

Ph.D. Course Coordinator:
Prof. Franco Chiaraluce

XXXV edition - new series

UNIVERSITÀ
POLITECNICA
DELLE MARCHE

PhD School in Engineering
PhD Course in Information Engineering

Measurement and processing of multimodal physiological signals in response to external stimuli by wearable devices and evaluation of parameters influencing data acquisition

Ph.D. Dissertation of:
Angelica Poli

Supervisor:
Prof. Susanna Spinsante

Ph.D. Course Coordinator:
Prof. Franco Chiaraluce

XXXV edition - new series

Polytechnic University of Marche
PhD School in Engineering
PhD Course in Information Engineering
Via Brecce Bianche – 60131 Ancona (AN), Italy

To My Loving Family,
To Those Who Supported Me and
Positively Influenced My Career

Abstract

In daily life, the presence of specific stimuli provokes reactions reflecting in unconscious experience-related changes of human physiological status. Variations in physiological signals may be obviously measured in ambulatory conditions by means of medical instruments, but the healthcare provisioning paradigm is shifting towards remote monitoring, and wearable devices have gained more and more popularity for daily collection of individual data. This research activity presents several contributions to this topic, including a set of studies conducted to develop and test measurement procedures, to validate signal processing techniques and wearable devices performance with the aim of quantitatively assessing human response to external stimuli, by means of physiological signals. Firstly, a study identified the elicitation of acoustic, visual, and physical stimulation on the Skin Conductance (SC) signal collected through a commercial wearable device (Empatica E4) by healthy individuals. Data analysis in time domain showed that listening to an unpleasant sound increases the subjective physiological response (higher number of SC peaks), especially when the sound duration is short. Additionally, SC proved to be a standalone reliable signal in recognising a driver's drowsy status (Random Forest classification: 84.10% accuracy) and different level of physical exertion when extracting features from frequency domain; in this regard, the proposed larger observation band (i.e. [0.40, 1.00] Hz) can avoid loss of potentially meaningful information about the skin activity, with respect to common investigated one [0, 0.40] Hz. These analyses were conducted also in collaboration with Prof. Florez-Revuelta (University of Alicante) to understand whether and how living environments design affects the individual emotional capacity. This research gives rise to a second study on cardiac activity measurements (i.e. Heart Rate Variability (HRV), Heart Rate (HR) and Blood Pressure (BP)) conducted by healthy subjects, while wearing the Empatica E4 at rest, during acoustic stimuli, and after physical exertion. Either in multimodal signals and cardiac activity standalone approach, Support Vector Machine algorithm identified the presence of acoustic and physical stimuli (accuracy: 72.62% and 97.50%, respectively). Regarding the physical exertion, it was shown that time-consuming data classification could be replaced by a simpler Fatigue-Related Index proposed to determine the fatigue-related stress levels. At rest, some commercial wearable devices were validated, by comparison with the related gold standards, including a metrological characterisation in terms of accuracy and precision (which are currently debatable, since unavailable or not reported with rigorous validations from the manufacturers). By using different wearable devices, results show different accuracy and precision of BP and HR values. This confirms the importance of these metrological figures, for a better device selection depending on the target application.

An additional analysis investigated the factors interfering the Photoplethysmographic sensor readings and consequently the quality of data processed, such as motion artefacts and skin-device contact pressure. Specifically, a new data artefacts correction method was implemented to improve the classification performance in external stimuli detection (accuracy from 48.81% to 66.67%) and the role of band tightness of a DIY wrist-worn device was investigated. All the overmentioned aspects are strictly related to the decision-making processes that need reliable and high-quality information to properly interpret results and take right decisions, especially in the healthcare domain (e.g., personalised care and remote monitoring). Within this context, the last relevant contribution concerns the issues in data acquisition from wearable devices focusing on acceleration signal. This study consisted in the measurement of wrist acceleration during daily human activities performed in a real-life setting, not-controlled environment, without instructions or guidance, by wearing wearable devices. Different hardware sensor technologies, playing a paramount role to determine the classification accuracy, are compared to quantify how this is affected by the nature and quality of the collected measurement data. Long-term data collection may expose users to privacy violations, due to leakage of personal details. To mitigate this aspect, a Multi Objective Evolutionary Algorithm (MOEA) approach was implemented to conceal subject's gender while maximising the accuracy on data recording. The proposed method strongly limited gender recognition (decreased of 25.00%), while maintaining high the activity recognition accuracy (decreased only of 5.45%). Research on how much the device choice and position, along with the measurement setup, affect data quality and variability has been conducted at the R&D Nestlé Orbe (Electronic Team) as well, during the training period abroad, to develop a new coffee machine with an embedded sensor. Here, the work focused on the sensor assessment and data analysis including pattern recognition for improving the quality of coffee thanks to continuous recording of meaningful coffee extraction parameters. Finally, to understand the users' requirements in terms of perceived benefits and barriers and to increase the awareness of issues associated with lifelogging, transversal activities were performed during an International Project (i.e. PAAL - Privacy-Aware and Acceptable Lifelogging services for older and frail people, More Years Better Lives JPI) in collaboration with other International Universities (i.e. University of Alicante, University of Toronto, Stockholm University and Aachen University).

Sommario

Nella vita quotidiana, la presenza di stimoli specifici provoca reazioni che si riflettono nei cambiamenti dello stato fisiologico umano legati all'esperienza inconscia. Le variazioni dei segnali fisiologici possono ovviamente essere misurate in condizioni ambulatoriali per mezzo di strumenti medici, ma il paradigma dell'assistenza sanitaria si sta spostando verso il monitoraggio remoto e i dispositivi indossabili hanno guadagnato sempre più popolarità per la raccolta quotidiana di dati individuali. Questa attività di ricerca presenta diversi contributi a questo argomento, tra cui una serie di studi condotti per sviluppare e testare procedure di misurazione, per validare tecniche di elaborazione del segnale e prestazioni di dispositivi indossabili con l'obiettivo di valutare quantitativamente la risposta umana a stimoli esterni, per mezzo di segnali fisiologici. In primo luogo, uno studio ha identificato l'elicitazione della stimolazione acustica, visiva e fisica sul segnale di conduttanza cutanea (SC) raccolto attraverso un dispositivo indossabile commerciale (Empatica E4) da individui sani. L'analisi dei dati nel dominio del tempo ha mostrato che l'ascolto di un suono sgradevole aumenta la risposta fisiologica soggettiva (numero maggiore di picchi SC), soprattutto quando la durata del suono è breve. Inoltre, SC si è rivelato un segnale affidabile autonomo nel riconoscere lo stato di sonnolenza di un guidatore (classificazione Random Forest: 84.10% di accuratezza) e il diverso livello di sforzo fisico se analizzato nel dominio della frequenza; a questo proposito, la banda di osservazione più ampia proposta (cioè [0.40, 1.00] Hz) può evitare la perdita di informazioni potenzialmente significative sull'attività cutanea, rispetto alla banda comunemente studiata [0, 0.40] Hz. Queste analisi sono state condotte anche in collaborazione con il Prof. Florez-Revuelta (Università di Alicante) per capire se e come il design degli ambienti di vita influisca sulla capacità emotiva individuale. Questa ricerca dà vita ad un secondo studio sulle misurazioni dell'attività cardiaca (cioè, Variabilità della Frequenza Cardiaca (HRV), Frequenza Cardiaca (HR) e Pressione Sanguigna (BP)) condotte da soggetti sani, mentre indossano Empatica E4 a riposo, durante stimoli acustici , e dopo uno sforzo fisico. Sia nei segnali multimodali che nell'approccio autonomo dell'attività cardiaca, l'algoritmo Support Vector Machine ha identificato la presenza di stimoli acustici e fisici (accuratezza: 72.62% e 97.50%, rispettivamente). Per quanto riguarda lo sforzo fisico, è stato dimostrato che la classificazione dei dati potrebbe essere sostituita da un più semplice indice proposto per determinare i livelli di stress da fatica. A riposo, alcuni dispositivi indossabili commerciali sono stati validati, per confronto con i relativi gold standard, inclusa una caratterizzazione metrologica in termini di accuratezza e precisione (che al momento sono discutibili: indisponibili o non validate rigorosamente dai produttori). Utilizzando diversi dispositivi indossabili, i

risultati mostrano una diversa accuratezza e precisione dei valori di HR e BP. Ciò conferma l'importanza di queste cifre metrologiche, per una migliore selezione dei dispositivi a seconda dell'applicazione target. Un'ulteriore analisi ha studiato i fattori che interferiscono con le letture del sensore fotopletismografico e di conseguenza la qualità dei dati elaborati, come gli artefatti da movimento e la pressione di contatto pelle-dispositivo. In particolare, è stato implementato un nuovo metodo di correzione degli artefatti dei dati per migliorare le prestazioni di classificazione nel rilevamento di stimoli esterni (precisione da 48.81% a 66.67%) ed è stato studiato il ruolo della tenuta della fascia di un dispositivo indossato al polso fai-da-te. Tutti gli aspetti sopra menzionati sono strettamente correlati ai processi decisionali che necessitano di informazioni affidabili e di alta qualità per interpretare correttamente i risultati e prendere decisioni giuste, soprattutto in ambito sanitario (es. cure personalizzate e monitoraggio remoto). In questo contesto, l'ultimo contributo rilevante riguarda le problematiche di acquisizione dati da dispositivi indossabili focalizzati sul segnale di accelerazione. Questo studio consisteva nella misurazione dell'accelerazione del polso durante le attività umane quotidiane eseguite in un ambiente reale, non controllato, senza istruzioni o guida, indossando dispositivi indossabili. Diverse tecnologie di sensori hardware, che svolgono un ruolo fondamentale per determinare l'accuratezza della classificazione, vengono confrontate per quantificare in che modo questo è influenzato dalla natura e dalla qualità dei dati di misurazione raccolti. La raccolta di dati a lungo termine può esporre gli utenti a violazioni della privacy, a causa della perdita di dati personali. Per mitigare questo aspetto, è stato implementato un approccio Multi Objective Evolutionary Algorithm per nascondere il genere del soggetto massimizzando l'accuratezza nella registrazione dei dati. Il metodo proposto limitava fortemente il riconoscimento del genere (diminuito del 25.00%), pur mantenendo alta l'accuratezza del riconoscimento dell'attività (diminuita solo del 5.45%). La ricerca su quanto la scelta e la posizione del dispositivo, insieme all'impostazione della misurazione, influenzino la qualità e la variabilità dei dati è stata condotta anche presso l'R&D Nestlé Orbe (Team elettronico), durante il periodo di formazione all'estero, per sviluppare una nuova macchina da caffè con sensore incorporato. Qui, il lavoro si è concentrato sulla valutazione del sensore e sull'analisi dei dati, incluso il riconoscimento dei modelli per migliorare la qualità del caffè grazie alla registrazione continua di parametri significativi di estrazione del prodotto. Infine, per comprendere le esigenze degli utenti in termini di benefici e barriere percepite e per aumentare la consapevolezza delle problematiche associate al lifelogging, sono state svolte attività trasversali durante un Progetto Internazionale ("PAAL" - Privacy-Aware and Acceptable Lifelogging services for old and fragile people, More Years Better Lives JPI) in collaborazione con l'Università di Alicante, Università di Toronto, Università di Stoccolma e Università di Aquisgrana.

Motivations and aim of the research

In daily life, individuals are constantly overexposed to stimuli, that can provoke unconscious experience-related changes of their physiological status. Thus, there is a need for effective methods to accurately measure the physiological responses for the early detection of modified conditions, which can affect both the mental and physical human health-related status. Although the medical instruments are the gold standard for physiological measurements, nowadays the healthcare provisioning paradigm is rapidly shifting towards remote monitoring. In this regard, wearable devices, especially wrist wearables such as smartwatches and wrist bands, have gained more and more popularity for daily collection of personal health-related data. In addition to the advantage of reduced size, comfortable use and relatively low cost, the majority of wrist wearables include multisensors system providing real-time continuous data about vital signs (e.g., cardiac and skin activity), along with further information (e.g., activity and physical performance) that can be collected in uncontrolled conditions and used for several purposes.
Among the others, a typical measurement manifestation of the sympathetic system to external stimuli are the mostly involuntary responses such as heart rate, sweating, breathing, and eye blinking. However, although a variety of sensor technologies are already available, some research questions are still open. The right collection and determination of such information from the huge amount of sensor data is a complex task in remote healthcare monitoring. Moreover, due to the complexity of the human's reactions and behaviours, extracting a meaningful knowledge for the context of the monitored person and detecting the health status represent open research challenges.

Within this context, the overall objective of this Ph.D. thesis is to quantitatively assess human response measurement to external stimuli, by means of physiological signals. This aim is achieved by developing and testing measurement procedures to validate wearable devices performance, and by investigating and testing proper signal processing techniques. As such, the research questions addressed in this Thesis work are the following ones:

- Can a wearable device accurately detect physiological changes due to stimuli?
- How can the accuracy performance be improved?
- Which factors can affect data collection from wearable devices or cause misinterpretation of data processed?

In this work, commonly-used physiological signals, such as blood volume pulse and the related derived parameters, are considered together with the skin conductance, specifically associated to involuntary human responses to different types of stimuli. Hence, starting from the analysis in time and frequency domain of the skin conductance signal (related to the skin sweating), different groups of healthy subjects were involved in several experimental tests: from listening to audio clips (Chapter 2.1), through performing a simulated driving path (Chapter 2.2), to engage in physical activities

(Chapter 2.3). Data were collected mainly through the commercial wearable device Empatica E4 (described in Chapter 1.5).

This first investigation gave rise to a second analysis on cardiac activity measurements (i.e. heart rate, heart rate variability and blood pressure), conducted by healthy subjects, while wearing the Empatica E4 at rest, during acoustic stimuli and after physical exertion. Being the gold standards for cardiac measurements, some commercial wearable devices were validated, including a metrological characterisation available in the market in terms of accuracy and precision (values currently debatable, since unavailable or not reported with rigorous validations from the manufacturers for wearables) to underline the importance of the device selection depending on the target application (Chapter 3.1). Further analyses investigated the factors interfering the photoplethysmographic sensor readings and consequently the quality of data processed, such as motion artefacts and skin-device contact pressure (Chapter 3.2). The overmentioned aspects are strictly related to the decision-making processes that need reliable and high-quality information for a proper interpretation of the results and for the right decisions, especially in the healthcare context (e.g., personalised care and remote monitoring). Within this domain, wrist acceleration signals from wearable devices can be affected by several factors. Therefore, data were recorded during daily human activities performed in a real-life setting, not-controlled environment, without instructions or guidance, by wearing wearable devices (Chapter 4.1). Different hardware sensor technologies, playing a paramount role to determine the classification accuracy, were compared to quantify how the nature and quality of the collected measurement data affect the performance. Moreover, since long-term data collection may expose users to privacy violations, due to leakage of personal details, a Multi Objective Evolutionary Algorithm approach was implemented to conceal subject's gender while maximising the accuracy on data recording (Chapter 4.2).

General discussion and conclusions of the Ph.D. thesis are reported in Chapter 5. As the concluding part of the studies performed, a specific Appendix section is dedicated to the transversal activities on assistive technologies (Chapter 6.1) and experiences performed abroad (Chapter 6.2).

A block diagram related to the structure of this thesis is reported below 1.

It is important to underline that the participants were always informed about the purpose and the methods of the study and they decided voluntary to be part of the research, by signing the informed consent compliant to the General Data Protection Regulation (GDPR) before starting the tests (which were performed following the principles outlined in the WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects [1]).

Some of the experimental studies described in this Thesis were affected by the spread of COVID-19 pandemic, representing a social and economic context posing some restrictions for involving a large test population.

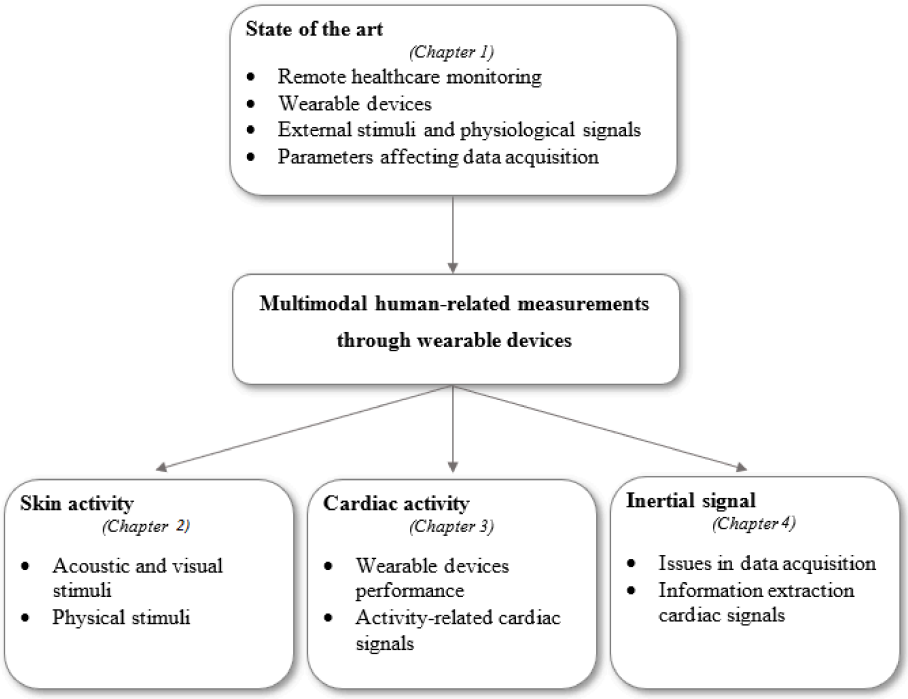


Figure 1: Block diagram related to the structure of the Thesis.

Contents

1 Introduction **1**

1.1 Remote healthcare monitoring 1

1.1.1 Wearable sensors for human-related measurements 2

1.2 Machine Learning 3

1.3 External stimuli response and physiological responses 7

1.4 Issues related to data acquisition 10

1.5 Empatica E4 Acquisition Device 11

2 Measurement of Skin Activity **13**

2.1 Case study 1: Physiological Response to Acoustic Stimuli 13

2.1.1 Acquisition protocol 14

2.1.2 Data Analyses 16

2.1.3 Experimental Results 17

2.1.4 Discussion and conclusions 20

2.2 Case study 2: Physiological Response to Visual Stimuli 21

2.2.1 Acquisition Protocols 23

2.2.2 Data Analyses 23

2.2.3 Experimental Results 26

2.2.4 Discussion and conclusions 27

2.3 Case study 3: Physiological Response to Physical Stimuli 28

2.3.1 Acquisition protocols 28

2.3.2 Data Analyses 29

2.3.3 Experimental results 31

2.3.4 Discussion and conclusions 35

3 Measurement of Cardiac Activity **37**

3.1 Case study 4: Assessment of wearable devices performance 37

3.1.1 Acquisition protocols 39

3.1.2 Data Analyses 45

3.1.3 Experimental results 46

3.1.4 Discussion and conclusions 53

3.2 Case study 5: Assessment of activity-related cardiac signals 57

3.2.1 Acquisition protocols 58

3.2.2 Data Analyses 59

3.2.3 Experimental results 64

3.2.4 Discussion and conclusions 67

4 Assessment of Inertial Signal **69**
4.1 Case study 6: Issues in data acquisition from wearable devices 70
4.1.1 Human activity recognition and factors of influence 70
4.1.2 Acquisition protocol 70
4.1.3 Data Analyses 72
4.1.4 Experimental results 74
4.1.5 Discussion and conclusions 79
4.2 Case study 7: Issues in information extraction from wearable devices . 81
4.2.1 Human activity recognition and privacy preservation 81
4.2.2 Acquisition protocol 82
4.2.3 Data Analyses 82
4.2.4 Experimental results 84
4.2.5 Discussion and conclusions 86

5 Discussion and conclusions **89**

6 Appendix: Additional Research Activities **93**
6.1 Assistive Technologies for Ageing 93
6.2 Experiences and collaborations abroad 97
6.2.1 Influence of affective environments in ageing at home 97
6.2.2 A smart sensor system to improve the beverage systems 99

Publications **121**

List of Figures

1 Block diagram related to the structure of the Thesis. xiii

1.1 General framework of the machine learning pipeline. 7
1.2 9-points SAM scale: valence (first row), arousal (second row) and dominance (third row). 9
1.3 Schematic representation of different influencing parameters which are part of the output function. Specifically, q_i and q_o represent the input and output quantity, respectively, while $d_1, \ldots, d_n$ correspond to the different influencing disturbances. 10
1.4 Empatica E4 wristband and sensors embedded. 11

2.1 General acquisition protocol divided in three main phases: rest, stimuli-based elicitation and physiological response after the stimuli. 15
2.2 Median filter based extraction of the tonic (SCL) and phasic (SCR) components from SC signal. 16
2.3 Portion of SC signal to pleasant sound for S2: a) SC signal, b) SCR component and peaks marked with a star. 18
2.4 Normalised PSD of acquired SC signals to a) neutral, b) pleasant and c) unpleasant acoustic stimuli. The vertical red lines indicate band limits: 0.25 Hz, 0.40 Hz and 1.00 Hz. 19
2.5 Experimental setup: driving simulator, monitor with overnight path, Empatica E4 device position and tablet with KSS scale. 23
2.6 Sample acquired BVP (left) and SC (right) signals during a whole session (40 min duration). 24
2.7 Comparison of SC mean values in time, over the test population for all the intensity levels. 31
2.8 *Discrepancy* for the normalised mean value features computed over the test population for all the intensity levels. 32
2.9 Example of FFT magnitude: tonic (left) and phasic (right) components (S2); a) at rest, b) medium, c) high intensity activity. 33
2.10 Confusion matrices for a) SVM classifier and b) BT classifier, applied on cross-domain physiological and acceleration data. 34
2.11 Confusion matrices for a) SVM classifier and b) BT classifier, applied on SC data. 35

3.1 HR measurement setup: smartwatches worn on both wrists, pulse-oximeter placed on the right index finger; BP meter cuff on the left arm; chest-strap monitor on the thorax, where also ECG electrodes were positioned (I lead); artery pulse detected through palpating the carotid. 40
3.2 BP measurement setup: the three smartwatches (1) Samsung Galaxy Watch3, (2) Asus VivoWatch SP and (3) Omron HeartGuide HEM-6411T-MAE worn on the left wrist one by one, while the BP meter cuff located at the arm level, same side; subject seated and relaxed for the whole test duration. 43
3.3 DIY wrist-worn wearable device. 44
3.4 DIY wrist-worn wearable device components. 44
3.5 Distribution of deviations (a) and Bland-Altman plot (b) related to HR measurement (Polar Vantage V2). μ: mean of the deviation; σ: standard deviation of the deviations. 47
3.6 Correlation between HR measured by Garmin HRM-Swim chest-strap monitor and reference instrument (ECG). 48
3.7 Distribution of deviations values for Samsung Galaxy Watch: systolic (SBP) and diastolic (DBP) blood pressure data in $mmHg$. μ: mean of the deviation; σ: standard deviation of the deviations. 49
3.8 Bland-Altman plots for Asus VivoWatch smartwatch: systolic (SBP), and diastolic (DBP) blood pressure data in *mmHg*. The black solid line indicates the bias μ, while the dotted grey lines indicate the confidence interval ($\mu \pm 1.96 \cdot \sigma$). 51
3.9 Correlation between blood pressure systolic (SBP) and diastolic (DBP) data in $mmHg$, acquired by Omron HeartGuide and the sphygmomanometer. 52
3.10 Tightening force intervals (mean±standard deviation) measured for *loose* (a), *medium* (b) and *tight* (c) levels, for the whole population. . 53
3.11 Distribution of the measured tightening force (a) and HR (b) in the tests with different tightening levels: *loose*, *medium* and *tight*, for the whole population. 54
3.12 General acquisition protocol divided in three main phases: rest, stimuli-based elicitation and physiological response after the stimuli. 58
3.13 Peak detection implemented on BVP signal and the related RR intervals, i.e. tachogram (from left to right). 62
3.14 BVP signal and the extracted BVA. 62
3.15 Comparison between tachograms obtained with Proposed Method (a) and Kubios Method (b). In red the corrected data. 64
3.16 Example of deviations distributions between the two methods - LF_{peak} parameter. 65
3.17 FRI values nW/bpm before and after the physical exertion. 66

4.1 Arduino-based wristband prototype and the X, Y, and Z axes relative to the accelerometer onboard. 71
4.2 Empatica E4 wristband and the X, Y, and Z axes relative to the accelerometer onboard. 71
4.3 Samples plot illustrating the single acceleration components and the resulting SMV. 73
4.4 Comparison of accuracy before and after applying the Information Gain - Empatica (a) and Arduino-based (b) device acceleration measurement data. 78
4.5 Accuracy and number of trees for RF classifier. 85
4.6 Confusion matrix related to the HAR (before MOEA). 86
4.7 Final set of solutions for the MOEA algorithm. Blue dot: best balanced solution; red dot: initial result with all features are equally considered. For the sake of clarity, 0.5 has been added to f_2, which is then the gender accuracy. 87
4.8 Confusion matrix related to the HAR (after MOEA). 87

6.1 Workflow of vINCI platform. 94
6.2 Structure of the online survey. 95
6.3 Environments and ageing: from physiological data collection to the customisation of living environment based on personal feelings. 98
6.4 Smart sensor system: OOH machine, sensor in the machine and Coffee Brewing Control Chart. 99

List of Tables

1.1 Machine Learning algorithms used in this Thesis work, along with the related classification methods 5

2.1 Evaluation (mean ± standard deviation) of the audio clips chosen to elicit emotions, based on gender and dimension 15
2.2 Number of SC peaks for the three acoustic stimuli for each subject . . 19
2.3 PSD percentage computed within the frequency bands for the three acoustic stimuli 20
2.4 Features extracted from the SC signal and its components 25
2.5 Features extracted from HRV 26
2.6 Classification performance of proposed approach 26
2.7 Confusion matrix for RF algorithm 26
2.8 Time- and frequency-domain features from SC signal 30
2.9 Relative performance evaluation of the SVM and BT classifiers on the three classes, with cross-domain data 34
2.10 Relative performance evaluation of the SVM and BT classifiers on the three classes, with SC data only 35

3.1 Acquisition devices and related technical specifications (*Fs*: sampling frequency) 41
3.2 Technical specifications of the BP acquisition devices and modalities (NA = not available) 42
3.3 HR deviations analysis between tested methods and reference (ECG): mean of the deviation (μ), standard deviation of the deviations (σ) and 95% confidence interval (CI95%) 47
3.4 Correlation between tested methods and reference (ECG): m and b coefficients of the fitting line $y = mx + b$, where x and y are HR measured by ECG and the tested methods, respectively; ρ, Pearson's coefficient 48
3.5 Validation of BP acquisition devices, according to the recommendation criteria of the AAMI and the BHS 50
3.6 Distribution of SBP and DBP deviations between tested smartwatches and reference device, in terms of bias (μ) and 95% confidence interval (CI95%) 50
3.7 Fitting line $y = mx + q$, where x: data measured with the sphygmomanometer; y: data acquired with the tested smartwatches; ρ Pearson's correlation coefficient 51

3.8 Inter-subject variability related to tightening force values obtained from the load cell, for the three different tightening levels (i.e. *loose*, *medium* and *tight*) . . . 52
3.9 Inter-subject variability related to HR obtained from PPG sensor, with the three different tightening levels (i.e. *loose*, *medium* and *tight*) . . . 53
3.10 Features extracted from HRV signals in different domains . . . 61
3.11 Features extracted from the BVP signal . . . 63
3.12 Classifiers performance metrics . . . 66

4.1 Features extracted in time and frequency domain, divided per computation type . . . 74
4.2 Datasets of features . . . 74
4.3 Average percent accuracy of tested classifiers on acceleration data from Empatica E4 and Arduino-prototype devices . . . 75
4.4 Spectral Energy and Entropy values (mean ± standard deviation) computed for each device and ADL . . . 76
4.5 Partial and average percent accuracy (Avg.) of the tested classifier (CL.) – Empatica E4 . . . 76
4.6 Partial and average percent accuracy (Avg.) of the tested classifier (CL.) – Arduino prototype . . . 76
4.7 F-measure of the tested classifiers on ADLs-related acceleration data . 77
4.8 Subset of features selected by applying Information Gain filter . . . 78
4.9 ADLs performed by the subjects . . . 82
4.10 Features extracted in time and frequency domain, divided per computation type . . . 83
4.11 Confusion matrix related to the gender recognition (before MOEA) . . 86
4.12 Confusion matrix related to the gender recognition (after MOEA) . . 88

Chapter 1

Introduction

1.1 Remote healthcare monitoring

In the last few decades, the healthcare system almost worldwide relied on centralised hospitalisation, where the individuals were supposed to move to the hospital for receiving own medical diagnosis. Nowadays, the remote health monitoring of individuals is largely required and has reached an unprecedented importance and popularity, also due to the global epidemics, such as COVID-19 pandemic [2]. Moving the medical care services from the hospital to the individual's home, thanks to the use of assistive remote technologies, offers significant opportunities to speed up the care, improve health outcomes and reduce the hospitalisations, along with the related reduction of the costs. In particular, today the individuals themselves or healthcare personnel can daily monitor health data and potentially receive an alarm when recorded measurements exceed a predetermined threshold. Undoubtedly, the continuous monitoring both patient's physiology (e.g., heart rate, blood pressure and skin activity) and behaviour (e.g., physical activity or fall events) from a distance [3] may allow the early detection or tracking of diseases and assessment of wellbeing; as an example, for the early diagnosis of COVID-19, together with blood oxygen saturation and physical activity level, the heart rate has been demonstrated to be a fundamental variable [4].

In this perspective, a paramount role is played by the miniaturisation of electronic components and the explosive growth of Information and Communication Technology (ICT) along with the Internet of Things (IoT) infrastructure, resulting in many commercial devices as "IoT-enabled" and, consequently, capable to share data from remote [5]. Indeed, the development of IoT-enabled wearable devices with wireless technology support has allowed and facilitated the transition from measurements conducted in laboratory settings, typically with bulky wired instruments, to minimally-invasive, comfortable and real-time recordings, in free-living conditions [6] with devices capable of streaming data to a cloud-based repository. Such remote healthcare monitoring system, known as telemedicine, results in the Internet of Medical Things (IoMT) paradigm [7, 8], which is leading towards the so-called Healthcare 4.0. Specifically, robust Artificial Intelligence (AI) algorithms are applied to manage the physiological data variability, especially for digital health applications [9]. To this aim, audio-, video- and sensor-based systems are explored in many application fields, such as Ambient Assisted Living (AAL), anomaly living patterns recognition, fall detection, fitness-

oriented purposes, disease detection, emotion recognition and pandemic emergency like COVID-19 [10, 11], just to cite some. Among the sensor-based systems, wearable devices are gaining more and more consensus for several applications (even if further actions and improvements are still needed to reduce the barriers for a wider adoption by older adults [12]); those of interest for this Thesis will be described in the following sections.

1.1.1 Wearable sensors for human-related measurements

Wearable sensors have been prominently featured in the healthcare domain, especially because of some advantages with respect to the standard equipment for monitoring the physiological parameters: user-friendly, comfortable, relatively low-cost, availability in different forms and quality, becoming capable to satisfy different customers, also thanks to user-experience oriented design of these devices [13, 14, 15]. Moreover, nowadays a single device can provide multiple physiological acquisition, such as heart rate, heart rate variability, energy expenditure, blood oxygen saturation, skin temperature, skin conductance, stress-related index, walking distance etc. An additional success of wearable devices is attributable to the possibility of easy remote health monitoring through a proper IoT architecture, generally working with WiFi and Bluetooth to connect with other devices and share individual-related data for example on a Cloud platform, making them remotely available and safely stored [15]. This aspect is particularly relevant in health monitoring applications, also to support the healthcare providers in decision-making processes; in fact, such wearable systems are capable to collect data 24 hours a day, seven days a week. This generates a huge amount of data, the so-called "big-data", that is potentially useful for AI algorithms training to provide useful information for the decision-making processes applied to different purposes. This offers a valid solution for well-being assessment [16], personal comfort measurement [17], stress level quantification [18], Activities of Daily Living (ADLs) recognition and classification [19], just to cite some. Among the existing sensors used for human-related measurements, the two categories of sensors used in this Thesis are following listed:

- **Biosensor**: biosensors detect a chemical substance, that combines a biological component with a physicochemical detector. However today, the term includes a wide type of sensors that are able to collect vital signs and physiological signals, encouraging a proactive approach to preserve individual's health through daily gathered data [20]. To this purpose, clinical-grade biosensors and wearable devices are expected to provide predictive guidance for clinical interventions. Examples of physiological signals collected from wearable sensors, which will be discussed in the Thesis, are the Heart Rate (HR), Blood Pressure (BP), Skin Temperature (SKT) and Skin Conductance (SC) [21].
- **Inertial sensor**: inertial sensors transduce inertial force into electrical signals to measure acceleration, inclination, and vibration of an object. Among the most frequently used for the human daily activity recognition, accelerometer

measures the value of gravitational force along one or more sensitive axes; as a result, such sensors, individually or in combination as Inertial Measurement Units (IMUs), provide descriptive features of body posture and user's movement [22]; hence, real-life applications include the monitoring of activities such as performing physical exercises, walking, standing, sitting, or walking upstairs and downstairs, from which several pathological states can be deducted (e.g., Parkinson's disease [23] and post stroke [24]).

Sensors embedded in wearable devices (e.g., wrist bands, watches, bracelets and belts) have the strong advantages to be minimally invasive, non-intrusive and non-bulky with respect to the hospital instruments; however, some limitations are evident. The placement of sensors on different body positions, or incorrect wearing, may result in different readings; combination of several sensors may lead to cumbersome feeling and low user's acceptance. Moreover, the overall system performance inevitably depends on the test protocol adopted and on the metrological characteristics of the wearable device used for acquiring the data [25], which need to be considered together with the intra- and inter-subject variability impacting on the system usability. Additionally, metrological properties of commercial wearable devices are often unavailable [26], and consequently data collected are often difficult to handle and the extracted indicators to process as well. Evaluation of accuracy and reliability of these devices require particular attention when using data gathered by wearable devices for healthcare purposes: to date, only a few wearable devices are classified as medical devices, e.g., Apple Watch Series 4-6 [27].

1.2 Machine Learning

As mentioned above, over the past few years, the monitoring and collection of individuals' and environmental features have been supported by the fast development of the sensing devices (e.g., optical sensors, accelerometers and cameras). However, most applications involve much more than simply gathering measurements from the quantities of interest. Indeed, healthcare systems may benefit from such technological advancements to ensure tailored services. In particular, AI along with Machine Learning (ML) techniques play a significant role in the effective diagnosis of monitoring patient's health condition, and in supporting the suitable measures for detected abnormality, chronic diseases or impairments [28, 29, 30]. The main idea is to create a machine that automatically builds the analytical models thanks to the designed algorithms, which iteratively learn from data without any human intervention. There are two ML approaches, namely supervised and unsupervised learning, which deal respectively with labelled and unlabelled data, depending on the association to a known actual class or not. Some other systems work with semi-supervised learning in which some data are labelled but most of them are unlabelled [31]. Most of the healthcare monitoring systems work firstly in a supervised context because the learning activity systems should return a label activity such as walking, stepping, hypertension

etc. Typically, the supervised learning algorithms learn a relationship between the input data attributes (features) and the target attribute (class label), by minimising a loss function determined by the pairs of input data and the corresponding target attribute. The result, that characterises the discovered relationship, is a model used for predicting the class label having the input data [32].

The classification approaches mainly used for healthcare monitoring aims may be grouped in three main categories, based on the estimation criterion for adjusting the parameters of the classification method:

- Discriminative, which create the decision boundaries in the feature space, thus learning the feature mapping to class labels. Such techniques face the problem of over-fitting and the large amount of labelled data required for training;
- Generative, which are statistical approaches to the pattern recognition problem. Although a huge amount of training data is needed, these techniques are flexible and able to deal effectively with the data uncertainty. Moreover, the generative techniques learn from both labelled and unlabelled data;
- Heuristic, which is a hybrid modality using a combination of generative and discriminative approaches. Generally, it was found to reach the best predictive performance with respect to its single use.

Among the overmentioned classification methods, the ML algorithms used in this Thesis are detailed in Table 1.1.

Such ML algorithms are generally fed with features extracted from time and/or frequency domain, representing the meaningful information computed from the gathered signals. Among the feature engineering techniques exploited in this Thesis, the Multi-Objective Evolutionary Algorithm (MOEA) is detailed in the next section. In almost all the studies that will be presented hereafter, the k-fold cross validation was implemented to train and test the models. This method randomly divides the entire set of features into k subsamples: in which $k-1$ subsamples are used as training data and a single subsample for testing the model, as validation set. After each subsample has been used once as the validation data, the accuracy percentage from the folds is the average over the k iterations, resulting in the cross-validated performance.

Multi-objective evolutionary algorithms

In multi-objective optimisation, the goal is to optimise simultaneously several objective functions [33]. These different functions have conflicting objectives, i.e. optimising one affects the others. Therefore, there is not a unique solution but a set of solutions. The set of solutions in which the different objective components cannot be simultaneously improved constitute a Pareto front. Each solution in the Pareto front represents a trade-off between the objectives. MOEA aims to [34]: (1) find a set of solutions as close as possible to the Pareto front (convergence); (2) maintain a diverse population that contains dissimilar individuals to promote exploration and to avoid poor performance

Table 1.1: Machine Learning algorithms used in this Thesis work, along with the related classification methods

Approach	**Algorithm**	**Classification method**
Discriminative	Decision Tree (DT)	• tree-like graph from the root to a leaf node; • attribute values with hierarchical architecture; • if node condition is satisfied, next node is checked until leaf containing classification result.
	Random Forest (RF)	• ensemble of decision trees; • random trees created varying at each step the features subset and descriptors of space; • outcome is an average of all the trees trained.
	Support Vector Machine (SVM)	• instances map in higher dimensional space; • to divide data with linear decision boundary (optimal hyperplane); • outcome is only from data closest to boundary.
	Artificial Neural Networks (NNs)	• human biological neural system model; • reception, processing and transmission of information replicated through network links;
	K-Nearest Neighbor (kNN)	• instances labelled as the most common class among k-nearest neighbors (measured by distance function - e.g., Euclidian);
Generative	Naïve-Bayes (NB)	• attributes conditionally independent given the class variable value;
Heuristic	/	• hybrid approach combining the generative and discriminative approaches.

due to premature convergence (diversity); and (3) obtain a set of solutions that spreads in a more uniform way over the Pareto front (coverage).

In our specific work the Non-dominated Sorting Genetic Algorithm II (NSGA-II) [35] was employed as wrapper algorithm. NSGA-II: (1) uses an elitist principle,

i.e. the elites of a population are given the opportunity to be carried to the next generation; (2) uses an explicit diversity preserving mechanism (Crowding distance); and (3) emphasises the non-dominated solutions. The algorithm will obtain a set of solutions, some of them optimising one over the other objective and vice versa. From these set of solutions, a specific solution fulfilling particular conditions could be selected.

Performance evaluation metrics

According to the literature, there is a wide range of metrics commonly used to evaluate the performance of classification algorithms. In the current Thesis, for each tested approach we focused on two evaluation measures that are the percentage accuracy and the F-measure. The former is defined as the number of correctly classified instances over the total number of instances considered [32], or:

$$accuracy = (1 - ErrorRate) \times 100 \tag{1.1}$$

where:

$$ErrorRate = \frac{|N_{cci} - N_{ti}|}{N_{ti}} \tag{1.2}$$

being N_{cci} the number of correctly classified instances, and N_{ti} the number of total instances considered by each classification algorithm. Contrarily, F-measure refers to a weighted harmonic mean of the recall and precision, as an indicator of the overall effectiveness of the activity classification. According to the definitions, the F-measure (also named F-score) can be calculated as per (1.3):

$$F - measure = \frac{(1+\beta)^2 \times Sensitivity \times Precision}{\beta^2 \times Sensitivity + Precision}. \tag{1.3}$$

Assuming that β is the weight coefficient ($\beta = 1$ in WEKA tool), *Sensitivity* (or equivalently *Recall*) is the ratio of correctly classified instances to the total number of identified instances that are true, calculated as:

$$Sensitivity = \frac{TP}{TP + FN} \tag{1.4}$$

whereas *Precision* (or *Specificity*) is the ratio of correctly classified instances to relevant instances, defined as:

$$Precision = \frac{TP}{TP + FP}. \tag{1.5}$$

Hence, F-measure value depends on the above-mentioned performance indicators: True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) rates. An F-measure, as well as the Area Under the ROC Curve (AUC), close to 1 means high values of accuracy and strong algorithm. Such indicators are also used to generate the Confusion Matrix, that is a table summarising how the activities were

classified. Each row represents the data points and the actual instance label, while each column indicates the data points and the predicted instance by the classifier. The main diagonal represents the number of instances correctly classified, hence those for which the predicted instances equal the actual ones. The values outside the diagonal identify the prediction errors.

The general pipeline for the implementation of ML models is depicted in Figure 1.1.

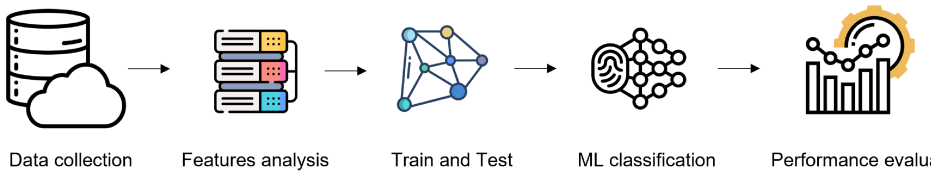


Figure 1.1: General framework of the machine learning pipeline.

1.3 External stimuli response and physiological responses

Recently, there has been an increasing interest in human emotion recognition on a global scale, particularly in both the healthcare [36] and neuromarketing [37] sectors. In 1989, Thoits [38] defined the emotion as "a combination of four main factors: subjective experience in assessing the current stimulus or context, emotion expressions with gestures, culture and physiological responses". Later on, Plutchick [39] defined a taxonomy to classify on a discrete scale eight distinct emotions: joy, trust, fear, surprise, sadness, disgust, anger, and anticipation. Moreover, multi-dimensional space-models have been developed, taking into account valence (pleasant/unpleasant), arousal (high/low) and also dominance (submissive/dominant), identifying different spaces characterising emotions.

Emotion recognition is most commonly based on speech analysis, since speech signals are the most natural, intuitive and fastest means of interaction for humans and are known to carry much information. Therefore, speech analytic, facial/body expressions and self-reports have been widely investigated and used for the human emotion detection [40, 41]. However, these approaches are not totally reliable for identifying emotions, especially if the subjects under test want to hide their own feelings. Although subjects may mask externally the feelings, undoubtedly the inner physiological fluctuations are under the control of the sympathetic nervous system, thus without voluntary control [42]. Generally, every emotion is accompanied by both physical and/or physiological reactions, and consequently changes in physiological values with respect to a baseline measurable at rest conditions and in the absence of any stimulation. Examples of such variations, produced by the human's nervous system, are changes of the body hormonal levels, reactions like sweat production, increased heart rate and muscle activation. Respiration becomes faster and the blood pressure increases, resulting in speech characteristics changed too. Skin temperature decreases together with hands and feet temperature and the heart rate variability

decreases. Moreover, pupil diameter can vary. Knowing all these potential physiological changes, a reliable alternative may be built around the use of physiological signals, from which objective measurements can be obtained, to identify the participants' unconscious emotional changes. If thinking to the emotions as subject's reaction to a stimulus, the monitoring of physiological signals might provide useful information on the presence/absence of stimuli influencing the subject's state. When individuals are exposed to daily circumstances and stimuli such as images, sounds, noises and physical, the sympathetic division of the Autonomic Nervous System (ANS), with no conscious control, induces emotions such as fear, anxiety, stress, joy and sweat reaction (even if sometimes imperceptibly). To this aim, several physiological signals have been exploited in the literature, such as the Electrocardiogram (ECG), the Electroencephalogram (EEG), the Electromyogram (EMG), the Photoplethysmogram (PPG), the Skin Conductance (SC) signal, and the Skin Temperature (SKT) signal [43]. Among the others, in this Thesis, skin- and cardiac-related changes are investigated and assessed.

SC signal, named also as Electrodermal Activity (EDA) or Galvanic Skin Response (GSR), is a biometric index reflecting changes in the electrical properties of the skin and, consequently, the sympathetic activity [44, 45]. Thanks to the availability of wearable devices for easily monitoring, the SC is currently one of the most involved signals in the emotion research. Changes of the skin electrical characteristics can be monitored by placing two electrodes on specific areas of the skin surface (e.g., fingers, hand, and foot palm [46]). Research studies have suggested two different methods to measure SC signals: the exosomatic approach (applying a direct current (DC) or an alternating current (AC) source) and the endosomatic approach (without applying any external current or voltage source) [47]. The former approach is the most commonly used, with an external constant voltage source that is applied on the human skin through electrodes.
Among the others, the wrist is not the most appropriate place to measure SC, nevertheless, it is considered a comfortable position for the user and also for the designer, to hold electrodes and drive the SC signals through the bracelet to the microcontroller onboard the device. The gathered information is double: the tonic component (i.e. Skin Conductance Level, SCL), which slowly changes over time being the individual's baseline characteristic, and the phasic component (i.e. Skin Conductance Response, SCR), which is related to fast-changing signal contributions and to events [48].

Several research studies already demonstrated the connection between heart rate variability (HRV) and ANS, so that HRV-derived features are typically exploited to detect human feelings, by recognising emotions, mental fatigue and also stress generated by physical activity as well [49, 50, 51]. Additionally, the HRV analysis, i.e. the physiological variability of the cardiac rhythm, along with Heart Rate (HR) and Blood Pressure (BP), according to the autonomic regulation, may be used to evaluate different health conditions, such as cardiovascular diseases, hypertension, diabetes, pregnancy. Recently, some commercial wearable devices, especially those acquiring electrocardiogram (e.g., Apple watch [52]), have been demonstrated to be a valid tool

for detecting the likelihood of abnormal cardiac activity (e.g., abnormal arrhythmia and hypertension), providing guidelines aimed at preventing cardiovascular diseases (e.g., heart failure), cognitive impairments diseases (e.g., dementia) and other health complications (e.g., stroke, blindness and kidney dysfunction), consequently promoting a better quality of life [53, 54]. Being tools daily used by individuals, it is essential to consider the metrological performance of such acquisition devices, to properly interpret results and make right decisions, being conscious of the accuracy required by the target application. Most probably, if using at the same time a wearable device (e.g., photoplethysmographic (PPG) - widely spread in the current smartwatches) and the related gold standard (e.g., electrocardiograph, ECG), the former will show a different metrological performance.

Together with the physiological signals, also the self-assessments have to be encoded somehow, representing the ground truth and labels for the related data collected. Questionnaires should be provided and filled out right after the stimuli elicitation and experiencing emotions [55]. According to the dimensional emotional models, e.g., Russell's circumflex [56], equivalent to valence-arousal-dominance model used in the very popular Self-Assessment Manikin (SAM) shown in Figure 1.2. The graphical pictures, shown to the participants of experimental tests, represent the three different domain of emotions: valence (first row) from a frowning figure to a smiling one, arousal (second row) from a sleepy to a widely awake with an incremental explosion at the centre and dominance (third row) from a very small to a very large figure, where the largest figure describes maximum control in situation.

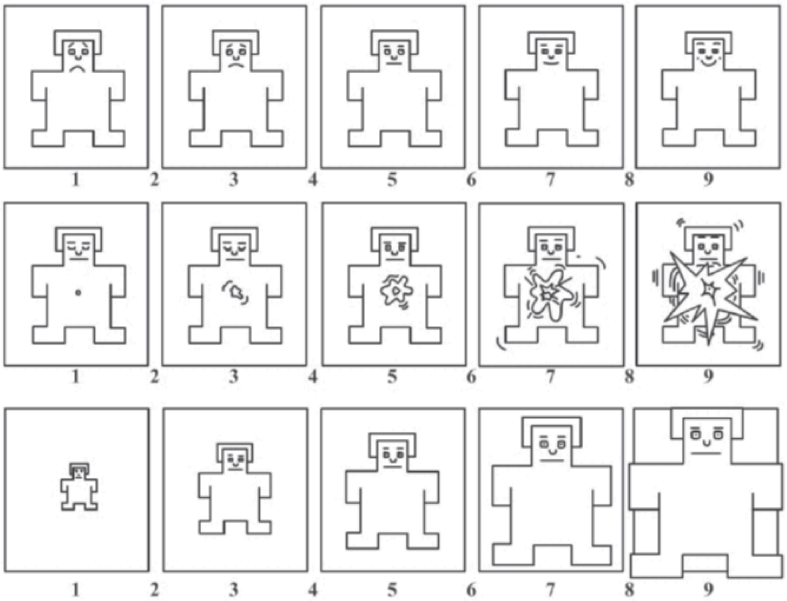


Figure 1.2: 9-points SAM scale: valence (first row), arousal (second row) and dominance (third row).

1.4 Issues related to data acquisition

As mentioned above, nowadays a continuously growing number of people wear a smartwatch or a smartband, both able to monitor a plethora of physiological parameters. These measurements can be useful for different purposes, from cardiovascular monitoring to emotion recognition, through activity assessment, fitness-oriented applications and blood pressure observation, just to cite some. Nevertheless, some critical aspects should be thoroughly considered, especially pertaining to health-related data privacy issues and measurement accuracy of these innovative wearable instruments [26], which undoubtedly play important roles in the era of personalised medicine and digital health [57, 58]. In order to develop robust models from physiological signals collected through wearable devices potentially 24 hours a day, 7 days a week, capable to provide reliable information, data quality is fundamental [59]. Wearable technology is characterised by a significant sensitivity to various influencing factors that affect acquisitions and results interpretation, especially in normal daily life conditions. Figure 1.3 highlights how the output signal is not a function of the single input quantity, but of a combination of multiple influencing factors, which in many cases cannot be controlled or isolated.

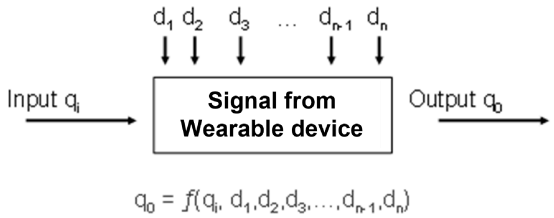


Figure 1.3: Schematic representation of different influencing parameters which are part of the output function. Specifically, q_i and q_o represent the input and output quantity, respectively, while $d_1, \ldots, d_n$ correspond to the different influencing disturbances.

The challenges are mainly due to the hardware and acquisition options (e.g., sampling frequency, signal-to-noise ratio (SNR), resolution, etc.), along with the software possibilities (especially data processing running in the device, such as artefacts removal algorithms) through the measurement procedure optimisation (e.g., protocol used, such as laboratory versus realistic data collection) that have a big impact; however, also the data variability, linked both to different sources to collect data [60], but also to the physiological variability itself [9] have a huge impact. In this perspective, J.A. Behar et al. [61] underlines also the limited standards to represent data (e.g., the amplitude resolution and sampling frequency), and also the frequent development of relatively small datasets failing to capture the variability across patients and time; this definitely reflects on reduced robust algorithms able to interpret the information contained in physiological time series collected. Indeed, the classification performance of AI algorithms surely depends on the variability observed in the data collected on the test population: although data variability hinders a perfect discrimination among

classes, it is necessary to test a wide population in order to include its variability (intra- and inter-subject) and avoid overfitting issues.

These aspects should be thoroughly considered in function of the target application requirements, especially for digital health applications, which cannot neglect physiological variability characterising the involved population, and consequently the measured data. Similarly, among the challenges which highly affect the quality of the human activity recognition, the most common ones are how to collect the data in the real-life conditions, and how to select and extract the features to be computed [62].

1.5 Empatica E4 Acquisition Device

Both the individual behaviours and physiological parameters were mainly measured by wearing the Empatica E4 device (shown in Figure 1.4 [63]). Empatica E4 is a wrist-worn multi-sensor device, with a medical certification (Class IIA Medical Device according to the CE Crt. No. 1876/MDD - 93/42/EEC Directive), designed for real-time, continuous and comfortable monitoring in free-living conditions. It is worthy to underline that when the experimental tests were executed the Empatica E4 smartband was on the market as the highest quality Empatica device; however, currently it is out of production to be soon replaced by a latest generation device.

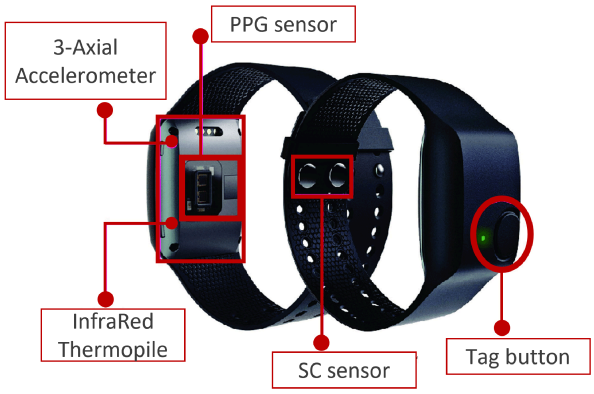


Figure 1.4: Empatica E4 wristband and sensors embedded.

In the datasheet, the manufacturer declares four sensors embedded in such device with their own technical specifications, as follows:

- PPG sensor, sampling at 64 Hz, provides the HR and the IBI signal based on the BVP detected. The digital sensor output, with a resolution of 0.9 $nW/Digit$, is generated by the light produced with 4 light emitting diodes (LEDs, 2 green and 2 red ones) and 2 photodiodes with a total sensitive area of 14 mm^2. The light during the green exposure mainly contains the information on the heartbeats, while the red exposure helps the reduction of motion artefacts that are dynamically compensated by firmware;

- SC sensor measures the changes in skin electrical conductance with a sampling rate of 4 Hz, in the range of [0.01, 100] μS and with a resolution of 1 digit$\sim$900 pS. Through 10 mm diameter silver-coated with copper underlay on brass electrodes placed on the ventral wrist, a small alternating current (8 Hz frequency – max 100 μA) is applied to the user's skin;

- 3-axial MEMS accelerometer to measure motion-related activity with a sampling frequency of 32 Hz and a resolution of 8 bit of the selected range. Each acceleration sample includes the three components, representing the three spatial directions (X, Y, Z axes). The conversion factor between raw samples and true acceleration values is attained by $g/64$ (where $g = 9.81\ m/s^2$), given that according to E4 manual a sample component value of 64 corresponds to 1 g, in the measurement range [-2g, 2g] m/s^2 ($\pm 4g$ or $\pm 8g$ can be selected with custom firmware);

- IR thermometer is configured with a sample frequency of 4 Hz with a resolution of 0.02 $°C$. SKT values are measured by an optical thermopile sensor. The reported accuracy within the range of human skin temperature (i.e. 36–39 $°C$) is $\pm$ 0.20 $°C$. Calibration is valid in the range [-40, 115] $°C$.

It is important to notice that the over-mentioned sampling frequencies, specified in the E4 wrist-worn device datasheet, are not customisable values. Prior to collecting data, the working modality can be chosen. Indeed, the device can operate either in streaming mode for real-time data visualisation using a Bluetooth Low Energy (BLE) interface and the E4 Realtime app from supported mobile devices, or in-memory recording mode storing temporarily data in the internal flash memory. The battery duration declared by the manufacturer is up to 24 hours in the former modality, more than 48 hours in the latter. In any case, the recording sessions associated to the serial number of the used Empatica E4 are saved into a secure cloud-based repository (i.e. E4 Connect) for data management. The cloud-based repository provides to each registered user the list of recorded sessions, and a dashboard for data visualisation; each session, saved in the form of a .zip archive, includes separate .csv files, each one pertaining to a specific sensor. After pressing the button to start the recording, the device takes around 15 s to calibrate, thus improving the accuracy of the sensor readings.

Chapter 2

Measurement of Skin Activity

This Chapter is focused on a set of studies aimed to identify the elicitation of acoustic, visual, and physical stimulation on the skin conductance signal collected through the commercial wearable device Empatica E4 by healthy individuals. Data analysis in time domain, performed in [64], showed that listening to an unpleasant sound increases the subjective physiological response (higher number of peaks raised on the skin conductance signal), especially when the sound duration is short. Additionally, in [65] SC proved to be a standalone reliable signal in recognising a driver's drowsy status (Random Forest classification: 84.10% accuracy; reaching 93.00% of accuracy when combined with BVP data [66]) and also in distinguishing different levels of physical exertion when extracting features from frequency domain [67]; in this regard, the proposed larger observation band (i.e. [0.40, 1.00] Hz) can avoid loss of potentially meaningful information about the skin activity, with respect to the common investigated one [0, 0.40] Hz, as demonstrated in [68]. These analyses were conducted also in collaboration with Prof. Florez-Revuelta, at the University of Alicante to understand whether and how living environments design affects the individual emotional capacity.

To examine the feasibility of the proposed approaches, data analysis was performed first in MATLAB environment and then, when needed, by using the WEKA tool for the ML performance evaluation.

2.1 Case study 1: Physiological Response to Acoustic Stimuli

By stimulating emotional responses with external stimuli, variations in physiological parameters (e.g., heart rate and skin conductance) can be measured. Among the stimuli that the individuals are daily facing, the acoustic, visual and physical ones are the most common. Regarding the auditory stimuli, the International Affective Digital Sounds (IADS) [69] database is among the most used collections of short sound clips (6 s), each one associated to classification labels obtained by using the Self-Assessment Manikin (SAM) [70] and the three basic-emotion rating scales (i.e. Arousal, Valence and Dominance). Although, the SC signal resulted reliable to assess the presence of cognitive stress in subjects [71], relatively few studies have investigated the SC response, either under several external stimuli either in standalone processing.

Pozzi et al. [72] suggested a features analysis by combining the physiological and audio signals, to understand the relationship between music and emotional feelings. Although reaching good results, some open issues were declared on both the reliability of ground truth data and the evaluation of prediction results due to the subjective perception of emotion - strongly affecting the classification performance. This happens in [73], where the overall accuracy percentage in recognising the emotion valence, from SC signal, shifted from 81.1% to 57.5%, for subject-dependent and subject-independent classification, respectively. As an example, the familiarity of music can influence the personal event, by inducing learned emotional responses rather than totally unconscious experience. Different engagement is reflected on bodily changes, such as the SC [74] and the HRV [75], sometimes combined to detect user's emotion response to music, considering also the personality and music preferences.

Whatever stimulus is used, it may affect participants' cognitive, and consequently physiological status. In order to consider both the user's perception of an emotional stimulus together with the physiological recordings, the self-assessment questionnaires are used in literature as ground truth data, such as the 9-points SAM scale. The overmentioned studies are mostly focused on the emotion recognition and classification, performed by extracting several features to feed and test different ML algorithms. However, in order to achieve high performances for automatically detecting the emotions, a detailed analysis of the measurement data properties is essential. Therefore, in this work, the characteristics of the SC physiological signals in response to acoustic stimuli, are performed, namely by analysing the event-related changes in SC curve morphology in both time and frequency domains. Such signals are measured in real-life contexts, i.e. out of a lab, thanks to the use of the Empatica E4 device, which may open new possibilities in terms of exploitation of physiological information generated from wearable devices.

2.1.1 Acquisition protocol

In order to measure the influence of auditory stimuli, a small dataset was collected from 7 healthy individuals, namely 2 males and 5 females of age (35.7 $\pm$ 17.9) years (mean $\pm$ standard deviation), average weight 64.4 kg, while being at rest condition and during the sound listening, wearing only the Empatica E4 on the dominant wrist. Subjects will be referred to as S1 to S7 in the text. To gather the physiological measurement data, each individual involved in the tests listened to three different auditory stimuli once, resulting in 7 recordings for each different acoustic stimulus (for a total of 21 recordings).The scheme of general measurement setup is shown in Figure 2.1.

Thanks to the event-marker button on Empatica E4, a real-time annotation was possible by pressing the button to label the three phases: "pre-stimulus", "during-stimulus" and "post-stimulus". Specifically, in the first and last part, SC data at resting condition were acquired for 5 minutes as a subject's physiological baseline (i.e. reference measurement), while during the central minutes the physiological variations

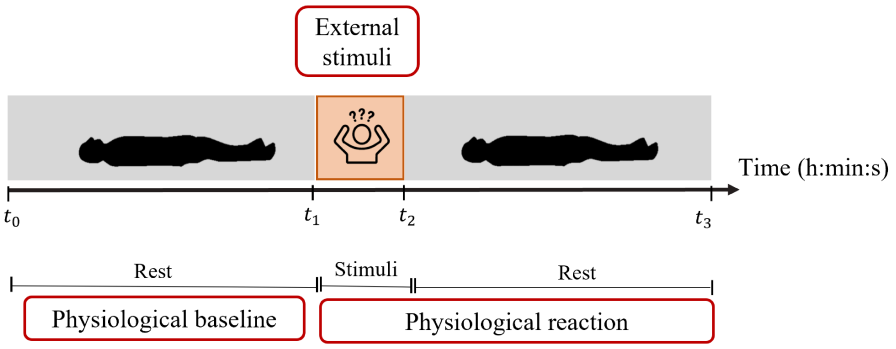


Figure 2.1: General acquisition protocol divided in three main phases: rest, stimuli-based elicitation and physiological response after the stimuli.

during acoustic stimuli were recorded; in the first session the external elicitation lasted 1 minute, while it was set to 2 minutes in the second one. In previous studies, stimuli lasting 1 to 5 minutes produced visible variations of physiological parameters, including the SC [76, 77]. Therefore, two playlists with different length were created for each sound to investigate the effect of the stimulus duration on the elicited SC physiological changes. Firstly, three audio clips were chosen from the IADS database, considering the associated valence score: pleasant (no. 815: 'RockNRoll'), neutral (no. 722: 'Walking') and unpleasant (no. 275: 'Scream') sound. For each audio clip, mean and standard deviation values of the three evaluation dimensions are reported in Table 2.1, based on the subject's gender.

Table 2.1: Evaluation (mean ± standard deviation) of the audio clips chosen to elicit emotions, based on gender and dimension

Gender	**Sound**	**Valence**	**Arousal**	**Dominance**
Female	RockNRoll	8.13 ± 1.41	6.75 ± 2.28	6.99 ± 1.99
	Walking	5.02 ± 1.19	4.87 ± 1.86	4.85 ± 1.41
	Scream	1.65 ± 1.16	8.35 ± 1.32	2.11 ± 1.74
Male	RockNRoll	7.56 ± 1.65	7.00 ± 1.77	6.67 ± 2.00
	Walking	4.61 ± 1.22	5.08 ± 2.00	4.45 ± 1.56
	Scream	2.49 ± 1.94	7.96 ± 1.67	3.04 ± 2.19

Secondly, the same clip was reproduced ten and twenty times to obtain 1 minute- and 2 minutes-long stimuli, respectively in the two different playlists. The 1 minute-long clip was added to investigate whether the habituation of the same sound may affect or not the subjects' reaction. During the acquiring sessions, the participants were left alone, lying on a bed with closed eyes avoiding arm movements, to reduce both the possible external distractions and the potential motion artefacts. Then, to combine the users' subjective perception of an emotional stimulus together with the physiological recordings, after each auditory elicitation, the participants were asked to identify

themselves with the five pictographs (scoring from 1 to 9) along the three dimensions of SAM scale. Obtained scores were compared with the standardised values listed in the IADS database to investigate whether the participants' experience was consistent or not with the standardised ranges.

2.1.2 Data Analyses

When an external stimulus is elicited, a peak appears quite evident in the SC signal, and consequently sparse peaks are considered as an evidence, in time domain, of the immediate response to stressful stimuli [78]. In order to properly study the phasic component, i.e. Skin Conductance Response (SCR), as a stimulus reaction, the SC raw data were analysed in time domain, not by resorting to automatic tools (such as LedaLAB[1]) but following the standard procedure described by iMotions [79]. First of all, data from the first and last 4 s of acquisition were discarded to remove the potential artefacts (e.g., transient noise due to the subject's movements, mostly at the beginning and at the end of each session). Secondly, through a sliding-window filter, the median SC was computed for each sample and the surrounding samples in a window of 4 s, centred on the current sample. This technique allowed to decompose both the SCR and Skin Conductance Level (SCL) components from the SC signal, as shown in Figure 2.2.

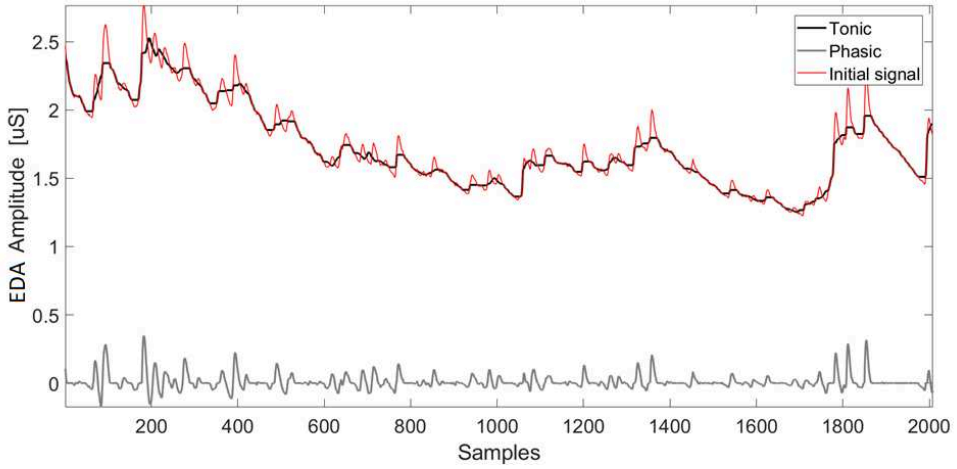


Figure 2.2: Median filter based extraction of the tonic (SCL) and phasic (SCR) components from SC signal.

The SCR was intended as the component representing the signal physiological content, and the number of peaks as a meaningful feature to describe the effects of an external stimulus [47], and thus to compare the reaction to different stimuli of a same subject, or to the same stimulus by different people. Therefore, according to [80], a peak-and-through detection algorithm has been implemented to identify two

[1] `http://www.ledalab.de/`

thresholds of the SCR curve: the onsets at $TH_{on} = 0.01\ \mu S$ and offset at $TH_{off} = 0\ \mu S$. An onset was marked when SCR $> TH_{on}$, while an offset when SCR $< TH_{off}$. Then, back to the original SC signal, the exact position of each corresponding peak was identified for each onset-offset interval, and counted as a peak. This is the reason why, along this Thesis, the peaks will be referred as SC peaks, as well as SCR peaks.

Regarding the frequency domain, the analysis was conducted by computing the Power Spectral Density (PSD) of SC signal in specific frequency bands to assess the psychophysical response to external stimuli. Commonly, bands up to 0.40 Hz are adopted for spectral analysis of HRV signal in the assessment of SNS activity [81]. More specifically, in [82] it is shown that the SC frequency band related to the SNS activity is defined in the range [0.045-0.25] Hz, with the maximum value (95%) contained below 0.25 Hz in the case of orthostatic, physical and cognitive stress tests [47].

Firstly, the collected SC raw signals were filtered through a high-pass Butterworth filter of the 8^{th} order at a cut-off frequency of 0.02 Hz. Then, the Welch's periodogram with a Blackman window (128 points) was applied to compute the power spectra of SC signals [83], and the Fast Fourier Transform was calculated on the periodogram. Finally, the PSD was computed as follows:

$$PSD = \lim_{T \to \infty} \frac{1}{2T} |X_T(f)|^2, \tag{2.1}$$

where $X_T(f)$ is the Fourier transform of the SC signal $x(t)$ in the time window T. Three specific frequency bands were considered: $[0.02, 0.25]$ Hz analysed in [84], $(0.25, 0.40]$ Hz investigated in [84] and $(0.40, 1.00]$ Hz proposed in our work [68].

2.1.3 Experimental Results

A preliminary analysis, performed in time domain, aimed to understand the event-related changes on SC signal in different conditions. Therefore, the *peaks frequency* (i.e. number of SC peaks per minute) was counted over the three phases (i.e. "pre-stimulus", "stimulus" and "post-stimulus") to understand the variation among them, irrespective of the absolute time duration of each one, i.e. one or two minutes long. By examining how much the pleasant audio affects the physiological changes in SC curve morphology, while the subjects S3, S4 and S5 are more sensitive to audio clips lasting 1 minute, the subjects S1, S2, S6 and S7 are more sensitive to pleasant clip 2 minutes long. Regarding the unpleasant clip, in 1-minute-long audio clips, six out of seven participants, except S3, show a number of peaks per minute greater or equal to the one recorded during the resting phase (i.e. ≥ 1). This clearly indicates an increase of the *peaks frequency* from the resting to the stimulating phase. Nevertheless, when the stimulating period was longer (i.e. 2 minutes), the *peaks frequency* decreases drastically, even reaching zero for S4. This means that the subjects S2, S5, S6 and S7 were more sensitive to unpleasant sounds of short duration. Finally, while listening to the neutral sound, four out of seven participants, namely S2, S4, S5 and S7, show an average peaks frequency that increases from 0 to higher values under longer stimuli.

Based on these findings, a second analysis was conducted by computing the number of SC peaks only on the 5 minutes following the acoustic elicited stimulus to investigate the variation caused by the listening of the three different audio clips (i.e. neutral, pleasant and unpleasant) for 1 minute. An example of a portion of raw SC signal to pleasant sound acquired by S2, where just two peaks are identified, is presented in Figure 2.3.

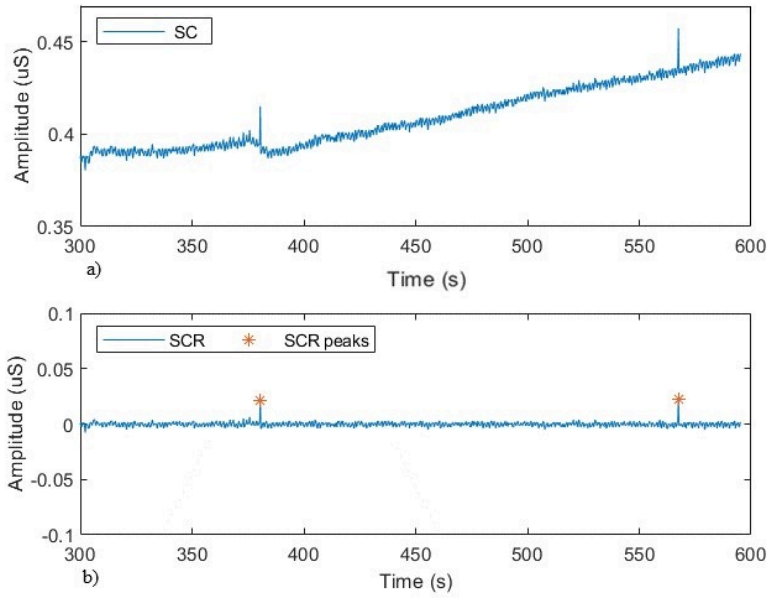


Figure 2.3: Portion of SC signal to pleasant sound for S2: a) SC signal, b) SCR component and peaks marked with a star.

As shown in Table 2.2, the majority of subjects exhibited more peaks to neutral and unpleasant stimuli, with respect to pleasant ones: in six out of seven subjects, namely S1, S2, S3, S4, S5, S7, the number of SC peaks detected after listening to neutral sound is greater or equal to the peaks detected after pleasant sound. Similarly, six subjects, namely S1, S2, S3, S5, S6, S7, exhibited higher or equal number of peaks while listening to unpleasant sound with respect to the pleasant one. Only S4 and S6 showed an equal number of peaks in the SC signals corresponding to pleasant and neutral sounds, and to pleasant and unpleasant sounds, respectively. This means that the involved subjects are more sensitive to neutral and unpleasant sounds than to pleasant sounds.

Regarding the frequency domain, in Figure 2.4 an example plot shows the normalised PSD of the SC obtained from three different subjects, for different acoustic stimuli.

As it is clear, the power spectra of SC signals are predominantly located at frequencies higher than 0.40 Hz. Concerning the pleasant (b) and the unpleasant stimuli (c), for S3 and S2 respectively, the SC spectrum is mainly contained in two sub-bands: below 0.25 Hz and between 0.40 and 1.00 Hz, then decreasing after this bound. The

Table 2.2: Number of SC peaks for the three acoustic stimuli for each subject

Subject	**No. of peaks**		
	Neutral	Pleasant	Unpleasant
S1	6	2	3
S2	3	2	5
S3	5	1	5
S4	4	4	3
S5	5	4	7
S6	3	5	5
S7	4	1	4

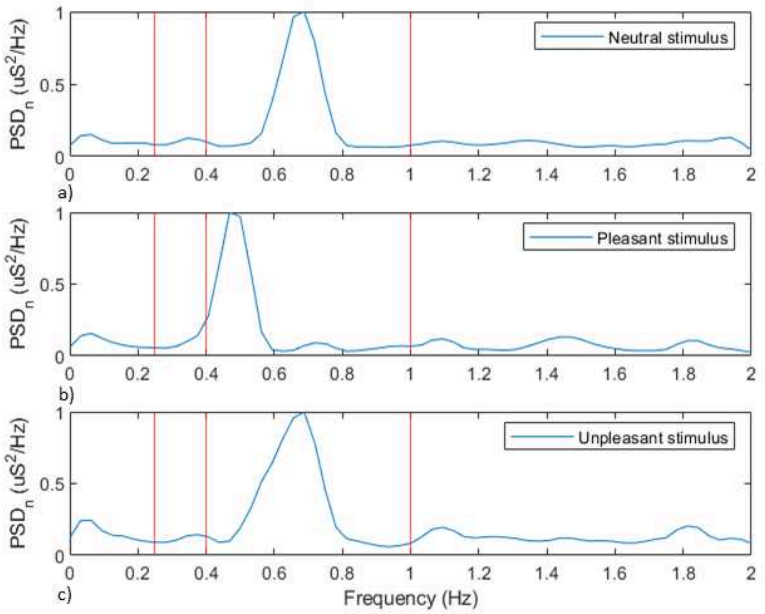


Figure 2.4: Normalised PSD of acquired SC signals to a) neutral, b) pleasant and c) unpleasant acoustic stimuli. The vertical red lines indicate band limits: 0.25 Hz, 0.40 Hz and 1.00 Hz.

frequency behaviour related to neutral stimulus (a) for S5 is mostly shifted to higher values, between 0.40 and 1.00 Hz, with a small amount of spectral power below 0.25 Hz and above 1.00 Hz. Since the spectral power above 1.00 Hz was minimal for several subjects at the best for all the stimuli, the overall SC bandwidth investigated in this study was limited to 1.00 Hz. In order to exactly quantify the spectral content of SC signals, the energy contained in the three frequency sub-bands (i.e. [0.02 - 0.25] Hz named B1, (0.25 - 0.40] Hz named B2 and (0.40 - 1.00] Hz named B3) was computed. In Table 2.3, the resulting percentages of power spectrum in each frequency sub-band are detailed, distinguishing the three different acoustic stimuli.

Table 2.3: PSD percentage computed within the frequency bands for the three acoustic stimuli

Subject	**PSD** (%)								
	Neutral			Pleasant			Unpleasant		
	B1	B2	B3	B1	B2	B3	B1	B2	B3
S1	19.44	52.37	28.19	60.01	21.31	18.67	12.74	6.67	80.59
S2	14.37	4.76	80.87	17.26	6.01	76.67	15.53	5.29	79.18
S3	14.80	19.64	65.55	14.60	5.47	79.93	15.83	6.39	77.78
S4	9.61	6.76	83.63	11.02	7.38	81.60	8.24	4.35	87.41
S5	13.35	6.07	80.58	93.10	2.37	4.53	58.42	6.24	35.34
S6	13.30	30.03	56.67	33.17	6.65	60.18	6.36	62.34	31.30
S7	14.00	5.11	80.89	11.84	6.73	81.42	10.54	5.38	84.07

Specifically, Table 2.3 illustrates that, computing the percent distribution of the PSD over the whole bandwidth acquired in the neutral acoustic stimulus, B1 ([0.02 - 0.25] Hz) contains always less than 20.00% of the total power spectrum. Indeed, the percent power spectrum located in B3 ((0.40 - 1.00] Hz) is higher than in the lower sub-bands, in six out of seven subjects (i.e. S2, S3, S4, S5, S6, S7), reaching the 83.63% in S4. Contrarily, S1 presents a higher value in the adjacent sub-band B2 ((0.25 - 0.40] Hz). For what concerns the pleasant acoustic stimulus, the sub-band B1 (below 0.25 Hz) contains the greatest amount of power spectrum for S1 and S5; contrarily, for the other subjects the percentages are lower than 34.00% in B1, followed by an increase in the last sub-band B3 (0.40 - 1.00] Hz, achieving values higher than 80.00% for S1 and S7. Similarly for the response to the unpleasant acoustic stimulus, in which only S5 reaches the 58.42% of PSD in the first frequency band B1. Almost all the remaining individuals (i.e. S1, S2, S3, S4, S7) achieved a percent of SC power spectra higher in B3 (0.40 - 1.00] Hz than in the first sub-band B1; contrarily, S6 featured the same condition in (0.25 - 0.40] Hz.

2.1.4 Discussion and conclusions

Regarding the analysis in time domain, the SC *peaks frequency* (i.e. number of SC peaks per minute) was computed being the event-related feature, and then compared among the different acoustic stimuli. An increase of SC peaks per minute during the unpleasant sound listening period in almost all individuals was presented especially for short sound, probably due to the track played: an unexpected and well-known annoying sound (i.e. 'Scream'). Probably, the negative emotion induced a high sweat reaction, and consequently evident physiological changes. However the same reaction, if the external stimulus is too long (i.e. 2 minutes), can be affected by the habituation phenomenon [85], resulting indeed in a lower number of peaks per minute on the SC curve. Contrarily, the findings from the pleasant and neutral stimuli are more randomly distributed, resulting in any specific visible physiological reaction to 'Walking' and

'RockNRoll' sounds, especially when presented the one-minute-long tracks. Concerning the frequency domain, although in the literature the most investigated frequency band for SC signal is generally confined to $[0, 0.40]$ Hz [71], we found that, by using acoustic elicitation, the extended range $[0.40, 1.00]$ Hz contains a higher amount of power spectrum with respect to lower frequency bands. Therefore, when acoustic stimuli are used to elicit human emotions, it could be useful to extend the SC spectrum analysis to a upper frequency bound (i.e. until 1.00 Hz), with respect to the case of other types of stimulus. A larger observation band can avoid the loss of potentially meaningful information. In general, the results confirm that the physiological changes in SC are visible, but subjective. Even though different individuals can share some emotional status or mental perception of the same sound track (as declared in SAM scale scores), their physiological features can have significant differences. This statement is evident for pleasant sound, where high perception of affective valence and intensity corresponds to a small number of peaks during the stimulation. Contrarily, the low valence and high arousal of 'Scream' sound can be strictly associated to the bigger number of peaks during the elicitation of an unpleasant experience. Moreover, from findings in frequency domain, it is evident that the reaction to acoustic stimuli is very subjective, showing different characteristics of SC power spectra among the subjects involved in the experiment. Moreover, results underline different reaction times under the three stimuli: the SC power spectra, mostly located at high frequencies for neutral acoustic stimuli, suggest a faster reaction than for the unpleasant and pleasant ones in human brain.

Although the results of these experiments are promising, some limitations are present. Future works to generalise these findings would intend to enrol a wider and more heterogeneous population in terms of gender and age, reducing the inter-subject variability. As an example, the different perception of an external stimulus (e.g., acoustic and visual), and consequently the resulting SC fluctuations can be compared among males and females, and different range age. Also selecting more sound tracks from IADS database and more driving paths, can allow to obtain more generalised and reliable findings. Moreover, additional physiological signals besides SC, such as HRV, can be taken into account to identify the frequency sub-bands related to emotional responses under acoustic stimuli. Also different cross validation techniques (e.g., Leave-One-Subject-Out Cross Validation, LOSO) may be employed to improve the ML detection accuracy, especially for distinguishing the two classes *alert* and *drowsy*.

2.2 Case study 2: Physiological Response to Visual Stimuli

Video stimuli are among the most widely delivered when trying to evoke strong psycho-physiological responses [86] for measuring and quantifying physiological changes. Dominguez et. al [87] collected SC signals alone while showing video clips to stimulate

sadness, amusement and neutral reactions; the target raised emotions were well-recognised from RF classifier (up to 100% of accuracy). Contrarily, in order to tackle the influence of subject's personal, cultural and cognitive aspects on physiological changes, Zhao et al. [88] recorded multi-physiological signals (i.e. SC, HRV and skin temperature). In this case, the average accuracy of the emotion recognition process dropped down to 75.56%. Further studies employed 2D visual stimuli selected from the International Affective Picture System (IAPS) database [89] or specific images, among which pleasant and unpleasant pictures for exciting feelings, and neutral ones for calm emotions [90] [91] to evaluate different emotion classification techniques. Within the domain of visual stimulation, this Thesis explored how the driving path might act as a visual stimulus affecting the daily psychophysical status of a driver. Based on the World Health Organisation estimates, nearly 1.35 million deaths each year [92] are due to driver drowsiness, occurring with decreasing levels of arousal. Although some safety technologies are already implemented in cars with on-board sensors to classify the driving behaviour (e.g., lane deviations, steering wheel rotation, along with camera-based systems for eye gaze location), recent investigations are focused on the drowsiness and the consequent ANS activity reflected by changes in physiological signals [93]. According to the literature, the BVP, leading an estimation of HRV, is a useful indicator for drowsiness conditions [8], while the SC signal is useful to detect changes in subjects' arousal [78]. Both the gold-standard measurement methods for BVP and SC require the direct contact between electrodes and the subject's skin in specific positions. This represents a major shortcoming in the automotive context [94]. Pushed by the growing trends of the IoT-enabled wearables market, several wrist-worn devices (or bracelets [95] and double-rings [96]) were tested to collect physiological data, then processed by classification algorithms to monitor drivers and detect a drowsy status [97]. For example, Lee et al. assessed the driving behaviour by utilising the built-in motion sensor of a smartwatch; SVM classifier achieved an accuracy of 98.15% [98]. Then, by combining the accelerometer signal and the PPG signal, they obtained an accuracy of 95.80% [99]. Contrarily, Leng et al. measured both PPG and SC through a wearable device on the fingers; SVM model reached an accuracy of 98.70% [100]. Choi et al. developed an intrusive system with a wrist-worn wearable device with PPG, SC, temperature, acceleration, gyroscope sensors and an additional PPG signal acquired on the ear. Although the SVM employed to distinguish the normal, stressed, and drowsy states achieved with an accuracy of about 85.00%, the uncomfortable system results less suitable for real-life driving applications [101]. Therefore, in this Case study, a dataset was collected within the automotive context to both explore the physiological human response to visual stimuli, and detect the driver's drowsiness by wearing only a smartband. In particular, firstly the SC signal is considered alone, and secondly, a multimodal system is proposed by combining the SC and the HRV signals.

2.2.1 Acquisition Protocols

As illustrated in Figure 2.5, the experimental tests were executed in a room hosting the driving simulator, showing an overnight three-lane highway with a length of around 80 km and no traffic.

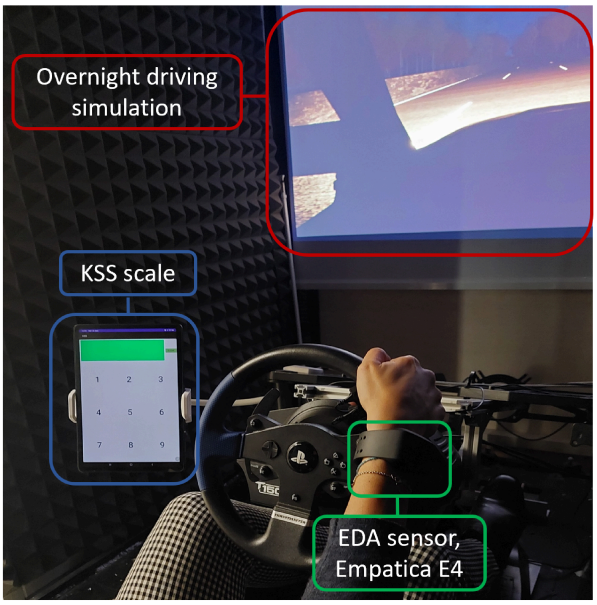


Figure 2.5: Experimental setup: driving simulator, monitor with overnight path, Empatica E4 device position and tablet with KSS scale.

In the room, the average temperature was maintained around 23 °C to reduce the impact of ambient temperature on user's skin (e.g., too cold or too hot), and consequently on signals recorded. The idea was to emulate the real-life condition, where an individual can drive wearing the own smartwatch, equipped with the sensor capability to monitor physiological parameters. The test population involved 9 healthy subjects, 4 males and 5 females; a male and a female were selected in each cohort of 10 years width (20-60 years of age) to cover a wide range of active drivers. During the whole driving recording session (lasting around 40 minutes), participants worn the Empatica E4 device on the dominant wrist to acquire simultaneously the BVP and SC signals. Samples of the acquired BVP and SC signals over a whole 40 min-long session are given in Figure 2.6. Moreover, individuals were asked to assess, on a tablet every 10 minutes, their own perceived alertness/drowsiness status through the 9-point Karolinska Sleepiness Scale (KSS) questionnaire [102], that connects the psycho-physical status experienced to verbal sentences.

2.2.2 Data Analyses

A proper approach has been conducted for data collected during visual stimuli, because natural movements during driving actions can affect the quality of SC signal acquired

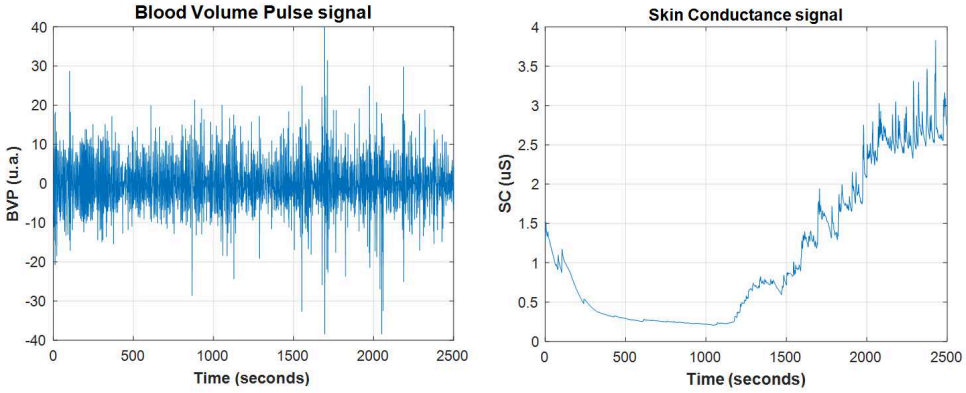


Figure 2.6: Sample acquired BVP (left) and SC (right) signals during a whole session (40 min duration).

from the wrist, and consequently the performance of driver's drowsiness detection by ML algorithms. Therefore, to identify and reduce the motion artefacts, the Stationary Wavelet Transform (SWT) denoising with *haar* mother wavelet (7 levels of decomposition) was implemented according to previous studies [103]. This approach firstly models N wavelet coefficients d_j as zero-mean Laplace distribution, with j representing the wavelet decomposition level. Then to identify the boundary between clean SC signal and motion artefacts, the T_{high} and T_{low}, high and low threshold respectively, are defined as follows [104]:

$$\begin{cases} T_{low} = (\frac{1}{N}\sum_{n=1}^{N}|d_j|) \cdot \log_e(\delta) \\ T_{high} = -T_{low}, \end{cases} \tag{2.2}$$

where δ is the proportion of motion artefacts in the original SC signal, that quantifies how much motion artefacts affect the signal by exploiting the 3-axial acceleration values simultaneously acquired with the SC samples through Empatica E4. Depending on the subject's wrist movement amplitude, the value of δ depends on the standard deviation (σ) of the acceleration samples collected from all the three directional components (i.e. acc_x, acc_y, acc_z), as follows:

$$\begin{cases} \delta = 0.01, & \sigma_{(acc_x, acc_y, acc_z)} < 0.04\ m/s^2 \\ \delta = 0.10, & otherwise \end{cases} \tag{2.3}$$

The limit on the value of σ, with the threshold of 0.04 m/s^2 (heuristically set by visual inspection of both acceleration and SC signals), has to be satisfied by each acceleration component: if just one is greater than 0.04 m/s^2, then the motion artefact is identified, and consequently removed. When wavelet coefficient exceeds these thresholds, they are set to zero. Then SWT is applied resulting in a denoised signal. At this point, the SC signals and their components (i.e. SCR and SCL) were divided

in short-time intervals with fixed size of 15 s (i.e. 60 samples); then, according to the KSS scale's scores reported by the participants, each window was labelled creating three groups, depending on the drowsiness level: scores between 1 and 5 in class 1 (labelled as *alert*), 6 and 7 in class 2 (labelled as *slightly drowsy*), 8 and 9 in class 3 (labelled as *drowsy*). Finally, a total of 23 features (listed in Table 2.4 - some used in previous studies [68], [105]) were computed, from each labelled window, in time and frequency domains to analyse the temporal and spectral information contents.

Table 2.4: Features extracted from the SC signal and its components

Type of signal	**Domain**	**Features (measurement unit)**
SC signal	Time	Mean (μS), standard deviation (μS), minimum (μS), maximum (μS), kurtosis (μS), skewness (μS), variance ((μS)^2), range (μS), median (μS)
	Frequency	Mean (μS/Hz), standard deviation (μS/Hz), minimum (μS/Hz), maximum (μS/Hz), kurtosis (μS/Hz), skewness (μS/Hz), variance ((μS/Hz)^2), range (μS/Hz), median (μS/Hz)
SC components	Time	SCR number of peaks (-), SCL mean (μS), SCL standard deviation (μS), SCL minimum (μS), SCL maximum (μS)

In order to improve the ML performances, the correlation coefficient (ρ) was computed to quantify the strength of the features relationship; features with $\rho > 0.90$ (i.e. strong correlation) were discarded to reduce the potential high correlation and redundancy between tested features, affecting the classification algorithms. As a result, five features, namely SC mean, SC maximum, SC median, SCL mean and SCL maximum, were discarded, resulting in a total of 18 features to use. Once selected the relevant features, three ML algorithms, namely RF, Bagging and Boosting , were trained and then tested through the 10-fold cross validation method [106]. The classification performances were compared by using accuracy, precision, recall and confusion matrix.

The added BVP signal was not filtered, due to an algorithm embedded in Empatica E4 firmware (details are not disclosed by the manufacturer) that removes the motion artefacts exploiting data measured during exposure to the red LED. Being the HRV a well-known drowsiness indicator [107], it was derived from the BVP signal, by quantifying the inter-beat intervals (i.e. IBIs, the distance between two consecutive signal peaks). Then, obtained signal was divided into time interval with 15 s, from

which 9 features were extracted (Table 2.5).

Table 2.5: Features extracted from HRV

Type of signal	**Domain**	**Features (measurement unit)**
HRV signal	Time	HR (bpm), SDNN (ms), RMSSD (a.u.), pNN50 (μS)
	Frequency	LF (a.u./Hz), HF (a.u./Hz), LFn (a.u./Hz), HFn (a.u./Hz), LF/HF (a.u.)

2.2.3 Experimental Results

After processing the SC signals collected during the visual stimulation of an overnight driving path on a highway, three ML algorithms were tested and compared depending on the classification performance. In Table 2.6, the resulting metrics are summarised. The average accuracy in identifying the three different drowsiness conditions (i.e. *alert* (class 1), *slightly drowsy* (class 2) and *drowsy* (class 3)) is equal to 84.10% for Random Forest, 83.20% and 82.80% for Bagging and Boosting, respectively. Precision and recall are equal to 84.20% and 84.10%, respectively, for RF, whereas 83.30% and 83.20% for Bagging, and both equal to 82.80% for Boosting.

Table 2.6: Classification performance of proposed approach

Classifier	**Accuracy (%)**	**Precision (%)**	**Recall (%)**
Random Forest	84.10	84.20	84.10
Bagging	83.20	83.30	83.20
Boosting	82.80	82.80	82.80

Among ML classifiers tested, RF proves to be the best one, according to all the performance metrics considered. Therefore, in Table 2.7 the confusion matrix related to the RF algorithm is detailed with percentage values.

Table 2.7: Confusion matrix for RF algorithm

		Predicted Instances		
		1	2	3
Actual Instances	1	85.7%	11.7%	2.6%
	2	6.1%	83.0%	10.9%
	3	3.9%	12.2%	83.9%

From the resulting Table, the classes 1 and 2, representing the *alert* and *slightly drowsy* conditions respectively, are well-distinguished (867 instances classified, out

of the 1045 actual instances), except for a few instances. This means that there are evident and meaningful physiological variations captured by the recorded SC signals, that uniquely characterise the *alert* and the *slightly drowsy* conditions. Thanks to such SC changes, the features extracted from these two conditions can be well-distinguished from the *drowsy* class. In the future, this capability may be exploited to design automatic systems to alert the driver, just wearing a smartwatch. On the contrary, class 2 (*slightly drowsy*) is often misclassified with class 3 (*drowsy*); most probably the SC features related to the *drowsy* status are quite similar to those computed over the data collected during the *slightly drowsy* ones. Anyway, considering a safety system applied in real-life, the most important challenge is the capability to detect the physiological variation in short-terms between the *alert* and the *slightly drowsy* conditions.

When adding the HRV features to the initial 23 ones referred to the SC signal, the RF algorithm (specifically the Ensemble algorithm on Weka tool) reached a higher performance with an average accuracy of 93.0 % in detecting the *drowsy* status.

2.2.4 Discussion and conclusions

When visual stimuli were exploited to detect the driver's drowsiness, 18 features were extracted from SC signals and an average accuracy of 84.1% was provided by RF in distinguishing the *alert*, *slightly drowsy* and *drowsy* classes, while it reached 93% when combining SC and HRV features. About the visual stimuli, the findings demonstrate the feasibility of detecting driver drowsiness exploiting either only SC signals, acquired from a single wrist-worn device or exploiting the joint SC and HRV information. Moreover, the classification performance is obtained with short-term time windows, essential for detecting short-term events as the natural drowsiness onset. This way, when abnormal changes are detected, a proper timely alert can notify the driver, for example suggesting to take a short break to rest.

Although the results of these experiments are promising, some limitations are present. Future works to generalise these findings would intend to enrol a wider and more heterogeneous population in terms of gender and age, reducing the inter-subject variability. As an example, the different perception of an external stimulus (e.g., acoustic and visual), and consequently the resulting SC fluctuations can be compared among males and females, and different range age. Also selecting more sound tracks from IADS database and more driving paths, can allow to obtain more generalised and reliable findings. Moreover, additional physiological signals besides SC can be taken into account to identify the frequency sub-bands related to emotional responses under acoustic stimuli. Also different cross validation techniques (e.g., Leave-One-Subject-Out Cross Validation, LOSO) may be employed to improve the ML detection accuracy, especially for distinguishing the two classes *alert* and *drowsy*.

2.3 Case study 3: Physiological Response to Physical Stimuli

The second Case study aims to investigate whether and how the SC signal changes in response to physical stimuli, by examining the morphological characteristics of this specific physiological signal. In the sport context, where the body is elicited by physical efforts, measuring SC in different training sessions, before and after a positive or negative performance, may support the evaluation of the most suitable and personalised training schedule for athletes [108], by assessing the sympathetic function during exercises [84]. These applications need to firstly recognise the type of stimulus associated to specific SC variations. To do this, recent studies involve the features extraction and the machine learning classification, such as in [109]. However, the SC signal is characterised by a great variability influenced by the data acquisition procedure and the physiological variability among the individuals. Such factor makes it difficult to find the proper meaning of the classification outcomes and to relate the results in the SC signal response to a specific kind of stimuli [110]. To improve the accuracy and reliability, an accurate feature selection and evaluation are required prior to classification, as described in [67]. Once selected the most meaningful features describing the SC physiological response to the physical exercise, the second study [111] aims to recognise the perceived physical exertion intensity by implementing two machine learning classifiers (i.e. Support Vector Machine and Bagged Tree - both supervised approaches), fed with features computed on heart-related parameters (from BVP signal), SKT, and wrist acceleration, all collected by wearing the Empatica E4. As a novelty, such outcomes are exploited to validate the use of the SC data only, to classify the user's perceived effort.

2.3.1 Acquisition protocols

In the first study, SC measurement data were acquired from four healthy subjects (2 males and 2 females aged between 20 and 59 years, with a Body Mass Index (BMI) between 21.07 and 25.24 kg/m^2), by wearing the Empatica E4 on the dominant wrist. Each individual was involved in three different recording sessions (at rest - R, after mild exercise - M, and after intense exercise - H) acquired in three different moments of the day (morning, afternoon and evening), obtaining nine sessions per each subject. This allowed to have different repetitions of the same type of Physical Activity (PA), irrespective to the moment of the day it was recorded. Depending on the activity to perform, the acquisition time was different. Specifically, the sessions at rest lasted 15 minutes, while the subject was lying on a bed. Then, subject was asked to perform almost 5 minutes of mild exercise (e.g., walking and climbing stairs). As soon as the exercise terminated, the subject was asked to lie again down on a bed for a 10-minute recording. The same procedure was repeated to acquire 10-minute recordings after almost 5 minutes of intense exercise (e.g., running stairs or repetitions of a vigorous free body exercise). Such exercises were selected based on the guidelines provided by

the Mayo Clinic [112]. It is important to note that all the acquisition sessions were conducted after the physical activity, in order to assess the physiological response after the physical effort. Contrarily, a second small experimental test, involving three healthy participants (2 females and 1 male, ageing between 25 and 29 years old with a BMI between 17 and 23 kg/m^2), was focused on the physiological response during the physical effort. In particular, participants performed two test sessions per day, in the morning and in the afternoon always at the same hour, for 5 consecutive days, resulting in a total of ten test sessions. Similarly to the data collection previously described, each session included three trials with different PA intensity levels (10 minutes for each): the first trial consisted in a sitting condition (as R level), the second one in a squatting period (as M level), then the third one in a squatting with a high frequency of execution (as H level). Two minutes of rest were included in the acquisition protocol, at the end of moderate activity, to ensure the vital signs could get back to the physiological baseline. The overall protocol lasted around 32 minutes. During each trial, the HR values displayed in the running E4 Realtime app were used as ground truth for the correct definition of PA intensity level. In particular, the HR values between 90 and 120 beats per minute (bpm) defined the moderate activity, while those between 120 and 140 bpm indicated the intense one [113]. Moreover, as soon as each trial terminated, participants rated their perceived exertion while performing the PA trial using the Borg Rating of Perceived Exertion (RPE) scale [114], where different levels of exertion are categorised into sedentary, moderate and vigorous intensity.

2.3.2 Data Analyses

Following the analysis in both time and frequency domain of SC signals, the extraction of meaningful features was explored. Generally, features mostly computed in time domain are the statistical descriptive ones (e.g., signal mean value, standard deviation, kurtosis, skewness, and variance) [109, 115]. Further metrics frequently extracted are related to morphological alterations in the signal (e.g., the area under the curve), along with the computation of the first derivative of the signal, considering its mean and the mean of the negative one [116]. A few researches focused on frequency-domain features, but since the transient characteristics of the SC signal affect mostly the time domain, the computation of features on the Fast Fourier Transformed signal is proposed as a promising approach. Statistical aspects such as signal magnitude area, range, kurtosis, skewness, mean value, energy and entropy have been proposed as well. In Table 2.8, both the time- and frequency-domain features computed are listed.

The overmentioned features were extracted from both the SC signal components, i.e. the tonic and phasic ones. A data reduction process was applied separately for the features computed on the two different components, for managing the dataset collected from the experiments and to reduce its complexity. First of all, the average of features related to the same subject and to the same physical activity, measured over different time sessions, was computed. Then, feature values across all the subjects

Table 2.8: Time- and frequency-domain features from SC signal

Domain	**Features (measurement unit)**
Time	mean (μS), standard deviation (μS), area under curve, variance (μS)^2, kurtosis (μS), skewness (μS), mean derivative (μS/s), negative mean derivative (μS/s)
Frequency	mean (μS/Hz), standard deviation (μS/Hz), signal magnitude area, range (μS/Hz), kurtosis (μS/Hz), skewness (μS/Hz), energy, entropy

were normalised by the maximum value to be included in the range $[-1, 1]$. Finally, a *discrepancy* metric was evaluated among the value assumed by the same feature over the two SC components, in order to understand in which component the specific feature had a highest weight and relevance. *Discrepancy* was defined as the difference (Δ) between the normalised value of the same feature F extracted from the tonic and phasic components related to the same SC signal:

$$\Delta_F = F_{tonic}^{norm} - F_{phasic}^{norm} \tag{2.4}$$

Only the features for which $\Delta_F \geq 0.5$, concerning the subject and physical activity they are related to, were considered significant. After finding the meaningful features describing the response to the physical exercise in terms of SC signal, other physiological signals were considered to recognise the perceived physical exertion.

Considering the second small dataset, a multimodal physiological signal system was proposed, including also the HR, SC, SKT and acceleration samples. According to [117], from each BVP signal collected, the IBI (measured in s) and the HR (measured in bpm), both related to heart rate variability, were extracted. Since SKT and SC data contain the greatest information content regarding the PA intensity, in terms of skin temperature changes and sweat secretion respectively, any filter was implemented, to avoid loss of information. To remove the motion artefacts from the raw acceleration samples due to potential loss of contact between the E4 device and the subjects' wrist, a 4^{th} order Butterworth bandpass filter, with a low- and a high-pass cut-off frequency of 0.5 and 1.5 Hz respectively, was applied. Then, signals were segmented with a fixed-size sliding window of 12 s, with 50% overlap (6 s), from which time- and frequency-domain features were computed to feed the ML algorithms. A total of 50 features were computed and standardised by Z-score normalisation to obtain a common interval by scaling the signal amplitude. In particular, 8 from each acceleration axis, 7 from the IBI, 4 from the HR signal, 11 from the SKT signal, and 7 features from the SC signal. Each time window was labelled as a 0 (Sedentary - R), 1 (Moderate -

M) or 2 (Vigorous - H) class activity, depending on the individual's perceived effort expressed by the Borg RPE scale. Moreover, the Synthetic Minority Over-Sampling Technique was implemented to re-sample the classes by adding synthetic data and generate a balanced dataset. Among the ML algorithms generally used to predict the PA intensity, the SVM and Bagging Tree (BT) were selected and compared to test the performance of a supervised learning approach.

2.3.3 Experimental results

By analysing SC time responses from each subject, there is a remarkable variability among them. Data shows quite constant amplitudes after medium exercise or at rest, than increasing after high-intensity activity. In Figure 2.7, showing the average of SC signals, this is evident for S1, S2 and S4, whereas for S3 all the traces have similar ranges.

Figure 2.7: Comparison of SC mean values in time, over the test population for all the intensity levels.

In particular, the SC amplitudes increased according to the exercise intensity, with exception of S3, who maintains really low mean values over any activity. S2 reports the highest amplitudes also in the medium exercise intensity, with respect to the other subjects.

Regarding the *discrepancy*, Figure 2.8 shows the comparison among normalised mean values of tonic (left) and phasic (right) components for all the subjects in the three different PA conditions. Significant discrepancies (≥ 0.5) are only found for S1 and S2 after performing high intensity physical effort.

Regarding the frequency domain, the mean FFT magnitude was computed. As an example, Figure 2.9 shows the FFT magnitude of the tonic and phasic components of the SC signal collected for S2, after executing the three different activity intensities. It is remarkable the difference between the tonic components (left column) at high intensity and at rest. In particular, the FFT magnitude of the tonic component takes maximum value at lower frequencies and decreases with the increase of the frequency.

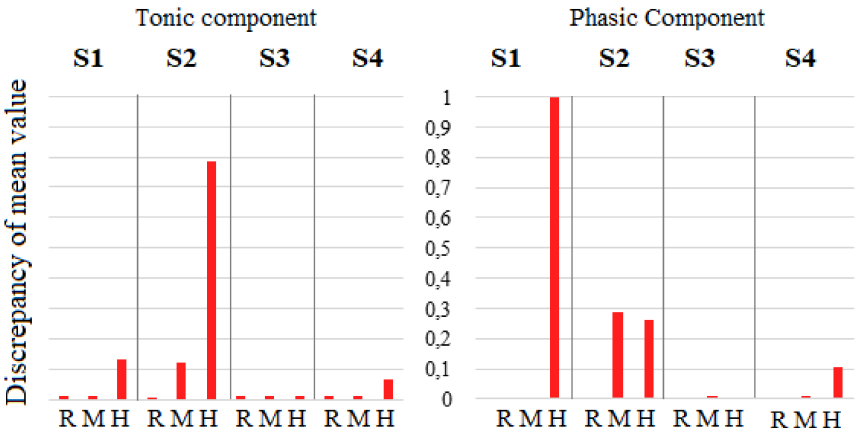


Figure 2.8: *Discrepancy* for the normalised mean value features computed over the test population for all the intensity levels.

On the other hand, the FFT magnitude of the phasic component takes maximum values at higher frequencies than those of the tonic component. In particular, both at medium and high intensity exercise levels, the maximum values of the phasic component are in the frequency band [0-0.50] Hz, while for the condition at rest the frequency content of the signal is clearly shifted to higher frequency band (i.e. [0.50-1.00] Hz).

Despite the great inter- and intra-subject variability in SC traces, all the involved subjects exhibited an increase in the signal response from rest to the highest physical effort. When exposed to PA, there are mainly two different response types: for S1, S3 and S4 the presence of SCR peak bursts can be underlined, while, for S2, at a high level of physical activity there is a greater increase in the signal amplitude values, with respect to the other subjects. Characteristics of the subject can be a reason of this trend, along with the different perception, among subjects, of the same exercise intensity. As an example, the S3 (59 years-old woman) presented smallest variation in the SC mean amplitude measured with the E4 device. From the time domain analysis, it was possible to observe that the phasic component, which is zero centred, corresponds to high frequencies information, while the tonic component resembles most the slowly varying component of the original signal, maintaining a variable offset. Once decomposed the SC signal, the different behaviour was observed to the tonic component when undergoing intense effort, such as in S2. On the other hand, the presence of distinct SCR peak bursts in the other subjects is in the morphology of the phasic component. Frequency domain analysis shows that the mean FFT magnitude may discriminate the PA levels if extracted from the tonic component, while could discriminate the presence/absence of PA if computed from the phasic component; so this feature might be helpful in classifying the different levels of physical effort.

A different analysis was conducted through features extracted from cross-domain

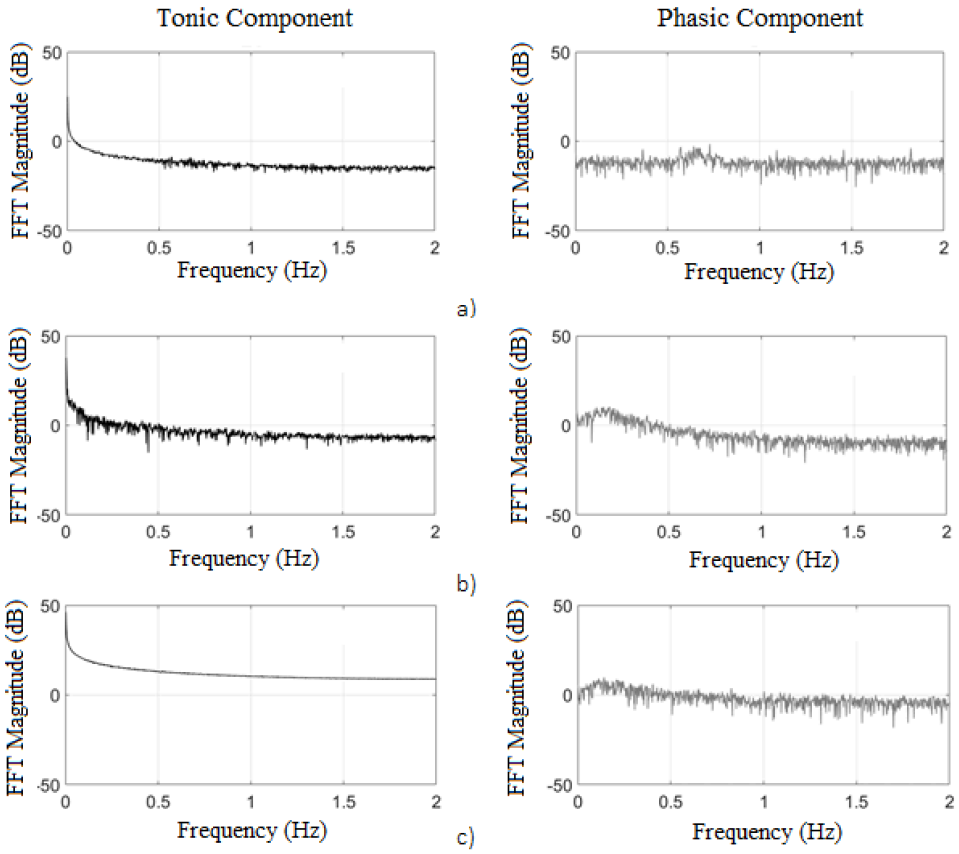


Figure 2.9: Example of FFT magnitude: tonic (left) and phasic (right) components (S2); a) at rest, b) medium, c) high intensity activity.

signals, namely the 3-axial acceleration signals, HR, IBI and SKT, with data labelled via the Borg scale rated by the participants.

In Figure 2.10, the confusion matrices obtained with the SVM classifier and BT classifier on both physiological and acceleration signals are shown. As mentioned before, there are three classes to be recognised: 0, 1 and 2 related to sedentary, moderate, and vigorous activity, respectively. The principal diagonal (i.e. blue cells) indicate the class correctly identified, that is the number of instances that obtained a predicted class equal to the true class. Contrarily, the values outside the diagonal (i.e. shade of pink cells) identify the prediction errors. The SVM classifier correctly predicted 5266 instances out of the 5344 actual instances in the class 0 (sedentary), 5219 instances in the vigorous activity (class 2), while only 4724 instances were predicted correctly in the moderate activity (class 1), as shown in Figure 2.10 (a). By looking at the Figure 2.10 (b), the BT classifier correctly predicted 5208 instances out of the 5344 actual instances belonging to the sedentary class; for the vigorous class, 5065 instances were predicted correctly, whereas in the moderate class, only 4823 instances. Considering

Figure 2.10: Confusion matrices for a) SVM classifier and b) BT classifier, applied on cross-domain physiological and acceleration data.

the overall accuracy, the SVM classifier reached the 94.5% in classification, versus the 93.9% of the BT. The relative performances (i.e. AUC, Specificity, Sensitivity and F-score) of the two ML classifiers over the three classes have been assessed according to the further metrics, and the results are summarised in Table 2.9.

Table 2.9: Relative performance evaluation of the SVM and BT classifiers on the three classes, with cross-domain data

Classifiers	**Class**	**AUC**	**Specificity** [%]	**Sensitivity** [%]	**F-score** [%]
SVM	0	1.00	98.67	98.54	98.15
	1	0.99	97.79	88.41	91.69
	2	0.99	95.23	96.83	93.88
BT	0	0.99	99.29	97.46	98.01
	1	0.99	95.90	90.27	90.95
	2	0.99	95.65	93.97	92.77

As previously mentioned, SC signals are affected by intra- and inter-subject variability, so their use in classification approaches is often reinforced by fusion with other signals, such as the HR [67]. However, in our study, we aimed for testing the classification performance obtained by using the SC signals alone; thanks to the synchronous collection with the acceleration and the physiological signals already discussed, we based the labelling of SC instances on the classification obtained from the previous experiments.

Figure 2.11 shows the confusion matrices obtained using SVM classifier and BT classifier on SC data. It is possible to notice that the SVM classifier correctly predicted 4977 instances belonging to the class 0 (sedentary activity) while the BT classifier correctly predicted 4232 instances of the same class. Regarding the moderate intensity

Figure 2.11: Confusion matrices for a) SVM classifier and b) BT classifier, applied on SC data.

(class 1), the SVM classifier correctly predicted 1121 instances, versus the 2976 instances correctly classified by the BT; whereas for the class 3 (vigorous activity), the SVM classifier correctly predicted 4380 instances while the BT was correct on 4727 instances. Similarly to the previous analysis, Table 2.10 reports the relative AUC, specificity, sensitivity and F-score values obtained by the two classifiers over the three classes. In terms of the overall accuracy, it was evaluated equal to 65.8% for SVM and 73.8% for BT. As reasonably, the results obtained exhibit lower performances than those obtained from the classification exploiting cross-domain physiological and acceleration data.

Table 2.10: Relative performance evaluation of the SVM and BT classifiers on the three classes, with SC data only

Classifiers	**Class**	**AUC**	**Specificity** [%]	**Sensitivity** [%]	**F-score** [%]
SVM	0	0.99	65.65	94.19	71.26
	1	0.98	93.55	21.30	31.66
	2	0.99	88.67	79.20	78.88
BT	0	0.99	87.17	80.09	77.64
	1	0.99	85.79	56.55	60.88
	2	0.99	85.43	85.48	82.37

2.3.4 Discussion and conclusions

From the the overall results discussed above, the time domain analysis of SC signals shows that most of physiological information are subject-dependent, especially because related to the variability of the responses to the same physical stimuli from different

subjects. Discrepancies in the signal behaviour may be due to intrinsic physiological characteristics of each subject (e.g., gender, age and sweat glands density distributed in the different locations of the body), and to extrinsic physiological characteristics (e.g., level of physical training and fitness conditions of the subject, which brings to a different perception on the effort requested to perform the physical activity). Since these aspects could reflect in the computed features, they may be useful to implement customised and subject-dependent classification algorithms, by considering user-related characteristics. On the other hand, frequency domain analysis demonstrated significance in the discrimination of different levels of activities (e.g., high information content at lower frequency for medium and high intensity activity, with respect to the resting condition), and in the association of this differences to a signal component rather than another.

When considering cross-domain data, namely the 3-axial acceleration, HR, IBI and SKT signals, from one hand the vigorous intensity class achieved the highest classification performance, being associated to evident differences in the signals (e.g., the amplitude of the acceleration data characterising the vigorous movements), with respect to those acquired during the moderate and the sedentary activities. On the other hand, as expected, the moderate intensity class was the most often misclassified (i.e. always lowest accuracy percentage), being an intermediate class that probably include feature values either low and/or high, hence attributable to the sedentary and vigorous PA intensity classes, respectively. This fact may also depend on the different perception of PA effort, that was assessed by each subject when evaluating the Borg's RPE scale. As an example, a hard trained subject may feel the vigorous activity as moderate and vice versa. For a potential real-life application of the proposed approach, the misclassification between moderate and sedentary classes may have a great impact on the information reliability received for the user's self-tracking evaluation, especially if thought for the healthcare operators' assessment [118].

Given that in the first study some analyses were not performed according to the purpose of the study, some of its limitations were investigated in the second study. In particular, to assess the best ML predictors for PA intensity, cross-domain signals were involved and, when required, filtered from motion artefacts. However, some limitations still need to be assessed in future works. For example, in both cases, to validate and generalise the proposed approaches, a wider population in terms of different physical training and different ages should be involved in the studies, along with additional PAs to be performed.

Chapter 3

Measurement of Cardiac Activity

This Chapter is focused on a set of studies on cardiac activity measurements conducted by healthy subjects, while wearing the Empatica E4. At rest, the validation of some commercial wearable devices collecting activity-related cardiac signals is described, by comparison with the related gold standards. The assessment is performed in terms of accuracy and precision (metrological characterisation is currently debatable, since unavailable or not reported with rigorous validations of commercial wearable devices from the manufacturers) to underline the importance of these metrological figures for a better device selection depending on the target application, as reported in [119, 120]. A further analysis investigated the factors interfering the Photoplethysmographic sensor readings and consequently the quality of data acquired and processed, such as motion artefacts in [117] and skin-device contact pressure in [121]. Following this analysis at rest, a new signal processing technique and an approach to reduce the complexity of current detection processes are proposed as in [8], for data collected during acoustic stimuli and after physical exertion, to improve the ML classification performance for external stimuli recognition systems.

To examine the feasibility of the proposed approaches, data analysis was performed first in MATLAB environment, then, when required, by using the WEKA tool for the ML performance evaluation.

3.1 Case study 4: Assessment of wearable devices performance

The daily measurement and continuous monitoring of cardiac data plays a pivotal role in many fields, from hospital care to sport applications, through robotics and affective computing, including the emotion recognition, just to cite some [122]. According to the target application, multiple devices (e.g., ECG, pulse-oximeter, PPG sensors largely widespread in wearable devices, e.g., smartwatches, chest-strap monitors, automatic blood pressure meters based on oscillometric method, ballistocardiograms) can be used to measure a subject's heart rate. Similarly, multiple BP measurement methods exist, namely cuff-based and cuff-less ones. The former category includes, among others, the auscultatory and oscillometric methods, both requiring a sphygmomanometer cuff. In the case of auscultatory method, a skilled operator takes the BP values using a stetho-

scope and jointly listening for the Korotkoff sounds [123]. In oscillometric methods, it is possible to relate intra-arterial BP to the oscillations of a sphygmomanometer cuff pressure, recorded during its gradual deflation [124]. Cuff-less measurement of BP has been made possible by exploiting several techniques, among which the Pulse Transit Time (PTT - time interval in which a BP wave, created by a heartbeat, travels between two points along an artery, [125]) obtained from a pulse wave measured at the wrist. Similarly, the Pulse Wave Velocity (PWV) - based technique, which considers the velocity of the BP wave propagating in the vessels [126]. Both the techniques involving two measurement sites, require an electrocardiogram; BP readings are provided following a proper calibration (by which a user's physiological baseline is established, considering her/his physiological characteristics, e.g. arterial walls rigidity) performed with a validated BP measurement instrument equipped with an inflatable cuff and, then, using the calibration results to properly configure the wearable device. Among the approaches to detect the BP changes from cuff-less wrist-worn devices, along with the HR values, the photoplethysmography has been widely investigated in last decades. A PPG sensor, composed by one or more Light Emitting Diode (LED) and a photodetector, records the pulsatile cardiovascular signal driven by the heartbeat [127]. The HR pulsations are associated to the arterial pressure pulse, then if the PPG signals are integrated with other modalities, such as ECG, it is possible to obtain features like PWV, PTT and Pulse Arrival Time (PAT) for BP measurement.

Although the ECG and the sphygmomanometer are both considered to be the most reliable devices for HR and BP, respectively, (not considering invasive techniques, which however require expert operators as well as specialised healthcare settings), they are not comfortable for use during the daily living activities, especially for long-term monitoring (e.g., an arm cuff inflated from an external source). On the other hand, the reliability of more comfortable commercial wrist-worn devices requires to be verified. Unfortunately, the wearable devices metrological characterisation is currently debatable because information are often unavailable or not reported with rigorous validations from the manufacturers. For this reason, the aim of the studies described in the following section is to extensively evaluate the metrological performance of two cardiac activity-related data (i.e. HR and BP) acquisition methodologies and devices. As such, different consumer technologies devices are simultaneously employed with respect to the gold standard instruments (i.e. ECG and sphygmomanometer, respectively) in order to validate them from a metrological point of view, following a well-defined testing procedure. Standard analyses for assessing measurement accuracy and precision, which are of particular interest in remote physiological monitoring applications, are used.

It is beyond doubt that the hardware characteristics of the used device influence its performance and the quality of measurement results [59]; also the correct positioning of the sensing device is important to collect reliable data, along with the band tightness that obviously influences the measurement results, since sensors are susceptible to the contact pressure with skin; as an example, when wearing a smartwatch, these factors affect particularly the PPG sensor, for which the measured signal depends on

the quantity and quality of light received by the photodetector after being emitted by the LED and having crossed or been reflected by skin tissues. However, to the best of the available knowledge, concerning the consumer wearable devices, neither manufacturers provide specific indications on the optimal band tightening level value that should be achieved in order to maximise the signal-to-noise ratio (SNR), while minimising the environmental light that can reach the photodetector, and maximising the capture of the light reflected/transmitted by the skin, nor data related to this type of investigation are available in literature. Therefore, a prototype of a wrist-worn device including both a PPG sensor and a load cell was realised in order to quantify the effect of different band tightening levels on the recorded signal, evaluating in particular its variability.

All the following tests were conducted according to the principles outlined by the World Medical Association (WMA) in the Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects [128].

3.1.1 Acquisition protocols

In the first study, a test population of 20 healthy subjects (12 females, 8 males), aged (27.5 ± 6.1) years (mean ± standard deviation) has been enrolled. HR data were acquired through the several devices (see Table 3.1 for technical specifications). Additionally, data from pulse-oximeter, automatic BP meter and palpatory method were manually recorded.

The overall dataset was acquired at rest, with subjects remaining as relaxed and still as possible in order to minimise the influence of movement artefacts [129]. Four repeated acquisitions were performed on each subject, for a total of 28 recordings per each device according to the measurement setup shown in Figure 3.1.

All the used acquisition devices were positioned before starting the tests in the following positions:

- Empatica E4 smartband was worn on the right wrist, while Polar Vantage V2 on the left one. Devices were connected to the E4 RealTime and Polar Flow applications, respectively, running on a smartphone;
- Nonin 9560 Onyx II pulse-oximeter was placed on the right index finger;
- cuff of TD-3128 automatic BP meter was positioned on the left arm (following the general recommendations for BP measurements according to the oscillometric method [130]);
- three electrodes for the acquisition of the ECG I lead were positioned according to the Einthoven's triangle (two electrodes on the shoulders, the third on the iliac crest - being neutral from an electrical point of view). The patient cable was connected to the DAQ system;
- chest-strap monitor was worn just below the sternum and made snug enough to ensure a good skin contact.

Figure 3.1: HR measurement setup: smartwatches worn on both wrists, pulse-oximeter placed on the right index finger; BP meter cuff on the left arm; chest-strap monitor on the thorax, where also ECG electrodes were positioned (I lead); artery pulse detected through palpating the carotid.

Table 3.1: Acquisition devices and related technical specifications (*Fs*: sampling frequency)

Acquisition device	**Data acquired**	**Technical specifications**	**Additional notes**
PowerLab 4/25T (DAQ board)	ECG signal	4 digital inputs Accuracy: $> 0.1\%$ Set *Fs*: 1 kHz Set full scale: 5 mV Data format: .txt	• ADInstruments MLA2540. • 5 Lead Shielded Bio Amp Cable (reference device).
Empatica E4 (smartband)	PPG signal	*Fs*: 64 Hz Resolution: 0.9 $nW/Digit$ HR storage *Fs*: 1 Hz Data format: .csv	• HR derived from BVP signal. • 2 red LEDs for motion artefacts removal.
Polar Vantage V2 (smartwatch)	PPG signal	Measurement range: [15,24] *bpm* HR storage *Fs*: 1 Hz Data format: .csv	• HR derived from BVP signal.
Garmin HRM-Swim (chest-strap monitor)	HR signal	HR storage *Fs*: 1 Hz Data format: .fit	• Not standalone device. • Combined with Garmin Venu Sq watch.
Nonin 9560 Onyx II (pulse-oximeter)	HR single value	LED Red (660 nm) LED InfraRed (910 nm) Measurement range: [20,250] *bpm* Precision: $\pm$ 3%	• Manually recorded data.
TD-3128 (BP automatic meter)	HR single value	Measurement range: [40,119] *bpm* Precision: $\pm$ 4%	• Manually recorded data.

Firstly, the ECG acquisition was started, as well as the beginning of the activity was set on wearable devices in order to start data recording. The measurement through the BP meter was launched and the operator started simultaneously to count pulses by palpating the carotid artery (for a time interval of 30 s - then the obtained value was doubled to provide HR in *bpm*). When the oscillometric-based method measurement was completed, the resulting value was annotated, along with the pulsations counted by palpatory method and the HR value readable on the pulse-oximeter display. The remaining data, namely those acquired by the two smartwatches, the chest-strap and the DAQ board acquiring ECG, were saved automatically. In order to obtain a temporal tag for the end of each acquisition, the Empatica E4's mark event button was pushed by the subject at the end of the measurement performed by the automatic

BP meter. In the second study, an automatic sphygmomanometer, the AND A&D Medical UA-767 plus BT-Ci [131], has been used as reference device to measure BP by an oscillometric technique. As summarised in Table 3.2, three smartwatches have been tested, namely the Samsung Galaxy Watch3, Asus VivoWatch SP and Omron HeartGuide).

Table 3.2: Technical specifications of the BP acquisition devices and modalities (NA = not available)

Device	**Acquisition Modality**	**Accuracy** $[mmHg]$	**Range** $[mmHg]$	**Medical Device**
A&D Medical UA-767 plus BT-Ci [131]	Oscillometric method	±3	BP: [20;280]	Yes (93/42/EEC)
Samsung Galaxy Watch3 [132]	PPG sensor and electrode	NA	NA	No
Asus VivoWatch SP (HC-A05) [133]	PPG sensor and electrode	NA	NA	No
Omron HeartGuide HEM-6411T-MAE [134]	Oscillometric method	±3	SBP: [60;230] DBP: [40;160]	Yes (ISO 81060-2:2013)

In this second study, 19 volunteer healthy subjects (10 males, 9 females), aged between 18 and 35 years with weight (70.84 ± 13.75) kg and height (173.74 ± 10.00) cm (mean ± standard deviation) were recruited. As shown in Figure 3.2, participants were sitting on a chair with their arms resting on the table and legs not crossed with both feet on the floor, and instructed to avoid speaking during data collection.

Environmental conditions (e.g., room temperature and ambient light) were maintained as much stable as possible to reduce their influence on the measurements. Before starting the BP data acquisition, the calibration procedure for each device was properly conducted following the related manufacturer's instructions. Also recommendations provided by the manufacturer in the datasheet were considered to setup the test protocols.

According to the Association for the Advancement of Medical Instrumentation (AAMI)/ESH/ISO universal standard for the validation of BP measuring devices, the test methodology included two reference BP measurements, one preceding and one following the device BP acquisition, for each device under test [135]. After 5 minutes of subject's rest, the acquisition protocol started with an initial measurement by using the automatic sphygmomanometer worn on the left arm, along with the calibration

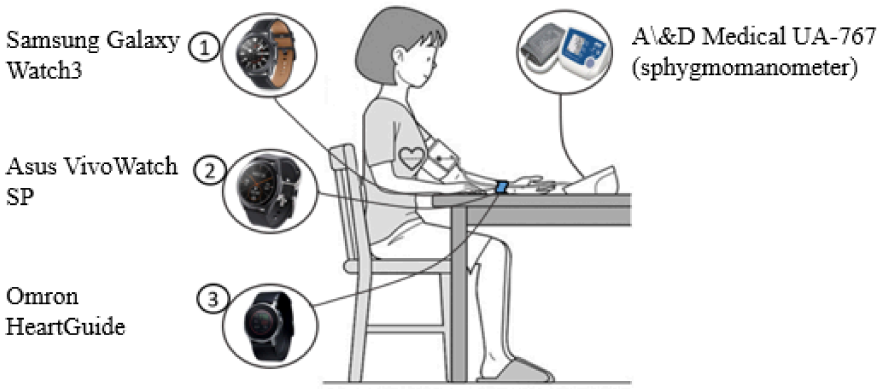


Figure 3.2: BP measurement setup: the three smartwatches (1) Samsung Galaxy Watch3, (2) Asus VivoWatch SP and (3) Omron HeartGuide HEM-6411T-MAE worn on the left wrist one by one, while the BP meter cuff located at the arm level, same side; subject seated and relaxed for the whole test duration.

of the Samsung Galaxy Watch3 worn on the right wrist. This step was supported by the Samsung Health Monitor smartphone app, where the user reports the BP values provided by the sphygmomanometer in three consecutive measurement sessions. Once the calibration terminated, three sequential recordings were taken, alternating the reference and the under-test device as described previously, all placed on the left arm or wrist. In particular, for both the DBP and SBP, the average of the two sphygmomanometric recordings was assumed as the reference DBP and SBP value, to be compared with those obtained through the tested device.

Similarly, the Asus VivoWatch SP was first calibrated with three measurements obtained from the sphygmomanometer. Then, the test procedure described above was performed. The acquisition with the device under test required to place the thumb and the index fingers on the PPG sensor and the electrode, respectively, so as the PTT could be assessed and the on-board AI-based algorithm fine-tuned, according to the subject's specific characteristics.

Finally, the Omron HeartGuide was placed on the left wrist. Calibration is not needed in this case; the acquisition procedure, as recommended by the manufacturer, requires to flex the left arm on the chest, and touch the right shoulder with the left hand. This way, the wrist is placed at the same height of the heart, reducing measurement errors (a difference of 10 cm in height below the level of the heart could lead to an increase in blood pressure up to 7.50 $mmHg$ [136]). In total, 19 recordings for both SBP and DBP were obtained per each investigated and reference devices, respectively. It is worthy to underline that the reference BP values were computed separately for each smartwatch and the devices were tested sequentially on each subject (with an elapsed time between two devices testing of approximately 5 minutes).

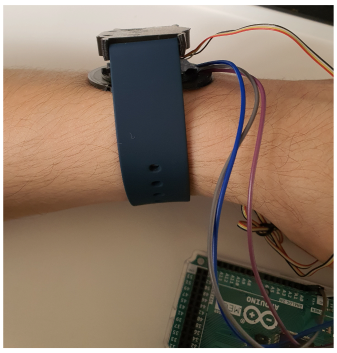


Figure 3.3: DIY wrist-worn wearable device.

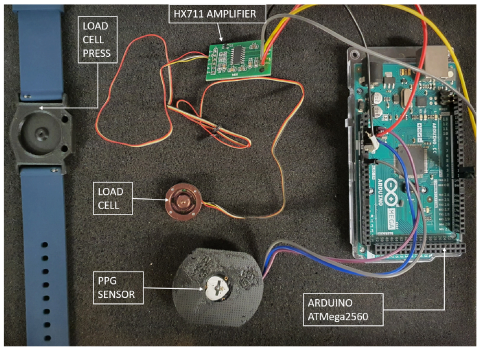


Figure 3.4: DIY wrist-worn wearable device components.

Regarding the third study, for assessing the effect of different band tightening levels and, consequently, of different tightening force values on the HR measurements, 10 healthy subjects (4 males and 6 females aged between 21 and 31 years, with a BMI between 19 and 26 kg/m^2 and a wrist circumference between 14 and 19 cm, with a skin colour classification of Type II (Fitzpatrick scale)) have been enrolled. The PPG data were recorded by using a Do-It-Yourself (DIY) wrist-worn wearable device, consisting of a PPG sensor (Keyestudio XD-58C Pulse Sensor, with a 515 nm green light LED), a button load cell (FX1901, Meas. Spec., Schaffhausen, Switzerland), an HX711 amplifier board and an Arduino ATmega2560 with a sampling rate of 9600 Hz as acquisition board. The assembled device is shown in Figure 3.3, whereas the separate components in Figure 3.4. Concerning the employed sensors, the PPG sensor was fixed in place by means of glue on a custom 3D-printed watch case, whereas the load cell was placed on a 3D printed casing, in order to be held in the right position every time the wrist band is worn.

To evaluate the metrological characteristics of the PPG sensor, the participants were submitted to nine acquisition sessions: three repetitions, lasting 30 s per each, for the three levels of band tightness (i.e. *loose*, *medium* and *tight*). The band tightening levels were defined according to the subject's wrist size. The wrist size was considered as the *loose* level (starting point), i.e with the band length equal to the wrist circumference (L_0, tightening of 0.0 cm). The DIY wrist-worn device was placed on the same wrist as the one where each of the subjects usually wears her/his watch (5 subjects on right wrist and 5 subjects on the left one). An automatic oscillometric digital monitor was employed to collect subjects' data related to BP and HR before starting the acquisitions; this device was positioned on the left arm, as generally recommended. Furthermore, the measurements were repeated after having changed the band tightness. In particular, the prototype wearable device worn at the loose level was tested according to the subject wrist circumference. After 30 s, the band was unfastened and then braced again to test the PPG measurement repeatability. The same procedure, repeated for three times, was adopted for the *medium* and *tight* levels, with +0.5 cm and +1.0 cm of band tightening, respectively. During the tests,

participants were required to avoid hand and/or arm movements to reduce the motion artefacts potentially compromising the measurements. Moreover, data collected by using both the PPG sensor and the load cell were graphically displayed with the Telemetry software to have a preliminary signals visual inspection. Indeed, data from the load cell helped to verify that the reached tightening level was comparable, over all the subjects, irrespective of the personal wrist circumference.

3.1.2 Data Analyses

MATLAB environment was used to process the collected values and perform the statistical analysis in each study presented in the previous section. Before the processing operations, data collected in the first study, were synchronised using the timestamps provided by the different devices (namely the two smartwatches, the chest-strap monitor and the DAQ board acquiring ECG). Then, data were analysed in 30-s long portions, ending with the tag provided by Empatica E4. While wearables directly supply HR data, ECG signal, acquired through the PowerLab 4/25 (i.e. reference instrument), needs to be processed to obtain the tachogram, from which HR values can be extracted. In particular, the reference Pan-Tompkins algorithm [137] was applied to detect R peaks on the ECG signal; then, RR intervals (RRs) were computed and tachogram derived. Average HR value was computed on the tachogram related to the 30 s before the tag. Likewise the mean HR was computed from the 30 s-portions of the signals acquired by means of wearable devices (i.e. the 2 smartwatches and the chest-strap monitor). This way, all the computed HR values are related to the same temporal span over which oscillometric-based method, palpatory method and pulse-oximeter provided their outcomes, so that a comparison can be properly made.

In the second study, the reference SBP and DBP values were computed by determining the average of the two reference measurements preceding and following each smartwatch recording. If the two SBP or DBP reference measurements differed by more than 5 $mmHg$ [138], the recording session was discarded and repeated, up to two additional attempts. In order to identify the best device for blood pressure measurements among the considered wearables, the comparison was performed by following both the British Society of Hypertension (BHS) [138] protocol and the AAMI standard [139]. According to these criteria, a device can be recommended if it receives at least grade B (under the BHS protocol, the highest agreement with the reference device corresponds to grade A, while the lowest one corresponds to grade D) for both systolic and diastolic pressures values, and if the absolute mean difference between standard and under-test device measurement is $< 5\ mmHg$, or the standard deviation is $< 8\ mmHg$, according to AAMI.

Then, in both these overmentioned studies, the deviations between tested methods and the reference instrument (i.e. ECG and sphygnomanometer, for HR and BP respectively) were evaluated; in particular, deviations were computed as reported in Equation 3.1:

$$deviation = data_{testedmethod} - data_{ref.device}[bpm] \tag{3.1}$$

The obtained deviations were at first analysed in terms of their distribution: histogram was plotted with mean and standard deviation values, i.e. the bias μ, and the 95% confidence interval (CI95%). Moreover, the Bland-Altman plots [140] were employed to visualise the measurement accuracy and precision. Finally, the correlation between the data resulting from each different acquisition device and the reference one was assessed with a scatter plot and related fitting line (i.e. $y = m \cdot x + q$), deriving the Pearson's coefficient (ρ) to estimate the strength of the correlation. A $\rho > 0.70$ was considered index of a strong correlation, based on literature findings [141].

Regarding the third study, both PPG and load-cell signals gathered with the DYI wrist-worn device were resampled at 1 kHz, by using the modified Akima piecewise cubic Hermite interpolation. Then, the signal peaks were searched for the computation of HR from PPG signal. Once the HR series were obtained, the statistical quantities were derived, namely the mean (μ), the standard deviation (σ) and the coefficient of variation (c_{v}, also known as relative standard deviation) computed as follows:

$$c_{\mathrm{v}} = \sigma/\mu \tag{3.2}$$

Same features were extracted from data collected through the load cell, to verify that the band tightening system was effective and that the contact pressure levels among different subjects were compatible, as well as repeatable on the same subject. Histograms were plotted to describe the distribution of the measurement results; the number of bins (K) was computed by means of the Sturges' rule formula, as follows:

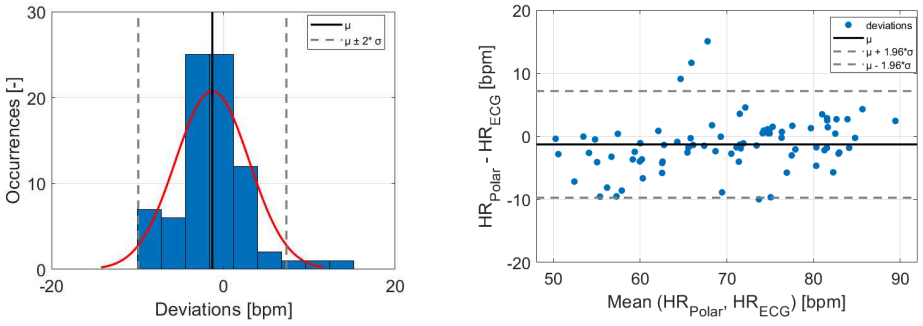
$$K = 1 + \frac{10}{3} log_{10}(N), \tag{3.3}$$

where N is the numerosity of the sample.

3.1.3 Experimental results

For what concerns the evaluation of wearable devices performance in measuring the HR data, an example of the distribution of deviations is shown in Figure 3.5(a) (related to Polar Vantage V2 smartwatch); the black vertical line reported in the graph indicates the mean deviation (i.e. bias μ: -1 bpm) and the grey dashed lines identify the 95% confidence interval CI95% defined as $\mu \pm 2{\cdot}\sigma$, where σ is the standard deviation of the population of deviations obtained from the comparison among devices. For the same smartwatch, in Figure 3.5(b) an example of Bland-Altman plot is reported; as expected, the bias coincides with the mean of the deviation value obtained from the histogram describing the distribution of deviations. Moreover, the mean of the deviation does not present any trend if HR varies within a CI95% of [-10, 7] bpm.

The results from the deviations analysis related to HR measurements performed by all the tested methods with respect to the reference ECG are summarised in Table 3.3. It is worthy to underline that the mean of the deviation (i.e. bias, μ), can be obtained



(a) (b)

Figure 3.5: Distribution of deviations (a) and Bland-Altman plot (b) related to HR measurement (Polar Vantage V2). μ: mean of the deviation; σ: standard deviation of the deviations.

Table 3.3: HR deviations analysis between tested methods and reference (ECG): mean of the deviation (μ), standard deviation of the deviations (σ) and 95% confidence interval (CI95%)

Tested methods	μ [*bpm*]	σ [*bpm*]	CI95% [*bpm*]
Empatica E4	1	4	[8, 10]
Polar Vantage V2	-1	4	[-10, 7]
Garmin HRM-Swim	0	4	[-8, 8]
Nonin 9560 Onyx II	0	4	[-7, 8]
TD-3128	3	12	[-21, 26]
Palpatory method	-1	6	[-12, 10]

both by the deviations distribution in a histogram and the Bland-Altman plot, as already mentioned for the examples related to Polar smartwatch device. According to the results obtained, the Garmin chest-strap results to be the most accurate device, with a bias of approximately 0 *bpm* (i.e. null mean deviation); similarly, the pulse-oximeter (Nonin 9560 Onyx II), with a bias of approximately 0 *bpm*, that turns also to be the most precise instrument, with the lowest standard deviation and the smallest CI95%, of 4 *bpm* and [-7, 8] *bpm* respectively, among the others. The chest-strap appears more accurate than smartwatches (both Empatica and Polar models), which is in line with what it is reported in literature [142]. The sphygmomanometer TD-3128, working according to the oscillometric-based method, is the least accurate device, even if the mean deviation is still acceptable (3 *bpm*); however, its measurements result to be very dispersed (σ = 12 *bpm* and CI95% = [-21, 26] *bpm*). If these results are compared to the device technical specifications, some discrepancies can be underlined; in fact, the pulse oximeter should have a precision of $\pm$ 3% of the measured value, which in the tested HR range should be approximately 2 *bpm* on average (considering HR = 70 *bpm*, the mean value measured in the tests, see Figure 3.5 (b)), versus the obtained σ = 4 *bpm* (hence higher deviation, meaning lower precision than that one

reported). The discrepancy results to be even higher for the BP automatic meter, with an assessed $\sigma = 12$ *bpm* versus the 3 *bpm* on average corresponding to the 4% of the reading reported in the datasheet. Regarding the correlation analysis, an example of the plot depicting HR values measured by a tested instrument (in particular, Garmin HRM-Swim chest-strap) compared to the reference ones is reported in Figure 3.6. It is possible to notice that data are well distributed along the fitting line, with the exception of few outliers, which are quite far from the line. Results of correlation

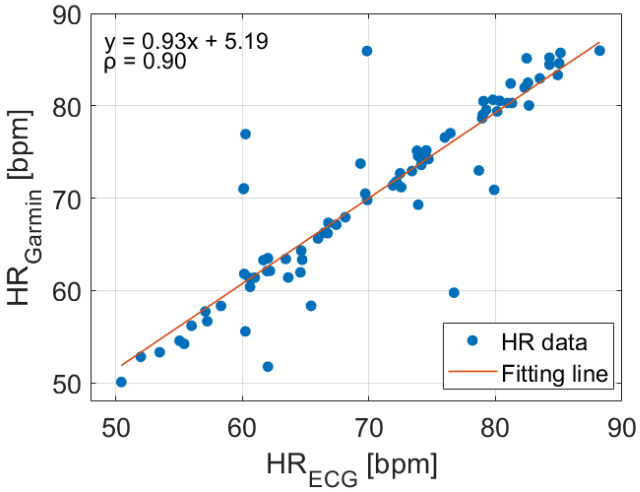


Figure 3.6: Correlation between HR measured by Garmin HRM-Swim chest-strap monitor and reference instrument (ECG).

analysis in terms of fitting line and Pearson's coefficient are reported in Table 3.4, for all the tested devices. It is possible to observe that all the measurement methods, with the exception of oscillometric-based one, show a strong correlation with respect to ECG (i.e. reference). On the other hand, pulse-oximeter provides the strongest correlation with respect to the reference, underlining the high reliability of such a device.

Table 3.4: Correlation between tested methods and reference (ECG): m and b coefficients of the fitting line $y = mx + b$, where x and y are HR measured by ECG and the tested methods, respectively; ρ, Pearson's coefficient

Tested methods	m	b	ρ
Empatica (smartband) E4	1	6	0.90
Polar Vantage V2 (smartwatch)	1	-4	0.92
Garmin HRM-Swim (chest-strap monitor)	1	5	0.90
Nonin 9560 Onyx II (pulse-oximeter)	1	3	0.93
TD-3128 (BP automatic meter)	1	32	0.44
Palpatory method	1	13	0.82

On the other hand, the wearable acquisition devices were evaluated according to the recommendation criteria of the AAMI standard and the BHS protocols. The distribution analysis between under-test device and reference measurements, in terms of bias and standard deviation, along with the Bland-Altman plot and the correlation analysis was performed. An example of distribution analysis is reported in Figure 3.7 for the Samsung Galaxy Watch. The bias is the black solid line, whereas the dotted grey lines indicate the confidence interval limits.

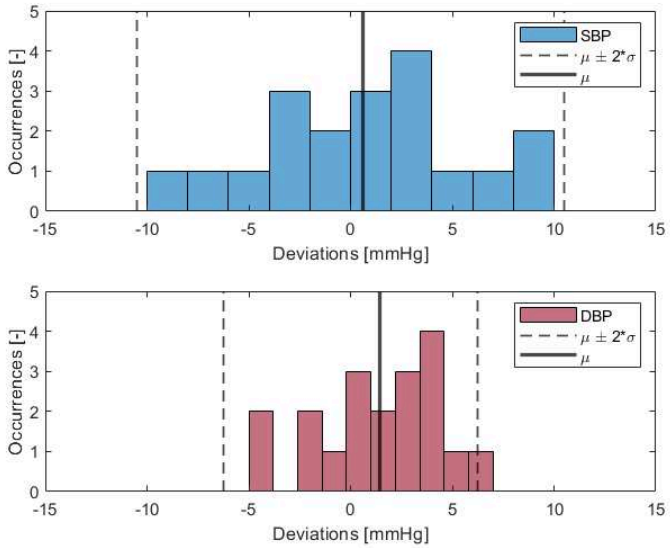


Figure 3.7: Distribution of deviations values for Samsung Galaxy Watch: systolic (SBP) and diastolic (DBP) blood pressure data in $mmHg$. μ: mean of the deviation; σ: standard deviation of the deviations.

According to the AAMI and BHS recommendation criteria, among the tested devices, the Samsung Galaxy Watch is the unique recommendable smartwatch for BP monitoring (Table 3.5). In fact, Omron HeartGuide met both the requirements only for DBP recordings, while Asus VivoWatch received grade D for both SBP and DBP values.

Besides the distribution of deviations analysis, the agreement between the sphygmomanometer and the tested smartwatches was assessed. In Table 3.6, the inter-subjects results related to the bias and the CI95% are reported for all the considered smartwatches. The desired values are represented by null bias and smallest CI95%. While the Samsung Galaxy Watch exhibits the best performance, with low values of both bias and CI95% (the combination of values closest to the desired performance, thus resulting the most precise smartwatch among those tested), and the Omron HeartGuide slightly higher, the Asus VivoWatch shows an evident increase of both the absolute value of the bias (μ_{SBP} = -14.89 $mmHg$; μ_{DBP} = -10.44 $mmHg$) and the interval of agreement (CI95%$_{SBP}$ = [-40.71, 10.92] $mmHg$; CI95%$_{DBP}$ = [-31.70, 10.91] $mmHg$),

Table 3.5: Validation of BP acquisition devices, according to the recommendation criteria of the AAMI and the BHS

	Criteria		
Tested device	AAMI (SBP/DBP)	BHS grade (SBP/DBP)	**Recommendation**
Samsung Galaxy Watch3	Passed/Passed	A/A	Yes
Asus ViVoWatch SP	Failed/Passed	D/D	Not
Omron HeartGuide	Passed/Passed	D/B	Not

resulting in the least accurate device with worst data agreement when compared to the reference device.

Table 3.6: Distribution of SBP and DBP deviations between tested smartwatches and reference device, in terms of bias (μ) and 95% confidence interval (CI95%)

	SBP		**DBP**	
Device	Bias $[mmHg]$	CI95% $[mmHg]$	Bias $[mmHg]$	CI95% $[mmHg]$
Samsung Galaxy Watch3	0.60	[-9.70, 10.91]	1.44	[-4.68, 7.57]
Asus VivoWatch SP	-14.89	[-40.71, 10.92]	-10.44	[-31.70, 10.91]
Omron HeartGuide	1.73	[-21.68, 25.15]	0.73	[-15.72, 17.19]

As an example, Figure 3.8 shows the Bland–Altman plots, which evaluate both the SBP and DBP values discrepancies between data from automatic sphygmomanometer and Asus smartwatch, and the stability across the different values of the blood pressure. The agreement limits are defined by the bias $\pm 1.96 \cdot \sigma$ of the quantified differences.

According to the agreement analysis, the Samsung Galaxy Watch is the most accurate device for the SBP ($\mu_{DBP} = 0.60\ mmHg$), while the Omron HeartGuide for the DBP ($\mu_{DBP} = 0.73\ mmHg$).

Regarding the correlation analysis, for both SBP and DBP values, in Table 3.7, the results underline a strong positive relationship ($\rho > 0.90$; $\rho_{SBP} = 0.93$, $\rho_{DBP} = 0.97$) between the Samsung Galaxy Watch and the automatic sphygmomanometer, a moderate relationship for the Omron HeartGuide smartwatch (as shown in Figure 3.9) and a weak relationship ($\rho < 0.50$) for the Asus VivoWatch.

Concerning the last study about the metrological evaluation of a wrist-worn device at different band tightening levels obtained from the load cell, it is possible to notice that the same band tightening level resulted in different tightening force values. This is probably due to the different subjects' wrist circumference and morphology (i.e.

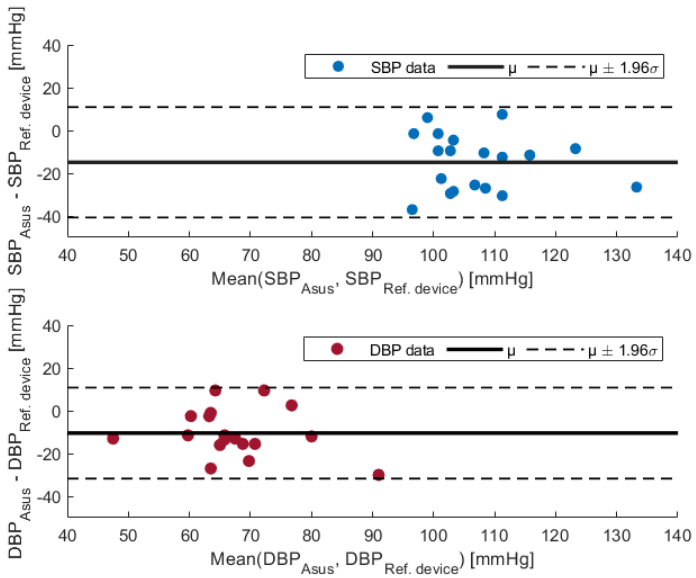


Figure 3.8: Bland-Altman plots for Asus VivoWatch smartwatch: systolic (SBP), and diastolic (DBP) blood pressure data in *mmHg*. The black solid line indicates the bias μ, while the dotted grey lines indicate the confidence interval ($\mu \pm 1.96 \cdot \sigma$).

Table 3.7: Fitting line $y = mx + q$, where x: data measured with the sphygmomanometer; y: data acquired with the tested smartwatches; ρ Pearson's correlation coefficient

	SBP		**DBP**	
Tested device	$y = mx + q$ $[mmHg]$	ρ	$y = mx + q$ $[mmHg]$	ρ
Samsung Galaxy Watch3	$y = 0.92x$+9.62	0.93	y = $0.88x$+9.72	0.97
Asus VivoWatch SP	y = $0.29x$+65.64	0.32	y = $0.38x$+34.17	0.46
Omron HeartGuide	y = $0.57x$+50.72	0.66	y = $0.66x$+25.91	0.78

physiological diversity), which means a different contact condition between the band, consequently the load cell, and the skin. However, considering the whole test population along with all the acquired force signals, the variability (quantified with the mean μ and the standard deviation, σ) among the subjects is quite low ($c_v < 1\%$, see Table 3.8).

On the other hand, analysing the variability within the same subject (i.e. intra-subject variability), higher standard deviation values are reported for some subjects with respect to others (e.g., subject no. 5). However, considering all the tightening levels, it is possible to observe a homogeneous increasing trend (from *loose* to *tight*

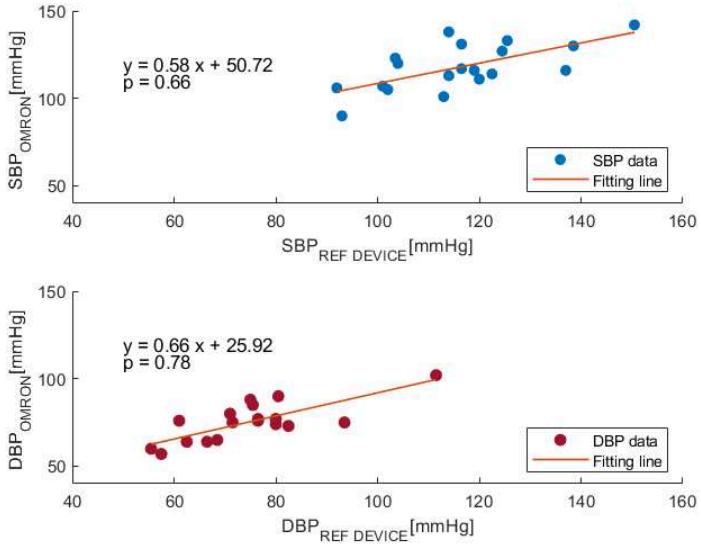


Figure 3.9: Correlation between blood pressure systolic (SBP) and diastolic (DBP) data in $mmHg$, acquired by Omron HeartGuide and the sphygmomanometer.

Table 3.8: Inter-subject variability related to tightening force values obtained from the load cell, for the three different tightening levels (i.e. *loose*, *medium* and *tight*)

Tightening level	**Tightening force**		
	μ [N]	σ [N]	c_v [%]
Loose	0.49	0.15	0.30
Medium	1.07	0.17	0.16
Tight	2.51	0.62	0.24

level) for all the subjects, even if different absolute values of force are reported, as it can be seen in Figure 3.10.

The force distributions, along with the corresponding HR distributions, related to the data acquired with different band tightening levels, are both reported in Figure 3.11.

As it is evident, the force distribution is unimodal (not normal), with a positive skew (i.e. the tail is on the right), whereas the distributions of HR values are approximately Gaussian-like. Regarding the inter-subject variability of PPG results, the values obtained with the three different band tightening levels are compatible (see Table 3.9), considering that HR parameter shows an intrinsic physiological variability, irrespective of the band tightness.

However, it is worthy to underline that the device wearing conditions undoubtedly influence the quality of the acquired data and, consequently, the reliability of the

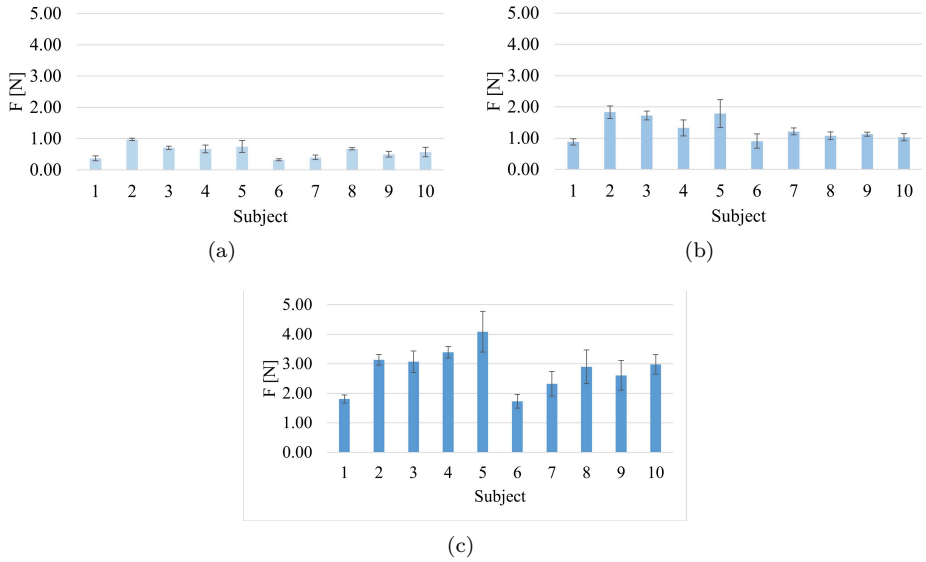


Figure 3.10: Tightening force intervals (mean±standard deviation) measured for *loose* (a), *medium* (b) and *tight* (c) levels, for the whole population.

Table 3.9: Inter-subject variability related to HR obtained from PPG sensor, with the three different tightening levels (i.e. *loose*, *medium* and *tight*)

Tightening level	**HR**		
	μ $[bpm]$	σ $[bpm]$	c_v [%]
Loose	81	15	18
Medium	80	13	16
Tight	81	14	17

measurement results. In particular, the measured tightening force is different depending on the wearing conditions of the device. It is possible to see that the mean value increases with tightening level, as expected (see Table 3.8 and Figure 3.10). However, the coefficient of variation shows a trend not coherent with the tightening level, suggesting that it is possible to obtain more stable results with a higher tightening level of the band. The lowest variation is obtainable with the medium tightening level (corresponding to a tightening of 0.5 *cm* with respect to the subject's wrist circumference).

3.1.4 Discussion and conclusions

The first study aimed at comparing the HR values acquired through several acquisition methodologies (i.e. smartwatch, chest-strap, BP automatic meter, pulse-oximeter and palpatory method) with respect to the gold standard instrument (i.e. ECG). Results show that different technologies provide diverse measurement accuracy and

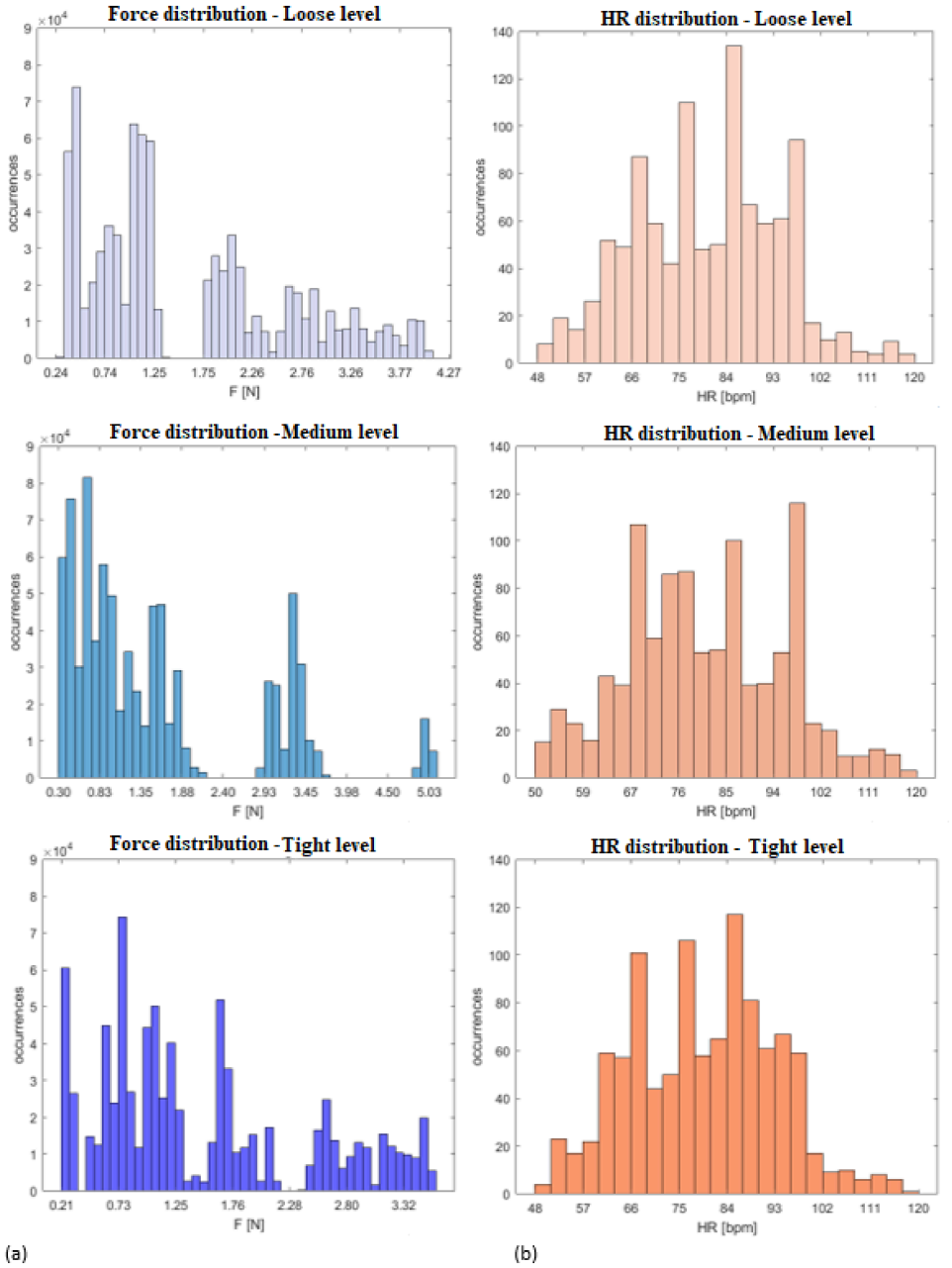


Figure 3.11: Distribution of the measured tightening force (a) and HR (b) in the tests with different tightening levels: *loose*, *medium* and *tight*, for the whole population.

precision. In particular, the pulse-oximeter, recently becoming very popular in relation to COVID-19 (both for SpO_2 and HR parameters), is the most precise ($\sigma = 4$ *bpm* and smallest CI95% = [-7, 8] *bpm*) among the tested devices, whereas the chest

strap (Garmin HRM-Swim) turns to be the most accurate one (bias, $\mu = 0$ *bpm*). Contrarily, the oscillometric-based method (i.e. BP automatic meter) results to be quite inaccurate (bias = 3 *bpm*, still acceptable but underestimated in the device technical specifications) and not precise ($\sigma = 12$ *bpm*), despite being classified as a medical device. A very strong correlation ($\rho > 0.70$) is proved for almost all the tested devices, indicating that the devices outcomes agree with the reference ones, except for BP automatic meter ($\rho = 0.44$).

For what concern the BP measurements from wrist-worn devices, findings confirm that also blood pressure can be monitored by using a PPG sensor. In this case, the sensor needs to be used in conjunction with an electrode for the initial setting and calibration of the data processing method, based on the acquisition of the subject's electrical ECG signal. BP values recorded by wearing the Samsung Galaxy Watch3 are linearly correlated to the values collected by the reference device (i.e. automatic sphygmomanometer), for both SBP and DBP ($\rho_{SBP} = 0.93$ and $\rho_{DBP} = 0.97$). Indeed, the same wearable device results to be the most accurate ($\mu_{SBP} = 0.60$ $mmHg$) and precise (CI95%$_{SBP}$ = [-9.70, 10.91] $mmHg$; CI95%$_{DBP}$ = [-4.68, 7.57] $mmHg$) among the three tested ones; whereas the Asus VivoWatch SP results in the least accurate device, showing the lowest data agreement and correlation when compared to the reference device. The clear advantage is to provide accurate and reliable HR and BP readings to the user, with higher comfort (especially for BP recordings thanks to the cuff-less methods), leading to a high adherence to self-recording sessions for control, hence increasing the user's awareness and commitment on cardiac-related health. Furthermore, the so-called "white-coat effect" may be reduced through wearable-based acquisition.

Later on, the analysis was moved on to whether and how different tightening levels of a wrist-worn band may affect the variability of collected data (from PPG and load cell sensors) and, hence, the reliability of the measurements. Over all the subjects, the results show that the different tightening levels produce an increasing tightening force when passing from *loose* to *tight* through *medium* level. Nevertheless, the coefficient of variation is minimum (i.e. $c_v = 0.16$ %) when the band tightening level is *medium*, according also to the subject's comfort conditions in commonly wearing the DIY wrist-worn wearable device; *tight* level sometimes causes discomfort, particularly in those subjects having a higher wrist circumference (i.e. > 17 *cm*). According to the results, the load cell could be avoided and replaced with a commercial watch wristband, used at different predefined tightening levels, starting from the subject's wrist circumference (i.e. *loose* level). Such replacement can be performed after a dedicated calibration of the different tightening levels on a wide test population, properly including the physiological variability in wrist morphology and circumference also to collect data better fulfilling the normality condition.

In the future, HR and BP reliability from wearables might be performed recruiting a larger and more heterogeneous cohort of subjects to generalise the analysis results and better fulfil the Gaussian data distribution, properly including the physiological variability depending on the skin morphology (e.g., different pigmentation) and age

(two important factors changing the absorption of PPG light). Moreover, since the main usage limitation is the influence of body posture and motion artefacts on the device reliability, future validation studies should investigate the wearable devices with a specific protocol to assess the stability of data acquired over long-term period, also by testing the performance of measuring devices when subject's physiological parameters are altered (e.g., after physical effort). Furthermore, it would be important to define guidelines on how to choose the acquisition device depending on the final application, making users (healthcare personnel but also common citizens) aware of the real performances, hence able to properly interpret the results, without drawing wrong conclusions. However, further interesting investigations can be conducted focusing on the optimal band tightening of wearables in real life, when the subjects perform activities of daily living.

3.2 Case study 5: Assessment of activity-related cardiac signals

As mentioned in the Case study 3, the PPG sensor is sensitive to several disturbance factors, including the motion disturbance. As a result, the use of a proper artefact correction method is of paramount importance. In the literature, three main strategies can be identified for this purpose: deletion, interpolation (with different methods, e.g., nearest neighbours, cubic spline or piecewise cubic Hermite [143, 144]), or filtering techniques. The correction of faulty inter-beat-intervals values can be performed by considering the surrounding IBIs over short windows [145]. At present, there is no consensus on the best way to edit artefacts, but in any case interpolation should be preferred to deletion, in order to avoid spurious frequency components linked to spikes, causing an increase of the high frequency (HF) power and a decrease of the low frequency (LF) one, causing an erroneous evaluation of sympathetic and vagal activities described by the two components, respectively.

The motivation of the studies described in the following subsections is to improve the performance of emotion recognition systems in which classification is performed by exploiting features extracted from HRV analysis. Therefore, this work proposes a new method for artefact correction in IBI time series, based on replacing the missing beats with the mean value of those previous and next to the data gap. In particular, IBI signals from the PPG sensor of Empatica E4 have been considered during tests eliciting emotions through audio stimuli chosen from the IADS-2 database. The effect of the artefact correction method on the results has been then evaluated by comparison to the artefact correction approach implemented in the Kubios toolbox [146]. The presence/absence of stimuli has been detected through the Support Vector Machine (SVM) classifier, considering as input features those identified with statistically significant differences through a preliminary Student's t-test.

Alternatively, when the IBI signal is not available among the data collected through the used device, the IBI signals may be derived from the PPG signal. This is the case of Empatica E4 device, that is able to generate the BVP signal as a primary output from the PPG sensor. Reduced Blood Volume Amplitude (BVA) in the BVP signal, i.e. vasoconstriction, along with limited HRV, increase HR, increase blood pressure and increase of energy expenditure are the physiological variation as consequence of intense physical exertion (that can be thought as a physical stress). Within a work context (e.g., industry), IoT-enabled wearable devices can easily monitor the intensity of physical exertion, and potentially suggest a correct level of performing activity (and, therefore, energy expenditure). This advice could improve the general status of a subject, since it has been demonstrated that interventions at level of physical activity can determine changes not only in physical but also in mental well-being. Therefore, it would be useful to recognise the fatigue-related stress conditions from a calm state, by starting from a few simple physiological parameters derived from BVP signals measured through an IoT wearable device.

3.2.1 Acquisition protocols

The acquisition protocols described in this section were conducted according to a similar general procedure, as reported in Figure 3.12, including a resting phase, a stimuli-based elicitation phase and the last one a second resting phase, since the effect is expected to last also a few minutes following the emotion elicitation. The duration of each phase changed depending on the external stimuli proposed and the target application.

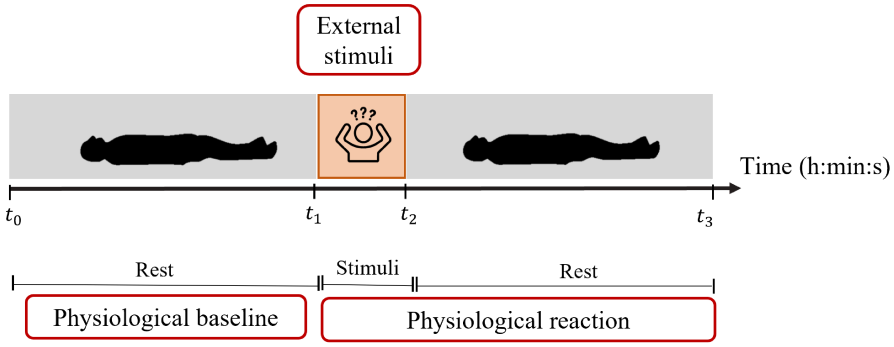


Figure 3.12: General acquisition protocol divided in three main phases: rest, stimuli-based elicitation and physiological response after the stimuli.

In the first study, tests were performed on a population of 7 voluntary healthy subjects, 5 females and 2 males, aged (36±18) years (mean ± standard deviation), with a BMI of (22.7±2.1) kg/m^2. Regarding the measurement setup, subjects laid in supine position and closed eyes in a quiet room, in order to be relaxed as much as possible and avoid interference on the test execution. Data were acquired through Empatica E4 wearable device, already described in detail in the section 1.5; this device requires an initial calibration time of approximately 15 s, therefore the samples in this time interval are discarded from further analyses. By using a Bluetooth speaker placed within the test room, three different audio clips from IADS-2 database [69] were randomly proposed to each subject to measure the physiological changes under acoustic stimuli. The three selected stimuli were different in terms of valence:

- A neutral stimulus, consisting in walk footstep sound (Walking sound no. 722);
- A pleasant stimulus, consisting in rock'n'roll sound (Rock'N'Roll sound no. 815);
- An unpleasant stimulus, consisting in scream sound (Scream sound no. 275).

The duration of each acquisition was equal to 10 minutes: the former 5 minutes in resting conditions, to acquire the baseline of IBI values, the latter including 1 minute of audio stimulus (repeated 10 times, since each sound in the IADS database lasts 6 s) and 4 minutes at rest. The two parts of the acquisition, of equal length, were then used for HRV analysis and comparison between absence (t_0, t_1) and presence (t_1, t_3 - including the acoustic stimuli t_1, t_2) of stimulus. Subjects were asked to press the

event-marker button of the wristband, in correspondence of the beginning and the end of each acoustic stimulus, in order to allow real-time annotations of the recording sessions.

In the second study, BVP data were acquired on a test population of 4 healthy volunteers (2 females and 2 males), for a total of 20 recordings (in particular, each of them repeated the test 3 or more times). The involved healthy subjects aged between 20 and 59 years, with a BMI between 21.07 and 25.25 kg/m^2. Before the test execution, all the subjects signed an informed consent, describing the study objectives and test modalities. The initial phase of each acquisition is conducted at a rest condition lasting 15 minutes (t_0, t_1) while the subjects lay on a bed with Empatica E4 on the dominant wrist. Then, subjects perform 5 minutes (t_1, t_2) of intense physical effort (e.g., running stairs or free body exercises), without wearing the Empatica device. The exercise to execute was selected from the guidelines provided by Mayo Clinic [112], depending on the intensity of the exertion subjectively perceived by each study participant. As soon as the exercise terminates, the subjects are asked to lay down for a 10-minute recording (t_2, t_3), again while wearing the Empatica E4 on the dominant wrist. Hence, each recording included two rest periods, before and after the physical effort (i.e. stimulus absence/presence, respectively). Furthermore, it is worthy to underline that only 10 minutes before the physical exertion were considered for the analysis, in order to have two comparable intervals of the same duration before and after the exertion.

3.2.2 Data Analyses

For the first study, only IBI signals gathered with Empatica E4 were considered. They consist in two columns of data, the former reporting the time instants (in seconds, s) of the sampled data, the latter representing the values of the recorded time separation (in seconds, s) from the previous beat (i.e. RR interval). However, the generated undisclosed tachogram (i.e. the succession of consecutive RR intervals) is not complete due to the artifact avoidance algorithm embedded in Empatica E4 that enables solely the registration of the IBIs obtained from a noise-free PPG signal, by exploiting data from accelerometer to detect motion disturbances. Discarded data cannot be recovered in any way. Hence, when data are corrupted by motion noise, samples are not reported in the .csv file; therefore, a proper algorithm for artefact concealment is needed, in order to be able to perform a reliable HRV analysis without spurious frequency components introduced by missing beats.

In order to avoid spikes that would introduce spurious frequency components, two different correction methods were considered in the present study: the interpolation method implemented within Kubios tool, and a new method proposed. The former (referred as "Kubios Method") corrects the detected abnormal beats with a linear/cubic spline interpolation, considering a time-varying T threshold computed from the distribution of consecutive IBIs. This threshold can be selected from the very low, low, medium, strong, very strong or customised options, depending on the settings chosen

by the operator. In this study, a medium threshold equal to 0.25 s was selected. On the other side, the algorithm proposed (referred as "Proposed Method" in our study [117]) consists in the following steps:

1. detection of missing beats, $T = 2 \cdot \sigma$;
2. computation of the corrected beat, $RR_i = \frac{RR_{i-1}+RR_{i+1}}{2}$;
3. reconstruction of the corrected tachogram

where σ is the deviation between consecutive sampling time intervals. Correction of more than three consecutive beats was rarely performed to avoid significant artificial distortion of the signal.

The signals resulting from both the analyses were assessed in terms of HRV through the Kubios toolbox. Obviously, when considering signals pre-processed with the Proposed Method, no further beat corrections were performed within Kubios toolbox. Concerning the HRV analysis, the parameters listed in Table 3.10 were considered. Then, the two artefact correction methods were compared in terms of deviations (considering their mean value and related standard deviation, with a coverage factor, k, equal to 2) and their distributions.

A preliminary investigation of the differences in HRV analysis parameters with/without an acoustic stimulus was carried out through the Student's t-test (5% significance level). Two sample populations were realised: one with the 5-minute signals before listening to the stimuli (t_0, t_1), the other with the 5-minute signals, including both during and after listening to the stimuli (t_1, t_3). The statistical analysis was performed on the HRV parameters computed by means of Kubios. Looking at the results obtained with the t-test, the parameters more sensitive to stimuli were identified; then, those reporting statistically significant differences between presence/absence of stimuli were used as input variables for a ML analysis. In particular, the SVM classifier was considered, since it is previously reported to have the best performance for HRV analysis [147]. To define the class identity of the proposed features, the classification was performed in WEKA toolbox. Regarding the testing approach used to assess the classification system, the 10-fold cross validation was selected. Finally, the performance of the two artefact correction methods was evaluated in terms of the ability of classifying the presence/absence of a stimulus; since the classification accuracy seems to be higher for arousal than for valence dimension of emotional state, only arousal was considered. Therefore, the classification was of binary type: presence of stimulus or not. The classification performance was evaluated in terms of accuracy, F-measure, sensitivity and precision.

An alternative is represented by the possibility to directly consider the raw BVP signals collected through the PPG sensor embedded in Empatica E4. In this case, a pre-processing phase to obtain the IBI is needed. The two main steps are shown in Figure 3.13.

Firstly, peaks are identified through an algorithm to find the local maxima of BVP pulses and considering the related information content. Secondly, the RR intervals (i.e.

Table 3.10: Features extracted from HRV signals in different domains

Domain	**Features (measurement unit)**
Time domain	mean (RR_{mean}, ms), standard deviation (RR_{std}, ms) of IBIs; mean (HR_{mean}, bpm), standard deviation (HR_{std}, bpm), minimum (HR_{min}, bpm), maximum (HR_{max}, bpm) values of HR; root mean square of successive RRs differences ($RMSSD$, ms); n. successive RRs whose difference exceeds 50 ms ($NN50$, $beats$) and their proportion ($pNN50$, %)
Frequency domain	very low frequency, VLF, 0-0.04 Hz; low frequency, LF, 0.04-0.15 Hz, high frequency, HF, 0.15-0.40 Hz, maximum amplitude (VLF_{peak}, LF_{peak}, HF_{peak}, Hz), absolute (VLF_{abs}, LF_{abs}, HF_{abs}, ms^2), relative powers (VLF_{perc}, LF_{perc}, HF_{perc}, %); band powers in normalized units (LF_{nu}, HF_{nu}, n.u.); total power (P_{tot}, ms^2); ratio between spectral power in LF and HF ranges (LF/HF)
Non-linear parameters	standard deviation of Poincaré plot parameter ($SD1$ and $SD2$, ms); balance between long- and short-term variability ($SD2/SD1$); approximate ($ApEn$) and sample ($SampEn$) entropy; short term ($alpha1$) and long term ($alpha2$) fluctuation slope

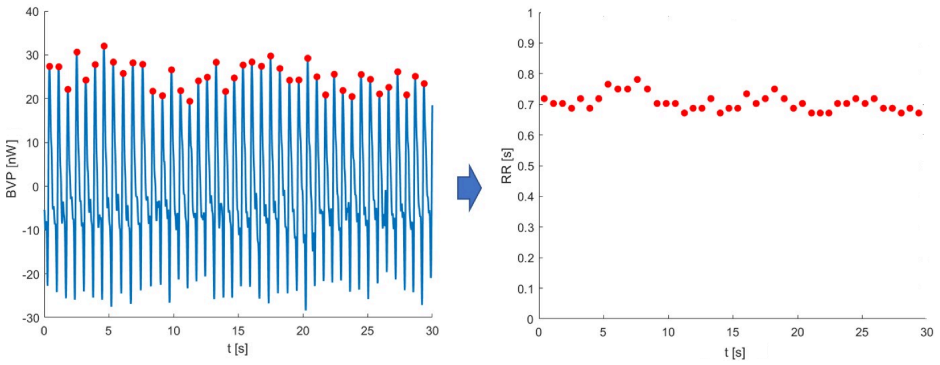


Figure 3.13: Peak detection implemented on BVP signal and the related RR intervals, i.e. tachogram (from left to right).

the inter-beat-intervals) are computed in order to obtain the corresponding tachogram for HRV analysis. In [117] the HRV analysis was performed through the Kubios toolbox [146].

Measured BVP signals were analysed only in time domain both in their original form (extracting the amplitude BVA - as shown in Figure 3.14), and as 1^{st} and 2^{nd} derivatives, considering both the average and the standard deviation of their amplitudes.

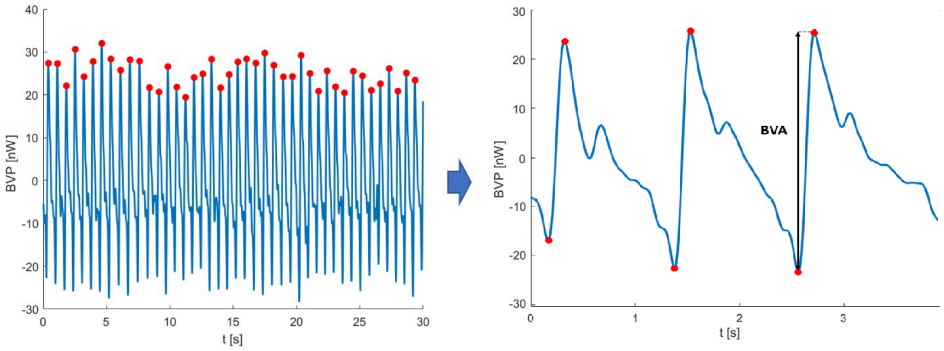


Figure 3.14: BVP signal and the extracted BVA.

HRV analysis provided features, summarised in Table 3.11, in both time and frequency domain, as well as non-linear and information theory measures.

In order to discriminate between physical stress/non-stress conditions, two ML classifiers were considered: Support Vector Machine (SVM) and Random Forest (RF), since they have been previously employed for stress detection, reporting good values of accuracy [148]. As input parameters for the classifiers, the authors considered the following ones: average and standard deviation of HR (HR_{avg} and HR_{std}), absolute powers in LF, and HF frequency bands (LF_{abs}, and HF_{abs}) and their balance (LF/HF),

Table 3.11: Features extracted from the BVP signal

	Type of signal	**Feature (measurement unit)**
MATLAB	BVP as acquired 1^{st} derivative BVP 2^{nd} derivative BVP	Average amplitude BVA, BVA_{avg} (nW) Standard deviation BVA, BVA_{std} (nW) Average amplitude BVA 1^{st} derivative, BVA'_{avg} (nW/s) Standard deviation BVA 1^{st} derivative, BVA'_{std} (nW/s) Average amplitude BVA 2^{nd} derivative, BVA''_{avg} (nW/s^2) Standard deviation BVA 2^{nd} derivative, BVA''_{std} (nW/s^2)
Kubios	Time Domain	Average RR intervals, RR_{avg} (ms) Standard deviation RR intervals, RR_{std} (ms) Average HR, HR_{avg} (bpm) Standard deviation HR, HR_{std} (bpm) Minimum value HR, HR_{min} (bpm) Maximum value HR, HR_{max} (bpm) Root mean square successive RR intervals, $RMSSD$ (ms) Number of RRs with difference > than 50 ms, $NN50$ $(-)$ Proportion of NN50 with respect to total RR, $pNN50$ (%)
	Frequency Domain	Absolute power in VLF band (0-0.04 Hz), VLF_{abs} (ms^2) Absolute power in LF band (0.04-0.15 Hz), LF_{abs} (ms^2) Absolute power in HF band (0.15-0.40 Hz), HF_{abs} (ms^2) Total power, P_{tot} (ms^2)
	Non-linear and information theory measures	HRV short-term variability, $SD1$ (ms) HRV long-term variability, $SD2$ (ms) Balance HRV short/long-term variability, $SD1/SD2$ $(-)$ Approximate entropy, $ApEn$ $(-)$ Sample entropy, $SampEn$ $(-)$ Short-term fluctuation analysis, $alpha1$ $(-)$ Long-term fluctuation analysis, $alpha2$ $(-)$

average and standard deviation of BVA, i.e. amplitude of BVP signal (BVA_{avg} and BVA_{std}). This way, it is possible to assess the significance of the considered physiological parameters in absence/presence of physical stress (and consequently in terms of well-being level).

The WEKA toolbox was employed to perform the classification, using a 10-fold cross validation method as testing approach. The average of each performance metric was computed over the 10 iterations, with a confidence interval of 95%. The performance

of classifiers was evaluated through standard figures, namely Accuracy, F-measure, Sensitivity, Precision and Confusion Matrix [32].

Starting from the same data used for the classification, the authors defined the FRI as:

$$FRI = \frac{(BVA_{avg})}{(HR_{avg}) \cdot (LF/HF)} [nW/bpm]. \tag{3.4}$$

The parameters used to compute the FRI were chosen based on the fact that stress is physiologically characterised by an increased HR_{avg} and a higher LF/HF, as well as reduced BVA_{avg} (i.e. vasoconstriction), hence this index should show lower values in the presence of stress, derived in particular from a physical exertion comparable to that performed in the workplace. As it can be noticed from variables used for FRI calculation (Equation 3.4), the index is proposed as a surrogate for energy expenditure with the final measurement unit homogeneous to energy. Hence, the correlation analysis between FRI and the traditional EE measurements may be evaluated to generalise the index meaning.

3.2.3 Experimental results

Results showed that the two artefact correction methods, namely Kubios Method and Proposed Method, are not equivalent. An example of the comparison between the two obtained tachograms is reported in Figure 3.15.

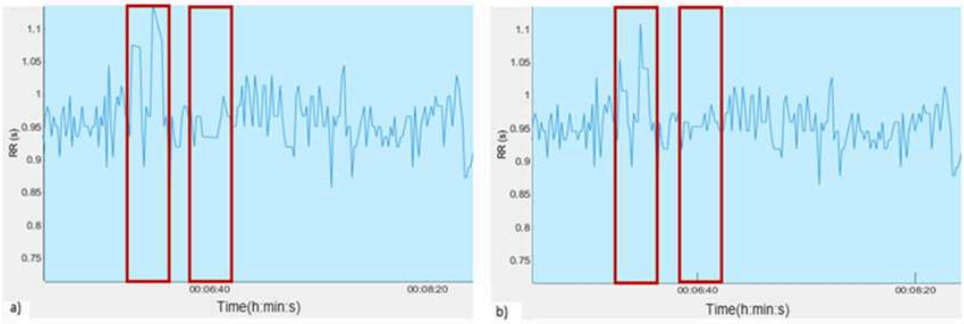


Figure 3.15: Comparison between tachograms obtained with Proposed Method (a) and Kubios Method (b). In red the corrected data.

It is important to underline that only the HRV parameters obtained for the data pre-processed with Kubios Method gave Not a Number (NaN) values sometimes. Differences were highlighted from the deviations analysis between the two methods. An example is reported in Figure 3.16 for the LF_{peak} parameter; it can be noticed that the distribution is Gaussian and its mean is approximately 0 Hz.

In particular, for the reported example, we have a deviation of (0.00±0.06) Hz between parameters obtained with Proposed Method and Kubios Method, considering a coverage factor k = 2. Based on this assumption, further analyses were conducted

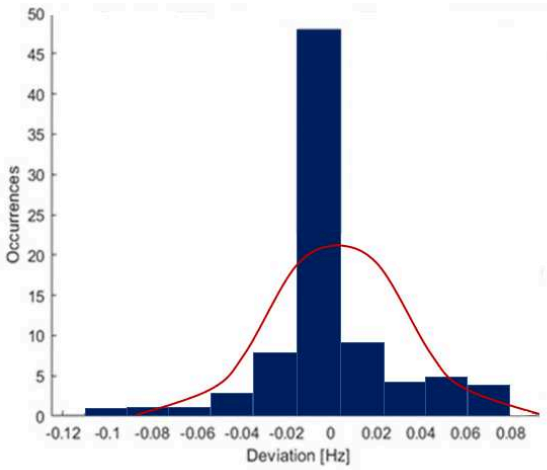


Figure 3.16: Example of deviations distributions between the two methods - LF_{peak} parameter.

to understand which one, among the two tested approaches, provides the best signal reconstruction, and consequently the best extracted features set in order to achieve the highest performance in SVM classification. By means of Student's t-test, parameters with statistically significant differences (significance level: 5%) between presence and absence of stimuli were identified for data pre-processed with the Proposed and Kubios Methods. In particular, 18 parameters were identified as input for the SVM classifier, namely RR_{avg}, RR_{std}, HR_{avg}, HR_{std}, HR_{min}, HR_{max}, $RMSSD$, VLF_{abs}, LF_{abs}, HF_{abs}, P_{tot}, $SD1$, $SD2$, $SD2/SD1$, $ApEn$, $SampEn$, $alpha1$ and $alpha2$, HR_{max}, $RMSSD$ and $SD1$. Since the two methods were demonstrated to be not equivalent (both through the analysis of deviations and t-test between HRV related parameters), it was expected to find different results concerning emotion recognition as well. The accuracy of the classification between the presence and the absence of a stimulus was equal to 66.67% for the data pre-processed with the Proposed Method, against the 48.81% with the Kubios Method. It is worthy to note that, considering the data pre-processed with the Proposed Method, almost all the signals with stimuli were correctly classified (i.e. 41 instances up to 42), whereas some of those without stimuli were not (i.e. 15 instances up to 42). On the contrary, for the data pre-processed with the Kubios Method, there were misclassified signals for both the presence and the absence of stimuli (i.e. only 50% of instances correctly classified). Besides the accuracy, also F-measure was computed in the two classes: presence of stimulus (0.74 and 0.49, for Proposed and Kubios Method, respectively) and absence of stimulus (0.52 and 0.48 for Proposed and Kubios Method, respectively). From findings, the use of the Proposed Method provides a better classification performance in terms of the detection of emotions elicited by an audio stimulus. However, the accuracy cannot be considered completely satisfactory, not even with the Proposed Method.

The good performance of SVM is confirmed in the second study for the identification

of stress presence (labelled as the trial after the physical exertion execution), with respect to RF. All the figures computed for the performance evaluation of both classifiers are reported in Table 3.12. The SVM classifier proves to be the best one according to all the metrics, including the related confusion matrix: the before and after physical exertion conditions, representing the non-stressed and stressed status, are well-distinguished (i.e. 19 and 18 instances out of 20, respectively), except for few instances.

Table 3.12: Classifiers performance metrics

Classifier	**Accuracy (%)**	**F-measure**	**Sensitivity**	**Precision**
SVM	92.50	0.92	0.92	0.93
RF	82.50	0.82	0.82	0.83

The values of FRI obtained before and after the execution of physical exertion are reported in Figure 3.17 for each recording.

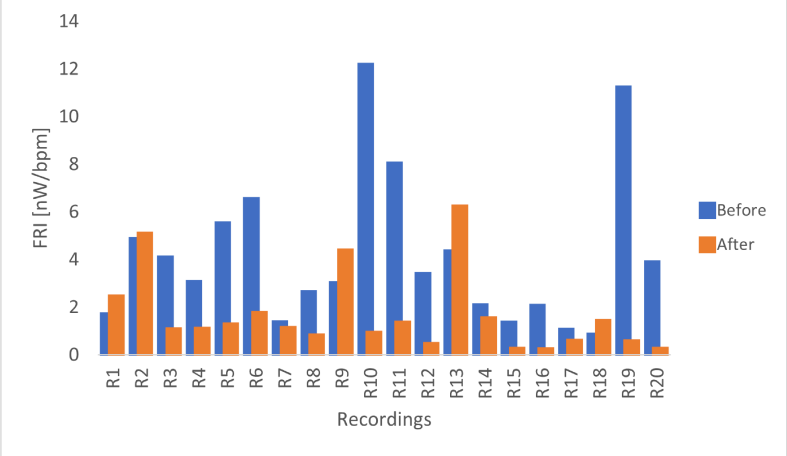


Figure 3.17: FRI values nW/bpm before and after the physical exertion.

FRI values of (4.03±3.19) nW/bpm and (1.51±1.67) nW/bpm (average ± standard deviation) are reported before and after the physical exertion, respectively. Median values are equal to 3.09 nW/bpm and 0.98 nW/bpm before and after the physical exertion, with inter-quartile range (IQR) values of 3.06 nW/bpm and 1.01 nW/bpm, respectively. As it is evident, there is a reduction of 62.47% in FRI after the exertion, which presumably causes a certain level of stress in the subject. Moreover, the high performance of the SVM classifier in discriminating between stress/non-stress conditions confirms that the parameters chosen for the definition of FRI are significant in terms of stress detection, and, hence, fatigue-related conditions like those outstanding in the workplace.

3.2.4 Discussion and conclusions

The first work proposed a new method for artefact correction in IBI signals acquired through the PPG sensor of a wearable device (Empatica E4). Such method was validated in comparison to linear/spline interpolation implemented by the Kubios toolbox used as a reference tool in the literature. Then, the performance in terms of emotion classification accuracy was evaluated, by playing three different audio stimuli to elicit emotion in 7 subjects. A total of 18 features, identified with a significance level $< 5\%$ through a preliminary Student's t-test, are employed as input variables for SVM classifier, in order to detect and classify the presence/absence of the audio stimulus, depending on the emotional response reflected in the changes of IBIs time series. According to the percentage of accuracy achieved by the SVM classifier, the Proposed Method (accuracy: 66.67%) for the artefact correction provides a better performance with respect to that one implemented in the Kubios tool (accuracy: 48.81%). This emphasises the importance of the artefact correction method and the strong impact on the results: the capability to recognise the presence of a stimulus (considering only its arousal) is significantly affected by the artefact correction method used to pre-process the acquired data.

After verifying the efficiency of the proposed artefact correction method, the same was implemented also in the second study, where two ML algorithms were assessed in distinguishing between non-stress and stress conditions, caused by physical effort. An index, called Fatigue Related Index, was proposed to assess the fatigue-related stress level in response to physical exertion, starting from physiological parameters measured with the commercial wearable device Empatica E4. From the measured BVP signal, being linked to stressed and relaxed statuses, the BVA_{avg}, HR_{avg} and LF/HF parameters were selected to define the FRI and calculated before and after the execution of a physical exertion. The choosen classifiers, namely SVM and RF, provide accuracies equal to 92.50% and 82.50%, respectively, in the discrimination between stress/non-stress conditions. Moreover, the proposed index show a decrease of 62.47% after the physical exertion because of the stress/fatigue-related reaction of the subject. Based on the results, the combination of BVP, HR, and HRV data measured through an IoT-enabled wearable device yields an accurate recognition of fatigue-related stress conditions. When applied to the Industry 4.0, the continuous monitoring of fatigue-related stress conditions in workers may both support the assessment of workers' well-being status and drive actions on changing processes and organisation aimed to improve the workers' perception of workload and overall working environment.

Future studies may improve the proposed method, by implementing a proper algorithm for the automatic identification of missing beats in IBI signals. Moreover, an idea may be to test a multimodal system to improve the classification performance, thanks to the broader fingerprint obtained at multiple levels, facilitating the emotion recognition task with respect to the use of a single physiological signal. This consideration might be useful also in tests performed on subjects during normal daily activities, somehow compensating for the influence of movement artefacts on the recorded signals, and consequently on the classification performance of ML classifiers.

In addition, further classifiers may be tested and compared in terms of classification performance. Finally, it would be interesting to increase the number of involved people, hence collecting enough data for the classifier to be able to reliably distinguish among stimuli with different intensities, which with the considered dataset results not possible. To this aim, also the inclusion of self-assessment questionnaires could be helpful for the ML algorithms training, and consequently classification.

Chapter 4

Assessment of Inertial Signal

Decision-making processes need reliable and high-quality information to properly interpret results and take right decisions, especially in the healthcare domain (e.g., personalised care and remote monitoring). Within this context, the last set of relevant studies concerns the issues in data acquisition from wearable devices focusing on acceleration signal. Compared to the majority of ADLs datasets available in literature that are collected in lab conditions, the following studies consisted in the measurement of wrist acceleration during daily human activities performed in a real-life setting (i.e. not-controlled environment), without instructions or guidance provided to the subject, by wearing wrist-worn devices (dataset collected available in [149]). Since in preliminary works the findings showed that both different hardware sensor technologies and the modality to collect data (real vs lab conditions) play a paramount role to determine the accuracy of automatic activity classification, in [150] and in [151] respectively, further investigations were conducted, as reported in [59]. In particular, data collected from Empatica E4 device and an Arduino-based wristband prototype are compared to quantify how the ML performance is affected by the nature and quality of measurement data. When processing long-term data, users may be exposed to privacy violations due to leakage of personal details. If gait has been proved to enable a subject's identification from own motion data and algorithms to sanitise this aspect are well-known in the literature, in our study [152] we demonstrated the risk of exposure of personal information, namely the subject's age, also when the collected acceleration data are not related to gait, but associated to different types of daily activities captured from data collected on the wrist. To mitigate these aspects, a Multi Objective Evolutionary Algorithm (MOEA) approach is proposed in [153] and implemented to conceal firstly the subject's gender while maximising the accuracy on data recording. Therefore, the aim of these studies is to evaluate the performance in HAR accuracy by implementing different learning algorithms, in relation to the sensing device, the measurement data accuracy and the privacy preservation approach.

To examine the feasibility of the proposed approaches, data analysis was performed first in MATLAB environment and Python, then, when required, by using the WEKA tool for the ML performance evaluation.

4.1 Case study 6: Issues in data acquisition from wearable devices

4.1.1 Human activity recognition and factors of influence

Among the quantities commonly assessed by wearable devices, many studies in literature related to HAR deal with the acceleration signal, since it quantifies the motion across the three spatial dimensions , and, at the same time, it can be acquired by means of small, lightweight, and inexpensive sensors, thus appearing particularly suitable for the monitoring, recognition and tracking of ADLs. It is worthy to note that the choice of device type [154] and the sensor positioning [155, 156] significantly influences the HAR performance, especially in terms of classification accuracy. A possible solution that can mitigate these problems is to collect a huge variety of data, hence improving the classification accuracy by fusion with data generated by other sensors [157], or multiple body-worn accelerometers [158]. Nevertheless, the user's comfort should always be accounted for, in particular to ensure the compliance in the prolonged use of wearable sensors for long-term remote monitoring. Concerning ML algorithms, different approaches have been extensively applied in the literature. Among the others, both Jiménez-Gómez et al. [159] and Xu et al. [160] confirmed the RF validity by reporting a performance greater than 90% in the classification of different activities measured with the wearable device. Kuncan et al. [161] tested the Motif Patterns approach, for extracting the features from accelerometer and magnetometer signals, with the Long-Short Term Memory (LSTM) method in the classification of 19 different activities, including daily and sport activities, reporting an accuracy of 98.42%.

However, the performance of the overmentioned algorithms is highly affected by the nature and quality of the collected dataset. Such awareness results in challenges during the study, especially in the phase of experimental setup and subsequent data analysis. Among the challenges which highly affect the quality of the recognition, the most common ones are how to collect the data in the real-life conditions, and how to select and extract the features to be computed. In this Chapter, following the preliminary findings presented in [150], the influence of the wearable sensing device characteristics and measurement uncertainty on the accuracy of ADLs classification is described. In particular, two different measuring devices have been considered (i.e. Empatica E4 smartband and an Arduino-based wristband prototype) to measure the acceleration signal at the dominant wrist of the subject performing different ADLs in real-life conditions. Even the prototype, despite the known limitations, might replace the Empatica E4 if coupled with properly selected classification algorithms.

4.1.2 Acquisition protocol

In order to measure and collect acceleration values from the subjects' wrist, an Arduino-based wristband prototype (shown in Figure 4.1) was assembled at the Electronic Instrumentations and Measurements Lab of the Università Politecnica delle Marche, using an Arduino UNO Microcontroller connected through a four wires-

connector to an I2C Grove 3D accelerometer and gyroscope (based on LSM6DS3 chip by ST^{TM}Microelectronics). The four different wires are used as follows: the red wire is connected to the 5 V output pin on the Arduino UNO board and provides the power supply; the yellow wire is connected to the SCL (Serial Clock) pin and carries the clock signal; the white wire is connected to the SDA (Serial Data) pin and carries the acquired data samples; the black wire must be connected to the GND (Ground) pin on the Arduino UNO board. For consistency with the acceleration data format of the E4, the sampling frequency was set to 32 Hz, in the full-scale linear acceleration measurement range set to ±2 g, with corresponding sensitivity values of 0.061 mg/LSB (Least Significant Bit). Moreover, the acceleration measurements acquired by Arduino are saved in a file with a 4-column format, containing the timestamp and the acceleration components' values (expressed as a multiple of g m/s^2) along the three axes (A_x, A_y, A_z).

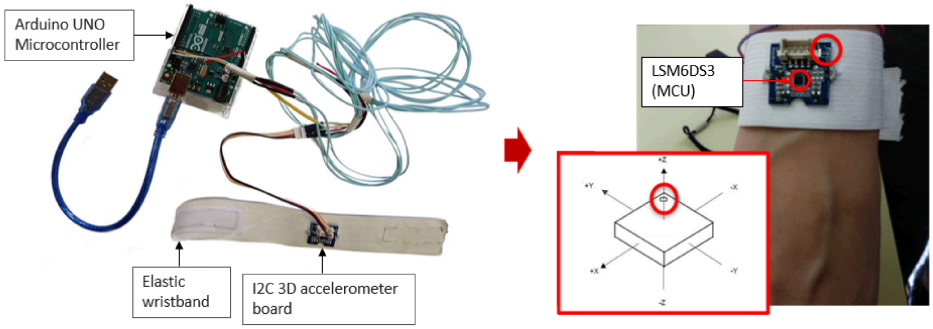


Figure 4.1: Arduino-based wristband prototype and the X, Y, and Z axes relative to the accelerometer onboard.

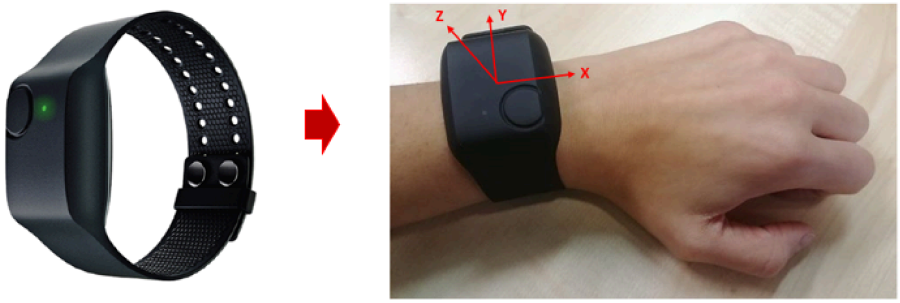


Figure 4.2: Empatica E4 wristband and the X, Y, and Z axes relative to the accelerometer onboard.

On the other hand, Figure 4.2 shows the 3-axes orientation of Empatica E4 device, already described in Chapter 1.5. Experimental tests were performed by 36 healthy subjects (18 females, 18 males), aged 29.5 ± 3.4 years (mean ± standard deviation);

11 subjects were left-handed and 25 right-handed. Data were simultaneously recorded through both the devices, i.e. the Empatica E4 and the Arduino-based prototype, placed on the dominant wrist. Each test subject performed a series of six ADLs, pertaining to personal hygiene and housekeeping related activities, specifically: washing hands (WH), brushing teeth (BT), brushing hair (BH), dusting (D), ironing (I), and washing dishes (WD). Each activity was continuously recorded for 5 minutes and repeated 3 times by each subject to have a quite long and large data collection, but also to facilitate the signal processing, such as the spectral analysis thanks to a better resolution). An important aspect is that the activities were all performed in free conditions and in a real-world scenario, namely at subject's home environment, in standard daily hours by means of real tools, without any instructions about the execution. Although the signals collected were realistic as much as possible, the actual activity duration was sometimes modified (e.g., for WH or BT), with respect to what usually happens in daily life, for the reasons mentioned above. In order to facilitate the labelling of data collected and to annotate reference information for the activity classification task, the participants were supervised during the recorded sessions.

4.1.3 Data Analyses

Pre-processing phase improves data quality and consequently increases the reliability of outcomes, since raw data are generally affected by noise, maybe linked to movement artefacts or a not adequate device-wrist contact. For refining noisy datasets, two filters were applied: a 4^{th} order low-pass Butterworth filter with a cut-off frequency equal to 15 Hz, to preserve human motion while eliminating noise, and a 3^{rd} order median filter to remove abnormal spikes [162, 163]. Regarding the sensor calibration, note that E4 calibrates itself during the initial 15 s of each session, which, for this reason are removed (the subject did not move during this interval). Contrary, the calibration of accelerometer in the Arduino-based device is performed before starting the measurements, by exploiting the acceleration due to gravity for 30 s (i.e. the accelerometer is fixed in three different positions and orientation to align each sensor axis with the axis of gravitational force). Before moving to the ADLs classification, the acceleration signals were windowed to fixed-sized parts, namely into 3 s non-overlapping windows, since the performed activities are not rapidly time-variant. For each window, time features strictly related to the changes in the acceleration signal were derived from the directional components ($A_{x,i}$, $A_{y,i}$, $A_{z,i}$, where i is the index of the sample) and the Signal Magnitude Vector (SMV) to reduce the impact of sensor orientation on the activity discrimination The SMV can be defined as a vector of N elements a_i, with $i = 1 \dots N$, being N the total number of samples in a given acquisition (i.e., the product between the acquisition duration and the sampling frequency):

$$SMV_i = \sqrt{A_{x,i}^2 + A_{y,i}^2 + A_{z,i}^2} \tag{4.1}$$

The samples plot shown in Figure 4.3 illustrates the single acceleration components and the resulting SMV.

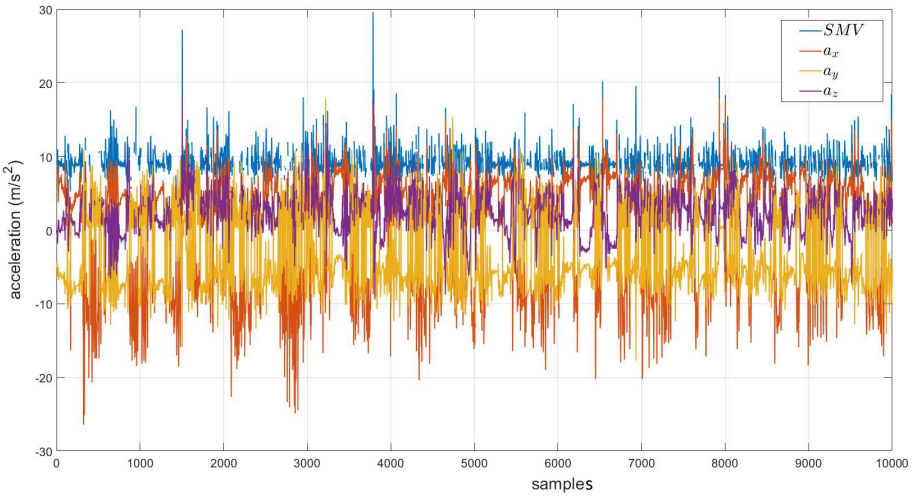


Figure 4.3: Samples plot illustrating the single acceleration components and the resulting SMV.

For what regards spectral features, they were extracted from the magnitude of the discrete Fast Fourier Transform (FFT) of each acceleration signal. Table 4.1 summarises the whole set of time- and frequency-related features considered.

From this whole set, five datasets of features were derived as shown in Table 4.2.

After organising the datasets for the signals collected from E4 and Arduino-based prototype, ADLs classification was performed. The performance of different ML classifiers was estimated by means of the Weka tool to compute the evaluation metrics.

In order to identify only the most discriminant features, by excluding the redundant ones, a feature selection method was used. Among those supported in Weka, the Information Gain (IG) approach [164] was selected, as validated in a previous work [162]. This approach evaluates the amount of information for each feature by measuring the information gain with respect to the class of activity, namely:

$$IG = H(Class) - H(Class|Feature) \tag{4.2}$$

where *H(Class)* is the entropy of the class of activity and *H(Class|Feature)* is the conditional entropy, which represents how the considered feature is consistent to identify a particular class of activity. All the ML approaches mentioned in Chapter 1, i.e. DT, RF, SVM, ANNs, kNN and NB, were selected to classify the considered ADLs, adopting the 10-fold cross-validation method. The performance was assessed through two validation metrics, namely the accuracy and F-measure. It is worthy to say that among the DT, the J48 algorithm was used since available in Weka tool.

Table 4.1: Features extracted in time and frequency domain, divided per computation type

Domain	**Features**	**Computation**
Time Domain	Mean, Median, Standard Deviation, Range Maximum, Minimum, Zero Crossing, Coefficient of Variation, Autocorrelation, Median Absolute Deviation, Skewness, Kurtosis	X, Y, Z axes, SMV
	Axes Correlation	XY, YZ, ZX axes
	Signal Magnitude Area, No. of Peaks, Percentiles (20^{th} - 50^{th} - 80^{th} - 90^{th}), Peak - Peak Amplitude	SMV
Frequency Domain	Spectral Entropy, Spectral Energy, Spectral Centroid	SMV

Table 4.2: Datasets of features

Dataset	**Features Domain**	**Computation**
D1	Time	SMV
D2	Frequency	SMV
D3	Time and Frequency	SMV
D4	Time	A_x, A_y, A_z and SMV
D5	Time and Frequency	A_x, A_y, A_z and SMV

4.1.4 Experimental results

In terms of accuracy, the performances of the six different algorithms tested with the same set of features for both the Empatica E4 and the Arduino-based prototype signals are summarised in Table 4.3.

Regarding the Empatica E4, the J48, RF, and ANNs classifiers achieve an accuracy $> 90\%$ for D1 and D3 datasets, whereas SVM and kNN only reach values in the range 70-80%; among the others, NB classifier reported the worst performance: correctly recognised around 50% of the activities (i.e. 56.78% and 48.65% for D1 and D3, respectively). Considering D4 and D5 datasets, similar results are reported; however the accuracy of all the classifiers is $< 40\%$ if D2 is considered, confirming that the

Table 4.3: Average percent accuracy of tested classifiers on acceleration data from Empatica E4 and Arduino-prototype devices

Device	Classifier	D1	D2	D3	D4	D5
Empatica E4	J48	99.55	31.36	99.51	99.39	99.23
	RF	98.55	30.77	95.53	99.12	97.67
	NB	56.87	20.80	48.65	50.66	54.20
	SVM	70.65	17.84	71.11	83.39	86.12
	ANNs	93.28	24.24	92.67	96.25	93.52
	kNN	80.65	27.78	81.41	89.82	89.75
Arduino-prototype	J48	83.49	31.63	80.37	90.04	87.05
	RF	89.86	39.08	88.70	95.06	93.81
	NB	54.37	16.73	55.93	66.54	67.23
	SVM	65.17	18.81	65.97	81.24	81.57
	ANNs	75.31	23.80	72.79	91.70	91.28
	kNN	79.03	40.03	78.28	92.75	92.23

frequency domain features used alone are not suitable for the classification of the ADLs considered, based on acceleration signals collected on the wrist. Similarly the Arduino-based prototype, the highest accuracy values ($> 80\%$) are obtained for D1 and D3 datasets processed by J48 and RF classifiers. Both D4 and D5 datasets (including SMV, along with the features over the three directions of the acceleration signal, i.e A_x, A_y and A_z) allow to achieve accuracy values $> 90\%$ for RF, ANNs, and kNN. Again, NB classifier reported the overall worst performance (i.e. 16.73%) and poor results are obtained when only features in frequency domain (i.e. D2 dataset) are chosen, irrespective to the classifier used. In this regard, a further investigation was conducted by computing the mean $\pm$ standard deviation of the Spectral Entropy and Spectral Energy. In Table 4.4, the results show that values related to the Spectral Energy are very similar either for different devices and activities. This explains why the use of such feature would not improve the accuracy of the classification algorithm in a significant fashion. On the other hand, the values of the Spectral Entropy are mostly higher and much more dispersed for the Arduino-based device than the E4. Being the Spectral Entropy a measure of uncertainty, the differences among the devices are based on the fact that E4 runs a proprietary firmware that identifies the unreliable or noisy samples and automatically remove them, while this capability is not available from the Arduino-based device. As a result, Spectral Entropy may be considered as a relevant feature in recognising the type of device used (E4 or Arduino), and not the activity (since values are similar among the overall ADLs investigated).

Being D4 the dataset allowing very good performance, both the partial and the average percent accuracy values provided on it by the six classifiers are detailed in Table 4.5 (Empatica E4) and Table 4.6 (Arduino-based prototype). In particular, the Tables show how accurately each activity performed by subjects (i.e., Washing Hands, Brushing Teeth, Brushing Hair, Dusting, Ironing, Washing Dishes) is classified,

Table 4.4: Spectral Energy and Entropy values (mean ± standard deviation) computed for each device and ADL

	Spectral Entropy		**Spectral Energy**	
ADL	**E4**	**Arduino**	**E4**	**Arduino**
WH	4.28±3.79	2.78±1.80	0.51±0.22	0.61±0.19
BT	3.91±2.84	4.74±3.05	0.44±0.15	0.55±0.16
BH	2.31±1.61	4.44±3.81	0.58±0.18	0.53±0.23
D	5.42±5.44	4.41±3.61	0.46±0.23	0.49±0.19
I	5.75±5.09	7.33±9.21	0.44±0.19	0.31±0.19
WD	2.44±1.27	3.12±2.01	0.55±0.14	0.54±0.17

depending on the ML classifier selected.

Table 4.5: Partial and average percent accuracy (Avg.) of the tested classifier (CL.) – Empatica E4

	ADLs						
CL.	WH	BT	BH	D	I	WD	**Avg.**
J48	99.09	98.55	100.00	99.81	99.45	99.40	99.39
RF	98.55	98.91	100.00	98.73	99.46	99.27	99.12
ANNs	96.19	97.46	99.45	96.90	96.24	91.84	96.25
kNN	90.39	90.57	99.45	92.39	88.94	77.17	89.82
NB	90.94	85.68	17.93	34.42	44.21	29.34	50.66
SVM	86.59	90.99	99.45	58.51	89.31	76.81	83.39

Table 4.6: Partial and average percent accuracy (Avg.) of the tested classifier (CL.) – Arduino prototype

	ADLs						
CL.	WH	BT	BH	D	I	WD	**Avg.**
J48	92.32	91.38	89.51	88.57	87.07	91.38	90.04
RF	94.38	96.82	96.63	94.38	93.82	94.38	95.06
ANNs	90.45	95.88	94.57	90.07	89.88	89.32	91.70
kNN	93.44	95.50	97.00	92.13	88.01	90.45	92.75
NB	71.72	82.21	82.78	52.06	64.80	45.70	66.54
SVM	83.52	91.76	85.39	76.03	84.64	66.10	81.24

Although both J48 and RF achieved 100% of accuracy for BH activity recorded by using the Empatica E4, their performance decreases when processing acceleration signals collected wearing the Arduino-based device. Similarly, ANNs and SVM were less accurate in recognising the ADLs activities from the Arduino-based tested device

than those recorded using the Empatica. Contrarily, kNN classifier reached an accuracy greater than J48 for five activities (except for WD activity) from Arduino-based tested device. When using NB classifier, the highest number of misclassifications happened providing a relatively low accuracy (lowest value: 17.93% for BH activity collected using the Empatica E4).

To validate the findings and compare the performance of the learning algorithms with the highest accuracy of the model (i.e. J48, RF and ANNs), further evaluations were carried out. Besides the accuracy, the F-measure was computed for each single dataset, from both Empatica E4 and Arduino-based devices, as reported in Table 4.7.

Table 4.7: F-measure of the tested classifiers on ADLs-related acceleration data

	Arduino - based device			**Empatica E4**		
Dataset	**J48**	**RF**	**ANNs**	**J48**	**RF**	**ANNs**
D1	0.83	0.90	0.75	0.99	0.98	0.93
D2	0.32	0.38	0.20	0.31	0.31	0.23
D3	0.99	0.98	0.89	0.99	0.96	0.92
D4	0.90	0.89	0.92	0.99	0.99	0.96
D5	0.87	0.94	0.91	0.99	0.98	0.93

Firstly, the high values of F-measure on D1, D3, D4 and D5 confirm the accuracy trend. Secondly, the low F-measure values in D2 (including frequency-domain features only) demonstrate that the features computed in the time domain are essential to reach a good classification performance. According to all the findings, RF clearly showed a performance better than the other classifiers in terms of accuracy measures, in classifying the six different activities working on data collected from different wearable devices.

To develop a successful automatic classification system based on ML, the appropriate features must be identified, since a high correlation among the features may strongly affect the classification performance [165]. Therefore, the Information Gain filter was applied on D3 dataset, because it includes the whole features set for SMV_i signals, from both Arduino and E4. Even if a list with different features was extracted for each device, a new optimal subset was composed including the 10 features with the highest IG ranking score to have the same number of features and according to a previous similar work [166]. Then, the classifiers performances were tested again. For both the devices used, the set of features is listed in Table 4.8, in descending order of their IG value.

It is interesting to notice how, for the two sensing devices, the IG filter selected almost the same features but with different rankings. The two couples of features uniquely characterising the specific device are median absolute deviation with interquartile, and signal magnitude area with variance for Arduino-based prototype and E4 device, respectively. Selected features were used to feed the six classifiers and compare the performances obtained in two different cases: when using all features versus a smaller

Table 4.8: Subset of features selected by applying Information Gain filter

Arduino - based device	**Empatica E4**
Autocorrelation	Signal Magnitude Area
Maximum	Mean
Mean	Percentiles 80^{th}
Median	Median
Percentiles 80^{th}	Autocorrelation
Percentiles 20^{th}	Percentile 20^{th}
Range	Maximum
Standard Deviation	Variance
Median Absolute Deviation	Standard Deviation
Interquartile	Range

carefully selected set, as shown in Figure 4.4.

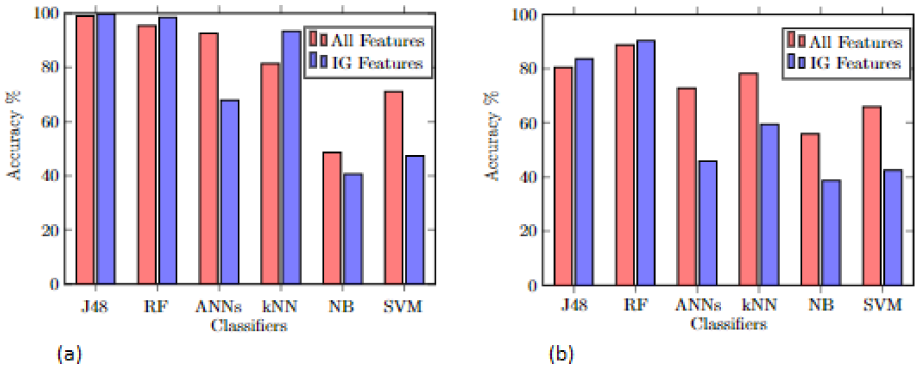


Figure 4.4: Comparison of accuracy before and after applying the Information Gain - Empatica (a) and Arduino-based (b) device acceleration measurement data.

By examining the E4 device results in Figure 4.4 (a), among the learning algorithms, the RF, kNN, and J48 reached the highest values, improving the performances. In details, after applying the IG selection, J48 provided the highest accuracy of 99.81%. On the contrary, the classification results from NB, SVM, and ANNs are not satisfactory compared to RF, kNN, and J48. Regarding the results obtained by analysing data collected with Arduino-base prototype (Figure 4.4 (b)), the highest accuracy of 89.39% is achieved by RF, whereas the performances of NB and SVM decreased from 56% to 40% and from 66% to 42%, respectively. Finally, ANNs contributed with a wider range of accuracy values, namely from 72% to 45%. Looking the overall performance obtained using all the features versus a specific subset, it is possible to notice that some algorithms slightly increased their performance when processing less features, namely J48 and RF. In contrast, the accuracy of other classifiers drastically decreased

when using the reduced subset of features, especially to classify data from Arduino-based device; this is the case of four out of six classifiers (i.e. ANNs, kNN, NB, and SVM), resulting in a substantial decrease in terms of accuracy (maximum decrease obtained: 27.02% for ANNs). By wearing E4, the results indicate that ANNs, NB, and SVM obtained lower accuracy percentages: 67.96%, 40.61%, and 47.34%, respectively. Classifiers characterised by different natures reasonably can be differently affected by irrelevant and redundant features.

4.1.5 Discussion and conclusions

The fluctuation of the classification performance may depend on the activity, signal processing and classifier selected (Table 4.5 and Table 4.6). Therefore, five datasets including different features were arranged and tested. From the results, the highest accuracy is obtained thanks to the combination of time and frequency domain features extracted from both the acceleration SMV and the signal components along the three directions (i.e. A_x, A_y, and A_z), corresponding to D4 and D5 datasets. Additionally, the performance of almost all the ML classifiers, in terms of accuracy, F-measure, specificity, and precision, results higher by using data collected through the Empatica E4 than those obtained using the Arduino-based prototype: Empatica E4 reached accuracy values in the range of 51%-99%, while the Arduino-based prototype in the range of 66%-95%; in both cases, the best performance is achieved by using J48 and RF classifiers. As a result, it is possible to infer that the measurement device characteristics affect the data quality and consequently the activities classification accuracy. This point is strongly important for the reliability of activity classification in critical application fields, such as health monitoring and assistive technologies for ageing people.

Another interesting point is the type of activity performed, and consequently the signal quality because the acceleration signals are sensitive to the movements of the subjects. Accordingly, some activities are classified with a good performance, whereas others are more difficult to discriminate. In particular, the quality of data declines very drastically in the case of non predictable physical activities: as an example, "Washing Dishes" and "Washing Hands", characterised by personal and casual movements, featured low classification rates with high misclassification, mostly due to the difficulty of distinguishing the highly random gestures (see Table 4.5). Contrarily, activities such as "Brushing Teeth", "Brushing Hair" and "Ironing" are performed quite similarly among the users because the movements are repetitive, linear and performed mainly along one spatial dimension.

The classifiers performance was assessed also by considering the most meaningful features. The number of relevant features for a good classification of ADLs changes depending on both the sensing device and the classifier, denoting the importance of an appropriate choice of both hardware and software components of a monitoring system, especially when high accuracy and reliability are required, such as in AAL and health-related applications, sometimes also supporting decision-making processes

in the definition of therapeutic strategies.

The results show the potential implications that measurement device and data quality may have on HAR classification accuracy. This study would analyse the consequences of making decisions in HAR system in real-life, investigating some possible solutions to improve the results. As a future work, the classification approaches herein considered can be trained and tested on data gathered from older adults, verifying the classification performance not only with respect to the choice of device and the selection of features, but also in terms of the influence of gender- and age-related patterns in the acceleration signals.

4.2 Case study 7: Issues in information extraction from wearable devices

4.2.1 Human activity recognition and privacy preservation

Depending on the nature of the data collected and also on the target item to classify (e.g., activity and gesture), different ML algorithms can be suitable for different scenarios of HAR. Generally, the algorithms exploit features computed on the acceleration data, either in time or frequency domain; however, some of the features may capture not only the information strictly needed to classify the human activity, but also the subject's personal details (e.g., age, height, gender and weight) can be exposed. In the literature, several mechanisms and approaches have been proposed in order to guarantee anonymous sharing and avoidance of privacy disclosure [167]. However, the same personal information could be disclosed also without the explicit consensus of the user, by processing motion-related data collected from wearable sensors, especially those generated by accelerometers; thus becoming a potential threat despite being thought to be harmless. As explained in [168], a potential privacy leakage exists about the user's identity from the sensed accelerometer data, because it is possible to infer gait characteristics depending on a user's muscle growth, height, and weight. Based on the above observations, motion data can be classified as a *quasi-identifier*, because it may allow the identification and tracking of a user. In fact, every individual has a distinctive way of walking, which is the reason why gait can be a key element of biometric techniques to authenticate and/or identify the user of a wearable device. As an example, Boutet et al. in their report [169] present a privacy-preserving tool to sanitise motion sensor data against unwanted sensitive inferences (thus improving privacy), while keeping an acceptable accuracy of the HAR (thus maintaining data utility). To do so, the tool builds several models to sanitise the motion data against the specified attribute (such as gender), by exploiting Generative Adversarial Networks (GANs). Authors present test results on available data collections, for which gender inference is reduced up to 41% while decreasing the HAR accuracy only by 3%. While acceleration signals related to gait has been proved to enable a subject's identification from user motion data in the literature, a second important aspect is the risk of personal details exposure also when the data acceleration is to associated to different types of gestures or daily activities (e.g., house cleaning or brushing teeth). Disclosure of personal details, namely the subject's age, has been investigated in [152]. This motivated the need for approaches to both privacy by design and by default in wearable devices, and for designing de-identification algorithms to be applied onto acceleration signals gathered by wrist-worn devices, to reduce the unintentional release of personal details. So, features computed from the acceleration data acquired with a smart wristband while performing different activities, and used by supervised classification algorithms, should be properly selected; so that HAR accuracy (i.e. the utility of the data) is not affected, while personal attributes (i.e. unnecessary private details in the processed data), in our specific study [153] the gender, are concealed.

A Multi-Objective Evolutionary Algorithm (MOEA) is applied to find appropriate weights for each feature, differently from the paper cited above which exploits GANs.

4.2.2 Acquisition protocol

A dataset was acquired with the wearable multi-sensor device Empatica E4, providing raw activity motion data in real-life conditions. The dataset and the related description is publicly available as PAAL ADL Accelerometry dataset [149], and it is continuously updated with new data collected. Among the signals collected by the sensors, only the acceleration has been extracted to monitor the users performing different activities of daily living. To promote the real-life acquisition procedure, subjects acted in their natural environment, with no instructions about how and for how long to perform each activity.
The dataset includes 24 different activities (listed in Table 4.9) performed using real objects with the common duration. Each activity was repeated between 3 and 5 times by 33 healthy subjects, 19 females and 14 males between 18 and 77 years (mean = 45.24 years and standard deviation = 18.24 years), to reach a gender balance and large age range. Information about the gender and age of each subject are included in the dataset.

Table 4.9: ADLs performed by the subjects

1. Drink water	13. Sit down
2. Eat meal	14. Stand up
3. Open a bottle	15. Write
4. Open a box	16. Phone call
5. Brush teeth	17. Type on keyboard
6. Brush hair	18. Salute
7. Take off jacket	19. Sneeze/cough
8. Put on jacket	20. Blow nose
9. Put on a shoe	21. Wash hands
10. Take off a shoe	22. Dusting
11. Put on glasses	23. Ironing
12. Take off glasses	24. Wash dishes

4.2.3 Data Analyses

As for the first study, each acceleration time series A was filtered by firstly applying a 4^{th} order low-pass Butterworth filter (cut-off frequency: 15 Hz) then with a 3^{rd} order median filter was applied to remove abnormal spikes. In order to infer the information contained in the human activity data, the features listed in Table 4.1 were reviewed according to the aim of this study. In particular, 62 features were extracted from both raw data of the three axes and SMV, each segmented by fixed-size sliding windows of 5 s (i.e. 160 samples), with 20% (i.e. 1 s) of overlapping between two adjacent

windows, being activities of short duration. The list of features used is presented in Table 4.10 [150, 170].

Table 4.10: Features extracted in time and frequency domain, divided per computation type

Domain	**Features**	**Computation**
Time	Mean, Median, Standard Deviation Maximum, Minimum, Range, Coefficient of Variation Median Absolute Deviation	X, Y, Z axes, SMV
	Axes Correlation	XY, YZ, ZX axes
	Signal Magnitude Area, Autocorrelation Skewness, Kurtosis, Root Mean Square Percentiles (20^{th} - 50^{th} - 80^{th} - 90^{th}) Interquartile range, No. of Peaks Peak - Peak Amplitude, Energy	SMV
Frequency	Mean, Standard Deviation	X, Y, Z axes, SMV
	Spectral Entropy, Spectral Energy Spectral Centroid Percentiles (25^{th} - 50^{th} - 75^{th})	SMV

The overall implementation was made using the Python library Scikit-learn. Among the ML classifiers commonly selected to train and test the HAR classification model, Random Forest was used thanks to its good performance reported in previous similar studies [171]. In this study, the whole set of features was used to build each tree of RF, measuring the quality of split with the Information Gain function. The number of trees was investigated by selecting different numbers of estimators (from 10 to 170) to verify whether such parameter can improve the model accuracy for activity and gender recognition. The k-fold cross validation was implemented with $k = 10$.

Although the above data analysis process was designed to reach an high accuracy in HAR, if the goal is to protect identity details of the users performing those activities, the same set of features also gets good results for gender recognition. Hence, it is important to understand whether and how the input features may be transformed so that HAR remains good but gender (or any other private details) recognition accuracy decreases. In this study conducted within the PAAL project, namely in collaboration

with the University of Alicante, the acceleration signal captured by the wristband was filtered trying to maximise the HAR accuracy and at the same time to minimise the recognition of personal characteristics of the subjects, such as gender or age. In particular, the MOEA with the wrapper approach (i.e. the Non-dominated Sorting Genetic Algorithm II (NSGA-II) offered by pymoo [172] in Python) was implemented to find appropriate weights for each feature, by using the following parameters selected experimentally: (i) size of the population: 50; (ii) new individuals (offsprings) created per generation: 10; and (iii) number of generations without changes in the best individual to stop the algorithm: 100. Other characteristics of the algorithm are as follow:

- individuals in the initial population are created with random real values between 0 and 1;
- binary tournament is used to select the parents to generate a new offspring;
- Simulated Binary Crossover (SBX) [173] with default parameters is employed to create each individual;
- each new individual is mutated by applying Polynomial Mutation [173] with default parameters;
- duplicates are eliminated after merging the parent and the offspring population: duplicates with respect to the current population or in the offsprings itself are removed and the mating process is repeated to fill up the offsprings until the desired number of unique offsprings is met.

An individual in the population (potential solution) is encoded as a real vector U whose elements $u_j, \forall j \in [1..62]$ represent the weight of a particular feature during the classification (62 being the number of features, see Table 4.10), i.e. each feature is multiplied by the appropriate weight before being input to the classifier. The fitness functions to be optimised are:

$$f_1 = HAR \tag{4.3}$$

$$f_2 = \left| GenderRecognition - \frac{1}{2} \right| \tag{4.4}$$

The objective is to maximise f_1 while minimising f_2, i.e. maximising HAR while taking gender recognition close to random. Therefore, for other private information, the second term in f_2 must be $\frac{1}{Number\ of\ categories}$.

4.2.4 Experimental results

Generally, in the HAR context, supervised learning approaches are widely exploited by using labelled data as inputs for predicting the classification of unknown data through the ML algorithms. Firstly, the performance of the RF supervised algorithm used to

investigate both the activity and gender recognition, separately, was verified. Then, the proposed approach was tested.

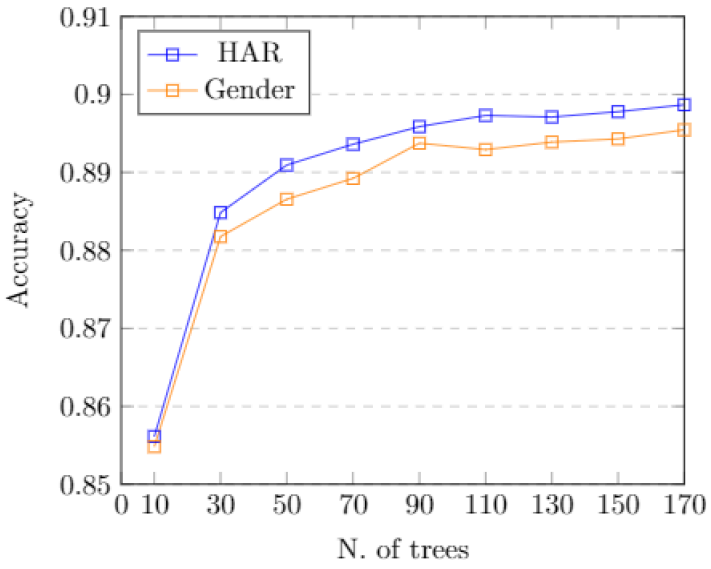


Figure 4.5: Accuracy and number of trees for RF classifier.

In Figure 4.5, the number of trees is associated to a value of accuracy percentage achieved by the RF algorithm. Contrary to what often may be thought, a large number of trees does not necessarily cause a better recognition performance for the RF algorithm: for both models, the accuracy percentages stabilise or slightly decrease as the number of trees increases, especially after a number of trees equal to 90. This means that the benefit in recognition performance when using more trees is lower than the cost in computation time for learning the further trees. At 90 estimators, the global accuracy percentage is 89.59% and 89.37% for human activity and gender recognition, respectively. Besides the accuracy, the performance of both the activity and gender recognition was assessed using the confusion matrix to summarise the relationship between the actual and predicted classes, either correctly or wrongly classified. Figure 4.6 shows the results obtained for the RF algorithm in classifying each activity. Although reaching good accuracy, some activities were highly misclassified, namely: open a box classified as open bottle, take off a shoe classified as put on a shoe, take off glasses classified as put on glasses, and stand up classified as sit down. On the other hand for distinguishing the gender of the participants, as it is clear from Table 4.11, male subjects were confused more than female ones, probably because there were more female users involved in the study.

As shown in Figure 4.7, the application of the MOEA algorithm obtains a set of solutions, optimising one over the other objective and vice versa. Among these, the solution closer (using Euclidean distance) to perfect HAR (value equal to 1) and random gender recognition (value equal to 0.5) was selected according to the overall aim of the study. The accuracy percentage obtained from this selected final

Actual class \ Predicted class	drink water	eat meal	open bottle	open a box	brush teeth	brush hair	take off jacket	put on jacket	put on a shoe	take off a shoe	put on glasses	take off glasses	sit down	stand up	writing	phone call	type on a keyboard	salute	sneeze/cough	blow nose	washing hands	dusting	ironing	washing dishes
drink water	90	1									3	3	1				1		1	2				
eat meal	2	55	5	4							5	3	6	7		3			5	5				
open bottle	1	4	44	30					1	1	1	3	3	1		1		1	1	9				
open a box	1	1	24	57						1	2	1		1				1	1	9				
brush teeth					95	0	0	0	0	1							1				0	0	2	0
brush hair	0				1	90	1	2	2	0				0				0	0		0	1	2	0
take off jacket						1	84	9	1	1								2				1		
put on jacket						1	3	87	4	0					0							2	1	1
put on a shoe			0		0	1	0	0	85	3			0				0			1	1	1	4	3
take off a shoe		1	3	1	1	0	1		14	51	0		4	2	0	0	0	1	3	3	2	3	4	3
put on glasses	5	3		1							70	11	1	3		6		1	1					
take off glasses	5	3	3	6						1	12	54	4	2		1		4	3	1				
sit down		0	2	1						1	0	1	72	16				1	1	1				1
stand up		4	2	1						2		1	18	69				1	1	1				
writing															98		1						0	
phone call		3		2		1	2			1	10	3	1	2		73		1		1				1
type on a keyboard									0						3		96			0			1	
salute				1		1	1	1		1	1	3				3		90						
sneeze/cough	3	6	6	8						2	8	5	5	6		1		6	41	3		1	1	
blow nose	3	4	4	5	1	1			1	1	1	1	2		1	1	1		4	67		1	1	2
washing hands		0					0	0	1	1			1	0							92	0	0	4
dusting				0	0	1	1	1	3	2						0		1		0	1	85	2	2
ironing					0	0			0	0	0		0		0		2		0	0		0	96	1
washing dishes					0	0	0		1	0										0	2	1	1	95

Figure 4.6: Confusion matrix related to the HAR (before MOEA).

Table 4.11: Confusion matrix related to the gender recognition (before MOEA)

Actual Instances \ Predicted Instances	Male	Female
Male	79	21
Female	4	96

solution (represented as blue dot) are 84.14% and 64.38%, for activity and gender recognition, respectively. Comparing with the results in which all the features were equally considered (represented as red dot), HAR accuracy was reduced 5.45 percent points while gender recognition worsened 24.99%.

This is confirmed in Figure 4.8 and Figure 4.6 showing similar results, although accuracy was reduced for most of the classes. Contrarily, gender classification was dramatically affected by applying the MOEA; indeed almost all the activities are now identified as being performed by a female user (Table 4.12).

4.2.5 Discussion and conclusions

Preliminary analyses performed in [174] showed that the accelerometer signals acquired by the elderly subject (86 years old woman) feature evident differences with respect to the signals collected by young volunteers (mean age 29.5 years old), at a parity of the ADL performed. With increasing age, physical limitations may appear or

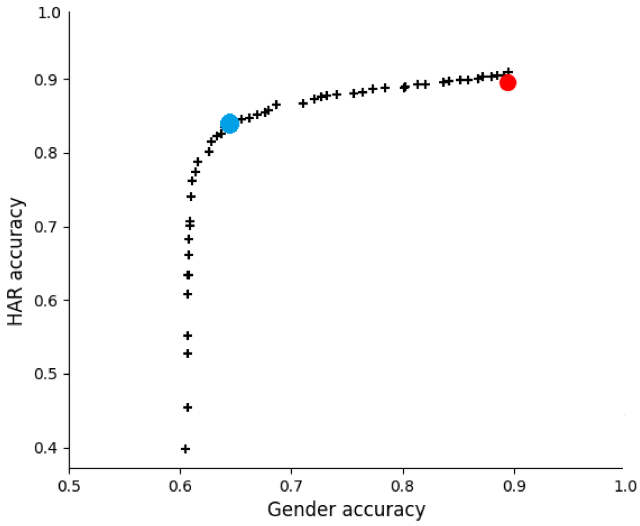


Figure 4.7: Final set of solutions for the MOEA algorithm. Blue dot: best balanced solution; red dot: initial result with all features are equally considered. For the sake of clarity, 0.5 has been added to f_2, which is then the gender accuracy.

		Predicted class																						
		drink water	eat meal	open bottle	open a box	brush teeth	brush hair	take off jacket	put on jacket	put on a shoe	take off a shoe	put on glasses	take off glasses	sit down	stand up	writing	phone call	type on a keyboard	salute	sneeze/cough	blow nose	washing hands	dusting	ironing	washing dishes
Actual class	drink water	83	1	1								7	4	1	1	1		1			1			1	
	eat meal	3	43	7	8					1	1	3	5	7	7	1	3		1	5	4			2	
	open bottle	1	5	41	26					2	2	3	1	3	1	2		1	1	3	5		1		
	open a box	1	1	27	50					1		2	1	3	3	1	1	1	1	2	3		1		
	brush teeth					91			0	0	0					1		2				0	0	4	1
	brush hair	0				3	79	0	2	2	0	0	0				0		0		0	0	1	7	3
	take off jacket					0	1	74	18	1	0								1			0	1	0	1
	put on jacket						1	1	84	2						1		0	0				3	1	5
	put on a shoe					2	1	0	0	68	1			1	0	1		1				2	1	6	16
	take off a shoe		1	5	0	7	1	1		9	16	0		3	3	9	1	1	2	2	0	5	1	11	21
	put on glasses	6	3	1	1							63	12	2			8		3	1				1	
	take off glasses	9	2	5	6							14	42	6	2	1	2	1	7	2				1	
	sit down		2	2	3		1			1	2	1	2	65	16	1	2		1	2			1	1	
	stand up		5	1	1						1	1	1	13	70	1	2		1	2			1	1	
	writing															98		2						0	
	phone call	1	3		1		1	2		1		7	3	2	1		67		5		1	1	4	1	1
	type on a keyboard															6		93						1	
	salute		1		1		1	1	1			1	4	1			6		82					1		
	sneeze/cough	2	3	5	5		1				1	10	5	5	6	3	1	1	8	36	3		1	3	
	blow nose	1	2	6	4	3	1				1	1	1	2	1	6	1	6		1	48		1	12	5
	washing hands		0			0			0	0	0			0	0	0					0	84	0	0	13
	dusting					3	1	2	3	0	1						0		0			1	72	5	13
	ironing					1				0	0					1	0	3		0	0	0	0	92	1
	washing dishes					1			0	0								0				1	0	1	96

Figure 4.8: Confusion matrix related to the HAR (after MOEA).

Table 4.12: Confusion matrix related to the gender recognition (after MOEA)

		Predicted Instances	
		Male	**Female**
Actual Instances	**Male**	10	90
	Female	0	100

increase, namely the movements become slower, the intensity and the energy involved to perform ADLs decrease; thus the acceleration signals from the wrist exhibit quite different characteristics in young and elderly subjects. This could lead to a disclosure of the observed subject's age range, somehow releasing a personal information that is not actually relevant for the aim of ADLs classification and automatic recognition.

Based on this preliminary investigation, the study focused on whether also gender information is included within the set of features extracted from accelerometer data. Therefore, a method for privacy preservation in non gait-related human motion data collection was proposed working on 24 different ADLs performed in realistic conditions by 33 healthy subjects, wearing the Empatica E4 wrist worn device. Adopting the RF algorithm with 90 trees, results show a global accuracy percentage of 89.59% and 89.37% for human activity and gender recognition, respectively, demonstrating that the user's gender can be easily detected from the wrist motion data. However, finding the appropriate weights by using a MOEA allows to conceal the gender of the user while HAR is not considerably affected, resulting in a global accuracy of 84.14% and 64.38% for human activity and gender recognition, respectively. Additionally, considering the set of solutions constituting a Pareto front, under some conditions instead of choosing the best balanced solution (as the overmentioned work proposes) any other of the obtained solutions could be chosen, for instance in the case that, for a specific user, activity recognition performance needs to be prioritised over privacy protection. In the future, a further development may examine the same algorithm for other private information which users would like to conceal, e.g., age. In that case, in which more than two objective functions should to be considered, the optimisation will be carried out with a many-objective evolutionary algorithms. Additionally, other sets of features extracted from the accelerometer data may be explored, for instance, by extracting deep features from the original data instead of using the handcrafted ones.

Chapter 5

Discussion and conclusions

The overall aim of this Ph.D. Thesis is to develop and test acquisition procedures in order to validate wearable devices performance and signal processing techniques with the aim of quantitatively assessing human response to external stimuli, by means of physiological signals. In particular, the studies are based on the evidence of wearable devices advantages experimented in several real health-related monitoring contexts, especially considering the ongoing shift towards the so-called Healthcare 4.0 through IoT infrastructure. Results show how, among the existing wearable devices used for human-related measurements, both biosensors and inertial ones, may be used to record human physiological data, to detect human patterns in response to external stimuli and eventually to send an alert for unusual data gathered or for promoted actions. In this perspective, the findings are significant with a specific accuracy depending on the target application, especially for future trials that may be conducted through IoT-enabled wearable devices involving a wider population, so as to approve both the efficiency of the proposed/tested approaches. However, it is confirmed that some factors may affect the data interpretation and consequently the system accuracy, such as the modality of data collection (e.g., real vs Lab), as well as the signal processing (e.g., filtering and features selection) and the device selection depending on the application targeted and user involved. Results concerning the wearable devices performance may be applied for the selection and also development of future commercial devices, dealing with appropriate specifications and technical recommendations related to the metrological requirement for the health-related monitoring, the illness preventing and the quality of life improving.

Chapter 2 is focused on the human physiological response, clearly reflected on SC signal, that can be recorded through a wrist-worn device. Results confirm that although different individuals may share emotional perception of the same audio clip, their intrinsic physiological characteristics may have significant differences due to gender, age and sweat glands density in the different locations of the body (or even due to extrinsic physiological characteristics, for example, the level of physical training and fitness conditions of the subject, which brings to a different perception on the effort requested to perform the physical activity). This statement is evident for pleasant sound, where high perception of valence and intensity corresponds to a small number of SC peaks arising during the stimulation. To have a wider fingerprint on acoustic elicitation, it is needed to extend the SC analysis until 1.00 Hz; this

larger observation band of the SC signal spectrum may avoid the loss of potentially meaningful information. Similarly, when designing an experiment with audio clips presented repeatedly, the habituation phenomenon of subjects to the acoustic stimuli may influence the SC response, by decreasing the affective reaction.
Concerning the visual stimuli, it is important to note that the results show the feasibility of detecting driver's drowsiness either with only SC signals or in combination with the HRV signal, gathered from a single wrist-worn device. This assumption is based on a population that is varied and balanced in terms of gender and age. Additionally, short-term time windows yield strong classification performance (84.1%) and are crucial for identifying short-term events like the onset of natural sleepiness. Thanks to this wearable-related approach, a good timely warning might alert the driver when abnormal variations in skin conductance are identified, for example, proposing that they take a little pause to relax. Based on the overall results discussed above about physical stimuli, the vigorous intensity class obtained the best classification performance, being associated to clear differences in the signals (e.g., the amplitude of the acceleration data), with respect to those acquired during the moderate and the sedentary activities. This fact may be influenced by how the participants perceived their PA effort while assessing the score on Borg's RPE scale. Additionally, in order to prevent features from being automatically selected by the classifier and not based on the information content, a correct weight should be assigned to each feature when classifiers are used for the distinction of PA levels exerted by a certain individual.

Results from the Chapter 3 underline how it is extremely important to properly choose the acquisition device, especially in the case of commercial wearables, depending on its metrological performance for reliable values assessment. In fact, the decisions made considering the measurement results will be reliable only if the measurement itself is adequately accurate and precise. The fundamental role of these aspects have become extremely clear during the COVID-19 related pandemic, which made individuals conscious of the importance of remote monitoring and, hence, of IoT-enabled devices and their metrological performance. Despite promising usage, experimental research, systematic testing procedures and a proper validation are needed to validate the metrological properties of BP and HR commercial wearable devices. For what concerns the correct modality to wear such devices, specifically the wrist-worn devices, the most suitable option to enhance reliable PPG sensor readings, is the *medium* tightening level. In this perspective, it could be interesting to develop custom wearable devices providing real-time alerts to the user whenever the device positioning or functioning are not correct, indicating what actions can solve the issue. Also the inclusion in workplace environment would be interesting, along with additional parameters (e.g., ambient noise, temperature and relative humidity) in order to consider the subject's conditions in a more general fashion, not limited to physiological parameters. This would provide a large view on the working conditions of people monitored through a simple wearable device, which would neither compromise the workers' capabilities or privacy, nor disturbing them in the execution of daily tasks. This way, when a too low FRI is obtained, proper actions can be undertaken or promoted in order to

improve the worker's stress perception, maybe through a mobile app suggesting what to do (e.g., changing a pose, taking a short break or starting a different activity to interrupt the routine). The reliability and accuracy of a physiological measurement system is highly affected by several choices starting from the measurement bench up to the classifier, passing through the used sensors and the following signal processing techniques used to elaborate data and obtain the final results supporting the ML decision-making actions.

Finally, as it is clear from the Chapter 4, the type of relevant features for a good classification of ADLs varies both with sensing device and classifier, denoting the importance of an appropriate choice of both hardware and software components for a monitoring system, especially when high accuracy and reliability are required, such as AAL and health-related applications, sometimes also supporting decision-making processes in the definition of therapeutic strategies. An accurate HAR system, reaching a compromise between performance and computation load, may be a valid tool in many applications. Especially for ageing people, this tool may support as long as possible independent and safe living, thus aiding the healthcare system in remotely managing decentralised care, in home environments. Moreover, by implementing a concealing strategy on-board the sensor based on the selected condition on the Pareto front, it is possible to obtain what is envisaged as a privacy-by-design approach by the regulations, hiding individual personal details and obtaining high performances on the targeted purpose.

More in general, this Thesis contributes to the advancement of scientific knowledge along two main axes. The first one deals with the usage of wearable devices to provide reliable and meaningful signals to be analysed for obtaining relevant health indicators, especially when moving in real-life applications. Indeed, some of the described studies highlight the advantages and limitations regarding some commercial wearable devices against the related reference instrument (e.g., for blood pressure measurements). Every time the device must be very carefully chosen among available devices and depending on the target application. As an example, E4 device limitations are clearly found when the device stops providing data due to motion artefacts. In this sense, an approach to detect and fill in the missing data has been proposed. Within this context, it is possible to state that frequency parameters of SC signal should be studied in a wider frequency range to provide more information.

The second axis deals with the proper selection and processing of features collected from wearable devices for AAL and health-related applications. Firstly, the need for reliable devices to have better classifiers for ADLs is confirmed and underlined. Secondly, insights are provided on what might be applied to ensure privacy of the individuals wearing those device, with a limited reduction of the classifier performances using features based on the signals collected.

Finally, this Thesis provides useful information for scientific researchers seeking for efficient and reliable wearable devices to run experiment in real-life conditions, avoiding misinterpretation of signals gathered.

Chapter 6

Appendix: Additional Research Activities

In the above Sections, diverse wearable technologies have been described according to advantages, disadvantages and applications in human daily life. However, not only wearable devices are gaining ground, but also sensor-based ambient devices and systems are increasingly perceived as beneficial in supporting users. So far, it is not known whether and to what extent these three specific sensor types are perceived and accepted differently by future users. This could lead to the risk that developed technologies are not used from the final user. Therefore, in the first Section, the importance of knowing the individual's perspective on using technologies in home environments is underlined. Then, experiences and collaborations performed abroad, namely at the University of Alicante and at R&D Nestlé Orbe, respectively, are described in the second Section. In particular, at the University, the topic of assistive technologies for the ambient assisted living was further explored, while at the company, the measurement chain (e.g., data collection setup, sensor calibration etc.) was investigated and improved for the purpose of the project.

6.1 Assistive Technologies for Ageing

According to the demographic change and an increasing ageing population, assistive lifelogging technologies are developed for supporting people to be independent and active as much as possible. In fact, lifelogging may help reaching awareness about individual's own Quality of Life (QoL), thanks to the broad spectrum of available technologies, reaching from audio- and video-based approaches [175], over ambient systems based on different sensors and technologies (such as microphones, pressure or vibration sensors) [176] to wearable technologies using acceleration and rotation rate sensors [59]. Sometimes multiple wearable and ambient devices work together to create an aggregated solution able to capture the various facets of events leading to the decrease in the perceived QoL, as associated with old age. To this end, within the International project vINCI (clinically-Validated INtegrated Support for Assistive Care and Lifestyle Improvement: the Human Link), the vINCI platform was designed as a clinically-validated ambient intelligence framework, including different system components engaged in the communication, interaction, and exploitation of sensed

data and how they interoperate over using specific open middleware and software services [177]. The added value of vINCI, with respect to other platforms focused on assisted living, is its clinically-supported validation of the algorithms enabling the detection of anomalous evolving conditions in older adults, based on the observations regarding their daily level of activity status, social behaviour, and health-related parameters. The typical workflow should include the steps in Figure 6.1.

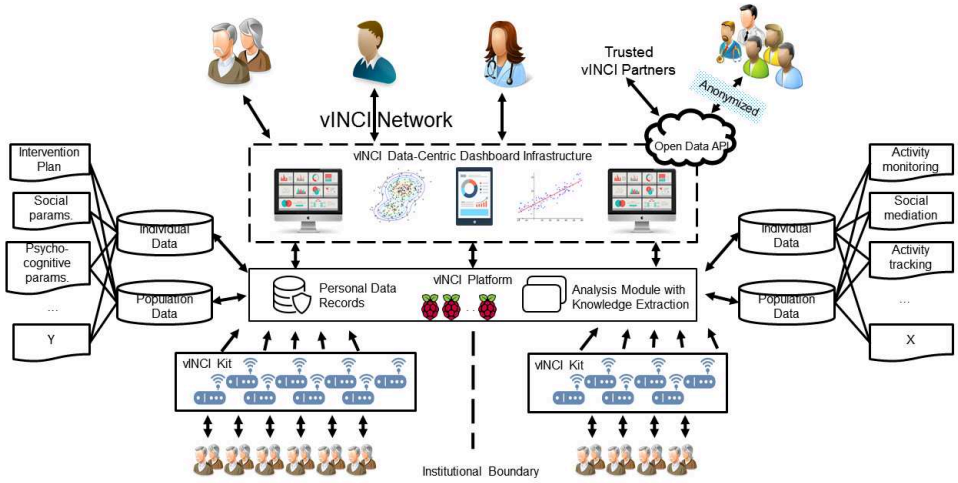


Figure 6.1: Workflow of vINCI platform.

All these technologies and systems have in common that they can be used to support people daily life, e.g., by detecting emergency situations, by identifying typical movement and behaviour patterns as well as anomalies, or by remembering functions. Beyond complex technical solutions, the usage of ambient room sensors represents a comparatively inexpensive and promising alternative. Besides the technical opportunities and functions (already discussed in the first part of the Thesis), it is currently not known whether and to what extent future users distinguish between different sensor-based lifelogging technologies; it is an open question how high future users estimate the costs for the acquisition and what costs they are willing to pay. Moreover, the current stage of research is insufficient when it comes to the ethical, legal, and social implications of using sensor- based technologies in home environments. There is a lack of a context-sensitive analysis of user acceptance and an empirical exploration of the users' requirements for the recording and managing of health-related user data referring to their privacy. Therefore, within the context of International PAAL project, namely in collaboration with the University of Aachen, in [178], the authors addressed the current gap in research and empirically investigated the users' attitudes towards acceptance and privacy in the context of sensor-based technologies in private environments. Moreover, [179] focused on the future users' perspective on different ambient sensor-based lifelogging technologies. Then, an online survey was

conducted to investigate both the perception and acceptance of participants ($N = 312$ German adults: 62.8% female and 36.9% male, 32.9 ± 18.7 years - mean ± standard deviation; $\text{Min}_{age} = 18$, $\text{Max}_{age} = 91$) about the three different sensor types as well as the assumed costs and the willingness to pay. Structure of the online survey is shown in Figure 6.2.

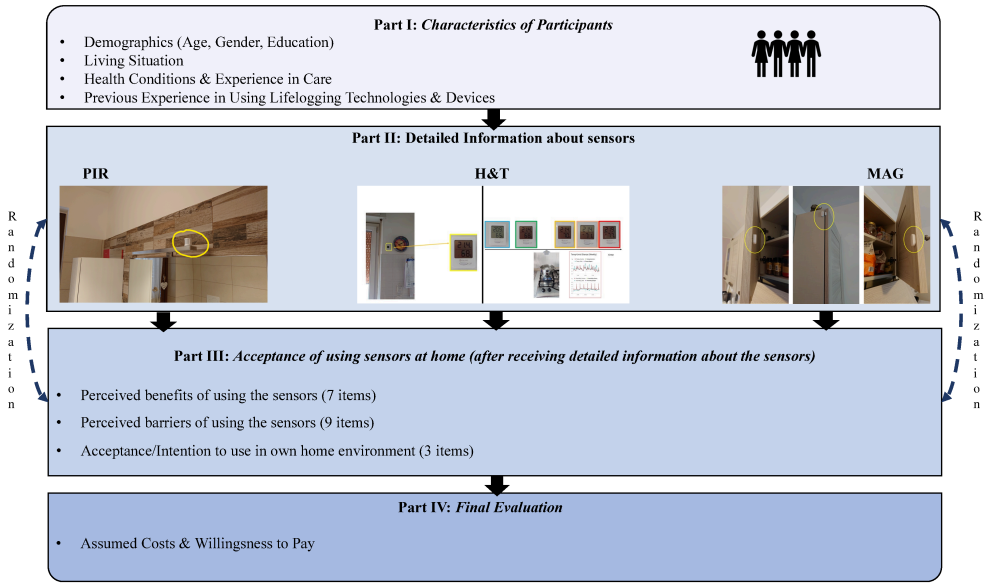


Figure 6.2: Structure of the online survey.

Figure 6.2 presents a schematic overview of the online survey's structure: part I, the participants indicated demographic characteristics, answered to questions regarding their living situation and health conditions, along with the experiences in care; part II, detailed and simplified information about three different ambient sensor-based lifelogging technologies (i.e. Passive Infrared - PIR, Humidity & Temperature sensor - H&T, and Magnetic sensors - MAG) are provided in a randomised order, using both descriptive text and illustrations; part III, the participants were asked to evaluate each time the perceived benefits using seven items, the perceived barriers using nine items, and the acceptance/intention to use the technology in the own home environment using three items; part IV, the participants were asked to assume the costs of sensor-based technologies, indicating also their willingness to pay for the acquisition of these room sensors.

Results and Discussion

Findings revealed significant sensors-related differences, especially regarding the perception of benefits, also perceiving these as useful and easy to use. Increased security was appreciated especially when using the H&T sensors, and PIR sensors were valued predominantly for the recognition of emergency situations. As to the perceived barriers, users primarily apprehended unintentionally triggered alarms and the invasion of their

privacy, which were especially true for the PIR sensors when proposed in the intimate spaces of domestic environments (e.g., bathrooms and bedrooms). Although, the participants perceived the feeling of being under surveillance in their own home environments, expressing also doubts about the reliability of the sensor-based technologies, they were also willing to provide data transparency in case of emergency (e.g., acute risk to their health). Regarding the cost estimations and the willingness to pay the sensors, on one side, the majority of the participants overestimated the costs for this technology with respect to the real one. On the other side, many respondents were willing to spend a realistic amount of money for a reasonable equipment of sensor-based technology integrated in their homes.

Results suggest that users' perceptions of personal privacy largely affect the acceptance and successful adoption of sensor-based lifelogging in home environments. Surely, these findings may enable to derive guidelines for both, the technical development and design as well as the communication and information of assisting sensor-based technologies and systems.

6.2 Experiences and collaborations abroad

6.2.1 Influence of affective environments in ageing at home

In 2021, after applying to the SHELD-ON COST Action call (CA16226 - Indoor living space improvement: Smart Habitat for the Elderly) for a Short Term Scientific Mission (STSM), I was selected to spend one month at the University of Alicante for collaborating with the Department of Computer Technology. Therefore, from the 1^{st} of March 2021 to the 31^{st} of March 2021, I was employed as a visiting Ph.D. Student, joining Prof. Florez-Revuelta's team. Due to the Coronavirus outbreak and the closure of Spanish universities, I had to interrupt my research activities at the University of Alicante on the 14^{th} of March. After getting to know that some universities in Spain were closing and more universities would have followed, Prof. Florez-Revuelta and I rescheduled the pending workplan, as described in the section below.

Introduction and aim

Literature suggests that older people prefer to live independently at home as long as possible [180, 181]. This requires providing care at home, and Ambient Assisted Living (AAL) solutions can represent a way to do so. Among other applications, AAL solutions can be used for monitoring and improving the emotional state and the well-being of an individual at home. For example, certain living environments can counteract negative psychological states (e.g., anxiety) and thus have an influence on the psychological and physical well-being. The main goal of the STSM was to investigate whether and how living environment design affects the individual emotional capacity. To do this, firstly, the physiological metrics (e.g., Skin Conductance, Heart Rate) which were measured by means of the Empatica E4 wristband device were evaluated and, secondly, relevant information from the signal were extracted to recognise individuals' emotions. The specific objectives follow:

1. To conceptualise the methodological approach for evaluating elderly's experience/emotion with different living environment designs, using non-intrusive devices;

2. To develop a system which is able to detect, process and interpret emotional state of elderly at home, by collecting individual physiological signals through wearable devices. Based on this, it would be possible to customise living environment designs to make living places more comfortable or changing depending on the emotional state.

A schematic representation is shown in Figure 6.3.

Results and discussion

Based on the discussions with the Research Group of the Department of Psychology, a conceptual framework was developed. From this, we can expect that the design of living

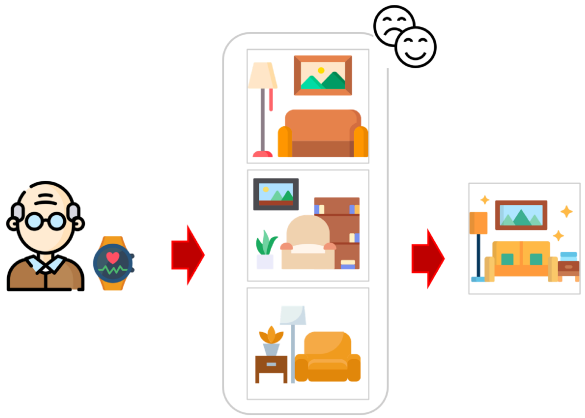


Figure 6.3: Environments and ageing: from physiological data collection to the customisation of living environment based on personal feelings.

environments and psychological states are strictly associated. Their link is crucial to increase the individual psychological and physical well-being. Correlates of emotion can be measured through non-invasive methods, detecting physiological indicators (e.g., HR and SC) for arousal to stimuli. In this respect, the underpinning pillar of the research is the data collection, employing two wearable devices (Empatica E4 and the ambulatory monitoring device, VU-AMS) for detecting the users' emotions by analysing their physiological signals. In order to develop a robust system which is able to interpret and associate the emotional state of older people with living environment design, an algorithms for processing the data was coded. Unfortunately, due to the closure of universities, data was not collected. Among the existing secondary public datasets with similar application, the following one was selected: WESAD [182]. This is a dataset for wearable stress and affect detection (neutral, stress, amusement). Such dataset features physiological and motion data, recorded from both a wrist-(Empatica E4) and a chest-worn (RespiBAN) device. Additionally, self-reports of the users from established questionnaires, are detailed. WESAD dataset allowed us to visualise data collected from E4 and to test some different processing approaches, especially for what concerns the SC signal which works as predictor of emotions. First, in order to eliminate artefacts, a pre-processing tool (i.e EDAExplorer) was used. Then, the signal was decomposed into phasic and tonic components using the cvxEDA tool. The most common features used in literature were selected and extracted from both the components. As the reliability of the device is essential for data quality, we decided to compare the two available E4s and check their performances in acquiring data. We collected a short session: 2 minutes resting and 10 minutes walking. The analyses included, first, a visual check on the GSR data to identify possible failed measurements, arrival signal delay and artefacts. Then, cross-correlation and statistical parameters were computed to determine the validity of the parameters. The results of these analyses showed that the two E4 wearables were correlated. Hence, they can be used interchangeably and their data can be merged.

6.2.2 A smart sensor system to improve the beverage systems

From the 1^{st} of March 2022 to the 31^{st} of August 2022, I was employed as an R&D Trainee by the Société des Produits Nestlé S.A, working at Nestlé System and Technology Center (NSTC) in Orbe, Switzerland. During this period, I collaborated in break-through beverage systems innovations, focusing my activities mainly on two projects: Coffee Sensor and Electronics Sustainability. Because of a patent submitted to protect the overmentioned inventions/topics, a brief and general presentation of the activities will be reported below.

Introduction and aim

Currently in the coffee market, the in-cup consistency is indirectly verified by the operator measuring the extraction time (measured in seconds – starting when the pump starts, ending when the pump ends). Time is strictly related to the coffee in-cup taste and quality. As an example, a too fast extraction can result in acidic notes, while a too slow extraction in bitter ones. In order to find the gold recipe, with a good compromise in taste, if the extraction time is shorter or longer than the target recipe originally validated in the factory, the grinder settings (i.e. coarser of finer grinded coffee) are modified to reach the target value and satisfy the taste. However, this change in grinder settings affects the extraction process, the extraction parameters and consequently the in-cup taste. For this reason, the Coffee Sensor project aims to record in real-time two extraction parameters commonly defining the coffee quality within the Coffee Brewing Control Chart [183], namely the Total Dissolved Solids (TDS - %) and Extraction Yield (EY - %), to continuously monitor and control the coffee extraction. In such a way, any deviation in the measured coffee extraction parameters can be detected with respect to the reference coffee cup, and then manually or automatically corrected. To this aim, a conductivity sensor was placed in an Out-Of-Home (OOH) R&G fully automatic coffee machines to directly measure the conductivity of coffee and estimate the TDS and EY values, through a specific algorithm. A schematic representation is shown in Figure 6.4.

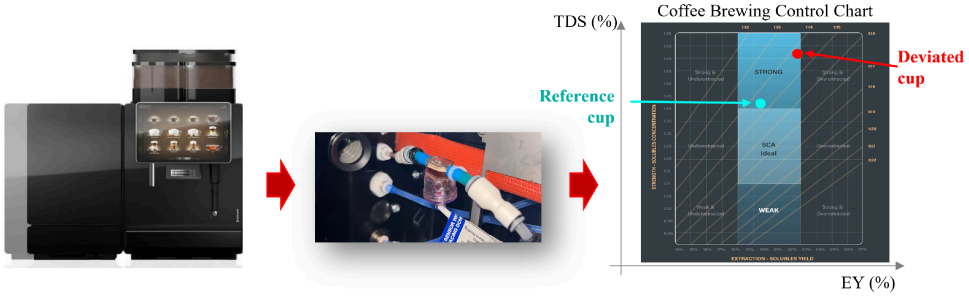


Figure 6.4: Smart sensor system: OOH machine, sensor in the machine and Coffee Brewing Control Chart.

Results and discussion

Being a new concept of sensor system for fully automatic coffee machines, knowledge/opinions from external and internal specialists about coffee and OOH machines were collected to identify the benefits reachable by applying this smart sensor system, as follows hereafter:

1. Operational costs:
    - speedup the first setup of the machine by the technician, ensuring same coffee cup characteristics everywhere;
    - optimise the number of onsite technical interventions thanks to remote monitoring of machine performance and coffee extraction quality.
2. Customer and consumer satisfaction:
    - improve the machine consistency;
    - reduce the claiming calls related to the beverage quality;
    - correlate the consumer/customer's quality feedbacks with the measurements of extraction parameters (make objective/quantify the subjective taste).
3. Sustainability:
    - improve the sustainability, optimising the number of onsite technical interventions.

After defining the business context of the project, a new standard in consistency of in-cup quality for coffee beverages was investigated. Firstly, by executing several experimental tests, the best setup (i.e., location and position in the coffee line) of the conductivity sensor was defined according to the best conductivity signal ($\mu S/cm$) collected in time. Then, the in-line conductivity readings (measured by the sensor embedded in the machine during the coffee extraction and averaged through a specific algorithm) along with the TDS (estimated through a specific algorithm) were compared with those measured in-cup after the coffee extraction. Values obtained in-cup are considered as reference values, being measured with reference lab instruments: refractometer for TDS (%) and conductivity meter for conductivity (mS/cm). Therefore, after calibrating both the in-line sensor and the reference device with a standard solution, a calibration factor was computed and integrated in the acquisition system. Such calibration phase highly improved the reading accuracy of the in-line conductivity, strongly reducing the delta difference between values read in-cup and those in-line. Once validated the measurements, the parameters affecting the perception of coffee taste were explored to understand if individuals might perceive the differences in coffee taste, depending on both coffee and machine changes. Three tests were conducted to compare two coffee machines, two coffee batches and two different espresso recipes. All the tests included a preliminary phase of sensory tasting and a second phase of analytic measurements, for investigating the correlation between the individual quality feedbacks and the lab measurements of TDS in-cup. From the overall sensory session

where 11 internal assessors were recruited, findings show that significant differences (95% confidence interval) in taste were perceived (around 10% of variation in both TDS and EY) when using two different machines equally manually calibrated, and when changing the coffee batch without re-calibrating the machine (representing a quite common scenario in real-life). In the overmentioned cases, the measurements of TDS and EY parameters would support the manual calibration phase indicating the exact target recipe to reach, and would also alert the operator for potential deviations due to re-calibration needed. Definitely, in order to reduce the taste difference between two coffee cups, it is important to include the TDS and EY measurements. An additional note can be reported for the third test, affirming that changes in coffee dosage are perceived more in flavour than in texture. At this point, data acquisition during coffee extraction was conducted to enlarge the dataset that will be used in the next steps to automatically correct the potential deviation detected in TDS from one-cup to another one.

Regarding the Electronics sustainability project, the activity was focused on implementing a strategy and identifying potential solutions to reduce the impact of electronics on the climate change. Firstly, the impact of electronics (i.e. for manufacturing and whole life of coffee machine - including power consumption) with respect to the overall impact of coffee machine was separately estimated for simple and complex coffee machines. Then, after identifying the most impactful electronics component in terms of CO_2, several innovative ideas were proposed to re-design such component, reducing the electronic impact on climate change for the next years.

Bibliography

[1] W. M. Association *et al.*, "Declaration of helsinki. ethical principles for medical research involving human subjects," *Jahrbuch Für Wissenschaft Und Ethik*, vol. 14, no. 1, pp. 233–238, 2009.

[2] V. Săndulescu, S. Pușcoci, M. Petre, M. Dumitrache, V. Bota, and A. Gîrlea, "mhealth application for remote health monitoring useful during the covid 19 pandemic," in *2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 1–6, 2021.

[3] L. P. Malasinghe, N. Ramzan, and K. Dahal, "Remote patient monitoring: a comprehensive study," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 1, pp. 57–76, 2019.

[4] G. Quer, J. M. Radin, M. Gadaleta, K. Baca-Motes, L. Ariniello, E. Ramos, V. Kheterpal, E. J. Topol, and S. R. Steinhubl, "Wearable sensor data and self-reported symptoms for covid-19 detection," *Nature Medicine*, vol. 27, no. 1, pp. 73–77, 2021.

[5] G. Coulby, A. Clear, O. Jones, F. Young, S. Stuart, and A. Godfrey, "Towards remote healthcare monitoring using accessible iot technology: state-of-the-art, insights and experimental design," *Biomedical Engineering Online*, vol. 19, no. 1, pp. 1–24, 2020.

[6] Y. S. Can, D. Gokay, D. R. Kılıç, D. Ekiz, N. Chalabianloo, and C. Ersoy, "How laboratory experiments can be exploited for monitoring stress in the wild: A bridge between laboratory and daily life," *Sensors*, vol. 20, p. 838, Feb 2020.

[7] K. Monteiro, É. Silva, É. Remigio, G. L. Santos, and P. T. Endo, "Internet of medical things (iomt) applications in e-health systems context," in *Emerging Technologies in Biomedical Engineering and Sustainable TeleMedicine*, pp. 1–12, Springer, 2021.

[8] G. Cosoli, G. Iadarola, A. Poli, and S. Spinsante, "Learning classifiers for analysis of blood volume pulse signals in iot-enabled systems," in *2021 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4.0&IoT)*, pp. 307–312, 2021.

[9] G. Cosoli, A. Poli, S. Spinsante, and L. Scalise, "The importance of physiological data variability in wearable devices for digital health applications," *ACTA IMEKO*, vol. 11, no. 2, 2022.

[10] C. M. Ranieri, S. MacLeod, M. Dragone, P. A. Vargas, and R. A. F. Romero, "Activity recognition for ambient assisted living with videos, inertial units and ambient sensors," *Sensors*, vol. 21, no. 3, 2021.

[11] M. L. Hoang, M. Carratù, V. Paciello, and A. Pietrosanto, "Body temperature—indoor condition monitor and activity recognition by MEMS accelerometer based on IoT-alert system for people in quarantine due to COVID-19," *Sensors*, vol. 21, no. 7, 2021.

[12] S. Farivar, M. Abouzahra, and M. Ghasemaghaei, "Wearable device adoption among older adults: A mixed-methods study," *International Journal of Information Management*, vol. 55, p. 102209, 2020.

[13] S. A. Mansi, G. Barone, C. Forzano, I. Pigliautile, M. Ferrara, A. L. Pisello, and M. Arnesano, "Measuring human physiological indices for thermal comfort assessment through wearable devices: A review," *Measurement*, vol. 183, p. 109872, 2021.

[14] D.-R. Izdrui, M. G. Hagan, O. Geman, O. Postolache, and R. Alexandre, "Chapter 11 - smart sensing systems for in-home health status and emotional well-being monitoring during covid-19," in *Biomedical Engineering Tools for Management for Patients with COVID-19* (V. E. Balas, O. Geman, G. Wang, M. Arif, and O. Postolache, eds.), pp. 173–186, Academic Press, 2021.

[15] F. John Dian, R. Vahidnia, and A. Rahmati, "Wearables and the internet of things (iot), applications, opportunities, and challenges: A survey," *IEEE Access*, vol. 8, pp. 69200–69211, 2020.

[16] S. Casaccia, G. Revel, G. Cosoli, and L. Scalise, "Assessment of domestic well-being: from perception to measurement," *accepted for publication, IEEE International Instrumentation and Measurement Magazine*, vol. 24, no. 6, 2021.

[17] A. Culić, S. Nižetić, P. Šolić, T. Perković, and V. Čongradac, "Smart monitoring technologies for personal thermal comfort: A review," *Journal of Cleaner Production*, vol. 312, p. 127685, 2021.

[18] A. Leonidis, M. Korozi, E. Sykianaki, E. Tsolakou, V. Kouroumalis, D. Ioannidi, A. Stavridakis, M. Antona, and C. Stephanidis, "Improving stress management and sleep hygiene in intelligent homes," *Sensors*, vol. 21, no. 7, 2021.

[19] M. Garcia-Constantino, A. Konios, M. A. Mustafa, C. Nugent, and G. Morrison, "Ambient and wearable sensor fusion for abnormal behaviour detection in activities of daily living," in *2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pp. 1–6, 2020.

[20] E. O. Polat, M. M. Cetin, A. F. Tabak, E. Bilget Güven, B. Uysal, T. Arsan, A. Kabbani, H. Hamed, and S. B. Gül, "Transducer technologies for biosensors and their wearable applications," *Biosensors*, vol. 12, no. 6, 2022.

[21] Y. Wang, F. Miao, Q. An, Z. Liu, C. Chen, and Y. Li, "Wearable multimodal vital sign monitoring sensor with fully integrated analog front end," *IEEE Sensors Journal*, vol. 22, no. 13, pp. 13462–13471, 2022.

[22] S. Balli, E. A. Sağbaş, and M. Peker, "Human activity recognition from smart watch sensor data using a hybrid of principal component analysis and random forest algorithm," *Measurement and Control*, vol. 52, no. 1-2, pp. 37–45, 2019.

[23] J. G. Habets, C. Herff, P. L. Kubben, M. L. Kuijf, Y. Temel, L. J. Evers, B. R. Bloem, P. A. Starr, R. Gilron, and S. Little, "Rapid dynamic naturalistic monitoring of bradykinesia in parkinson's disease using a wrist-worn accelerometer," *Sensors*, vol. 21, no. 23, p. 7876, 2021.

[24] G. Schifino, V. Cimolin, M. Pau, M. J. da Cunha, B. Leban, M. Porta, M. Galli, and A. Souza Pagnussat, "Functional electrical stimulation for foot drop in post-stroke people: Quantitative effects on step-to-step symmetry of gait using a wearable inertial sensor," *Sensors*, vol. 21, no. 3, p. 921, 2021.

[25] L. Scalise and G. Cosoli, "Wearables for health and fitness: Measurement characteristics and accuracy," in *2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, (Houston, TX, USA), pp. 1–6, 2018.

[26] G. Cosoli, S. Spinsante, and L. Scalise, "Wrist-worn and chest-strap wearable devices: Systematic review on accuracy and metrological characteristics," *Measurement*, vol. 159, p. 107789, 2020.

[27] G. Cosoli, S. Spinsante, F. Scardulla, L. D'Acquisto, and L. Scalise, "Wireless ecg and cardiac monitoring systems: State of the art, available commercial devices and useful electronic components," *Measurement*, vol. 177, p. 109243, 2021.

[28] A. Paderno, F. Gennarini, A. Sordi, C. Montenegro, D. Lancini, F. P. Villani, S. Moccia, and C. Piazza, "Artificial intelligence in clinical endoscopy: Insights in the field of videomics," *Frontiers in Surgery*, vol. 9, 2022.

[29] M. Ortiz-Barrios, E. Järpe, M. García-Constantino, I. Cleland, C. Nugent, S. Arias-Fonseca, and N. Jaramillo-Rueda, "Predicting activity duration in smart sensing environments using synthetic data and partial least squares regression: The case of dementia patients," *Sensors*, vol. 22, p. 5410, Jul 2022.

[30] A. Chang, *The Role of Artificial Intelligence in Digital Health*, pp. 71–81. Cham: Springer International Publishing, 2020.

[31] J. Han, M. Kamber, and J. Pei, "Data mining: concepts and techniques," in *Data Mining (Third Edition)*, The Morgan Kaufmann Series in Data Management Systems, Boston: Morgan Kaufmann, third edition ed., 2012.

[32] D. J. Cooke and N. C. Krishnan, *Activity Learning*, ch. 4, pp. 41–73. John Wiley & Sons, Ltd, 2015.

[33] C. A. C. Coello, G. B. Lamont, D. A. Van Veldhuizen, *et al.*, *Evolutionary algorithms for solving multi-objective problems*, vol. 5. Springer, 2007.

[34] A. Trivedi, D. Srinivasan, K. Sanyal, and A. Ghosh, "A survey of multiobjective evolutionary algorithms based on decomposition," *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 3, pp. 440–462, 2017.

[35] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[36] M. Poh, T. Loddenkemper, N. C. Swenson, S. Goyal, J. R. Madsen, and R. W. Picard, "Continuous monitoring of electrodermal activity during epileptic seizures using a wearable sensor," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 4415–4418, 2010.

[37] Y. Wu, Y. Liu, N. Su, S. Ma, and W. Ou, "Predicting online shopping search satisfaction and user behaviors with electrodermal activity," in *Proceedings of the 26th International Conference on World Wide Web Companion*, (Republic and Canton of Geneva, CHE), p. 855–856, International World Wide Web Conferences Steering Committee, 2017.

[38] P. A. Thoits, "The sociology of emotions," *Annual Review of Sociology*, vol. 15, no. 1, pp. 317–342, 1989.

[39] R. Plutchik, "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *American scientist*, vol. 89, no. 4, pp. 344–350, 2001.

[40] S. Prasomphan, "Detecting human emotion via speech recognition by using speech spectrogram," in *2015 IEEE International Conference on Data Science and Advanced Analytics on Proceedings*, pp. 1–10, IEEE, 2015.

[41] P. Ekman, R. W. Levenson, and W. V. Friesen, "Autonomic nervous system activity distinguishes among emotions," *Science*, vol. 221, no. 4616, pp. 1208–10, 1983.

[42] A. Martínez-Rodrigo, R. Zangróniz, J. M. Pastor, and M. V. Sokolova, "Arousal level classification of the aging adult from electro-dermal activity: from hardware development to software architecture," *Pervasive and Mobile Computing*, vol. 34, pp. 46–59, 2017.

[43] J. A. Domínguez-Jiménez, K. C. Campo-Landines, J. C. Martínez-Santos, E. J. Delahoz, and S. H. Contreras-Ortiz, "A machine learning model for emotion recognition from physiological signals," *Biomedical signal processing and control*, vol. 55, p. 101646, 2020.

[44] W. Boucsein, *Electrodermal activity*. Berlin, DE: Springer Science & Business Media, 2012.

[45] B. Ultan Cowley and J. Torniainen, "A short review and primer on electrodermal activity in human computer interaction applications," *arXiv*, pp. arXiv–1608, 2016.

[46] Y. Topoglu, J. Watson, R. Suri, and H. Ayaz, *Electrodermal Activity in Ambulatory Settings: A Narrative Review of Literature*, pp. 91–102. Cham, Switzerland: Springer, 2020.

[47] H. F. Posada-Quintero and K. H. Chon, "Innovations in electrodermal activity data collection and signal processing: A systematic review," *Sensors*, vol. 20, p. 479, Jan 2020.

[48] E. Smets, P. Casale, U. Großekathöfer, B. Lamichhane, W. De Raedt, K. Bogaerts, I. Diest, and C. Van Hoof, "Comparison of machine learning techniques for psychophysiological stress detection," in *MindCare 2015: Pervasive Computing Paradigms for Mental Health*, vol. 604, (Cham, Switzerland), pp. 13–22, Springer, 2016.

[49] Y. Yue, D. Liu, S. Fu, and X. Zhou, "Heart rate and heart rate variability as classification features for mental fatigue using short-term ppg signals via smartphones instead of ecg recordings," in *2021 13th International Conference on Communication Software and Networks (ICCSN)*, pp. 370–376, 2021.

[50] T. Chen, H. Yin, X. Yuan, Y. Gu, F. Ren, and X. Sun, "Emotion recognition based on fusion of long short-term memory networks and svms," *Digital Signal Processing*, vol. 117, p. 103153, 2021.

[51] U. Pluntke, S. Gerke, A. Sridhar, J. Weiss, and B. Michel, "Evaluation and classification of physical and psychological stress in firefighters using heart rate variability," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2207–2212, 2019.

[52] P. Jariwala and K. P. Jadhav, "Remote detection of arrhythmias using apple watch: A useful wearable during covid-19 pandemic," *IHJ Cardiovascular Case Reports (CVCR)*, vol. 5, no. 2, pp. 119–122, 2021.

[53] R. A. de Heus, C. Tzourio, E. J. L. Lee, M. Opozda, A. D. Vincent, K. J. Anstey, A. Hofman, K. Kario, S. Lattanzi, L. J. Launer, *et al.*, "Association between blood pressure variability with dementia and cognitive impairment: A systematic review and meta-analysis," *Hypertension*, pp. HYPERTENSIONAHA–121, 2021.

[54] L. Yu, S. Tao, W. Gao, and L. Yu, "Self-monitoring method for improving health-related quality of life: data acquisition, monitoring, and analysis of vital signs and diet," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 24–31, 2021.

[55] S. Saganowski, B. Perz, A. Polak, and P. Kazienko, "Emotion recognition for everyday life using physiological signals from wearables: A systematic literature review," *IEEE Transactions on Affective Computing*, pp. 1–1, 2022.

[56] J. A. Russell, "A circumplex model of affect.," *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[57] M. L. Millenson, J. L. Baldwin, L. Zipperer, and H. Singh, "Beyond dr. google: the evidence on consumer-facing digital tools for diagnosis," *Diagnosis*, vol. 5, no. 3, pp. 95–105, 2018.

[58] G. Cosoli, S. Spinsante, and L. Scalise, "Wearable devices and diagnostic apps: Beyond the borders of traditional medicine, but what about their accuracy and reliability?," *IEEE Instrumentation & Measurement Magazine*, vol. 24, no. 6, pp. 89–94, 2021.

[59] A. Poli, G. Cosoli, L. Scalise, and S. Spinsante, "Impact of wearable measurement properties and data quality on adls classification accuracy," *IEEE Sensors Journal*, vol. 21, no. 13, pp. 14221–14231, 2021.

[60] C. Sáez, N. Romero, J. A. Conejero, and J. M. García-Gómez, "Potential limitations in covid-19 machine learning due to data source variability: A case study in the ncov2019 dataset," *Journal of the American Medical Informatics Association*, vol. 28, no. 2, pp. 360–364, 2021.

[61] J. A. Behar, C. Liu, Y. Zigel, P. Laguna, and G. D. Clifford, "Editorial on remote health monitoring: from chronic diseases to pandemics," *Physiological Measurement*, vol. 41, no. 10, p. 100401, 2020.

[62] S. Slim, A. Atia, M. M.A., and M.-S. Mostafa, "Survey on human activity recognition based on acceleration data," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, Jan. 2019. Accessed on: May 21, 2020, DOI: 10.14569/IJACSA.2019.0100311, [Online].

[63] Empatica Inc., MI, IT, *E4 WristBand from Empatica User's Manual*, 2018.

[64] A. Poli, A. Brocanelli, S. Cecchi, S. Orcioni, and S. Spinsante, "Preliminary results of iot-enabled eda-based analysis of physiological response to acoustic stimuli," in *International Conference on IoT Technologies for HealthCare*, pp. 124–136, Springer, 2020.

[65] A. Amidei, A. Poli, G. Iadarola, F. Tramarin, P. Pavan, S. Spinsante, and L. Rovati, "Driver drowsiness detection based on variation of skin conductance from wearable device," in *2022 IEEE International Workshop on Metrology for Automotive (MetroAutomotive)*, pp. 94–98, 2022.

[66] A. Poli, A. Amidei, S. Benatti, G. Iadarola, F. Tramarin, L. Rovati, P. Pavan, and S. Spinsante, "Exploiting blood volume pulse and skin conductance for

driver drowsiness detection," in *accepted in 9th EAI International Conference on IoT Technologies for HealthCare (EAI HealthyIoT)*.

[67] S. Cecchi, A. Piersanti, A. Poli, and S. Spinsante, "Physical stimuli and emotions: Eda features analysis from a wrist-worn measurement sensor," in *2020 IEEE 25th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, pp. 1–6, IEEE, 2020.

[68] G. Iadarola, A. Poli, and S. Spinsante, "Analysis of galvanic skin response to acoustic stimuli by wearable devices," in *2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 1–6, 2021.

[69] W. Yang, K. Makita, T. Nakao, N. Kanayama, M. Machizawa, T. Sasaoka, A. Sugata, R. Kobayashi, H. Ryosuke, S. Yamawaki, M. Iwanaga, and M. Miyatani, "Affective auditory stimulus database: An expanded version of the international affective digitized sounds (IADS-E)," *Behavior Research Methods*, vol. 50, pp. 1415–1429, 2018.

[70] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49 – 59, 1994.

[71] H. F. Posada-Quintero, J. P. Florian, A. D. Orjuela-Cañón, and K. H. Chon, "Electrodermal activity is sensitive to cognitive stress under water," *Frontiers in Physiology*, vol. 8, p. 1128, Jan 2018.

[72] C. B. G. Pozzi, A. Sarti, "Music emotion detection. a framework based on electrodermal activities."

[73] N. Saffaryazdi, Y. Goonesekera, N. Saffaryazdi, N. D. Hailemariam, E. G. Temesgen, S. Nanayakkara, E. Broadbent, and M. Billinghurst, "Emotion recognition in conversations using brain and physiological signals," in *27th International Conference on Intelligent User Interfaces*, IUI '22, (New York, NY, USA), p. 229–242, Association for Computing Machinery, 2022.

[74] D. Williams, C.-Y. Wu, V. Hodge, D. Murphy, and P. Cowling, "A psychometric evaluation of emotional responses to horror music," in *Audio Engineering Society: 146th International Pro Audio Convention* (AES, ed.), 2019.

[75] X. Hu, F. Li, and T.-D. J. Ng, "On the relationships between music-induced emotion and physiological signals," in *19th International Society for Music Information Retrieval Conference*, 2018.

[76] D. Ayata, Y. Yaslan, and M. E. Kamasak, "Emotion based music recommendation system using wearable physiological sensors," *IEEE Transactions on Consumer Electronics*, vol. 64, no. 2, pp. 196–203, 2018.

[77] A. Joshi and R. Kiran, "Gauging the effectiveness of music and yoga for reducing stress among engineering students: An investigation based on galvanic skin response," *Work*, vol. 65, pp. 1–8, 02 2020.

[78] G. Iadarola, A. Poli, and S. Spinsante, "Reconstruction of galvanic skin response peaks via sparse representation," in *2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 1–6, 2021.

[79] iMotions, "Galvanic Skin Response (GSR): The Complete Pocket Guide." Last accessed Sept 2022.

[80] A. Gautam, N. Simoes-Capela, G. Schiavone, A. Acharyya, W. de Raedt, and C. Van Hoof, "A data driven empirical iterative algorithm for gsr signal pre-processing," in *2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 1162–1166, IEEE, 2018.

[81] M. Malik, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *Circulation*, vol. 93, pp. 1043–1065, Mar 1996.

[82] H. F. Posada-Quintero, J. P. Florian, A. D. Orjuela-Cañón, T. Aljama-Corrales, S. Charleston-Villalobos, and K. H. Chon, "Power spectral density analysis of electrodermal activity for sympathetic function assessment," *Annals of biomedical engineering*, vol. 44, no. 10, pp. 3124–3135, 2016.

[83] Z. Visnovcova, L. Bona Olexova, N. Sekaninova, I. Ondrejka, I. Hrtanek, D. Cesnekova, S. Kelcikova, I. Farsky, and I. Tonhajzerova, "Spectral and nonlinear analysis of electrodermal activity in adolescent anorexia nervosa," *Applied Sciences*, vol. 10, p. 4514, Jun 2020.

[84] H. Posada-Quintero, N. Reljin, C. Mills, I. Mills, J. Florian, J. VanHeest, and K. Chon, "Time-varying analysis of electrodermal activity during exercise," *PLOS ONE*, vol. 13, p. e0198328, Jun 2018.

[85] E. Gatti, E. Calzolari, E. Maggioni, and M. Obrist, "Emotional ratings and skin conductance response to visual, auditory and haptic stimuli," *Sci Data*, vol. 5, p. 180120, Jun 2018.

[86] M. K. Uhrig, N. Trautmann, U. Baumgärtner, R.-D. Treede, F. Henrich, W. Hiller, and S. Marschall, "Emotion elicitation: A comparison of pictures and films," *Frontiers in Psychology*, vol. 7, no. 180, 2016.

[87] J. Domínguez-Jiménez, K. Campo-Landines, J. Martínez-Santos, E. Delahoz, and S. Contreras-Ortiz, "A machine learning model for emotion recognition from physiological signals," *Biomedical Signal Processing and Control*, vol. 55, p. 101646, 2020.

[88] B. Zhao, Z. Wang, Z. Yu, and B. Guo, "Emotionsense: Emotion recognition based on wearable wristband," in *2018 IEEE SmartWorld, Ubiquitous Intelligence*

Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation, pp. 346–355, IEEE, 2018.

[89] M. M. Bradley and P. J. Lang, "The international affective picture system (IAPS) in the study of emotion and attention," *Series in affective science. Handbook of emotion elicitation and assessment*, pp. 29–46, 2007.

[90] T. Dumitriu, C. Cîmpanu, F. Ungureanu, and V. Manta, "Experimental analysis of emotion classification techniques," in *2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing*, pp. 63–70, IEEE, 2018.

[91] B. Myroniv, C.-W. Wu, Y. Ren, A. Christian, E. Bajo, and Y.-c. Tseng, "Analyzing user emotions via physiology signals," *Data Science and Pattern Recognition*, vol. 2, 12 2017.

[92] "Health topics hypertension."

[93] M. Q. Khan and S. Lee, "A comprehensive survey of driving monitoring and assistance systems," *Sensors*, vol. 19, no. 11, p. 2574, 2019.

[94] A. Amidei, P. G. Fallica, S. Conoci, and P. Pavan, "Validating photoplethysmography (ppg) data for driver drowsiness detection," in *2021 IEEE International Workshop on Metrology for Automotive (MetroAutomotive)*, pp. 147–151, IEEE, 2021.

[95] "Steer: Wearable device that will not let you fall asleep. available online:."

[96] "Stopsleep: The best solution against drowsiness available online:."

[97] A. A. Saleem, H. U. R. Siddiqui, M. A. Raza, F. Rustam, S. Dudley, and I. Ashraf, "A systematic review of physiological signals based driver drowsiness detection systems," *Cognitive Neurodynamics*, pp. 1–31, 2022.

[98] B.-L. Lee, B.-G. Lee, and W.-Y. Chung, "Standalone wearable driver drowsiness detection system in a smartwatch," *IEEE Sensors journal*, vol. 16, no. 13, pp. 5444–5451, 2016.

[99] B.-G. Lee, B.-L. Lee, and W.-Y. Chung, "Smartwatch-based driver alertness monitoring with wearable motion and physiological sensor," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6126–6129, IEEE, 2015.

[100] L. B. Leng, L. B. Giin, and W.-Y. Chung, "Wearable driver drowsiness detection system based on biomedical and motion sensors," in *2015 IEEE SENSORS*, pp. 1–4, IEEE, 2015.

[101] M. Choi, G. Koo, M. Seo, and S. W. Kim, "Wearable device-based system to monitor a driver's stress, fatigue, and drowsiness," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 3, pp. 634–645, 2017.

[102] T. Åkerstedt and M. Gillberg, "Subjective and objective sleepiness in the active individual," *International journal of neuroscience*, vol. 52, no. 1-2, pp. 29–37, 1990.

[103] W. Chen, N. Jaques, S. Taylor, A. Sano, S. Fedor, and R. W. Picard, "Wavelet-based motion artifact removal for electrodermal activity," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6223–6226, IEEE, 2015.

[104] J. Shukla, M. Barreda-Ángeles, J. Oliver, and D. Puig, "Efficient wavelet-based artifact removal for electrodermal activity in real-world applications," *Biomedical Signal Processing and Control*, vol. 42, pp. 45–52, 04 2018.

[105] A. Islam, J. Ma, T. Gedeon, M. Z. Hossain, and Y.-H. Liu, "Measuring User Responses to Driving Simulators: A Galvanic Skin Response Based Study," in *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pp. 33–337, 2019.

[106] S. Mehta, P. Mishra, A. J. Bhatt, and P. Agarwal, "AD3S: Advanced Driver Drowsiness Detection System using Machine Learning," in *2019 Fifth International Conference on Image Information Processing (ICIIP)*, pp. 108–113, 2019.

[107] H. Iwamoto, K. Hori, K. Fujiwara, and M. Kano, "Real-driving-implementable drowsy driving detection method using heart rate variability based on long short-term memory and autoencoder," *IFAC-PapersOnLine*, vol. 54, no. 15, pp. 526–531, 2021. 11th IFAC Symposium on Biological and Medical Systems BMS 2021.

[108] K. Oweis, H. Quteishat, M. Zgoul, and A. Haddad, "A study on the effect of sports on academic stress using wearable galvanic skin response," in *2018 12th International Symposium on Medical Information and Communication Technology (ISMICT)*, pp. 1–6, 2018.

[109] J. Shukla, M. Barreda-Angeles, J. Oliver, G. C. Nandi, and D. Puig, "Feature extraction and selection for emotion recognition from electrodermal activity," *IEEE Transactions on Affective Computing*, pp. 1–1, 2019.

[110] F. Al Machot, A. Elmachot, M. Ali, E. Al Machot, and K. Kyamakya, "A deep-learning model for subject-independent human emotion recognition using electrodermal activity sensors," *Sensors*, vol. 19, no. 7, 2019.

[111] A. Poli, V. Gabrielli, L. Ciabattoni, and S. Spinsante, "Cross-domain classification of physical activity intensity: An eda-based approach validated by wrist-measured acceleration and physiological data," *Electronics*, vol. 10, no. 17, 2021.

[112] Mayo Clinic, "Exercise intensity: How to measure it." `https://www.mayoclinic.org/healthy-lifestyle/fitness/in-depth/exercise-intensity/art-20046887`.

[113] K. Ellis, J. Kerr, S. Godbole, J. Staudenmayer, and G. Lanckriet, "Hip and wrist accelerometer algorithms for free-living behavior classification," *Medicine and science in sports and exercise*, vol. 48, no. 5, p. 933, 2016.

[114] G. Borg and H. Loellgen, "Borg's perceived exertion and pain scales," *Human Kinetics*, vol. 52, 01 2001.

[115] A. K. Chowdhury, D. Tjondronegoro, V. Chandran, J. Zhang, and S. G. Trost, "Prediction of relative physical activity intensity using multimodal sensing of physiological data," *Sensors*, vol. 19, no. 20, 2019.

[116] D. Girardi, F. Lanubile, and N. Novielli, "Emotion detection using noninvasive low cost sensors," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 125–130, 2017.

[117] G. Cosoli, A. Poli, L. Scalise, and S. Spinsante, "Heart rate variability analysis with wearable devices: Influence of artifact correction method on classification accuracy for emotion recognition," in *2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 1–6, 2021.

[118] 55th World Health Assembly, "Diet, physical activity and health: report by the secretariat," *Institutional Repository for Information Sharing*, 2002.

[119] G. Cosoli, A. Poli, L. Antognoli, S. Spinsante, and L. Scalise, "What is my heart rate right now? comparing data from different devices," in *2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 1–6, 2022.

[120] A. Poli, G. Cosoli, G. Iadarola, S. Spinsante, and L. Scalise, "Feasibility of blood pressure measurement through wearable devices: Analysis of smartwatches performance," in *2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 1–6, 2022.

[121] A. Poli, G. Cosoli, L. Verdenelli, F. Scardulla, L. D'Acquisto, S. Spinsante, and L. Scalise, "Diy wrist-worn device for physiological monitoring: Metrological evaluation at different band tightening levels," in *IoT Technologies for Health Care* (S. Spinsante, B. Silva, and R. Goleva, eds.), (Cham), pp. 214–229, Springer International Publishing, 2022.

[122] S. Chakrabarti, N. Biswas, L. D. Jones, S. Kesari, and S. Ashili, "Smart consumer wearables as digital diagnostic tools: A review," *Diagnostics*, vol. 12, no. 9, 2022.

[123] P. Bonnafoux, "Auscultatory and oscillometric methods of ambulatory blood pressure monitoring, advantages and limits: a technical point of view.," *Blood pressure monitoring*, vol. 1, no. 3, pp. 181–185, 1996.

[124] G. Geršak, M. Schiebl, M. Nawotka, E. Jugo, M. do Céu Ferreira, A. Duffy, D. M. Rosu, P. Pavlásek, V. Sedlák, and D. Pražák, "Physiology-based patient simulator for blood pressure meter testing," *Measurement: Sensors*, vol. 18, p. 100260, 2021.

[125] N. Agham and U. Chaskar, "Prevalent approach of learning based cuffless blood pressure measurement system for continuous health-care monitoring," in *2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 1–5, 2019.

[126] R. K. V., N. P.M., J. Joseph, and M. Sivaprakasam, "Methodological and measurement concerns of local pulse wave velocity assessment," in *2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 1–6, 2019.

[127] A. Chakraborty, D. Goswami, J. Mukhopadhyay, and S. Chakrabarti, "Measurement of arterial blood pressure through single-site acquisition of photoplethysmograph signal," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, 2021.

[128] W. M. Association, "World medical association declaration of helsinki: ethical principles for medical research involving human subjects," *JAMA*, vol. 310, p. 2191–2194, Nov 2013.

[129] M. Zhou and N. Selvaraj, "Heart rate monitoring using sparse spectral curve tracing," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pp. 5347–5352, 2020.

[130] T. G. Pickering, J. E. Hall, L. J. Appel, B. E. Falkner, J. Graves, M. N. Hill, D. W. Jones, T. Kurtz, S. G. Sheps, and E. J. Roccella, "Recommendations for blood pressure measurement in humans and experimental animals: part 1: blood pressure measurement in humans: a statement for professionals from the subcommittee of professional and public education of the american heart association council on high blood pressure research," *Hypertension*, vol. 45, no. 1, pp. 142–161, 2005.

[131] "UA-767 plus BT-Ci Telemedicine Medical Products A&D," Last accessed 01-11-2022. https://www.aandd.jp/products/medical/bluetooth/ua767pbt_ci.html.

[132] "Galaxy Watch3 Bluetooth Samsung Italia," Last accessed 01-11-2022. https://www.samsung.com/it/watches/galaxy-watch/galaxy-watch3-45mm-mystic-black-sm-r840nzkaeub/.

[133] "ASUS VivoWatch SP (HC-A05) Smartwatch e Indossabili ASUS," Last accessed 01-11-2022. https://www.asus.com/us/Mobile/Wearable-Healthcare/VivoWatch/ASUS-VivoWatch-SP-HC-A05/.

[134] "HeartGuide," Last accessed 01-11-2022. https://omronhealthcare.com/products/heartguide-wearable-blood-pressure-monitor-bp8000m/.

[135] A. Vybornova, E. Polychronopoulou, A. Wurzner-Ghajarzadeh, S. Fallet, J. Sola, and G. Wuerzner, "Blood pressure from the optical aktiia bracelet: A 1-month validation study using an extended iso81060-2 protocol adapted for a cuffless wrist device," *Blood Pressure Monitoring*, vol. 30, pp. 000–000, 2021.

[136] J. Solà, A. Vybornova, S. Fallet, E. Olivero, B. D. Marco, O. Grossenbacher, N. Ignjatović, B. Ignjatovic, M. Favre-Bulle, N. Levinson, N. Siutryk, V. Chapuis, M. Bertschi, and B. Alpert, "Are cuffless devices challenged enough? design of a validation protocol for ambulatory blood pressure monitors at the wrist: the case of the aktiia bracelet," *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 4437–4440, 2020.

[137] J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, 1985.

[138] E. O'brien, B. Waeber, G. Parati, J. Staessen, and M. G. Myers, "Blood pressure measuring devices: recommendations of the european society of hypertension," *Bmj*, vol. 322, no. 7285, pp. 531–536, 2001.

[139] W. B. White, A. S. Berson, C. Robbins, M. J. Jamieson, L. M. Prisant, E. Roccella, and S. G. Sheps, "National standard for measurement of resting and ambulatory blood pressures with automated sphygmomanometers.," *Hypertension*, vol. 21, no. 4, pp. 504–509, 1993.

[140] D. G. Altman and J. M. Bland, "Measurement in medicine: the analysis of method comparison studies," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 32, no. 3, pp. 307–317, 1983.

[141] H. Akoglu, "User's guide to correlation coefficients," *Turkish Journal of Emergency Medicine*, vol. 18, no. 3, pp. 91–93, 2018.

[142] G. Cosoli, S. Spinsante, and L. Scalise, "Wrist-worn and chest-strap wearable devices: Systematic review on accuracy and metrological characteristics," *Measurement*, vol. 159, p. 107789, 2020.

[143] S. Begum, M. Islam, M. U. Ahmed, and P. Funk, "K-nn based interpolation to handle artifacts for heart rate variability analysis," in *2011 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pp. 387–392, 2011.

[144] A. Bizzego and C. Furlanello, "Dbd-rco: Derivative based detection and reverse combinatorial optimization to improve heart beat detection for wearable devices," 03 2017.

[145] J. Rand, A. Hoover, S. Fishel, J. Moss, J. Pappas, and E. Muth, "Real-time correction of heart interbeat intervals," *IEEE transactions on bio-medical engineering*, vol. 54, pp. 946–50, 06 2007.

[146] Kubios, "Biosignal Analysis and Medical Imaging Group - Kubios HRV." Last accessed Sept 2022.

[147] Z. Cheng, L. Shu, J. Xie, and C. L. P. Chen, "A novel ecg-based real-time detection method of negative emotions in wearable applications," in *2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*, pp. 296–301, 2017.

[148] R. Ahuja and A. Banga, "Mental stress detection in university students using machine learning algorithms," *Procedia Computer Science*, vol. 152, pp. 349–353, 2019.

[149] P. Climent-Pérez, Ángela M. Muñoz-Antón, A. Poli, S. Spinsante, and F. Florez-Revuelta, "Dataset of acceleration signals recorded while performing activities of daily living," *Data in Brief*, vol. 41, p. 107896, 2022.

[150] A. Poli, L. Scalise, S. Spinsante, and A. Strazza, "ADLs monitoring by accelerometer-based wearable sensors: effect of measurement device and data uncertainty on classification accuracy," in *2020 IEEE international symposium on medical measurements and applications (MeMeA)*, pp. 1–6, IEEE, 2020.

[151] S. Fioretti, M. Olivastrelli, A. Poli, S. Spinsante, and A. Strazza, "Adls detection with a wrist-worn accelerometer in uncontrolled conditions," in *International Conference on Wearables in Healthcare*, pp. 197–208, Springer, 2020.

[152] A. Poli, A. Strazza, S. Cecchi, and S. Spinsante, "Identification issues associated with the use of wearable accelerometers in lifelogging," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12207 LNCS, pp. 338–351, 2020.

[153] A. Poli, A. M. Muñoz-Antón, S. Spinsante, and F. Florez-Revuelta, "Balancing activity recognition and privacy preservation with a multi-objective evolutionary algorithm," in *Smart Objects and Technologies for Social Good* (I. M. Pires, S. Spinsante, E. Zdravevski, and P. Lameski, eds.), (Cham), pp. 3–17, Springer International Publishing, 2021.

[154] I. Khokhlov, L. Reznik, J. Cappos, and R. Bhaskar, "Design of activity recognition systems with wearable sensors," in *2018 IEEE Sensors Applications Symposium (SAS)*, pp. 1–6, 2018.

[155] I. Cleland, B. Kikhia, C. Nugent, A. Boytsov, J. Hallberg, K. Synnes, S. McClean, and D. Finlay, "Optimal placement of accelerometers for the detection of everyday activities," *Sensors (Basel)*, vol. 13, pp. 9183–200, Jul. 2013.

[156] C. Jobanputra, J. Bavishi, and N. Doshi, "Human activity recognition: A survey," *Procedia Computer Science*, vol. 155, pp. 698–703, 2019. The 16th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2019),The 14th International Conference on Future Networks and Communications (FNC-2019),The 9th International Conference on Sustainable Energy Information Technology.

[157] M. Garcia-Constantino, C. Orr, J. Synnott, C. Shewell, A. Ennis, I. Cleland, C. Nugent, J. Rafferty, G. Morrison, L. Larkham, S. McIlroy, and A. Selby, "Design and implementation of a smart home in a box to monitor the wellbeing of residents with dementia in care homes," *Frontiers in Digital Health*, vol. 3, 2021.

[158] E. Fullerton, B. Heller, and M. Munoz-Organero, "Recognizing human activity in free-living using multiple body-worn accelerometers," *IEEE Sens. J.*, vol. 17, pp. 5290–5297, Aug. 2017.

[159] G. Jiménez-Gómez, D. Navarro-Escorcia, D. Neira-Rodado, and I. Cleland, "Determination of the most relevant features to improve the performance of rf classifier in human activity recognition," in *Computer Information Systems and Industrial Management* (K. Saeed and J. Dvorský, eds.), (Cham), pp. 28–42, Springer International Publishing, 2021.

[160] H. Xu, Y. Pan, J. Li, L. Nie, and X. Xu, "Activity recognition method for home-based elderly care service based on random forest and activity similarity," *IEEE Access*, vol. 7, pp. 16217–16225, 2019.

[161] F. Kuncan, Y. Kaya, Z. Yiner, and M. Kaya, "A new approach for physical human activity recognition from sensor signals based on motif patterns and long-short term memory," *Biomedical Signal Processing and Control*, vol. 78, p. 103963, 2022.

[162] A. Poli, S. Spinsante, C. Nugent, and I. Cleland, "Improving the collection and understanding the quality of datasets for the aim of human activity recognition," in *Smart Assisted Living*, pp. 147–165, Springer, 2020.

[163] I. Cleland, M. P. Donnelly, C. D. Nugent, J. Hallberg, M. Espinilla, and M. Garcia-Constantino, "Collection of a diverse, realistic and annotated dataset for wearable activity recognition," in *Proc. PerCom Workshops*, (Athens, Greece), pp. 555–560, 2018.

[164] G. H. John, R. Kohavi, and K. Pfleger, "Irrelevant features and the subset selection problem," in *Machine Learning Proceedings 1994* (W. W. Cohen and H. Hirsh, eds.), pp. 121–129, San Francisco (CA): Morgan Kaufmann, 1994.

[165] D. Hemavathi and H. Srimathi, "Effective feature selection technique in an integrated environment using enhanced principal component analysis," *J. Ambient Intell. Human. Comput.*, pp. 1–10, Jan. 2020.

[166] S. Fan, Y. Jia, and C. Jia, "A feature selection and classification method for activity recognition based on an inertial sensing unit," *Information*, vol. 10, p. 290, Sep. 2019.

[167] F. Liu and T. Li, "A clustering k-anonymity privacy-preserving method for wearable iot devices," *Security and Communication Networks*, vol. 2018, 2018.

[168] F. Xiao, M. Lu, Y. Zhao, S. Menasria, D. Meng, S. Xie, J. Li, and C. Li, "An information-aware visualization for privacy-preserving accelerometer data sharing," *Human-centric Computing and Information Sciences*, vol. 8, no. 1, 2018.

[169] A. Boutet, C. Frindel, S. Gambs, T. Jourdan, and R. C. Ngueveu, "Dysan: Dynamically sanitizing motion sensor data against sensitive inferences through adversarial networks," in *Inria Informatics Mathematics*, pp. 1–26, INRIA, 2020.

[170] D. Bautista-Salinas, J. González, I. Méndez, and O. Mozos, "Monitoring and prediction of mood in elderly people during daily life activities," in *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, vol. 2019, pp. 6930–6934, IEEE, 2019.

[171] W. Gomaa, *Comparative Analysis of Different Approaches to Human Activity Recognition Based on Accelerometer Signals*, pp. 303–322. Springer, 2021.

[172] J. Blank and K. Deb, "Pymoo: Multi-objective optimization in python," *IEEE Access*, vol. 8, pp. 89497–89509, 2020.

[173] K. Deb, K. Sindhya, and T. Okabe, "Self-adaptive simulated binary crossover for real-parameter optimization," in *Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation*, GECCO '07, (New York, NY, USA), p. 1187–1194, Association for Computing Machinery, 2007.

[174] A. Poli, A. Strazza, S. Cecchi, and S. Spinsante, "Identification issues associated with the use of wearable accelerometers in lifelogging," in *International Conference on Human-Computer Interaction*, pp. 338–351, Springer, 2020.

[175] S. Aleksic, M. Atanasov, J. C. Agius, K. Camilleri, A. Cartolovni, P. Climent-Peerez, S. Colantonio, S. Cristina, V. Despotovic, H. K. Ekenel, E. Erakin, F. Florez-Revuelta, D. Germanese, N. Grech, S. G. Sigurðardóttir, M. Emirzeoglu, I. Iliev, M. Jovanovic, M. Kampel, W. Kearns, A. Klimczuk, L. Lambrinos, J. Lumetzberger, W. Mucha, S. Noiret, Z. Pajalic, R. R. Peerez, G. Petrova, S. Petrovica, P. Pocta, A. Poli, M. Pudane, S. Spinsante, A. A. Salah, M. J. Santofimia, A. S. Islind, L. Stoicu-Tivadar, H. Tellioglu, and A. Zgank, "State of the art of audio- and video-based solutions for aal," 2022.

[176] J. Clemente, F. Li, M. Valero, and W. Song, "Smart seismic sensing for indoor fall detection, location, and notification," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 524–532, 2020.

[177] S. Spinsante, A. Poli, J. Mongay Batalla, P. Krawiec, C. Dobre, L. Băjenaru, C. X. Mavromoustakis, C. S. Costantinou, G. Molan, A. M. Herghelegiu, *et al.*, "Clinically-validated technologies for assisted living," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–22, 2021.

[178] W. Wilkowska, J. Offermann, S. Spinsante, A. Poli, and M. Ziefle, "Analyzing technology acceptance and perception of privacy in ambient assisted living for using sensor-based technologies," *PLOS ONE*, vol. 17, pp. 1–21, 07 2022.

[179] J. Offermann, W. Wilkowska, A. Poli, S. Spinsante, and M. Ziefle, "Acceptance and preferences of using ambient sensor-based lifelogging technologies in home environments," *Sensors*, vol. 21, no. 24, 2021.

[180] S. Aleksic, M. Atanasov, J. C. Agius, K. Camilleri, A. Cartolovni, P. Climent-Peerez, S. Colantonio, S. Cristina, V. Despotovic, H. K. Ekenel, E. Erakin, F. Florez-Revuelta, D. Germanese, N. Grech, S. G. Sigurðardóttir, M. Emirzeoglu, I. Iliev, M. Jovanovic, M. Kampel, W. Kearns, A. Klimczuk, L. Lambrinos, J. Lumetzberger, W. Mucha, S. Noiret, Z. Pajalic, R. R. Peerez, G. Petrova, S. Petrovica, P. Pocta, A. Poli, M. Pudane, S. Spinsante, A. A. Salah, M. J. Santofimia, A. S. Islind, L. Stoicu-Tivadar, H. Tellioglu, and A. Zgank, "State of the art of audio- and video-based solutions for aal," 2022.

[181] A. P. Europe, "The voice of older persons at eu level." Last accessed Sept 2022.

[182] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, ICMI '18, (New York, NY, USA), p. 400–408, Association for Computing Machinery, 2018.

[183] S. C. A. (SCA), "Towards a New Brewing Chart." Last accessed Sept 2022.

Publications

My research activity during the three years of Ph.D. resulted in the following publications:

International Journals

- A. Poli, G. Cosoli, L. Scalise and S. Spinsante, "Impact of Wearable Measurement Properties and Data Quality on ADLs Classification Accuracy," in IEEE Sensors Journal, doi: 10.1109/JSEN.2020.3009368.
- Di Nardo F., Spinsante S., Pagliuca C.; Poli A., Strazza A., Agostini V., Knaflitz M. and Fioretti, S. "Variability of Muscular Recruitment in Hemiplegic Walking Assessed by EMG Analysis". Electronics 2020, 9, 1572. doi: = 10.3390/electronics9101572.
- Cosoli, G., Scalise, L., Poli, A. and Spinsante, S. "Wearable devices as a valid support for diagnostic excellence: lessons from a pandemic going forward". Health Technol. 11, 673–675 (2021). https://doi.org/10.1007/s12553-021-00540-y.
- Spinsante S., Poli, A. Mongay Batalla, J. et al. "Clinically-validated technologies for assisted living". J Ambient Intell Human Comput (2021) https://doi.org/10.1007/s12652-021- 03419-y.
- Cosoli G., Poli A., Scalise L. and Spinsante S. (2021). "Measurement of multimodal physiological signals for stimulation detection by wearable devices". Measurement, 184, 109966.
- Poli A., Gabrielli V., Ciabattoni L. and Spinsante S. "Cross-Domain Classification of Physical Activity Intensity: An EDA-Based Approach Validated by Wrist-Measured Acceleration and Physiological Data". Electronics 2021, 10, 2159. https://doi.org/10.3390/electronics10172159.
- Offermann J., Wilkowska W., Poli A., Spinsante S. and Ziefle M. "Acceptance and Preferences of Using Ambient Sensor-Based Lifelogging Technologies in Home Environments". Sensors (2021), 21, 8297 https://doi.org/10.3390/s21248297.
- P. Climent-Pérez, Á. M. Muñoz-Antón, A. Poli, S. Spinsante and F. Florez-Revuelta, "Dataset of acceleration signals recorded while performing activities of daily living", Data in Brief (2022), Volume 41, 107896, https://doi.org/10.1016/j.dib.2022.107896.

- G. Cosoli, A. Poli, S. Spinsante and L. Scalise, “The importance of physiological data variability in wearable devices for digital health applications”, Acta Imeko (2022), Vol 11, 2, http://dx.doi.org/10.21014/acta.v11i2.1135.
- Wilkowska W, Offermann J, Spinsante S, Poli A and Ziefle M (2022) “Analyzing technology acceptance and perception of privacy in ambient assisted living for using sensor-based technologies”. PLOS ONE 17(7): e0269642. https://doi.org/10.1371/journal.pone.0269642.
- W. Wilkowska, J. Offermann-van Heek, L. Colonna, F. Florez-Revuelta, P. Climent-Perez, A. Mihailidis, A. Poli, S. Spinsante and M. Ziefle, “Interdisciplinary Perspectives on Privacy Awareness in Lifelogging Technology Development”, SUBMITTED to Journal of Ambient Intelligence and Humanized Computing (2022).

International Conference Proceedings

- Poli A., L. Scalise, S. Spinsante and A. Strazza, "ADLs Monitoring by Accelerometer-Based Wearable Sensors: Effect of Measurement Device and Data Uncertainty on Classification Accuracy," 2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Bari, Italy, pp. 1-6. doi: 10.1109/MeMeA49120.2020.9137265.
- Poli A., Strazza A., Cecchi S. and Spinsante S. “Identification Issues Associated with the Use of Wearable Accelerometers in Lifelogging”. In: Gao Q., Zhou J. (eds) Human Aspects of IT for the Aged Population. Technologies, Design and User Experience. HCII 2020. Lecture Notes in Computer Science, vol 12207. Springer, Cham. https://doi.org/10.1007/978-3-030-50252-2_26.
- Cecchi S., Piersanti A., Poli A. and Spinsante S., “Physical Stimuli and Emotions: EDA Features Analysis from a Wrist-Worn Measurement Sensor”. Online. 2020 IEEE International Workshop on Computer-Aided Modeling Analysis and Design of Communication Links and Networks (CAMAD).
- Fioretti S., Olivastrelli M., Poli A., Spinsante S., Strazza A. “ADLs Detection with a Wrist-Worn Accelerometer in Uncontrolled Conditions”. In: Perego P., TaheriNejad N., Caon M. (eds) Wearables in Healthcare 2020. ICWH 2020. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 376. Springer, Cham. https://doi.org/10.1007/978-3-030-76066-3_16.
- Poli A., Brocanelli A., Cecchi S., Orcioni S., Spinsante S. “Preliminary Results of IoT-Enabled EDA-Based Analysis of Physiological Response to Acoustic Stimuli”. In: Goleva R., Garcia N.R.C., Pires I.M. (eds) IoT Technologies for HealthCare. HealthyIoT 2020. Lecture Notes of the Institute for Computer Sciences, Social

Informatics and Telecommunications Engineering, vol 360. Springer, Cham. https://doi.org/10.1007/978-3- 030-69963-5_9.

- Gioacchini L., Poli A., Cecchi S., Spinsante S. “Sensors Characterization for a Calibration-Free Connected Smart Insole for Healthy Ageing”. In: Goleva R., Garcia N.R.C., Pires I.M. (eds) IoT Technologies for HealthCare. HealthyIoT 2020. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 360. Springer, Cham. https://doi.org/10.1007/978-3-030-69963-5_3.

- G. Cosoli, A. Poli, L. Scalise and S. Spinsante, "Heart Rate Variability Analysis With Wearable Devices: Influence of Artifact Correction Method on Classification Accuracy for Emotion Recognition," 2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), pp. 1-6, doi: 10.1109/I2MTC50364.2021.9459828.

- G. Iadarola, A. Poli and S. Spinsante, "Analysis of Galvanic Skin Response to Acoustic Stimuli by Wearable Devices," 2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA), pp. 1-6, doi: 10.1109/MeMeA52024.2021.9478673.

- G. Iadarola, A. Poli and S. Spinsante, "Reconstruction of Galvanic Skin Response Peaks via Sparse Representation," 2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), pp. 1-6, doi: 10.1109/I2MTC50364.2021.9459905.

- G. Cosoli, G. Iadarola, A. Poli and S. Spinsante, "Learning classifiers for analysis of Blood Volume Pulse signals in IoT-enabled systems," 2021 IEEE International Workshop on Metrology for Industry 4.0&IoT (MetroInd4.0&IoT), pp. 307-312, doi: 10.1109/MetroInd4.0IoT51437.2021.9488497.

- A. Poli, S. Cecchi, S. Spinsante, A. Terenzi, and F. Bettarelli, "A Preliminary Study on the Correlation between Subjective Sound Quality Perception and Physiological Parameters," AES E-Library, Paper 10495, (2021 May).

- A. Poli, A. M. Munoz-Anton, S. Spinsante and F. Florez-Revuelta “Balancing activity recognition and privacy preservation with a multi-objective evolutionary algorithm” 7th EAI International Conference on Smart Objects and Technologies for Social Good (GOODTECHS 2021), September 15-17.

- Dourou, N., Poli, A., Terenzi, A., Cecchi S. and Spinsante, S. “IoT-Enabled Analysis of Subjective Sound Quality Perception Based on Out-of-Lab Physiological Measurements”. In: Spinsante, S., Silva, B., Goleva, R. (eds) IoT Technologies for Health Care. HealthyIoT 2021. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 432. Springer, Cham. https://doi.org/10.1007/978-3-030-99197-5_13.

- Casaccia, F., Iadarola, G., Poli, A. and Spinsante, S. “CS-Based Decomposition of Acoustic Stimuli-Driven GSR Peaks Sensed by an IoT-Enabled Wearable Device”. In: Spinsante, S., Silva, B., Goleva, R. (eds) IoT Technologies for Health Care. HealthyIoT 2021. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 432. Springer, Cham. https://doi.org/10.1007/978-3-030-99197-5_14.

- Poli, A. et al. “DIY Wrist-Worn Device for Physiological Monitoring: Metrological Evaluation at Different Band Tightening Levels”. In: Spinsante, S., Silva, B., Goleva, R. (eds) IoT Technologies for Health Care. HealthyIoT 2021. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 432. Springer, Cham. https://doi.org/10.1007/978-3-030-99197-5_17.

- G. Iadarola, A. Poli and S. Spinsante, "Compressed Sensing of Skin Conductance Level for IoT-based wearable sensors," 2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), pp. 1-6, doi: 10.1109/I2MTC48687.2022.9806516.

- G. Cosoli, A. Poli, L. Antognoli, S. Spinsante and L. Scalise, "What is my heart rate right now? Comparing data from different devices," 2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), pp. 1-6, doi: 10.1109/I2MTC48687.2022.9806601.

- A. Poli, G. Cosoli, G. Iadarola, S. Spinsante and L. Scalise, "Feasibility of Blood Pressure Measurement through Wearable Devices: Analysis of Smartwatches Performance," 2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA), pp. 1-6, doi: 10.1109/MeMeA54994.2022.9856533.

- A. Amidei, A. Poli, G. Iadarola, F. Tramarin, P. Pavan, S. Spinsante and L. Rovati "Driver Drowsiness Detection based on Variation of Skin Conductance from Wearable Device," 2022 IEEE International Workshop on Metrology for Automotive (MetroAutomotive), pp. 94-98, doi: 10.1109/MetroAutomotive54295.2022.9854871.

- A. Poli, A. Amidei, S. Benatti, G. Iadarola, F. Tramarin, L. Rovati, P. Pavan and S. Spinsante “Exploiting Blood Volume Pulse and Skin Conductance for Driver Drowsiness Detection”, Accepted to HealthyIoT 2022.

National Conference Proceedings

- S. Spinsante, A. Poli and L. Scalise, “Impact Of The Measurement Device On Human Activity Machine Learning-Based Classification”, Memoria breve per GMEE 2020, Linea di ricerca A2 – Measurement signals and data, ISBN 978-88-9456-130-2, pp. 99-100.

- A. Poli and S. Spinsante, “Frequency Domain Analysis Of The Galvanic Skin Response For The Evaluation Of Sound Quality Perception”, Memoria breve per GMEE 2021, ISBN 9788894561326.
- G. Cosoli, A. Poli, S. Spinsante and L. Scalise, “Wearable Devices And Digital Health: The Role Of Data Variability In The Perspective Of Artificial Intelligence Algorithms”, Memoria breve per GMMT 2021, ISBN 9788894561326.
- A. Poli and S. Spinsante, “Classification of Physical Activity Intensity based on Eda and Validated by Wrist-measured Acceleration and Physiological Data”, Memoria breve per GMEE 2022, Linea di ricerca A3 – Measurements for quality and innovation management, ISBN 9788894561357, pp. 209-210.

Other Publications

- Pau Climent-Pérez, Ángela María Muñoz-Antón, Angelica Poli, Susanna Spinsante, & Francisco Florez-Revuelta. (2021). PAAL ADL Accelerometry dataset v2.0 (2.0) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.5785955.
- Aleksic, Slavisa, et al. "State of the art of audio-and video-based solutions for AAL" arXiv preprint arXiv:2207.01487 (2022) (White paper within COST Action).