1 2 9 0

UNIVERSIDADE D
COIMBRA

José Marcelo da Silva Lopes Fernandes

PEOPLE4.0 - IOT SENSING AND HUMAN-IN-THE-LOOP INTERACTIONS IN SMART ENVIRONMENTS

Tese no âmbito do Doutoramento em Engenharia Informática, Especialidade em Arquiteturas, Redes e Cibersegurança orientada pelo Professor Doutor Jorge Sá Silva, e pelo Professor Doutor Fernando Boavida, e apresentada ao Departamento de Engenharia Informática da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Outubro de 2022

Department of Informatics Engineering
Faculty of Sciences and Technology
University of Coimbra

People4.0 - IoT Sensing And Human-In-The-Loop Interactions In Smart Environments

José Marcelo da Silva Lopes Fernandes

PhD in Informatics Engineering
PhD Thesis submitted to the University of Coimbra

Advised by Prof. Dr. Jorge Sá Silva
and Prof. Dr. Fernando Boavida

October, 2022

Departamento de Engenharia Informática
Faculdade de Ciências e Tecnologia
Universidade de Coimbra

People4.0 - Sensoriamento IoT e interações com o Humano no loop em ambientes inteligentes

José Marcelo da Silva Lopes Fernandes

Doutoramento em Engenharia Informática
Tese de Doutoramento apresentada à Universidade de Coimbra

Orientado pelo Prof. Dr. Jorge Sá Silva
e pelo Prof. Dr. Fernando Boavida

Outubro, 2022

This work was partially supported by the Portuguese Foundation for Science and Technology (FCT) under the PhD grant SFRH/BD/147371/2019; Additionally, through projects: PTDC/EEI-SCR/2072/2014 and CENTRO-01-0246-FEDER-000014 both co-financed by COMPETE 2020, Portugal 2020 - Operational Program for Competitiveness and Internationalization (POCI), European Union's ERDF (European Regional Development Fund) and the Portuguese Foundation for Science and Technology (FCT).

FCT Fundação para a Ciência e a Tecnologia COMPETE 2020 PORTUGAL 2020 UNIÃO EUROPEIA Fundo Europeu de Desenvolvimento Regional

Acknowledgements

THIS PhD thesis would not have been possible without the help and support of many people. It was a tiring and hard *road*, but one that I am grateful I take because it gave me the opportunity to know some amazing people.

First of all, I would like to thank my supervisors, Prof. Jorge Sá Silva and Prof. Fernando Boavida, for their unconditional support, patience and guidance during all these years. I am glad I had the opportunity to work with both of them, as they always supported my work in every possible way, and without them this PhD thesis would not be possible.

I would also like to extend my thanks to the Centre for Informatics and Systems of the University of Coimbra (CISUC) and especially the LCT group for allowing me to carry out my research activities in their facilities and for supporting this work. A special thanks to Prof. André Rodrigues for all of his support in the development of my research, and all the discussion of ideas that helped to enrich it. I would also like to thank Prof. Edmundo Monteiro and Prof. Marilia Curado for their warm welcome to the LCT group and for being the promoters of many of the activities that helped our group grow closer.

I would also like to thank all of my current and former colleagues from the CISUC, but especially the ones with whom I worked more closely. A special thanks to my colleagues from the lab: Duarte, Jorge, Mabe, Oscar, Ngombo, Soraya, Ricardo, Sandra, Mateus and Rui Almeida. Which have not only helped me during this PhD, but also made the day–to–day more fun and enjoyable, and I am glad to call you all friends. A special thanks to my friends and the other members of the LCT *family*: David, Karima, Ricardo, Proença, Rui, Victor, Joca and Paulo. For the many sports activities, adventures and experiences that we have lived together and for the many more that I believe are still to come.

I am thankful for the unconditional support of my family, my parents and my brothers and sisters, for supporting me in any way possible. A special thanks to my sister Teresa and my brother Emanuel, for being a constant in my life, helping me and believing in me. You are without a doubt my two best friends.

Finally, a special thanks to Érica, for believing in me even when I had my doubts and for supporting me during all of this experience. Without your love and support this would not have been possible.

Abstract

TECHNOLOGY is changing the way we live and the way we interact with the world. We are witnessing advancements in several fields, from Internet services to quantum computing, personal assistants or even artificial intelligence. Internet of Things devices that surround us also increase in number every day. Furthermore, these devices have increasing capabilities, with more processing power, faster and more reliable communication mechanisms and more sensors to interact with the physical world. Internet of Things devices also allow us to create *"smarter"* systems. However, no matter how many devices and sensors make up a system, truly smart environments can only be developed by integrating humans into the overall system. Humans' actions, emotions, intents, and desires can largely affect the way systems behave and perform, and they are able to either improve a system or make it perform worse. As such, we must design our systems with the human factor in mind, not only as an end-user but, above all, as a fully integrated *"component"* of the system.

We now have at our disposal new means to perceive humans and better integrate them into emerging systems. Most humans now carry, at all times, smartphones that have several sensors, able to communicate with other devices and services, and even have the necessary processing power to make inferences about the human states. At the same time with the proliferation of social networks, humans themselves can act as sensors (i.e., social sensors). Advances in radio-frequency and image processing are also making possible the development of new unobtrusive sensing technologies, that can be used to gather information on human activity and states, without the need for any human action. Additionally, we now have interactive agents/chatbots that allow for human–like interaction of systems with humans.

In this thesis, we target all the aforementioned technology advancements. As a starting point, we identify, analyse and discuss state-of-the-art solutions for the Human–In–The–Loop–Cyber–Physical Systems paradigm. We then present a novel model for the Human–In–The–Loop–Cyber–Physical Systems paradigm, that considers humans in all phases of the cyber-physical loop. Additionally, we explore advancements in unobtrusive sensing of humans and propose a taxonomy for the classification of these novel sensing techniques. This thesis also describes concrete implementation examples of our Human–In–The–Loop paradigm. As such, we include details of our research work and associated case studies of Human–In–The–Loop–Cyber–Physical Systems, and present the respective results.

Keywords: Internet of Things, Human–In–The–Loop–Cyber–Physical Systems, Mobile Phone Sensing Systems, Unobtrusive Sensing.

Resumo

Os avanços na tecnologia estão a mudar a forma como vivemos e a forma como interagimos com o mundo. Estamos a ver o surgimento de avanços em vários campos, desde serviços de Internet, computação quântica, assistentes pessoais ou até mesmo inteligência artificial. Os dispositivos da Internet das Coisas que nos cercam também aumentam em número a cada dia que passa. Estes dispositivos são cada vez mais capazes, com mais recursos de processamento, capacidades de comunicação mais rápidas e confiáveis e com um maior número de sensores para interagir com o mundo físico. Através destes dispositivos podemos criar sistemas mais *"inteligentes"*. No entanto, os sistemas não podem ser verdadeiramente inteligentes e completos até que integrem o humano como parte do sistema. Os seres humanos e as suas ações, emoções, intenções e desejos podem afetar amplamente a maneira como os sistemas se comportam e funcionam, sendo capazes de melhorar um sistema ou torná-lo pior. Como tal, devemos projetar os nossos sistemas com o humano em mente, não apenas como usuário final, mas como um *"componente"* totalmente integrado do sistema.

Com os avanços da tecnologia temos à nossa disposição, novos meios para tentar entender o ser humano e integrá-lo melhor nos nossos sistemas. Hoje em dia a maioria dos humanos carrega, a todo momento, smartphones que possuem vários sensores, capazes de se comunicar com outros dispositivos e serviços, e que têm o poder de processamento necessário para fazer inferências sobre os estados dos humanos. Ao mesmo tempo com a proliferação das redes sociais, os próprios humanos podem atuar como sensores (i.e., sensores sociais). Avanços em radiofrequência e processamento de imagem também possibilitam o desenvolvimento de novas tecnologias de sensoriamento não obstrutivo, que podem ser usadas para captar informação dos humanos sem ser necessária qualquer interação. Adicionalmente, agora temos agentes interativos/chatbots que permitem a interação humana de sistemas com humanos.

Nesta tese, estudamos os avanços tecnológicos mencionados anteriormente e o paradigma do Humano–no–Loop dos Sistemas Ciber–Físicos. Como ponto de partida, revisamos o estado da arte de soluções já existentes para o paradigma Humano–no–Loop dos Sistemas Ciber–Físicos. Posteriormente, apresentamos um novo modelo para o paradigma Humano–no–Loop dos Sistemas Ciber–Físicos, que considera o humano em todas as fases do ciclo Ciber–Físico. Além disso, exploramos os avanços na deteção não intrusiva de humanos e propomos uma taxonomia para a classificação dessas novas técnicas de deteção. Esta tese também descrevemos exemplos concretos de implementação do nosso paradigma Humano-no-Loop. Como tal, incluímos detalhes dos casos de estudo, associados ao Humano–no–Loop dos Sistemas Ciber–Físicos, que implementados e exploramos os resultados deles.

Palavras-chave: Internet das Coisas, Humano–no–Loop dos Sistemas Ciber–Físicos, Sistemas de Sensoriamento com Smartphone, Sensoriamento não Intrusivo.

Foreword

THE work detailed in this thesis was developed at the Laboratory of Communications and Telematics (LCT) of the CISUC, in partnership with the Institute for Systems Engineering and Computers of Coimbra (INESC), within the context of the following projects and grants:

SOCIALITE - Social-Oriented Internet of Things Architecture, Solutions and Environment (PTDC/EEI-SCR/2072/2014). The project aimed at developing and exploring a generic CPS/IoT architecture that could be used to support both People2People interaction and People2Thing interaction. The idea was that communication between People2People and People2Thing would be smarter, more intelligent, transparent and unobtrusive to the user, and adapted to the context of the users. This would contribute to bridge the gap between CPS/IoT, in what the devices integration and management were concerned, and the beneficiaries of technologies.

ResiliScence 4 Covid-19 - Social Sensing & Intelligence for Forecasting Human Response in Future Covid-19 Scenarios, towards Social Systems Resilience.(Research 4 Covid-19 - Project nº 439) This project proposed a social sensors framework aimed at monitoring social perceptions of systemic risks (derived by evaluations of the perceived demands posed by the pandemic and the perceived resources to cope with these), self-reported protective behaviours against SARS-CoV-2 contagion and their predictors, and subjective and objective indicators of exposure to risks associated with SARS-CoV-2 contagion. This concept was extended in the proposed framework to include not only digital communications (e.g., mediated by social media), but also Information and Communication Technologies.

CentroAdapt - Centro de Vanguarda em Adaptações às Alterações Climáticas (CENTRO-01-0246-FEDER-000014) This project aimed at creating a platform for the transfer of scientific and technological knowledge between the entities of the R&ID system and companies, in order to enhance the adaptability to climate changes. In addition, it also aimed at promoting new business areas, with an increased potential in the country's central region so that it could distinguish itself and become a reference at an international level. This project aimed to establish the region as a national leader in adapting to climate change through the creation of new products and services resulting from R&D activities. At the same time, its mission was to establish an open and inclusive platform for innovation that incorporates the skills necessary for an efficient and effective transfer of scientific and technological knowledge to the business sector.

iFriend - Supervisión inteligente del estado de salud en personas mayores con insuficiencia renal mediante dispositivos inalámbricos (Supported by Fundación CSIC, Interreg Portugal-Espanha). This project had the goal of providing technological solutions within the reach of the majority of the population, allowing the early detection of major complications developed by individuals with complex chronic diseases, and favouring the early implementation of appropriate therapeutic measures, reducing the hospitalization rates in these individuals. This would improve not only the clinical evolution in short and medium term, as well as the quality of life. This project aimed to assess the health and general physical condition of a person and provide clues to the diagnosis of possible diseases by tracking the vital signs of the human body using the Wi-Fi network and human behaviour recorded on mobile phones, relying on recent advances in signal processing, mathematical models, artificial intelligence, and machine learning.

PhD grant - Foundation for Science and Technology (FCT) (SFRH/BD/147371/2019).

The outcome of the design, experiments, and assessments of several mechanisms carried out in the course of this thesis resulted in the following publications:

Journal papers:

- Fernandes, J., Silva, J. S., Rodrigues, A., Sinche, S., and Boavida, F. (2022b). Automatically assessing students performance with smartphone data. Submitted to IEEE Access (ISSN 2169–3536) (**Q1**)
- Fernandes, J., Silva, J. S., Rodrigues, A., Boavida, F., Gaspar, R., Godinho, C., and Francisco, R. (2022a). Social sensing and human in the loop profiling during pandemics: the vitoria application. Submitted to Pervasive Mobile Computing (ISSN 1574–1192) (**Q1**)
- Fernandes, J. M., Silva, J. S., Rodrigues, A., and Boavida, F. (2022c). A survey of approaches to unobtrusive sensing of humans. *ACM Computing Surveys (CSUR)*, 55(2):1–28. ACM New York, NY (**Q1**);
- Fernandes, J., Raposo, D., Armando, N., Sinche, S., Silva, J. S., Rodrigues, A., Pereira, V., Oliveira, H. G., Macedo, L., and Boavida, F. (2020). Isabela–a socially-aware human-in-the-loop advisor system. *Online Social Networks and Media*, 16:100060. Elsevier (**Q1**);
- Sanchez, O. T., Fernandes, J. M., Rodrigues, A., Silva, J. S., Boavida, F., Rivadeneira, J. E., de Lemos, A. V., and Raposo, D. (2022). Green bear - a lorawan-based human-in-the-loop case-study for sustainable cities. *Pervasive and Mobile Computing*, page 101701 (**Q1**);
- Armando, N., Almeida, R., Fernandes, J. M., Silva, J. S., and Boavida, F. (2021). End-to-end experimentation of a 5g vertical within the scope of blended learning. *Discover Internet of Things*, 1(1):1–12. Springer;
- Sinche, S., Hidalgo, P., Fernandes, J. M., Raposo, D., Silva, J. S., Rodrigues, A., Armando, N., and Boavida, F. (2020). Analysis of student academic performance using human-in-the-loop cyber-physical systems. In *Telecom*, volume 1, pages 18–31. MDPI;

Conference papers:

- Fernandes, J., Raposo, D., Sinche, S., Armando, N., Silva, J. S., Rodrigues, A., Macedo, L., Oliveira, H. G., and Boavida, F. (2019b). A human-in-the-loop cyber-physical approach for students performance assessment. In *Proceedings of the Fourth International Workshop on Social Sensing*, pages 36–42. ACM New York, NY;
- Fernandes, J., Raposo, D., Armando, N., Sinche, S., Silva, J. S., Rodrigues, A., Pereira, V., and Boavida, F. (2019a). An integrated approach to human-in-the-loop systems and online social sensing. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 478–483. IEEE;
- Rivadeneira, J. E., Silva, J. S., Colomo-Palacios, R., Rodrigues, A., Fernandes, J. M., and Boavida, F. (2021). A privacy-aware framework integration into a human-in-the-loop iot system. In *IEEE INFOCOM 2021-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 1–6. IEEE;
- Armando, N., Fernandes, J. M., Rodrigues, A., Silva, J. S., and Boavida, F. (2020). Exploring approaches to the management of physical, virtual, and social sensors. In *IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 954–959. IEEE;
- Armando, N., Fernandes, J., Sinche, S., Raposo, D., Silva, J. S., and Boavida, F. (2019). A unified solution for iot device management. In *2019 22nd International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pages 1–6. IEEE;
- Sinche, S., Polo, O., Raposo, D., Femandes, M., Boavida, F., Rodrigues, A., Pereira, V., and Silva, J. S. (2018). Assessing redundancy models for iot reliability. In *2018 IEEE 19th International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)*, pages 14–15. IEEE; and
- Armando, N., Raposo, D., Fernandes, M., Rodrigues, A., Silva, J. S., and Boavida, F. (2017). Wsns in fiware–towards the development of people-centric applications. In *International conference on practical applications of agents and multi-agent systems*, pages 445–456. Springer;

In parallel with the execution of the tasks related to this thesis, participation in different projects during this PhD course led to discussions and exchange of ideas that resulted in some publications listed above, which are framed within the context of this research and enriched the performed work. The cooperation underlying these papers encompassed several aspects, namely, the discussion and conception of ideas, the design and implementation of the experiments, and the analysis of data and results.

Contents

List of Figures

List of Algorithms

List of Tables

Acronyms

AP Access Point

API Application Programming Interface

AUC Area Under Curve

ANN Artificial Neural Network

BLE Bluetooth Low Energy

BR Breathing Rate

CISUC Centre for Informatics and Systems of the University of Coimbra

cm centimetres

CPU Central Processing Unit

CSI Channel State Information

CDF Cumulative Density Function

C-FMCW Correlation based FMCW

CPS Cyber–Physical System

DGS Direção Geral de Saude

dB decibel

EMA Ecological Momentary Assessment

ECG Electrocardiogram

EPoSS European Technology Platform on Smart Systems Integration

EU European Union

FMCW Frequency Modulated Continuous Wave

GE Generic Enabler

GPS Global Positioning System

GPA Grade Point Average

HR Heart Rate

HITL Human–In–The–Loop

HITLAI Human–In–The–Loop Artificial Intelligence

HITLCPS Human–In–The–Loop–Cyber–Physical Systems

IoT Internet of Things

INESC Institute for Systems Engineering and Computers of Coimbra

ISABELA IoT Student Advisor and BEst Lifestyle Analyser

Lasso Least Absolute Shrinkage and Selection Operator

LCT Laboratory of Communications and Telematics

LOOCV Leave–One–Out–Cross–Validation

MHApp Mental Health Applications

MIMO Multiple–Input Multiple–Output

MPSS Mobile Phone Sensing Systems

NLP Natural Language Processing

OSN Online Social Networks

PCA Principal Component Analysis

PPG Photoplethysmography

PSD Power Spectral Density

RFID Radio Frequency IDentifiers

RSS Received Signal Strength

ROC Receiver Operating Characteristic

ROI Region of Interest

SAM Self–Assessment Manikin

SSID Service Set Identifier

SMS Short Message Service

SNR Signal–to–Noise Ratio

SDK Software Development Kit

SVM Support Vector Machine

UX User Experience

Chapter 1

Introduction

Contents

Nowadays, technology is everywhere and is an important part of our lives, with humans being increasingly dependent on it. Concepts like the Internet of Things (IoT) and Smart Systems are usually used to describe the applications and systems that surround us. These concepts commonly describe applications that are able to sense their environment, collect data, communicate or process that data, and extract information from it. However, most of the systems that currently exist do not consider humans as part of the system, nor account for their actions and intents and how that can affect the system. Furthermore, even the systems that can perceive humans are lacking the proper technologies to be able to sense them, and interface with them.

This chapter provides the background and motivation for this research, which is aimed at improving the use of the HITLCPS paradigm, to create applications that can incorporate the user as part of the system. Furthermore, the objectives of this thesis and its contributions are also addressed. Lastly, we explain how the rest of the document is structured.

1.1 Background and Motivation

Now, more than ever, we have extremely large amounts of data at our disposal. For instance, a supermarket chain can deal with hundreds of thousands of products equipped with Radio Frequency IDentifiers (RFID), and use RFID readers to scan these items every second, generating about 12.6 GBytes of data per second and about 544 TBytes of data per day (Bin et al. [2010]). Building on the available data, the number of smart systems is also increasing. In on Smart Systems Integration [2013] the European Technology Platform on Smart Systems Integration (EPoSS) defines smart systems as systems that *"are able to sense, diagnose, describe, qualify and manage a given situation,…They are able to interface, interact and communicate with users, their environment and with other Smart Systems"*.

Smart systems should be able to perform and incorporate functions of sensing, actuation, and control, in order to describe and analyse a specific situation and make decisions based on the available data. In most cases the smartness of the system can be attributed to autonomous operation based on closed loop control, energy efficiency, and networking capabilities. However, despite the fact that current systems are *"smart"* in many ways, the majority of these systems ignore the human factor or treat humans as an external factor (Nunes et al. [2015]). Nevertheless, in order to achieve a truly smart system, we need to work on the integration of humans as part of the closed-loop process. That is, humans need to be considered in every phase of the loop, be it the acquisition of data, processing/inference, or the actuation phase.

At the same time, we are witnessing the emergence of new sensor techniques, such as virtual and social sensors (Armando et al. [2018]), that show the importance of humans within the sensing system. In our opinion, only a system that is able to extend its capabilities to perceive and adapt to human actions, intentions, and emotional states, can be considered a truly smart system. Only then we can enter the realm of HITLCPS, where humans and machines interact, cooperate, coexist, and enhance our current systems.

In order to achieve this goal, we first need to endow our systems with sensory capabilities that are tailored to perceive humans. One way to do that is by using the smartphones that most of us carry on ourselves. Smartphones have become a constant in humans' lives, as most of us carry at least one smartphone at all times during the day, and we even sleep with these devices next to us. Smartphones have evolved tremendously in the last few years. They now have high processing power, their batteries last for several days, their networking capabilities allow for fast and reliable communications, and they are packed with a large number of sensors (e.g., accelerometer, gyroscopes, barometers, GPS). Furthermore, they even incorporate virtual sensors (i.e., software entities that virtualize sensor capabilities), for instance step counter sensor values which are acquired from inference of the accelerometer values. We are also able to run complex models on these devices to capture context information, such as changing from one vehicle to another or reaching a certain location. These devices also offer Application Programming Interfaces (APIs) that allow us to create custom code to run on them. Additionally, these APIs allow us to even retrieve data that can be directly correlated with human health or physical states, such as activity levels and sleep patterns. As such, we can now leverage these devices to create systems that better account for humans, that is, to create HITLCPSs.

Although smartphones are an *entrypoint* into humans' lives, there are certain aspects of humans that can not be perceived by only using this type of technology. Human actions are, most of the time, unpredictable when analysed by a random observer. However, trained observers can perceive certain indicators that give them information to classify those actions. These indications are often accompanied by involuntary physiological reactions (e.g., fluctuations in Heart Rate (HR), or an irregular Breathing Rate (BR)). As such, by creating and incorporating sensors that are able to gather information about those physiological responses in our systems, we are, in turn, moving towards empowering them with the ability to perceive and understand humans.

Most of the work and effort done towards the development and integration of human–specific sensors was carried out in the field of wearable sensors, like, for instance, the works in Tapia et al. [2007], Parkka et al. [2006], Atalay et al. [2017] and Lee and Chung [2009]. These devices allow for long–term monitoring that, alternatively, would require long–term hospitalizations and/or ambulatory environments that are much more expensive to set up and maintain (Bonato [2003]). This represents a great advance not only in clinical terms but also when it comes to gathering more information about human lives. The success of this type of technology paved the way for the emergence of several commercial solutions (e.g., Jawbone [2018], Fitbit [2022], Zephyr [2022]). Other approaches like Body Sensor Networks or Body Area Networks, have also largely advanced in the last few years, improving medical solutions, the training of professionals (e.g., military, sport athletes), or even other aspects of our daily lives like virtual reality gaming (Chen et al. [2011], Almashaqbeh et al. [2014]). However, even when considering the advances in the last decade when it comes to miniaturization of devices and wireless technologies, these devices can still be bothersome to use, and most of them require some cooperation from the user.

Despite being around for more than one decade, wearable devices and associated sensing techniques have yet to be explored in real clinical scenarios and have not yet been approved for medical use (Piwek et al. [2016]), with even the most recent versions of commercial solutions having considerable measurement error when compared to a straightforward, traditional Electrocardiogram (ECG) (Benedetto et al. [2018]). Recent surveys showed that 32% of wearable device users stopped using the device after 6 months, and 50% of users stopped using it after the first year (Ledger and McCaffrey [2014]). Additionally, it has also been shown that people that tend to possess and purchase wearable devices, are the ones that are already leading a healthy lifestyle (Bhas [2013]). This shows that we still need to explore different approaches to perform long–term monitoring of humans.

In addition to the lack of proper sensing approaches, even the systems that aim to monitor humans, fail to fully integrate them as part of the system. Most systems that are developed to monitor humans focus mainly on their physical well–being, and ignore their emotional states. A true HITLCPS should consider the emotions and emotional states of humans. These states are not as easily perceived as the physical ones, especially when considering the use of electronic devices. Humans have studied the psyche and have tried to understand how other humans behave and feel for many centuries, and found several ways to monitor this through well–validated forms and questionnaires. We believe that in order for HITLCPSs to completely integrate the emotional states of humans, we must integrate human knowledge in the loop of the inference and decision–making.

Additionally, when considering the inference phase of the HITLCPS paradigm, most systems rely on machine learning models to create predictions about the humans' states and behaviours. Machine learning models have evolved tremendously in the past years, with the advances in technology allowing for the creation of faster and more reliable models. However, most machine learning models are constrained by the amount and quality of data that they have available for training. As such, machine learning systems must rely on humans to overcome these limitations, by applying the concept of Human–In–The–Loop Artificial Intelligence (HITLAI). HITLAI refers to the concept of using human knowledge and expertise to perform labelling and validation of instances and give feedback to the system, many times while the system is already deployed, which leads to faster production times for machine learning models and more reliable systems. Additionally, HITLAI also refers to the use of machine–learning models as decision–helping mechanisms for human–controlled systems, where humans have the final decision about a given operation. We believe, that HITLAI is also an important factor to address, in order to create more reliable HITLCPS.

Another relevant aspect to explore when considering the HITLCPS paradigm is the way that humans interact with the system and the way that the system interacts with humans. Like we stated before, humans should interact with the control system in all of its phases and, as such, systems should be designed from the ground up to be intuitive for humans. However, the interaction of systems with humans should also be designed to be as efficient as possible, especially when considering the actuation phase of the control system. The acceptance of feedback provided by the system will directly affect its performance. As such, the actuation phase of HITLCPSs needs to be designed with that in mind.

HITLCPS are undoubtedly the way to create more reliable and better performing systems. However, we must first address the aforementioned limitations, in order to refine and develop this paradigm.

1.2 Objectives

The main goal of this research is to evaluate and propose models, adapted to our novel reality, that allow for the creation of Human–In–The–Loop–Cyber–Physical Systems by using IoT and new sensing techniques. To achieve this goal the following objectives have been established:

- Review and evaluate the state-of-the-art of existing works in the field of HITLCPS, by focusing on what defines a HITLCPS system and the possible challenges and open issues to be address in this field.
- Propose a new model for HITLCPSs that targets the found open issues. Evaluate the possible implementations of this new model and evaluate it.
- Evaluate new sensing techniques, that can help the systems to better perceive humans, namely, by extending the systems with capabilities to passively perceive physical and emotional states.
- Design and develop case studies to evaluate our new HITLCPS paradigm, while proposing possible architectural implementations. Implementing systems that show new approaches to tackle the open issues of the field.
- Evaluate possible results from the case studies, and create models that can reason the acquired data to perceive human context.

We would like to also highlight that privacy is also one of the major concerns in the HITLCPS paradigm. However, studying and creating privacy mechanisms is a challenging separate task by itself and, as such, will not be addressed in the scope of this thesis. In fact, privacy is the focus of another doctoral thesis, from which part of the collaborative work listed in the Forward emerged.

While working on the aforementioned objectives, several contributions were produced. These contributions are listed in the next section.

1.3 Contributions

Taking into consideration the goals described above, this thesis has produced the following contributions:

- **Review of HITLCPS applications and approaches**. To better understand the challenges and propose advances in the HITLCPS paradigm, we reviewed the existing work in this field. This is covered in Chapter 2.
- **Creation of a new model for HITLCPSs**. In order to advance the field of HITLCPS, we raise the open issues that were found during the study of previous approaches, and propose a new model for the creation and implementation of this type of system. This is covered in Chapter 3.
- **Creation of Taxonomy for unobtrusive sensing solutions**. Unobtrusive sensing is an emerging and relatively new field of study. Furthermore, due to its nature, it is a fairly broad field of study, which includes a variety of solutions and approaches. In order to better deal with the heterogeneity of this field we propose a new taxonomy to help classify these approaches. This is covered in Chapter 4.
- **Review of the state-of-the-art of existing unobtrusive sensing solutions and approaches**. We evaluate several of the existing solutions in the field of unobtrusive sensing and classify them under the proposed taxonomy. This is covered in Chapter 4.
- **Development of several HITLCPS case studies**. We developed several case studies to validate the HITLCPS paradigm and present approaches to tackle the open issues in this field. We focused mostly on the development of HITLCPS by using smartphone applications. However, some case studies also include web-based applications and unobtrusive sensing techniques. This is covered in Chapter 5.
- **Development of real–world trials and analysis/profiling of human behaviour**. Using the applications developed in the implemented case studies, we developed several real–world trials to perform data collection. We analyse this data and identify different patterns between the users, namely, distinct behaviours between groups of different students (under-performing, and over-performing students). Additionally, we also present different pattern behaviours of humans during different epochs of the Covid-19 pandemic. This is covered in Chapter 6.
- **Creation of several machine–learning models to infer states in HITLCPS**. One of the situations in which the HITLCPS paradigm can be useful is to monitor students and their academic performance. In this context, one of our primary case studies was developed with that objective. Using the data collected in the developed trials, we created several machine–learning models to infer different human states, namely, models to predict sociability, sleep periods, sleep quality and academic performance of students. This is covered in Chapter 6.

- **Creation of a pipeline for automatic assessment of students' performance**. As a result of our work, we proposed a pipeline that uses a decision level median voting algorithm to further improve the models' performance, by using historic data from the students to further improve the prediction. Using this pipeline, it is possible to further increase the performance of the models, with some of them obtaining an accuracy greater than 90%. This is covered in Chapter 6.

In addition to the referred contributions, we would like to highlight the dissemination of our work throughout the international scientific community, through several publications in international journals and conferences. The rest of this document presents the work that underlies these contributions. In the next section, we present the outline of the rest of this document.

1.4 Outline of the Thesis

This thesis is organized into seven chapters, and the remainder of this thesis is organized as follows.

Chapter 2 offers the research context, describing the HITLCPS paradigm and how it can benefit most of the already existing systems. The chapter also presents a review of existing solutions that can fit under the HITLCPS paradigm, and that are closely related to the case studies presented in chapter 5.

In Chapter 3, we start by covering the open issues that we found in the related work of HITLCPS. Additionally, we present our proposal for a new HITLCPS paradigm that is able to account for humans, their actions, emotions and intents. In this model, we cover the inclusion of the human in all phases of the Cyber–Physical System, namely in the data acquisition, state inference and actuation phases.

Chapter 4 presents our proposal for a taxonomy for unobtrusive sensing. In this chapter we also present a review of the state-of-the-art and classify relevant existing work on unobtrusive sensing under the proposed taxonomy. Additionally, in this chapter we review all the solutions based on data multimodality and computing architecture.

Chapter 5 presents the implementation of several case studies and trials developed to implement and test the proposed HITLCPS paradigm. These case studies include the evaluation of university students' performance, the evaluation of risk and general well-being during the Covid-19 pandemic, the use of a recommendation system to match challenges related to climate changes and people to solve those challenges, and a case–study to evaluate the implementation of unobtrusive sensing in a HITLCPS.

Chapter 6 provides an analysis of the data collected from the different case studies presented in this document. Namely, we present the results from the trials performed in Ecuador with the ISABELA application, the results from using the data acquired from the ISABELA trials to create a model to automatically predict the performance of students, and the results from trials during the

Covid-19 pandemic, using the VITORIA application to monitor the behaviour of users. Lastly, we present the results from preliminary tests with the proposed unobtrusive sensing techniques to monitor HR and BR.

Chapter 7 concludes this document, offering a synthesis of the thesis, a compilation of the resulting contributions, and identification of possible research paths for continuing the work presented in this thesis.

Chapter 2

Review of HITLCPS approaches

Contents

In HITLCPS, technology is driven by humans and considers human intents, actions (present and foreseen), and mental states (beliefs, desires, emotions), thus opening the way to adaptable, intelligent advice systems (Nunes et al. [2015]). HITLCPS enable us to develop applications that convey emotions, psychological states, actions, and drives of humans, as part of larger–scale systems, where sensing and actuation approaches go beyond physical, electronic–based devices (Armando et al. [2018]). In fact, sensing may also comprise software–based entities like agents, as well as human beings, by considering human–originated activities from Online Social Networks (OSN) as raw data, for instance.

There has been an increase in the use of applications that are related with HITLCPS and related areas, namely smartphone sensing systems, online social networks sensing, and applications used in the context of the Covid-19 pandemic. Furthermore, these systems can also be used to perform data collection and inference of certain aspects of the humans' lives. One such case study is the prediction of students' performance, which can be achieved by leveraging data collected from smartphones and other devices. In this chapter we introduce the concept of HITLCPS and we present an overview of relevant related work in each of the aforementioned areas.

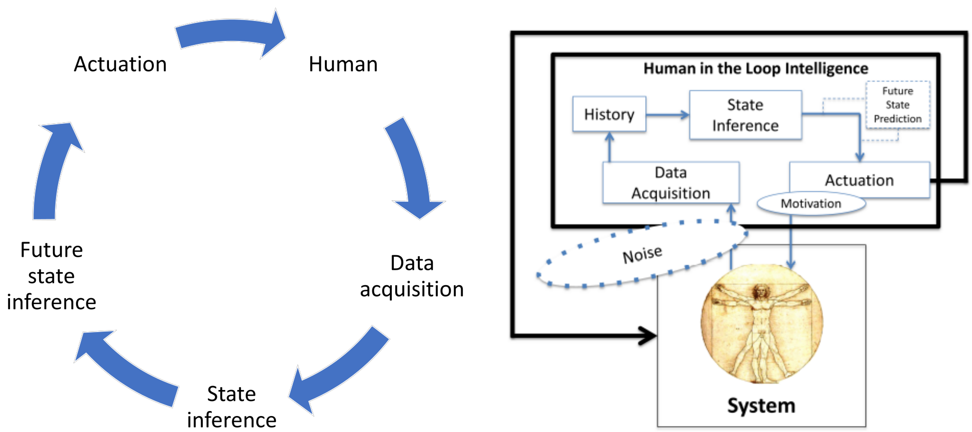
2.1 IoT, Control loops and Humans

The term *"Internet of Things"* was proposed in 1999 by Kevin Ashton. He was then referring to Linking the new idea of RFID with the supply chain of a company (Ashton et al. [2009]). Nowadays, the concept evolved, and no longer refers to RFID, instead now the term IoT is more than just a concept, we are surrounded by "Smart Things", with recent statistics point out that there are roughly 12 billion IoT devices, with an estimated for the number of these devices to more than double in the next few years (Statista [2022]). Therefore, everyday devices are now improved with sensors, connected to the Internet, and can communicate between them, which allows them to share sensing data. However, this sensing data needs to be understood, modelled and reasoned with, creating context–aware applications that can perceive everything around them (Perera et al. [2014]).

Combining the unprecedented amount of information that the IoT offers us and the functionalities of a Cyber–Physical System (CPS), we can now create new systems that are aware and are able to control the physical world. Technology is always created to improve humans lives, and the same is true for IoT applications and CPSs. By applying these concepts to human–centred applications we can enter the realm of HITLCPS. Where technology accepts the human as part of the system, and considers the human intents, actions (present and future), emotions and psychological states to create systems that better care for humans needs (Sousa Nunes et al. [2015]).

In Figure 2.1a we can see a simplistic diagram of the HITLCPS paradigm. As can be seen from the figure, it resembles a normal CPS, in which there are the 3 principal phases, namely the data acquisition phase, the state inference or state modelling phase and the actuation phase (Khaitan and McCalley [2015]). If we consider a human–centred application the human would be the focus of the application as we would only monitor the human, and all the actuation and state inference would be done with information about the human itself. However, in a HITLCPS we must consider a much larger system where several phenomena are being monitored and controlled and in which, the human is just one more of those phenomena. In turn, every human–centric application is a HITLCPS by definition, but a HITLCPS can be a set of CPS where the human–centric application is one of the sub-sets. In this last scenario the human–centred application can regulate itself, but can also provide information from the different phases for the larger system loop.

Additionally, as can also be seen from Figure 2.1a, in the HITLCPS there is one more state, that tries to deal with the unpredictability of the humans by inferring the humans' futures states. This is especially useful when dealing with humans, since the multitude of metrics that are monitored most of the time have correlations with each other (e.g., a depressive psychological state can indicate a future inactive physical state). By trying to predict the future we are endowing the system with more useful information.



(a) (b)

Figure 2.1: (a) Simplistic diagram of the HITLCPS concept; (b) HITLCPS scheme proposed in Sousa Nunes et al. [2015].

This is not just truth to human–centric application, as every system is influenced, directly or indirectly, by humans, and they can disrupt the proper functioning of a system or enhance it. For example, in a critical system, such as the one of a power plant, a malfunctioning machine is dangerous, but a "malfunctioning" worker can cause an even worst outcome. This concept could also be employed in less critical scenarios, such as the use of IoT to monitor transportation in cities. Most of the proposal systems now–a–days take only into account events, the flux of vehicles, transit signs, Global Positioning System (GPS) in Vehicles and rush hours (Mohanty et al. [2016]), but if we think about the human behaviour we can easily see that they could influence this scenario. For example, in a city with a coastline, a person with a more positive state of mind may tend to choose a path that passes near the beach even if it is not the fastest path, while a more negative state of mind could indicate that he will choose the more direct path to his destination. In this scenario, is also valid to envision a human–centric application that, with the aim of improving the psychological state of a person, may suggest less stressful routes, avoiding congested ones.

The HITLCPS enables to build an Internet that conveys emotions, psychological states, actions and drives of humans as part of larger–scale systems. Hence, both sensing and actuation approaches go beyond physical electronic-based devices. Indeed, they also comprise software–based entities like agents and human beings, for instance we can even consider raw data from human–originated activities on social media as a virtual sensor. Like the embedded chips in electronic-based sensors, the "human–based sensors" also need a mechanism to transform the collected raw data into values which can be interpreted, conducting to information and knowledge. To our mind, both sentiment and emotion analysis, and the opinion mining tools, such as Topic Modelling (Arora et al. [2018]) are one of these mechanisms to apply between what humans produce in multimedia activities and what can be leveraged to infer the context of their surroundings. By multimedia activities, we mean any content daily produced by users via image, audio, video and written digital support through communication devices such as smartphones, tablets and personal computers. Solutions in the community have not yet successfully joined Human–In–The–Loop (HITL) and online social media (Sousa Nunes et al. [2015]). In this regard, the attempt to overcome this challenge is also one of the contributions of this thesis. We believe that in the near future, by joining our personal devices, other IoT technologies and new sensing approaches will be possible to monitor seamlessly every aspect of the human live, obtain context–aware information and create smarter systems that take the human into account.

Although there are several advantages, for most systems, by including the Human in the loop of the system, apart from the e–heath area, whose primary objective is the monitoring of human beings, the majority of the systems still ignores the humans or treats them as external elements to the system. This happens in part because of the difficulty to predict humans, that are regarded as unpredictable most of the time. Other authors have also proposed models for HITLCPS, as can be seen from Figure 2.1b. However, we believe that these models are not complete and do not consider all the necessary conditions to build a HITLCPS. As such, the advance of the integration of the HITLCPS can only be done by the creation and use of models that can represent the humans behaviours, psychological states and physiological parameters in a precise and reliable way.

2.2 Smartphone-based Human-in-The-Loop Systems

Advances in the technology have also made it possible for almost every human to carry a computer in the palm of their hand or in their pocket. Smartphones now–a–days have a processing power and memory that surpasses by far the capacity of computers that in a past occupied an entire room. Furthermore, even the most modest of smartphones, now–a–days, is empowered with sensors such as accelerometers, gyroscopes, GPS, microphones and cameras. While other more high-end smartphones have more capabilities like biometric sensors that allow someone to unlock their smartphone with their fingerprint or even face identification. Smartphones have the capabilities necessary to run complex calculations, store data and have communications capabilities, which allow them to share sensed information. This provides us, with a new way to look at humans lives in an unobtrusive way. We can now monitor humans in almost every situation as they now carry these portable computers with them to every place. While there is still some controversy about the effects that these devices create in our lives (Ozdalga et al. [2012]; Park and Lee [2012]), they are, without doubt, a source of data that can be used to monitor people's lives.

These devices are increasingly being used to gather a variety of data on users and their environment, and this trend is expected to keep growing in the near future. One example of a study that used smartphones as a sensing source is *"StudentLife"* (Wang et al. [2014]), which investigated the use of smartphones to infer students' academic performance. The goal was to correlate students' academic performance with psychological and physical well-being, through a mobile application and self-reports. To monitor the students, their activity levels, sleep quality, and conversations, the authors reused some of their previous work, namely a mobile version of Ecological Momentary Assessment (EMA), that allowed them to capture students' states such as stress and mood (Shiffman et al. [2008]). Using their system, they conducted a study in Dartmouth's University, where a class of 48 students was monitored during a 10-week period. From the obtained dataset, the authors were able to extract some correlations between the progression of the term and the students' academic performance, self-assessed mental state, and behavioural trends.

Some of their results included a significant inverse correlation between sleep duration and pre- and post-depression, most findings being in line with existing literature and studies in the field. The same was found to be true for conversations frequency (students that have fewer conversations are more likely to be depressed), and overall social activity (students that are more social and are surrounded by people are less likely to get depressed). It is also worth noting that the authors mention that some of their data was also obtained from Facebook. However, they did not present any information about social sensors nor explain how this data was obtained. Moreover, the collected data is not available for public use, due to a lack of anonymization.

Another relevant contribution is the work presented in Wang et al. [2015], in which it was proposed a prediction model for the students' Grade Point Average (GPA), using data inferred from smartphones. Data such as, activity data, conversational data, class attendance, and social interaction. In this study, the authors tackle the problem of predicting the students' GPA as a regression problem, exploring some alternatives and choosing to base their model in a Least Absolute Shrinkage and Selection Operator (Lasso) regularized linear regression model (Tibshirani [1996]). The results obtained from their model have ± 0.179 mean absolute error from the ground truth. The study shows that there are clear relations between the students' behavioural patterns and respective grades. Although, in this study the authors leverage the previously obtained dataset and do not present a way in which the model can autonomously work and give feedback to the students on how to improve their performance, the proposed model can be useful in the state inference phase of a HITLCPS system.

Additionally, the work by Harari et al. [2017], used the dataset obtained in Wang et al. [2014] to detect the patterns in the behaviours of students and how they relate with health problems (e.g., mental health, physical health). In this study, the results were presented for weekly time windows, and the dataset was treated as a population where the mean value represents all the individuals. Moreover, the authors focused on two types of behaviour patterns, namely the physical activity behaviour patterns and the sociability behaviour patterns. The *"StudentLife"* study captured ambient sound in its dataset and used it to infer the time spent in conversation. According to the authors, this is representative of a sociability trait (i.e., the general tendency to affiliate with others). For the physical activity behaviour, the authors used accelerometer data to infer the status of the user, e.g., stationary, walking, running, driving or cycling. The study found interesting correlations between the term rhythm (e.g., deadlines, classes, midterm etc.) and students' mood, sociability, physical activity, and sleep periods. Subsequently, the authors derived some correlations with the mental health, depression, stress, and academic performance. As future work, the authors proposed the development of applications that can give personalized health information and context to students, although they alerted to potential privacy issues. Despite the fact that this study does not constitute an application of the HITLCPS concept because it is open-loop, it is applicable to the state inference phase of HITLCPS systems.

Other studies, such as the one by Eskes et al. [2016a], also proved that passive data from smartphones could be used to infer user sociability. In this study, the authors presented a smartphone application called "BeHapp", which used not only smartphone sensors, such as GPS and Bluetooth, but also other smartphone capabilities, such as mobile phone usage records, to infer sociability. Specifically, the sociability model also took into consideration SMS and calls activity. They based their sociability formulation on the work of Cheek and Buss [1981], and they defined sociability as a mix of social exploration and communication. The authors proposed a sociability score that used the frequency at which the user sent SMS or made calls, the diversity of those phone calls, and the duration of calls. Additionally, the score also considered the user's social exploration, by monitoring the user's movements pattern, the distance that he/she travelled having his/her home as a reference point, and the density of people around the user by capturing the number of surrounding Bluetooth devices.

Other authors also aimed to implement HITLCPS systems by using IoT devices, as happened in "HealthyOffice" (Zenonos et al. [2016]). This study was developed in office environment with the main objective of detecting the mood of people in the workplace. For this, the authors used two variants of the Toshiba

Silmee$^{\text{TM}}$, the Bar Type and Wristband (Miyamoto et al. [2020]). A smartphone application was used as a means to obtain the ground-truth. This application was repurposed from the application created for the IES Cities project (Vatsikas et al. [2017]). The Toshiba devices were able to collect physiological data, namely heart rate, pulse rate and skin temperature. Additionally, these devices were also able to capture 3-axial acceleration from an embedded accelerometer. A case study was also implemented where four subjects in an office context were monitored, during eleven working days. The study participants had to use the wearable devices and report their ground-truth mood every two hours. The self-report consisted of scoring eight different possible moods, with a numeric score ranging from 0–100. Two types of models were considered, namely individual models and generalized models, with the former leading to better results. The authors also proposed a framework comprising all the phases of the HITLCPS concept when implemented in its entirety, namely a data collection phase through wearable devices, a state inference phase, and an actuation phase during which individual feedback was given to the workers. Additionally, the framework also included a mechanism to determine the aggregated mood for the office. This aggregated mood value could then be used in deciding on actions to improve the overall mood in the office.

Additionally, the study in Yang et al. [2015] showed how the use of sensor–enabled smartphones is a promising solution to large-scale urban data collection. Powerful computing and communication capacities, massive market proliferation, and inherent mobility make Mobile Phone Sensing Systems (MPSS) a much more flexible and cost-effective sensing solution than traditional static sensor networks. Aligned to this assumption, the authors suggested a finite horizon stochastic solution for continuous data collection in MPSS using both hybrid cellular and short-range communications.

Lastly, one of the less explored phases of the HITLCPS is the actuation phase (i.e., closing the loop with contextualized feedback). In Lane et al. [2014], the authors developed the “BeWell” system, which was able to collect data regarding physical activity, sociability and sleep. In this study they made a close loop approach where they monitor the people’s day-to-day activities passively, infer their well-being and inform them about the assessment made. The feedback was given to users in one of two ways: first through a widget in the smartphone where they showed the users their stats on the sleep, activity and sociability measures; second, with a web portal where the user can see more information. This work implemented all the phases of the HITLCPS concept. However, the feedback was limited to numeric scores for the physical activity, social interactions and sleep. This is still valuable information that the humans could use to improve their well-being. However, in order to create a true HITLCPS the system should be able to give context–aware actuation, and feedback that is tailored for each specific user.

2.3 Online Social Sensing

OSNs are communication channels that can be used to infer the context of physical and virtual spaces (Armando et al. [2018]). In OSNs, people describe the real–world and beyond, including their emotional statuses. This integration of people and their social data into sensing systems increases scale and reduces costs (Bachiller et al. [2015]), as some sensors are virtual and do not require specific hardware. Intelligent, socially-aware HITLCPS leverage the capability of OSNs for providing contextualized data, which can be interpreted, leading to information and knowledge.

The explosion of IoT technologies and sensor deployments, and the emergence of social networking, can be combined to offer vast added value services for smart applications and services (Psomakelis et al. [2016]). In this regard, social sensing has emerged as a new paradigm of data collection, where groups of individuals, mainly using OSN channels, gather and share observations on their surroundings (Wang et al. [2019]). In Bachiller et al. [2015], the authors define a way of including users in distributed applications, enabling data to be acquired from both users' and mobile phones' sensors. In this sensing approach, the participants contribute with observed values and triggered actions. Three communication channels are used, namely Dedicated application, web application, and OSNs. In Lee et al. [2015a], the authors leverage the concept of social sensing for the development of an analytical model of dengue vector. They collected online sensing data, and then combined historical data as training datasets for analytical computation. Since environmental sensors are not as ubiquitously deployed in the environment as needed for situational awareness, messages of social sensors and real-time tweets were associated with possible events of dengue outbreaks, produced by Twitter users. The authors in Giridhar et al. [2017] leverage data from both Twitter and Instagram to design a system capable of detecting events with the support of an unsupervised approach.

One well-known issue with most systems that rely on data collected from OSNs and/or smartphones is user incentivization (Liu et al. [2018]; Alsheikh et al. [2017a,b]). In Yang et al. [2015], it was also suggested an incentive approach that relies on imposing taxes or providing subsidies for each phone user, depending on the impact on the rest of the phone users within the MPSS. Commonly, many systems tackle the incentivization problem with reward-based approaches, sometimes with financial incentives, which, overall, may work for specific, limited cases, but are not sustainable (Armando et al. [2018]). One plausible solution for the incentivization issue is making users both actors and beneficiaries of the system, as is generally the case for HITLCPS systems.

The proposed ISABELA platform combines and extends the approaches taken in the previously mentioned cases, namely leveraging smartphone sensing and smartphone usage data. Moreover, it gathers OSN data from both Facebook and Twitter, using NLP techniques, and generates insights on users' mood. Finally, by including the users in the sensing/actuating loop and making them the main beneficiaries, it effectively deals with the incentivization problem.

2.4 Human Sensing During Pandemics

In the recent past years the world has been affected by the Covid-19 pandemic, HITLCPS applications can also be used in this context. Understanding how people are affected and behave during pandemics could be very valuable, for the development of more efficient and personalized plans and measures, for tackling future Covid-19 waves or even future pandemics. In this section we will cover works of the state-of-the-art that focused on the same aspects of understanding human behaviour as well as the effects of the Covid-19 pandemic on people's well-being. In addition, we will also cover contact tracing applications, as these mobile applications were one of the most commonly used approach, to deal with the Covid-19 pandemic.

As stated, the ongoing pandemic required the creation of social distancing and in some periods even lockdown measures. This in turn created several changes in our daily lives routines, for example on children and their school attendance. The occurrence of lockdowns, imposed a shift in school *"normality"*, making most children change to a paradigm of virtual classes. In Limone and Toto [2021], the authors evaluated the effects that this shift had on children, particularly the effects of the increase in digital technology use, due to this change. In this work it was also highlighted, that the increase in the usage of technology during the pandemic had both positive and negative effects, depending on the usage. Digital technologies allow us to bypass the separation created by the social distance allowing us to connect with others. However, they may also be a source of risks related to psychological well-being and contribute to depression, anxiety, sleep problems, irritability, and cognitive impairment. In Francisco et al. [2020], the authors also found significant changes in lifestyle habits of children and adolescents (e.g., increased screen-time, reduced physical activity, more sleep hours per night), associated with an increase in their psychological and behavioural symptoms, such as anxiety, mood, food intake, cognitive and behavioural changes.

Other works have also explored the effects of the pandemic on the mental health of people. In Talevi et al. [2020] several works were surveyed, namely works that evaluated the mental health of people during the pandemic. They concluded that the ongoing pandemic has had a huge psychological impact on individuals, from common people, to people that were infected, and also people previously diagnosed with mental health issues, and even health-care workers. In Greenberg [2020], it was also explored how health-care workers were negatively affected by the ongoing pandemic, due to working extremely long hours in high-pressure environments or even the exposure to trauma and being faced with moral dilemmas when trying to deliver high-quality care.

Additionally, some studies evaluated the influence of the Covid-19 pandemic on physical activity levels (Caputo and Reichert [2020]). With evidence pointing towards a decrease in physical activity levels, due to home confinement and social distancing measures. Furthermore, it was also pointed out the positive correlation between physical activity and mental health.

Other studies have also used questionnaires to evaluate the perception of threat of the Covid-19 pandemic (Pérez-Fuentes et al. [2020]). Showing, for instance that it was possible to assess the general perception of the Spanish population using a short questionnaire. However, this type of study can only evaluate the momentary perception as the use of longer questionnaires, constrains the possibility of doing a longitudinal study.

Mobile phones and wearables have also been leveraged to respond to the public health crisis imposed by the Covid-19 pandemic. In Almalki et al. [2021] a systematic review of health apps for Covid-19 was provided, and five main purposes for the included apps were identified. The most common were apps developed by governments or national health authorities to promote users to track their personal health, including self-assessment for the identification of possible infection, symptoms monitoring, mood tracker, medication trackers, diagnosis recorders used during self-quarantine, etc. The second most common purpose of developed apps was to raise awareness about Covid-19, including providing basic health information and advice, presenting statistics, the latest news and updates about the pandemic, providing information or interactive maps of active cases and medical facilities, etc. The least common purposes of apps were for the management of Covid-19 exposure risk, providing health monitoring by healthcare providers (e.g., making medical appointments, virtual medical consultations, remote monitoring, helpline, etc.) and conducting research studies.

Another common line of work, when considering mobile phones and their use in relation to the management of Covid-19 exposure risk, is the one of the contact tracing applications. The main goal of these applications is trying to limit the propagation of the SARS-CoV-2 virus, by detecting in a short period of time the past contacts of a person that tested positive for the virus. By knowing the past contacts of a positive diagnosed person, it is possible to enforce preventive measures such as, for instance, putting those past risk contacts in a preventive quarantine to mitigate the spread of the virus. With the outburst of the virus, there was an emergence of several of these apps, in part due to the fact that most countries tried to implement their own contact tracing mobile applications. In Portugal, where our trials occurred, the government took this approach as well, with the creation of the app "STAYAWAY Covid" (StayWayCovid [2022]). This type of application, normally uses the Bluetooth capabilities of the smartphone to transmit and receive individual hashes that can be then used to identify the contacts of each person and notify users who were within close proximity of someone who had tested positive for the virus.

However, this type of application is merely intended to detect possible past contacts and is not able to offer more information about their users, nor provide feedback to users with relevant information about their risk of exposure to Covid-19. For example, the number of contacts they have had or the time they spent out of their house, which have been proved to be one of the most relevant features of human behaviour self-regulation (Carver and Scheier [1981]). Tailored feedback represents the most individualised type of feedback and research has pointed to its potential for increasing the effectiveness of behaviour change interventions, over providing generic or targeted information (DiClemente et al.

[2001]). However, tailored feedback requires personal information provided by some kind of assessment procedure, which can be burdensome for the user if it is only acquired by active data acquisition procedures, involving, for example, the user to fill in some measurements. In this regard, mobile phones present an effective way of unobtrusively sensing human behaviour and delivering tailored feedback.

In recent years, several applications have been built and studied to help promote positive well-being and Mental Health Applicationss (MHApps) in the general population. Despite promising results for positive findings for reducing mental health symptoms and/or promoting well-being or emotion regulation, the lack of generalization of the findings and low participant adherence to MHApp usage are commonly reported as limitations (Eisenstadt et al. [2021]). On the other hand, several MHApps were also developed specifically during the pandemic of Covid-19 to help reduce mental health problems in the general population (Jaworski et al. [2021]) and in healthcare professionals (Fiol-DeRoque et al. [2021]), but also with reduced engagement.

Additionally, there are other works that use mobile applications or IoT devices to monitor people, some of those works are explored in the survey (Kondylakis et al. [2020]). However, those works focus mainly on contact tracing or data sharing from patients to healthcare professionals.

As far as we know, the Vitoria system, developed in the context of this thesis, is the only system that leverages smartphones and smartwatches to obtain a complete overview of the users' internal (psychological) and external (contextual), as well as give them tailored feedback. The main goal of the Vitoria application is to monitor several aspects of humans' lives, both in their emotional and physical states. In order to detect and identify possible profiles and behaviours that are common in this and possibly in future pandemics. Additionally, it provides tailored feedback to the users, which is vital for them to self-regulate their behaviour in order to better manage their risk of exposure to SARS-CoV-2 contagion. Furthermore, in terms of privacy concerns, our application also differs from the aforementioned applications, as we do not intend and do not need to identify specific people. As such, all of our data is anonymized during its collection. Making it impossible for our system or for third-parties systems to identify the users. The implementation of the Vitoria system, as well as the results of the tests conducted with this application, are presented in the next chapters.

2.5 Students' Performance Assessment Approaches

One possible HITLCPS use case is that of a system for assessing students' academic performance and providing guidance. Students' performance can be affected by a variety of factors, from social-economic factors to psychological or even environmental factors (Hijazi and Naqvi [2006]). As such, it is hard to predict which students will have lower or higher performance. Additionally, novel educational mechanisms which rely on online tools and materials have been shown to improve students' performance (Strelan et al. [2020]). However, since a growing percentage of the learning process happens outside the classroom, evaluating the students' progress and needs is getting harder for teachers. In this respect, any system that can identify and/or predict poor performance can be invaluable.

We have covered in section 2.2, the use of smartphones as a way to collect data. We gave a special focus to the work in Wang et al. [2014, 2015]; Harari et al. [2017] which use the data collected from the "StudentLife" system to evaluate students' performance and their behaviours. In Harari et al. [2017] several interesting correlations were presented, between students' behaviour and their daily routines. Furthermore, these behaviour patterns can be tied to the final students' performance. However, in this work the authors do not propose the use of this data, nor founded correlations in order to help predict the academic performance of students.

Additionally, the work in Wang et al. [2015] proposed a regression model to predict student GPA. As we explored in section 2.2, the proposed model used data retrieved from smartphones, such as activity data, conversational data, class attendance, and social interaction. The results obtained from their model have ± 0.179 mean absolute error from the ground truth. The study shows that there are clear relations between the students' behavioural patterns and their respective grades. However, the focus on GPA can be limiting since this metric can only be applied to ungraded students. Furthermore, this metric is not normative across different countries. Additionally, external factors can influence the results of students (e.g., the Covid-19 pandemic), and even between editions of the same school subject there can be a variety of different conditions that impact the overall performance of students (e.g., changes in the lecturer or evaluation methods). As such, we believed that a classification scheme that can cope with varying conditions while determining students' performance is needed.

Other works also tried to predict the students' performance by using machine learning models. This is the case of Osmanbegovic and Suljic [2012]. However, in this work, instead of smartphone data, the authors suggested a model that uses demographic information, as well as past academic information, to predict the students' future performance. The authors present a dataset of 257 students and explore the use of three different models, based on the Naive Bayes algorithm, the Multilayer Perceptron, and the J48 algorithm, to classify student performance into one of two classes (students that fail, and students that pass). The results show that the tested models have an average recall of 85% when detecting the students that pass. However, the models present very poor results when classifying the students that fail. The use of demographic information and past academic information can be interesting and valuable in this kind of model. However, this can also raise several privacy concerns, especially for a smaller dataset, since this type of information can be used to infer the students' identities.

The Covid-19 pandemic has also affected people's daily lives in several ways in recent years, as several studies have pointed out (Afonso [2020]; Buzzi et al.

[2020]). This is also true for students, which had their routines changed, were confined to their homes and had to change to a remote class system. In Nepal et al. [2022], a study was conducted to assess behavioural changes of 180 undergraduate college students, during the pandemic, using mobile phone sensing and behavioural inference. This study was divided into two acquisition phases. The first phase occurred one year prior to the pandemic and was used as baseline. The second phase occurred during the first year of the pandemic. The authors used Principal Component Analysis (PCA) to reduce the dimensionality of data and be able to cluster the students into different groups. Specifically, it was possible to identify two distinct groups. It was also observed that one of the groups was mostly composed of students with higher self-reported Covid-19 concerns and higher levels of stress and anxiety. This shows that in fact, different groups of students experience the pandemic differently. Additionally, the authors found that there was a positive correlation between levels of depression, anxiety and stress with self-reported Covid-19 concern. Additionally, a deep learning model was implemented to classify student Covid-19 concerns, with an Area Under Curve (AUC) score of 0.70 and a F1 score of 0.71. Mental health and stress can be directly correlated with student performance. As such, evaluating these metrics can be important. Furthermore, this study reveals interesting insights into students' behaviour changes due to the Covid-19 pandemic. However, this study does not directly address the prediction of students' performance.

The early detection of students with poor performance can also be used to reduce the dropout rates. In Sandoval-Palis et al. [2020], a study was conducted to create a model to predict which students were at risk of dropping out. Additionally, this study is also relevant since it was carried out in the same academic institution, in which the ISABELA datasets were acquired, namely the Escuela Politecnica Nacional, in Quito, Ecuador. The model uses students' demographic and academic information to classify students into one of two classes (students who have dropout and students who did not). Furthermore, the authors tested two models, namely a Neural Network and a Logistic activation function. Although, the authors present a 76% accuracy for the Neural Network model, the model presented had a very high false positive rate. This is, due to the fact that the dataset is highly imbalanced (72.8% dropout, 27.2% no dropout) and the authors did not address that. However, the authors defend the results by saying that for their specific case study, it is better to raise alarms for interventions in students who do not need them (false positives) than to fail to identify students who do need them (false negatives). Additionally, it was possible to identify some patterns in students that are more probable to drop out, namely a low application grade, a low vulnerability index, enrolment in the month of March (the second of two available enrolment epochs), and being enrolled in bachelor's degrees. Although, this work does not address performance prediction, we believe that some of its findings can give some insights into the constraints and patterns of students from that university.

Although there has been some work in the field of students' performance prediction, none of the concerned studies led to results that can be widely used independently of academic level or region/country–dependent metrics (i.e., GPA).

Furthermore, as far as we know there is not any work that is able to predict the performance of students before and during the Covid-19 pandemic. As such, we believe that there is still a need for the creation of models that are able to automatically detect student performance in these scenarios.

In this thesis, we covered the exploration of trials performed with the ISABELA application, which has the main aim of predicting the students' performance and trying to modulate it by giving contextualized feedback. Specifically, we cover the analysis of the data collected from three real–world trials, namely, the analysis of the data collected OSNs (i.e., Facebook) from students in Portugal, and the data collected from the use of the application by students in Ecuador, in two distinct trials. For the first trial in Ecuador, a longitudinal analysis of some collected metrics is presented, namely, an analysis of the levels of activity, time spent studying, and the time spent in the university. Furthermore, we present three different machine learning models that leveraged the collected data, namely a model to infer sleep periods, a model to classify sleep quality and a model to classify sociability. These metrics are highly correlated with student performance and as such are of the highest importance. Additionally, by using the data from the two trials in Ecuador, we propose a new model to automatically infer students' performance. More specifically, a model capable of classifying the students belonging to the same school class into two groups, the students that perform *above the median* and the students that perform *below the median*. These results are explored in chapter 6.

2.6 Chapter Summary

In this chapter, we have introduced the technological context that enables the HITLCPS, namely, the IoT, CPS, smartphones, and social sensors. By merging all of these contexts we can create systems that better account for humans, and that are able to fully integrate them in the loop of the system, enabling the HITLCPS paradigm. However, even though we have seen several advancements in the technologies that enable the HITLCPS, most systems still ignore the humans in any aspect other than them being the end-user.

On the other hand, there are already some works in this field. As such, in this chapter we also review several of the approaches taken in the state–of–the–art. More specifically, we explore approaches based on technologies that enable the HITLCPS, namely, smartphones, and social sensing based on OSNs. Furthermore, we explore the use of HITLCPSs in relevant contexts, such as solutions used during the Covid-19 pandemic, and solutions to monitor and predict the performance of students.

Additionally, from the review of the state-of-the-art several open issues were found. In the next chapter, we will explore those open issues and present a novel model for the HITLCPS paradigm to address those issues.

Chapter 3

A new model for the HITLCPS

Contents

In the last chapter we reviewed the related work of HITLCPS in several fields of study. However, most of the reviewed work still does not fully integrate the human as part of the system. Furthermore, many other systems ignore the human completely.

HITLCPSs should account for humans' emotions, psychological states, actions and intents. These systems should also integrate the human as a part of the larger–scale systems, by including them in all phases of the control system process. Humans can interact with all phases by being part of the sensing, modelling and inference and even actuation phases. It is important to design novel systems that take all of this into account, to thus reach the HITLCPS paradigm. In this chapter, we will cover the open issues found in the related work and present our proposal for a new HITLCPS model.

3.1 Open issues and challenges

As previously stated, most of the systems now–a–days do not take the human into account as part of the system. They are, most of the time, ignored or treated as an external factor. Humans can influence a system in several ways, they could improve an existing system, but they can also be the reason that a system loses performance. As such including the humans in the loop of the system and accounting for their actions, emotions and intents can benefit most systems, even those that are not human–centric systems/applications.

Humans can interact with all the phases of the system's loop. They can be part of the acquisition process as a virtual sensor (e.g., when considering OSN activities), or they can be the object of the sensing phase. Additionally, humans can be part of the state inference and future state inference, humans can better comprehend the physical world and reason the acquired data, which can be helpful to label acquired data or feed context information to systems. Additionally, they can help on the decision–making procedure by having "the last word" on a classification mechanism. Furthermore, on one hand, humans can interact with the physical world and systems in ways that are not possible for electronic or virtual entities. As such, they can be a mean to actuate over a given system. On the other hand, actuation can be applied to humans themselves, in order for them to change their behaviour, or to influence their mental states (e.g., changing music or lights on a smart home to accommodate the humans' mental health).

The concept of HITLCPS has been around for several years, and as shown by the reviewed work there have been several approaches to tackle its implementation. However, there are still several areas and open issues that need to be addressed in order to achieve the HITLCPS paradigm.

We believe that most of the works explored in this review try to advance the field of HITLCPS. However, most of these works can fit inside the human–

centric classification, as the human is not only part of the system, but also the main focus of data acquisition, state inference and actuation. This is a valid implementation of the HITLCPS paradigm, however, we believe that in order to achieve a truly HITLCPS, we must apply this paradigm to a larger system where the human is only a part of the *"equation"*. For instance, when considering the example of a factory as a system, we should consider the whole factory as the system with the main objective of improving the production levels/performance of the factory. In that example the performance of humans (i.e., workers) it's a factor, but there are many other factors such as, the performance of machines, resource optimization and the distribution chain. We believe that the implementation of a larger scoped HITLCPS is an open issue to be addressed in this field, since it would lead to other challenges such as the orchestration between different smaller systems or challenges with data fusion due to the existence of many distinct data sources and data types.

Another open issue, is the difficultly to retrieve data related to the emotional states of people, since they are not as easily perceived as the humans' physical states. Additionally, the works reviewed that explore people's emotional states do so mainly by using questionnaires. Although these questionnaires are validated by years of use in social sciences, they require the humans/users cooperation, which by itself is another challenge of the HITLCPS. As such, there is a need for new sensing techniques that can help systems to perceive the physical and emotional states of humans, without cooperation or in a more passive manner.

Although the emotional states are harder to perceive than the physical states, they are usually accompanied by small physiological changes in our bodies, such as changes in the HR and BR. If we are able to capture these physiological changes, we can then better perceive the changes in the Human emotional states. In past years there has been the emergence of new unobtrusive techniques that are able to capture these physiological signals without the humans' participation, which also tackles the cooperation issue faced by many systems. As such, the use of these unobtrusive solutions could be a useful approach to endow the HITLCPS with better sensing capabilities. In chapter 4 we explore this type of solutions and present a review of the available state of the art.

As previously stated, another aspect of the HITLCPS that can include the human action is the inference of states (present and future). In recent years, many authors have pointed out the benefits of fusing machine learning systems with human intelligence, by having the Human–in–the–Loop of machine learning (Monarch [2021]). This approach is normally done with the goal of: achieving higher accuracy of the machine learning models; having faster times to reach a machine learning target; or increasing the efficiency of human tasks with the assistance of machine learning models. Nowadays, most of the machine learning models, used in decision systems, use supervised learning ($\approx$90% according to Monarch [2021]), as such most of these systems can benefit from the insertion of the Human–in–the–Loop of the system. We believe that this is another open issue to be explored in the HITLCPS paradigm.

Lastly, when considering the actuation phase of the HITLCPS, we can see from the reviewed work that most of the approaches either do not present any actuation or present a very limited actuation, by only showing numeric values to the user. In the HITLCPS paradigm, the human can be the actuator over the system, or even over themselves by changing their own actions. However, in order for this to happen the system should be able to give rich and detailed feedback to the humans that are in the loop. We believe that an approach based on interactive agents (i.e., chatbots) can be beneficial for the adoption of HITLCPS. The use of interactive agents allows for a more human–like interaction between the system and the humans, which can help with the participation/engagement of users of the system. Furthermore, these agents are usually context–aware, as such they are able to use context information to give better feedback to the users, and they are able to follow up on past questions or information requests from the users. We believe that the adoption of a more context–aware approach (e.g., interactive agents) for actuation in HITLCPS is also another challenge of these systems.

Additionally, as we have stated previously, privacy is also an important issue to be addressed in the context of HITLCPSs. However, in order to narrow the scope of this thesis we will not address these issues and left them for future work.

In order to achieve the paradigm of the HITLCPS we believe that first, we need to tackle the aforementioned open issues and challenges. In the next section, we will propose a new model for HITLCPS that aims to tackle these issues. Furthermore, in the next chapters, we present the implementation of several case studies, as well as the results from different experimental trials that aim to tackle these challenges.

3.2 New model of HITLCPS

As we have stated, we believe that there are still several open issues to be addressed with current systems in order to achieve the HITLCPS paradigm. Furthermore, we also believe that there is still room to improve on other proposals for the HITLCPS paradigm. A HITLCPS should go beyond a human–centric application where data about the humans is collected to create classification or statistic models. These systems should account for humans' actions, emotions, intents and desires. Furthermore, they should be aware of the context of the humans and the environment that surrounds them. And they should include the human in all phases of the control loop.

In Figure 3.1 we present a representation of what we believe should be a complete HITLCPS. It is not a requirement that all the components of this model be included in every possible system, but we firmly believe that most systems would benefit from the use of this model. We can see from the Figure 3.1 that a HITLCPS is similar to a normal concept of cyber-physical system by having a control loop constituted of 3 main phases. Namely, a data acquisition phase, an inference phase and an actuation phase. However, by including the human in

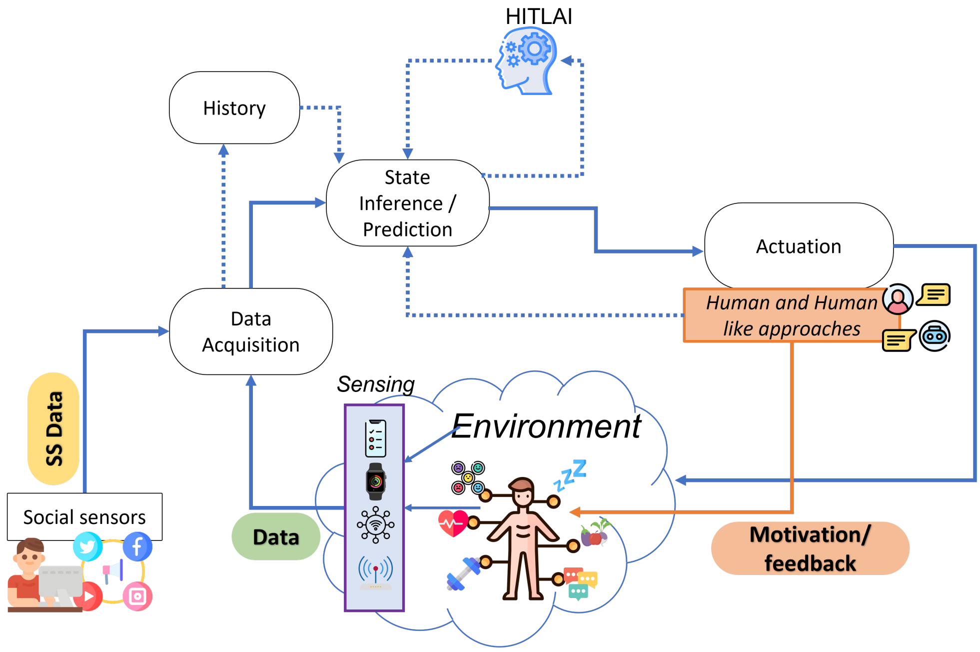


Figure 3.1: Representation of the proposed Human–In–The–Loop–Cyber–Physical Systems model.

the loop of a CPS we increase the complexity of that system exponentially. And as such, several changes should be made to the normal concept of control loop in order to achieve the HITLCPS paradigm.

3.2.1 HITL of data acquisition

Concerning the data acquisition phase of a control loop, we can see from the figure, that there are two main focus for the data acquisition process. Namely, the environment and the humans. The environment represents all other metric external to the human that concern the system. For instance, in a system that monitors a factory, the environment would comprise all other elements of that factory, such as machines, temperature, quality of air, supply chains, processes, etc. The environment is of course a valid source of data for any system since most of the time it is directly related to the performance of the system.

As for the humans, as we represent in the diagram, we should go beyond the visible and try to monitor all the aspects that comprise the humans' physical and emotional states. For instance, their activity levels, location and movements, sleep patterns, emotions, sociability and interactions with other humans, physiological signals, dietary patterns or even interactions with other elements of the environment. Humans are very complex beings, and as such, we should try to acquire as much data as possible in order to make sense of them.

Additionally, some aspects of human beings are even more difficult to evaluate such as their emotional states. For several centuries many researchers have aimed to gasp and understand human emotions and quantify them. As an output of that research efforts, we have at our disposal tools that have been validated to perceive emotions and other psych elements (e.g., questionnaires). As such, we should also include human knowledge, which goes beyond the technological one, in our HITLCPS. For instance, by including adaptations of the questionnaire used in psychology and sociology to perceive human emotions.

As we can also see from the figure, in order to convert the information from the physical world into data for our systems we must use the sensing techniques available to us (i.e., physical, virtual and social sensors). One way to perform sensing about the humans and the environment is by using the IoT devices that are available to us. The number of *"smart"* devices is increasing, and as such, we can leverage them to acquire data. IoT devices have sensors and communication capabilities that make them perfect candidates for being used on the HITLCPS paradigm. Additionally, the concept of IoT has long gone beyond just including constraint devices and wireless sensor networks, it now also includes wearables and smartphones.

Smartphones especially are one of the best candidates to perform sensing of humans and their surroundings. Almost every human carries at least one smartphone with them at all times. Humans are also beginning to use these devices at evermore younger ages, and although many researchers point to the negative impacts of this trend it also offers a way to monitor humans even at young age. Smartphones have several sensors that allow us to capture information from the physical world. Additionally, due to their nature we can also acquire context information such as the number of conversations/interactions with other users, location of the user or even usage statistics of applications. As such, we believe that the use of smartphones as a sensing techniques is one of the approaches to be taken towards the development of HITLCPS.

As we stated before, we believe that HITLCPS should also be able to monitor the physiological signals of humans. Since these signals are not only a way to perceive humans physiological well–being but can also be a way to monitor one's emotions. Wearable devices are a way to capture these signals, since more recent devices are able to capture HR, oxygen saturation or even ECG. However, there have been approaches to develop other types of technologies that are able to capture these types of signals in an unobtrusive and passive manner. In Figure 3.1 we represent these types of devices with an access point figure. However, the unobtrusive techniques are very heterogeneous in their nature and approach to the sensing task. In chapter 4 we will better introduce this type of sensing approach and present several examples of human monitoring, to be used in a HITLCPS, through a review of the state-of-the-art.

Additionally, in the HITLCPS paradigm sensing entities can go beyond physical and virtual entities. In fact, humans can be seen as sensors themselves, this concept is also known as social sensors. Humans can interact with the physical world, perceive events and give both subjective and objective points of view about occurrences, they can report their feelings and emotions, and perceive the state of other humans around them. Furthermore, we can interact with technology to transform this information that we are able to gather into data. Due to the mass use of OSN, human–originated data is probably one of the bigger sources of data. As such, social sensors can be used in a HITLCPS to obtain both information about the environment or the human. In Figure 3.1 we represent the social sensors detached from the environment for ease of view. In fact, these sensors can be both inside the environment and outside. For instance, in the context of a system that monitors the performance of a factory, a social sensor can be a worker that reports their emotions online, but can also be a consumer of any product of that factory that vents their frustration online for a missed delivery deadline.

In summary, we must leverage all the sensing techniques at our disposal to better perceive humans and the environment. Additionally, the data acquisition process of HITLCPSs must go beyond the physical world and capture the emotions and mental states of humans as well. In order to achieve this, we must leverage the knowledge of other areas of expertise, such as psychology and sociology. Furthermore, humans are not only a subject of the sensing process but also an active agent of it.

3.2.2 HITL of state inference

The other phase of the control loop is the state inference phase, which could be seen as the brain of the system. This phase comprises any inference made about the humans states or the environment. In traditional control systems, this is what makes the decision about whether to perform the actuation or not, or the parameters of the actuation. For instance, in a system that controls a factory process, the inference state would include models to determine if a given machine needs to be cooled down and if so, what temperature it should reach. When dealing with humans the systems become more complex, with the inference of state being for instance to determinate the quality of sleep of workers or their sociability values. Due to their complexity, when dealing with HITLCPS the models used for inference of state go beyond mathematical systems representations. In fact, they comprise complex machine learning models, that could be applied to perform classification, regression, or pattern recognition of the several human aspects of the system.

As we can see from Figure 3.1 data acquired from the data acquisition phase is passed directly to the inference phase. Which can include data from all sensors present in the system (physical, virtual, social) and both from the humans and the environment. As we explored in the last section, the data acquired can be from several distinct devices, from different aspects of the human life and come in distinct formats. As such, this phase of the HITLCPS is also responsible for dealing with the heterogeneity of data sources and data types.

Additionally, as can also be seen from Figure 3.1, models to perform state inference will not deal only with present data. HITLCPSs must also have awareness of past history of the system, the environment and the human. Humans use their past actions, events, and past outcomes from said actions to define their present actions. For HITLCPSs to be able to modulate the human behaviour, these systems must also use past data to perform inferences.

Furthermore, humans are very volatile beings, as such HITLCPSs must not only modulate the system for present actions, intents and emotions, but also for the future. Due to the volatility of humans, when the system accounts for a present action, it is possible that the human, might have already followed a different course of action. For instance, it is not uncommon for a human to enter a division of their home with a given intent, and forget what was the finality, moving then to another task. For that reason, the inference of state in a HITLCPS should go beyond the present or near future and use predictive models to predict future actions, intents and emotions of the human.

As can be seen from Figure 3.1, another aspect to be explored in the state inference phase of the HITLCPS paradigm is the use of HITLAI. In past years, machine learning models have seen a big evolution, with the advances in technology allowing for the creation of faster and more reliable models. However, there are some tasks in which no machine learning model can perform as well as humans. Furthermore, most machine learning models are constraint by the quality and amount of data available to them. HITLAI refers to the concept of using human knowledge to overcome these limitations. When considering the HITLAI approach there are two main ways in which humans can interact with inference models. On one hand, humans can perform labelling and validation of instances and give feedback to the system about wrong classifications or predictions. This can be used to have the system deployed in real scenarios even when if there is not a large amount of data available for training, since the system can improve their performance and increase their training dataset overtime with the human feedback. One the other hand, HITLAI refers to the use of machine learning models as decision–helping mechanisms for human-controlled systems. That is machine learning models, are used to give more information or context to humans that will make an inference based on their knowledge and the models' information. The humans in the HITLAI can be either external to the system (humans with the knowledge base for the specific model) or even the same humans that are included in the system (subjects of the data acquisition and actuation).

Another source of information for the state inference phase can be the retro–feedback of the human about the performed actuation. As we will explore in the next section, in the HITLCPS paradigm the actuation is unidirectional. In fact, the feedback of the humans about a given actuation in the system can be used as historic information for future predictions. As can be seen from the flux represented in Figure 3.1, we propose that the feedback given by humans about a given actuation, over the humans or over the system, should be used as retro–feedback to the state inference phase.

In summary, in order to have a system that can modulate human actions, intents, behaviours and emotions, we need to empower our systems with machine learning models and artificial intelligence. We must deal with the heterogeneity of the available data and data sources and reason with it. The models present in the state inference phase should be able to use past information to predict not only the present but the future as well. Additionally, humans should be included in the loop of the artificial intelligence

3.2.3 HITL of actuation

The last phase of any controlled-loop system is the actuation phase, where the system makes the changes necessary to increase its performance. In traditional CPS, the actuation could be for instance regulating the temperature of a given chemical process. In a HITLCPS we should go beyond performing changes to the system and to the environment, that is the actuation should be performed on the humans as well.

Human-beings are mindful and have control over their own actions, as such we cannot force them to change their behaviours or perform a given action. The actuation over the humans should be given in the form of motivation and feedback. For instance, in the context of a system that monitors the performance of a factory, if the system detects that a given worker is continuously performing better than the rest it could give him an incentivization message to keep up the *"good work"*. The feedback and motivation given by the system, should not only address positive states but also negative ones, that is the system should perform both positive and negative reinforcement actuation.

As we explored previously, humans are very volatile as such if the actuation is poorly designed it could have the contrary effect to the one desired. As such, in the paradigm of HITLCPS, the design of the systems should not only evaluate what actuation to perform but also how it could be performed.

In a HITLCPS, humans should be fully integrated in the system, as such they are not only a focus of the actuation, but they can also be a mean to actuate over the physical and virtual world, and over other humans. Especially, when considering actuation over other humans, this could be perhaps the most efficient way of actuation. As we stated, actuation over the humans its *"suggestive"*, as such we should try to maximize the acceptance of the given feedback and motivation. We believe that human interaction is perhaps the best approach to maximize the chances of a given actuation being accepted by a human, that is using humans to relay the message of the system to other humans that need actuation. For instance, in the context of a system that monitors the performance of university students, if a system detects that the performance of a given student is falling, the system could action the human resources department of the university or a professor to convey the situation to the student.

However, this type of approach is not possible in all systems, since there could be systems where there are not any humans to perform actuation, or systems where sharing information could have privacy and security concerns (e.g., medical field). In those cases, we must resort to available technologies. With the advances in machine–learning and especially in the field of NLP it is now possible to develop interactive agents, also known as chatbots, that are able of understanding human text and speech. These agents are contextually aware and can give personalized responses depending on the user. Additionally, they can also give responses based on past conversations, allowing them to perform long and contextualized conversations. These functionalities, allow us to create systems that can perform human–like interaction with their users.

Another advantage of this approach based on interactive agents, is the fact that the registry of conversations between users and agents can be stored and analysed. Some third–party libraries already offer the option to not only perform NLP but also sentiment analysis based on the text collected. As such, these agents offer us another way to not only perform actuation in a more automated and scalable way, but they also can be an additional source of data about the humans.

Additionally, we believe that in the HITLCPS paradigm the actuation over the humans should not be unidirectional. Humans can not only act but also react, as such in a HITLCPS the system should be able to receive the feedback of the humans about the performed actuation. This is represented in Figure 3.1 as a flux of data from the human/human–like actuation back to the state inference phase of the system. This could also be seen as an extension of the HITLAI, since the humans are giving a retro–feedback about the misclassified states. This allows the system to adjust rapidly to any misclassified state, without having to wait for further information acquisition in order to evaluate the impact of its actuation.

In summary, in the HITLCPS paradigm, the actuation should go beyond the environment and also include humans. However, humans have minds of their own and actuation should be performed only as feedback and motivation. We must also design our systems to maximize the acceptance of this feedback by the humans. Using other humans or interactive agents/chatbots is the more reliable approach to increase actuation acceptance. Lastly, in HITLCPSs actuation should not be unidirectional, the systems should be able to learn from the interactions with the humans.

3.3 Chapter Summary

In this chapter, we start by presenting an overview of the open issues found in our review of the state-of-the-art of HITLCPS approaches. In the HITLCPS paradigm humans should be part of all the phases of the control loop, as such there are open issues related to the integration of the human in each of these phases. Some of these issues are, for instance, the systems not considering significant sources of data (e.g., social sensors), lacking the data acquisition capabilities to monitor emotions, not including the human in the inference phase, or even lack of appropriated mechanisms of actuation.

In order to achieve the HITLCPS paradigm we believe that first, we need to tackle these open issues and challenges. As such, in this chapter, we also propose a new model for HITLCPS that aims to tackle these issues. This model proposes the use of social sensors, IoT (smartphones and other devices), using context–specific human knowledge (e.g., psychology–validated questionnaires), and novel unobtrusive sensing techniques to solve the issues related with the data acquisition phase. Additionally, this model proposes the use of HITLAI mechanisms to further enhance the state inference phase and deal with the open issues related with it. Lastly, we believe that the actuation in the HITLCPS paradigm must be either human–based or human–like, as such the presented model proposes the use of interactive agents to tackle the issues of this phase.

One of the open issues found was the lack of proper solutions to monitor humans' vital signals and emotions. As such, we study the use of novel unobtrusive solutions to obtain these metrics. In the next chapter, we present more details about the state-of-the-art of unobtrusive sensing and propose a taxonomy to better classify it.

Chapter 4

A Survey Of Unobtrusive Sensing

Contents

In order to achieve the HITLCPS paradigm, we first need to endow our systems with sensory capabilities to perceive the humans beings. Human actions, most of the time, can be unpredictable when analysed by a common observer. However, trained observers can perceive certain indicators that give them information to classify those actions or even predict future states. These indications are often accompanied by involuntary physiological reactions (e.g., fluctuations in the HR, or an irregular BR). As such, by creating and incorporating sensors that are able to gather information about those physiological responses in our systems, we are, in turn, moving towards empowering them with the ability to perceive and understand humans.

Most of the efforts towards creating sensors and systems able to capture these reactions of the Human body have been done in the field of wearable sensors, and body sensor networks. However, despite their advantage, this approach also carries drawbacks, since these devices can be bothersome to use, and most of them require some cooperation from the user. Even with solutions like the one in Atalay et al. [2017], which rely on wearable sensors that are embedded in clothing, we still have to take into account that the user has to wear that specific piece of cloth and have some special care in order to maintain the system in good condition. Furthermore, recent surveys showed that 32% of wearable device users stopped using the device after 6 months, and 50% of users stopped using it after the first year (Ledger and McCaffrey [2014]). Additionally, although being around for some decades, these devices have still to be accepted by the medical community, as many of them still present a higher error rate than the standard medical monitoring devices.

On the other hand, there has been the emergence of several approaches to unobtrusive sensing inside the scientific community. These approaches propose, novel ways to monitor physiological signals (i.e., HR and BR), that do not require the user cooperation. In this section we explore these approaches, by defining a taxonomy for this type of technology, reviewing the existing approaches and classifying these approaches.

4.1 A Taxonomy for Unobtrusive Sensing

Before we proceed to identify and present existing approaches to unobtrusive sensing of human beings, it is essential to define a general taxonomy for human sensing that can set the terminology and serve us as a guide for the discussion that follows. The proposed taxonomy can be seen in Figure 4.1. In this taxonomy we propose a three-level division, with a fourth common level. These are briefly presented below.

In the first level, we divide sensing into obtrusive and unobtrusive. On one hand, unobtrusive sensing solutions are able to monitor the users in a continuous way or during large temporal windows, in a contactless way and without requiring

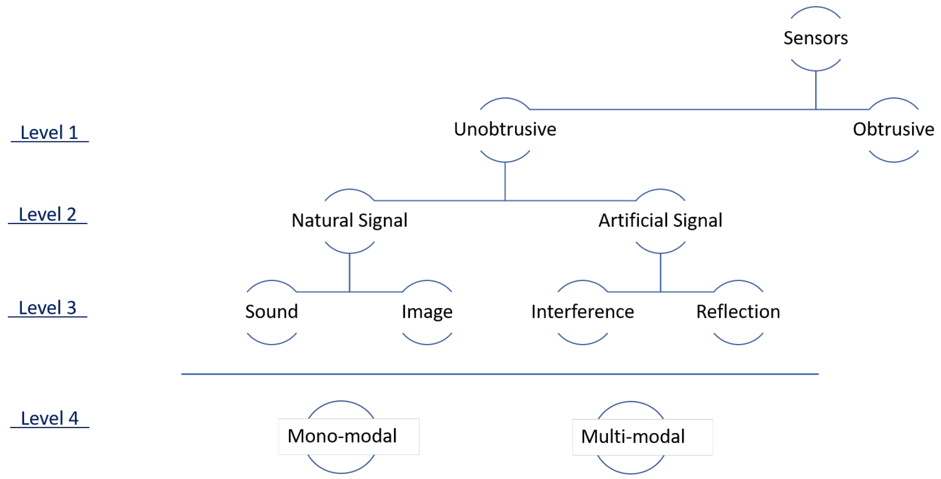


Figure 4.1: Proposed human sensing taxonomy

any specific actions from the users. On the other hand, unobtrusiveness cannot be defined as just being contactless, or user-interaction free, as the negative emotional effects of a user's perception of the sensing technique must also be considered obtrusive. For instance, some studies have addressed the negative effects of surveillance, and how it correlates with anxiety (O'Connor and Jahan [2014]). However, in this review we specifically address physical obtrusiveness. As such, we consider any solution that requires physical contact or explicit human actions as being obtrusive (e.g., wearables, ECG, etc.). In this paper we will focus mainly on solutions that set as their goal to monitor humans' physiological signals (e.g., heart rate, respiratory rate, activity detection) and emotional states (e.g., happiness, sadness, etc.).

The second level branch considers the nature/origin of the signal being used for sensing. Two clear patterns were noticed during the review of existing techniques, corresponding to two possible origins of the sensed signal. The signal can be natural or it can be artificial. Natural signals are originated by human beings themselves (e.g., heart-beat sound, respiration sound, human body motion images) and are then captured by the sensing technique. Artificial signals are generated for the purpose of monitoring humans. They are created by the sensing technique itself, interact with their target, and are then captured by sensing system (e.g., radar waves).

At the third level, for the natural signals branch, the division is on how the signal is being captured. Much like human beings perceive information from the environment through their senses, systems have been empowered with capabilities that emulate those senses, such as cameras to emulate the vision, microphones to emulate our hearing, or pressure sensors that emulate the sense of touch. For the purpose of unobtrusive sensing of physical and emotional states, only image and sound are used.

Still at the third taxonomy level, when considering signals of artificial nature, the signal is always external to the human body. When that signal interacts with the human body it is transformed and then perceived by the used technology.

Upon analyzing the state-of-the-art, two physic phenomena seem to be the most commonly used: signal interference, and signal reflection. Therefore, those were the considered classes for the Artificial Signals branch.

Last but not least, every third level node will be evaluated in terms of how many modalities it can work on, that is, if the technique can capture more than one physiological signal or if it can be used to perceive one or more states (e.g., physiological, emotional). For example, a technique that can only be used to monitor breathing rate, will be mono-modal while one that can monitor both the breathing rate and the heart rate will be considered multimodal. The same happens for one technique that can capture only the heart rate, against one that captures both the heart rate and stress level.

4.2 Existing Approaches to Unobtrusive sensing

This section provides an overview of the state-of-the-art of unobtrusive sensing of humans. For this, the taxonomy proposed in section 4.1 is used. In sub-section 4.2.1, we present and discuss approaches that rely on natural signals, whereas in sub-section 4.2.2 approaches that use artificial signals are addressed. Subdivisions of these sub-sections are also done according to the taxonomy.

4.2.1 Natural Signals

We consider a signal to be natural if it is directly produced by the human body (e.g., heat, breathing sound, body's image, speech, etc.). In this regard, sensing techniques may explore two types of signals, namely sound and image.

4.2.1.1 Sound-Based Techniques

The work by Jia et al. [2016] uses a geophone (sm 24 [0 10]) to capture the changes in a person heart rate during sleep. geophone is more commonly used to monitor earthquakes, but these devices can also generate a noticeable response to sounds such as those created by the sound of a beating heart. Other authors have proposed solutions based on custom bed or altered bed parts to monitor humans' heart rate, such as the works of Chee et al. [2005] and Watanabe et al. [2005]. However, these solutions require specific alterations to someone's bed and are not suited for wide adoption.

Geophones are insensitive to low frequency movements. Due to this feature, it automatically filters out any signal caused by low frequency movements, such as respiration. However, geophones are still able to capture other movements, such as rolling on the bed or even other people walking next to the bed. As such, these movements need to be filtered out in order to obtain a clear signal that can be compared to the heart rate signal. In Jia et al. [2016], the authors were able to obtain the number of sound peaks in a time window and, by extrapolation of that value to a one-minute window, the heart rate was obtained.

Two experiments were also conducted, one in a controlled environment and a second one in real apartments. The first experiment comprised 34 healthy

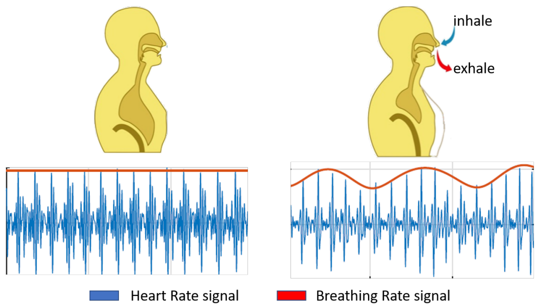


Figure 4.2: Signal modulated in amplitude by the respiratory rate. The signal at the left was obtained while the subject held his/her breath. The right-hand side signal was obtained while normally breathing. Image adapted from Jia et al. [2017].

subjects, including 26 males and 8 females, with ages between 22 and 65 years. An average heart rate estimation error of 1.30% was obtained when the subjects were lying still, as opposed to an average estimation error of 3.87% when the subjects were asked to perform some movements during the experiment. In the second scenario, the system was installed in the homes of 9 different subjects during a period of 25 nights, and a mean estimation error of 8.25% was obtained. All of these results were obtained by comparing the results with those from a finger oximeter, and correlating the data with a camera footage in the second experiment.

Some of the authors of the previous work decided to extend their work in Jia et al. [2017], by demonstrating that this technique was able to estimate the breathing rate and, furthermore, that it was able to monitor two persons sleeping in the same bed at the same time.

Concerning the first problem, due to the geophone being insensitive to low-frequency signals, it cannot normally detect the breathing rate. Nevertheless, after some experimentation and observations, the authors concluded that respiration modulates the geophone signal in amplitude. This can be seen in Figure 4.2, where the graphic on the left corresponds to the signal obtained while the subject held his breath, and the graphic on the right corresponds to the signal obtained while the subject breathed normally. This can be explained because breathing changes the amount of air in the chest and, in turn, its stiffness, and, consequently, the amount of energy that the heartbeat sound loses while crossing it. This phenomenon can be leveraged, in turn, to estimate the breathing rate.

The second problem addressed in Jia et al. [2017] was the use of the technique developed in Jia et al. [2016] to perform the detection of the heart rate and breathing rate for 2 persons while sharing a bed. The used system representation can be seen in Figure 4.3. By using two synchronized geophones, and by knowing each geophone's relative location to each subject (that is, the geophone G1 is closer to H1 than to H2, and the inverse situation happens to geophone G2), it was possible to separate the signals into two distinct signals, making it possible to estimate both the heart rate and the breathing rate of each of the two individuals.

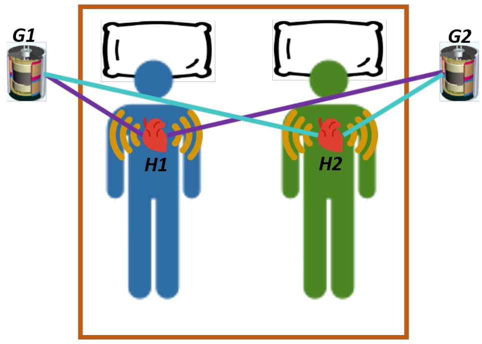


Figure 4.3: Representation of the system used in Jia et al. [2017] for two people in the same bed. With two geophones (G1 and G2) and two heartbeats sources H1 and H2.

In this work, tests were also performed with 86 participants, and a breathing rate estimation error of 0.38 breaths per minute was obtained, for a single person. For tests with two persons, sharing the same bed, an average estimation error of 1.90 beats per minute was obtained for the heart rate estimation, and 2.62 breaths per minute for the breathing rate estimation.

Smartphones are now ubiquitous devices, carried by most people and kept near them even while sleeping. Commercial sensors that aim to monitor sleep quality, such as Jawbone [2018] and Fitbit [2022], also leverage smartphones to store and process data. The work by Ren et al. [2015, 2019] discards the use of wearable sensors and proposes a new framework for breathing rate and sleep quality monitoring. In this framework, the authors used a common off-the-shelf smartphone and an earphone to monitor the sleep activity and estimate the breathing rate. Although, smartphones have a microphone, earphones have several advantages: firstly, earphone microphones have better audio quality; secondly, some users are reluctant to keeping smartphone close to them due to radiation; thirdly, earbuds can be used as microphones as well, thus increasing the recording capabilities and allowing them to record in stereo; lastly the use of earphones can increase the distance at which the device works. The authors also argue that despite the fact that earphones are additional devices to the system, they are a very common device that normally comes with the smartphone itself, and, as such, they are only reusing existing devices.

The framework presented in Ren et al. [2019] comprises three modules, namely, noise reduction, breathing rate detection, and sleep event detection. Although, most people sleep in a relatively quiet environment, there are always some sources of background noise, such as other electronic devices, pets, or even outside noises like cars. The first step to noise filtering is to apply a band-pass filter to remove high and low frequencies that are not in the range of the breathing frequencies. The biggest difference between ambient noise and respiration noise is stability, as the amplitude of ambient noise does not vary during small periods of time. As such, by computing the variability of the signal, it is possible to detect the frames that only contain ambient noise. The final step of this procedure is to subtract the noise from the signal that contains the breathing information and obtain a clean signal. After cleaning the signal, the authors first extracted the envelope from the acoustic sound. They then used the strong correlation relationship between the breathing cycles to capture the time length between the breathing cycles and, from there, they derived the breathing rate. The system was tested with nine people throughout a period of six months, and the average estimation error for the breathing rate was 0.5 breaths per minute.

In Ren et al. [2019], the system was also applied in a case study to detect and classify the level of sleep apnea. The results of the system correlate with the truth, since the system was able to correctly classify the three subjects involved in the study. The focus on the clinical area also shows another possible applicability of this type of technique.

In addition, to using sound-based techniques to detect physiological states, there is some work that explores the use of sound for detecting emotions. Apart from visual clues, one of the most reliable ways for humans to detect emotion is through speech. The goal of automatically detecting emotion through voice has also been around for quite some time, with some works, such as McGilloway et al. [2000], showing results in this field. Other authors have applied sound-based techniques to detect fear-type emotions, in dangerous or unpleasant situations for human beings, showing that these systems can also be used for increasing and promoting human safety (Clavel et al. [2008]).

This area has also seen some developments in recent years with the emergence of deep neural networks (Huang et al. [2019a]). In the latter work, the authors focused on emotion recognition not only from speech but also from non-verbal sounds. Additionally, there is also the approach of converting speech to text and processing the text in order to infer sentiment (Cambria [2016]). However, with this approach we lose information, such as non-verbal sounds and sentiment, and we consider that the sensing is not direct. As such, we chose not to include this area in this review of existing works.

4.2.1.2 Image-Based Techniques

Several studies exploit image and video techniques in order to monitor human activities, showing that this technique is quite effective (Pareek and Thakkar [2021]). Nevertheless, there is still an ongoing effort to use images and videos in order to monitor more than just activities.

The widespread use of camera-equipped devices (e.g., laptops, cameras in the dashboard of cars, smartphones), opens up opportunities to leverage the physiological phenomenon of skin color changes in order to unobtrusively capture the human heart rate. Some studies have already explored the possibility of exploiting cameras and flash as sensors, by analyzing variations of light reflection with the change of blood volume in fingers (Scully et al. [2011]). However, this requires the user to hold his/her finger against the camera's flash and thus falls under the category of obtrusive sensing.

Because cardiac pulsation leads to subtle skin color changes, a Photoplethysmography (PPG) signal can be measured through video analysis.

One of the body parts which is more susceptible to color changes is the face, and some studies try to exploit this in order to capture the heart rate signal. Works such as the one from Kwon et al. [2012], already proved that smartphones can indeed be used to monitor the heart rate through video recording. In that work, the author developed a mobile application that by recording 20 seconds of human face could estimate the heart rate of a person with an average error rate of 1.04% for the 10 trial participants. However, this technique has limitations as an unobtrusive technique, since the person has to stay still for the duration of the monitoring. Furthermore, high frequency parameters such as the heart rate variability could only be extracted from smartphones that supported high frame rates. However, current day smartphones have drastically evolved, especially in terms of hardware. Nowadays, smartphones can have up to 4 rear cameras with more than 40 Megapixels, record videos in 4k, and even capture slow-motion at more than 480 frames per second, which could lead to even better results.

The studies in Kwon et al. [2015] and Li et al. [2014] both use mounted cameras in order to explore the detection of the heart rate based on skin color changes in realistic scenarios, i.e., scenarios with movement, bad illumination, and noise. Both works try to detect the Region of Interest (ROI) for facial detection. The results from Kwon et al. [2015] show that the forehead and both cheeks are good candidates for computationally efficient ROI, while chin and nose are less suitable. These findings can address some of the limitations of Kwon et al. [2012], as the fact that the cheeks are a suitable ROI allows to monitor the heart rate not only in a frontal facial position but also from both facial profiles. In Li et al. [2014], the authors also explored the possibility to monitor the heart rate of a person while playing a video game. This could be used to monitor user experience throughout the game, and help game developers in their design work.

In Massaroni et al. [2018], the authors focused on detecting the breathing rate and patterns by using a camera to monitor the movements of a pit in the neck. The authors present results for breath-by-breath respiratory rate, which is estimated from the processed breathing pattern. In addition, the effect of image resolution on monitoring breathing patterns and respiratory rate is also addressed, by comparing different camera resolutions. This system was tested on a group of twelve participants which were all healthy. The system showed a mean average error of 1.53 breaths per minute with the worst resolution, and of 0.55 breaths per minute with the best resolution.

Other than image techniques that use the visual spectrum, there have also been efforts towards using Infrared imaging to monitor people physiological signals. In Pereira et al. [2015], the authors employed a long wave infrared camera to capture the breathing rate of people. In this work the authors explored the previously discussed technique of ROIs detection to segment the patient nose. Secondly, the authors segmented a second ROI in the region of the nostrils. The temperature around the nostrils fluctuates during the respiratory cycle (inspiration and expiration), making it possible to monitor the breathing rate from the subtle temperature changes (≈0.3ºC). Experimentation of this technique was also performed with 11 healthy subjects, obtaining a mean average error of 0.71 breaths per minute.

Previous work had also proved that thermography could be used to detect the Heart Rate (Garbey et al. [2007]). In this work, the authors based their work in the fact that the variance of skin temperature is strongest along the superficial blood vessels. The authors were able to detect the heart rate changes from major superficial vessels, such as those on the face, the carotid artery and the radial vessel complex. After segmenting the area of interest of the vessels and filtering the image, the authors computed the Fast Fourier Transform and found out the most relevant frequency between 0.67Hz and 1.67Hz (40-100 bpms). The system was tested with 34 subjects obtaining a performance of 90.33%. Other authors have also employed the use of thermal cameras to detect the heart rate and breathing rate of people during their sleep (Hu et al. [2018]). This system used an array of infrared lighting and a thermal imaging camera, and a custom-made deep learning model to detect ROI and infer people heart rate and breathing rate. The created system was tested with 26 and 25 sleeping subjects for the breathing rate and heart rate estimation, respectively, obtaining a mean estimation error of 1.865 breathing per minute and a mean estimation error of 4.293 beatings per minute. This shows that this kind of system can also be used in sleep monitoring, where normal imaging techniques cannot be used due to a lack of proper illumination.

Other authors focused on the practicality of these techniques in the medical area. Currently, to monitor newborns' vital signs, such as heart rate, breathing rate or oxygen saturation, it is necessary to have sensors and electrodes sticking to the skin. This can bruise their vulnerable skin, cause infections, cause stress or even pain. The use of unobtrusive techniques could therefore bring several advantages to the current state-of-the-art. In Klaessens et al. [2014] and Abbas et al. [2011], in addition to the use of video recording to measure the heart rate, the previously discussed infrared technique of estimating the respiration rate by capturing the temperature around the nostrils was also used to monitor infants' breathing cycles. However, in Villarroel et al. [2014] the authors were able to monitor newborns' heart rate, breathing rate, and oxygen's blood saturation with just a normal video camera mounted over their respective incubators. They were able to achieve this by filtering the image signal in the frequency domain. In this work, the authors went even further into the clinical realm, and were able to detect bradycardia accompanied by a major desaturation.

The use of unobtrusive techniques relying on natural signals can also be applied to emotional state inference. Some features extracted from physiological signals, such as heart rate variability, can be used to infer emotional states, as several studies proved (Appelhans and Luecken [2006]; Lane et al. [2009]). However, some works focus on directly detecting the emotional state using the facial expression of people. This happens in Fernández-Caballero et al. [2016], where the authors use a multi-modal framework for smart homes, constituted by three different components to perceive emotions from the users. The used modules enable facial emotion detection through video, behaviour detection through video, and valence/arousal detection with physiological wearable sensors. The

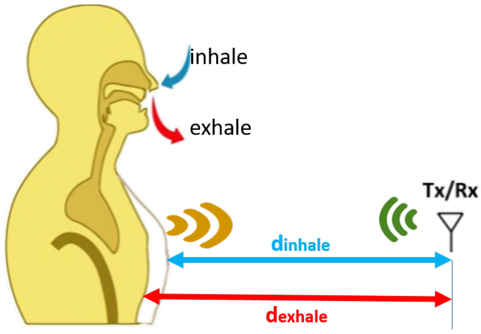


Figure 4.4: Scheme adapted from Adib et al. [2015]: Representation of the inhale and exhale events in the respiration cycle, and the respective distance from the chest to the radio antenna in both events.

fusion of the three modules happens at decision level, that is, each module works separately and a decision rule system then processes each output and gives a final classification. Although, the latter module does not fit our classification of unobtrusive sensors, since the fusion happens at decision level, we can envision a framework constituted by the other two modules only, which would be unobtrusive. Additionally, the aim of this work is not only to detect emotions, but also to influence them, through music and color/light, thus providing an example on how unobtrusive sensing techniques can open ways to improve our daily lives and mindsets.

4.2.2 Artificial Signals

Contrary to the techniques presented in the previous subsection, there are techniques that resort to signals that are external to the Human body, in order to monitor humans. Those techniques fall into the “Artificial Signals” branch of our taxonomy. There are two types of signal-body interactions used by these unobtrusive techniques, namely signal reflection and signal interference. These are addressed below.

4.2.2.1 Reflection-Based Techniques

One of the studies in the area of unobtrusive sensing that uses the phenomenon of signal-body reflection is presented in Adib et al. [2014]. In this study, a Frequency Modulated Continuous Wave (FMCW) radio wave was used to monitor the location of people in a 3D setup, by using the reflection effect of radio waves and T-shaped antenna arrays. By modulating the signal in frequency and capturing its reflection, the authors were able to capture the time of flight of the wave, which directly correlates to the distance to that object. Furthermore, by using one radio and 3 antennas, they were able to triangulate the signal and pin-point the location of an object or person in 3 dimensions. Due to the fact that radio waves can cross walls, in this work they were also able to detect people location and movements even when the device had no line of sight to the subject.

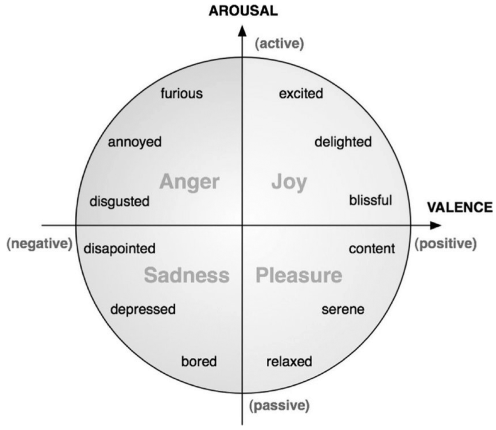


Figure 4.5: Two dimensional emotional model, figure adapted from Chakladar and Chakraborty [2018]

Building on the previous work and on the fact that the respiratory cycle creates chest movements that affect the distance from the subject to the radio source, as can be seen in Figure 4.4, the authors developed the work in Adib et al. [2015]. Using the system in Adib et al. [2014], they transmitted a low power wireless signal and measured the time it took for the signal to travel to the human body and reflect back to its originating antenna. With this, they were able to easily compute the distance to the subject as well as the distance fluctuations caused by breathing cycles and, thus, could easily determine the breathing rate. The output signal of their system corresponds to the phase of the signal that returned to the radar after ricocheting in the human body. In this signal, there is information about the breathing rate and the heart rate, and by resorting to signal theory to filter it in the frequency domain, they were able to extract this information. In this work, the authors used the system to monitor humans, while they stood still or performed actions that did not require considerable movement (e.g., work on their computer, read, browsing the web on their phone). The system has quasi-static requirements in order to function properly, due to the fact that the movements of other parts of the body, other than the chest, caused by moving around or doing a certain activity (e.g., walking, exercising, etc.), cause too much distortion in the signal. The measures were correlated with results from wearable sensors that were attached to the users. The results show that it is possible to obtain an accuracy of 99.4% and 99% for the breathing rate and heart rate, respectively.

In Zhao et al. [2016] the authors, build on the work of Adib et al. [2014] and Adib et al. [2015] by using the information gathered by the system to also monitor human emotions. Using the previously explained technique, they were able to construct a signal for the heart rate and breathing rate in the time domain, from which they then derived the features necessary to perform emotion detection. The features extracted from both signals were based on the work of Kim and André [2008]. They then classified the subjects in a two-dimensional model, whose axes were valence and arousal[1], based on four emotions spaces: sadness (negat-

[1]both of the definitions for arousal and valence can be found in Posner et al. [2005] along

ive valence and negative arousal), anger (negative valence and positive arousal), pleasure (positive valence and negative arousal) and joy (positive valence and positive arousal). An example of these model can be seen in Figure 4.5. The obtained results were compared with two different techniques: ECG-based and vision-based emotion recognition. They obtained results slightly worse than those of the ECG-based technique, and better results than the ones from the vision-based technique, for three of the four emotions under consideration.

Other studies, like the work of Lee et al. [2015b], and Lu et al. [2009], focused on the detection of the breathing rate and heart rate, respectively, by employing microwave sensors to perform contactless monitoring. The characteristics of the microwaves allow them to work at some distance way from the human body, and even go through clothes, being then reflected by the human body. The work in Lu et al. [2009] also had as its primary goal the measuring of heart rate variability based on this technique. Several recordings with 5-minutes duration were performed, for both the microwave sensor and for an ECG. The tests were made with sixteen different male subjects aged between 19 and 27, and no significant difference between the two techniques was found in both the frequency and time domains, as well as in non-linear dynamic analysis of heart rate variability measurements. Although, the results were promising, all the tests were performed under controlled research conditions, and all subjects were healthy. The authors claim that this work can prove to be a practical alternative to ECG for heart rate variability analysis, as it eludes the negative aspects of wires and body sensors.

Another work that addressed the use of signal reflection for determining a person's breathing rate is the work of Nandakumar et al. [2015]. In this work, the authors present a solution for detecting sleep apnea events with smartphones. To achieve this, they developed a solution where an off-the-shelf smartphone is turned into an active sonar system, that emits frequency-modulated sound waves and captures reflections. In this work they use the same approach employed in Adib et al. [2015], that is, the detection of the minute chest movements caused by breathing. The speaker of the smartphone is used to emit a FMCW signal with frequencies between 18 kHz and 20 kHz, and the reflection of these signals is then caught by the Smartphone's microphone. This range of frequencies is very close to the threshold of the human hearing and, as such, this work also addressed the issue of whether the sound was audible. They found out that the majority of people were not able to hear it, while a small minority were able to hear a small noise in quiet environments. This aspect, is also important when considering an unobtrusive solution, since an acute sound can generate discomfort and/or become bothersome and intrusive.

Additionally, in Nandakumar et al. [2015], the FMCW technique was used for monitoring more than one person at the same time, as long as they were separated by at least 20 cm. The authors claim that this technique works even when the subject uses a blanket, and that the system can maintain its accuracy even when the blanket is some centimetres (cm) thick. The system obtained a

with more information about the 2D model mean estimation error of 0.11% for one person with the device up to a meter away from the subject. The error increased with the distance to the device and when monitoring more than one person. This work also presented the results of a clinical trial with 37 subjects, where the solution was used for monitoring the breathing rate, as well as for detecting and classifying apnea events. The system correctly classified 32 out of the 37 subjects, for four types of sleep apnea. An average detection error of 1.9 apnea events per hour was achieved. This shows that this type of technique can also be leveraged in clinical context.

Other authors also proposed similar solutions based on commonly available devices. In Wang et al. [2018], the authors proposed a system based on a normal microphone and a computer speaker. Additionally, in Wang et al. [2021], the authors also proposed a similar system based on Smart Speakers. These devices are becoming more common as we witness an increasing trend in the use of smart devices and in the deployment of smart home solutions. This also addresses the fact, that we may not need new devices to power humans' unobtrusive sensing, as many of our existing devices can be used or modified to work as sensing solutions.

Additionally, there are also some commercial solutions that use this sensing technique, such as Pierce [2017]. While this solution primarily focuses on detecting objects through walls, the development version of this solution already offers an off-the-shelf breathing API that is able to use this board to detect people breathing, using FMCW. The company responsible for the device in I. Socionext America "Socionext CMOS 24-GHz Radar Sensor" [Sunnyvale, 0 10(@], which is mainly used for radar applications such as presence detection and security, also proposed that this device could be used for monitoring the heart rate and breathing rate. This is quite remarkable, since the device is roughly the size of a coin. Furthermore, even smartphone companies are starting to see these radar solutions as an opportunity. One of the recent version of Google's smartphone flag ship, the Google Pixel 4, has an incorporated radar sensor (innovative XENSIV™ 60 GHz radar chip enables things to see and revolutionizes the Human Machine Interface [0 10]). Although, details on the availability of APIs for developers have not yet been disclosed, this could pave the way for new mobile applications for healthcare.

4.2.2.2 Interference-Based Techniques

Several techniques use artificial signals and the signal-body interference phenomenon in order to capture humans' physical or emotional states. The human body allows high frequency signals to pass through it. However, the signal that enters the body is different from the one that leaves it. In Figure 4.6 we can see a representation of this phenomenon, where the signal that enters the body is modulated by its minute movements. The signals can be attenuated and suffer interference, which affects one or more frequency components. Different signal processing techniques can then be applied to detect these changes in frequency.

In Kaltiokallio et al. [2014], the authors leverage this phenomenon, by using a

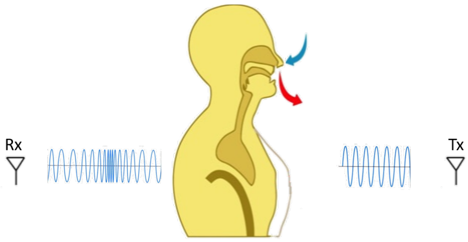


Figure 4.6: Illustrative Scheme the interference phenomena caused by minute movement of the human body in radio frequency signals.

single commercial off-the-shelf transmitter-receiver pair to monitor people respiratory rate, through the interference caused in the Received Signal Strength (RSS). In this work, the authors used a single transmitter node and a single receiver node, where the receiver antenna is also connected to a real-time spectrum analyser to obtain the baseline. The tests where conducted in a sleeping scenario where one person was lying on a king-size bed, and the antennas where placed at each side of the bed, two meters apart, the antennas were also placed at 0.2 meters above the chest. This careful positioning of the equipment creates the best signal-to-noise ratio scenario. Additionally, pre-filtering was also exploited to increase the signal-to-noise ratio in the RSS measurements. The obtained results were compared with those of a real-time spectrum analyser, to prove that the system could obtain comparable results to those of a system that costs 3–4 times more. The system was able to achieve a mean absolute error as low as 0.12 breaths per minute. In this work, however, the authors obtained the breathing frequency in the frequency domain using Power Spectral Density (PSD) and, therefore, they were unable to obtain information in the time domain. Furthermore, in instances of time where the subject is moving, breathing estimation cannot be performed. The authors believe that the work can still be enhanced by exploiting channel diversity to improve the breathing detection ability.

Additionally, in Kaltiokallio et al. [2014], the authors stated that they were building on the work done in Patwari et al. [2013], since they believed that by only using two sensors this would reduce the system complexity and increase it's feasibility. However, due to the complexity of the system in Patwari et al. [2013], other sensing opportunities emerge. In this work, in one of the experiments the authors used a 33-node wireless sensor network to monitor an apartment of 7 by 8 meters. Although, the obtained results were less precise than the results in their follow-up work, the use of several nodes allowed them to perform breathing estimation and location of people in two dimensions. They were able to detect the location of a breathing person with a mean error of 2 meters. Although, this value seems high for applications that require a precise location, these techniques can be useful in conditions such as the ones in search and rescue missions after an earthquake or a tsunami, where information about a relative location (even with a 2 meters error) can be useful.

Other authors have also based their approach in the fact that the human body minute movements interfere with radio signals, in a way that is related to vital signs. In one of those studies, the authors proposed the use of off-the-shelf Wi-Fi devices to track human vital signs (Liu et al. [2015]). In this case, the system is based on the use of the CSI of Wi-Fi signals, which is more suitable for this task than RSS-based approaches. The reason behind this is the fact that RSS is already an aggregation of all of the sub-carriers' signal strength in order to mitigate interference in the signal. The sub-carriers are affected differently depending on their frequency and, as such, there are some sub-carriers that suffer more visible interference caused by human movements. Furthermore, recent comparative studies prove that Wi-fi CSI measurements provide more robust estimations of breathing rates when compared to other radio frequency measurements (Hillyard et al. [2018]).

In this work only one laptop and one Access Point (AP) were used. In a first instance, the CSI data is collected, by using a CSI tool in the laptop (Halperin et al. [2011b]). That data is then filtered, and the system runs an algorithm that takes into account moments during which the person moves and moments during which the person is almost static. Similarly, to what happens in the studies that use the interference phenomena to detect human movements, this technique requires the Signal–to–Noise Ratio (SNR) to be high. As such, moments during which the subject moves around or moves a part of the body prevent the detection of the breathing rate and heart rate. Although, the authors claim that this system can be used in any scenario where quasi-static moments occur, the system was only tested with people in their sleep. In this particular scenario, events like turning over in the bed, or getting in and out of the bed, can contain precious information about the quality of one's sleep and, as such, the authors use this information as well. After filtering the movements, the system uses the filtered signal to estimate the breathing rate and heart rate.

In the mentioned study, the sub-carriers with greater variance in signal strength were selected, as those are the ones that are more affected by the human body in the frequency domain. After the signal is filtered, the peaks of the signal are detected, for each of the selected sub-carriers. The mean value for the location of each peak in all sub-carriers is computed and the respiration rate estimation is then given by calculating the number of peaks in one minute. In this work, it was also demonstrated that it is possible to detect the respiration rate for two people while sleeping in the same bed, without increasing the number of devices. In order to achieve this, the authors used the PSD technique. A strong sinusoidal signal, such as the respiration cycle, generates a frequency peak corresponding to the period of the sinusoidal PSD signal. When two people are being monitored, two strong frequency peaks will appear in the PSD, corresponding to the breathing rate of each person. By applying this technique to each of the selected sub-carriers, the authors used the K-means technique in order to find the two peak clusters that correspond the breathing rate of each person.

Using the same approach, it is also possible to detect the heart rate of a person. The movements caused by the heartbeat are smaller than those caused by breathing and, as such, are more difficult to detect. Nevertheless, the heart rate is higher than the breathing rate. This means that in the frequency domain the heart rate will be represented in higher frequencies than those of the breathing rate and, thus, it is possible to separate both signals in the frequency domain. The heart rate also generates a strong sinusoidal signal, which means that the PSD technique can then be applied to find the stronger component in all of the sub-carriers' signals. The mean value for all of those components can be calculated, which corresponds to the heart rate. Although, in this work the authors do not demonstrate it, they also propose that the same approach used for the breathing rate can be used to obtain the heart rate for two people at the same time. However, contrary to what happens with the respiratory rhythm module, is not possible to obtain a signal that corresponds to the cardiac rhythm in the temporal domain.

Most of the techniques that use the interference phenomenon are only applied to monitor one or two persons at the same time, as it is difficult to interpret interference without knowing the signal propagation path. The work in Wang et al. [2017], however, leverages CSI phase difference data to intelligently estimate the breathing rate of several people at once. The proposed system applies tensor decomposition, namely canonical polyadic decomposition, to obtain multi-persons breathing rates. In this study they demonstrate that, while normal CSI cannot be used to detect more than two persons' breathing rate at the same time with accuracy, the proposed technique can be used to estimate the breathing rate of five people with accuracy. The system was tested for different temporal size windows and sampling rates, obtaining better results when both attributes increase. The system was also tested in three different situations with several line of sight and non-line of sight situations, since the technique works through walls and objects as well. The system obtained an absolute estimation error of 0.9 breaths per minute even through walls for one person, with that value increasing for 2 breaths per minute for five people within the same confined space.

The CSI technique can also be leveraged for detecting emotion, as happens in Gu et al. [2018], where a system called EmoSense is presented. Contrary to what was done in Zhao et al. [2016], where physiological signs unobtrusively extracted were used to perform emotion detection, in this work the physical expression of the subject was captured in order to determine emotions, through CSI measurements. There were three major findings for this study: firstly, CSI was indeed able to capture emotional expression; secondly, the performance of the system depends on the experimental setup; and thirdly, the performance is person-dependent. The system was based on a data driven architecture, where the CSI measures were sent to a server. Classification models were then used to infer one out of four basic emotions, namely happiness, sadness, anger, and fear. The created system was compared to the sensor–based approach and was not able to achieve the same performance. Specifically, the sensor–based approach elicited an accuracy of 95.83%, while EmoSense only reached an accuracy of 80.48%.

4.2.3 Overview

Table 4.1: Distribution of the Reviewed Unobtrusive Sensing Solutions/Approaches per Taxonomy Branch and Application Field

	Natural		Artificial	
	Sound	Image	Reflection	Interference
Physical States	Jia et al. [2016], Jia et al. [2017], Ren et al. [2019]	Kwon et al. [2012], Kwon et al. [2015], Massaroni et al. [2018], Li et al. [2014], Klaessens et al. [2014], Pereira et al. [2015], Garbey et al. [2007], Hu et al. [2018], Abbas et al. [2011], Villarroel et al. [2014]	Adib et al. [2015], Lee et al. [2015b], Lu et al. [2009], Nandakumar et al. [2015], Wang et al. [2021]	Kaltiokallio et al. [2014], Patwari et al. [2013], Liu et al. [2015], Wang et al. [2017]
Emotional States	McGilloway et al. [2000], Huang et al. [2019a], Alonso-Martin et al. [2013], Yoshitomi et al. [2011]	Fernández-Caballero et al. [2016], Alonso-Martin et al. [2013], Yoshitomi et al. [2011]	Zhao et al. [2016]	Gu et al. [2018]
Multimodality	Jia et al. [2017]	Klaessens et al. [2014], Abbas et al. [2011], Villarroel et al. [2014]	Adib et al. [2015], Zhao et al. [2016]	Liu et al. [2015]
Medical Field	Ren et al. [2019]	Villarroel et al. [2014]	Nandakumar et al. [2015]	—
Sleep Monitoring	Jia et al. [2016], Jia et al. [2017], Ren et al. [2019]	Hu et al. [2018]	Nandakumar et al. [2015]	Kaltiokallio et al. [2014], Liu et al. [2015]

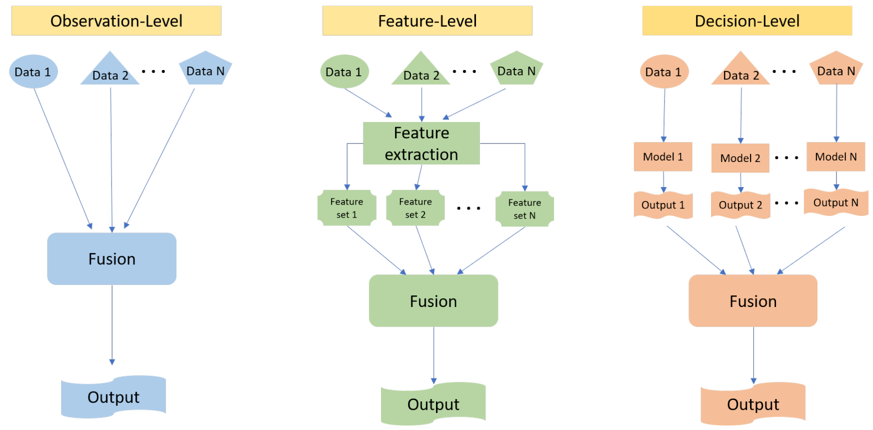


Figure 4.7: Types of data fusion.

The distribution of the various solutions/approaches presented in the previous section can be seen in Table 4.1. In this table, we identify not only the taxonomic branch that applies to the work, according to the taxonomy proposed in Section 4.1, but also the application field.

Most techniques aim at sensing the physical states of the subjects, such as heart rate, breathing rate, position, etc., either by using natural or artificial signals. A smaller number of studies/proposals address the assessment of emotional states.

Furthermore, when considering multimodality, only seven of the reviewed studies explored it, and most of these use multimodality for the monitoring of HR and BR. Moreover, only Zhao et al. [2016] explores multimodality for monitoring both physiological signals and emotional states.

As we can see in Table 4.1, only three proposals focus on the goal of using unobtrusive solutions for clinical purposes. Furthermore, one of them focuses on bradycardia detection in neo-natal monitoring, while the other two focus on sleep apnea monitoring. Sleep monitoring is, in fact, one of the major focuses on this field of research. This is partly due to the lower intensity and frequency of movements of people while they sleep. Most of the reviewed techniques are highly sensitive to noise generated by large movements. By concentrating on sleep monitoring, we are, in fact, avoiding the problems, difficulties, and limitations caused by such movements.

4.3 Data Multimodality

All of the work presented in the previous sections explores mono-modal sensing, that is, the use of a single source of data to infer the humans' physical or emotional states. However, in order to achieve a more robust system, work that fuses/mixes different sources of data should also be considered. As can be seen in Figure 4.7, when considering data fusion techniques there are three main levels of fusion, namely: observation-level, feature-level and decision-level (Hall and Llinas [1997]). Observation-level fusion is made directly on the input of the system, that is, the raw data are directly combined. On the other hand, feature-level fusion explores the preliminary extraction of several representative features from each input. Lastly, decision-level fusion occurs when an output or pre-output from independent models is first obtained and then fused to derive a final output. In this area of unobtrusive sensing most work is still experimental and, as such, focuses only on developing the sensing technology or exploring new algorithms. In this field of research, work that uses multi-modal approaches is scarce. However, if we consider pieces of work that do not focus primarily on detecting physiological and emotional states we can find several approaches to data fusion, that could also be reused for this purpose. As such, in this section we explore available multi-modal approaches to unobtrusive sensing, as well as relevant works to data fusion with the explored types of data sources.

One study used a multi-modal approach with sound and image-based techniques to predict humans' states. In Alonso-Martin et al. [2013], the authors propose a multi-modal system to predict emotion in human-robot interaction. The proposed system is based on decision level fusion. In this case, emotion detection trough voice and trough facial expression detection work as independent systems, and each of these gives a prediction for the human emotion. These predictions and their respective confidence intervals are then processed by a decision rule mechanism that elicits a final emotion classification. This work, also demonstrates another dimension to the applicability of this kind of systems, showing that they can also be used in emerging and fast-evolving fields such as robotic systems. The authors tested their work on a real case scenario where a robot interacted with 16 people, one at a time, and asked them to express several emotions. The tests demonstrated that the results obtained with the multi-modal approach outperformed both classifiers when considered as standalone solutions. Furthermore, other studies proved that some emotions such as anger, happiness, surprise, and dislike, are detected more easily by visual appearance, while other emotions such as sadness and fear were more evident through speech (De Silva et al. [1997]). These results indicate that multi-modal approaches can lead to better emotion detection approaches.

Although, in Alonso-Martin et al. [2013] the multi-modal approach was only applied to infer emotional states, we can envision the use of decision level fusion in order to combine the approaches taken in two or more mono-modal systems, such as the ones presented in the previous subsections. In decision level fusion, the sensing mechanisms work independently until they reach a classification or inference. At that point the output of each sensing mechanism is fed into a decision system that weights each output and reaches a final classification or inference.

Other authors have created a mixed approach of speech and thermal cameras to infer humans' emotions during speech. In Yoshitomi et al. [2011], the authors proposed a system that infers human emotion through vowel judgment and facial recognition, using thermal images. The main objective of the proposed system was to empower applied robotics with the capability of detecting humans' emotions as they speak even under varying lighting conditions. In this work, data fusion is made at feature-selection level, as the authors mainly use speech recognition to track the best frames from the thermal feed to perform emotion inference. The authors tested their system with three subjects with different genders and ages and also tested the system in the presence of glasses. The test results demonstrated a 79.8% accuracy while detecting facial expressions of the three subjects.

As already mentioned, work exploring multi-modal approaches is still scarce. Nevertheless, it is possible to find approaches that fuse the same types of data with different purposes. This is the case of Sánchez-Rodríguez et al. [2020], where RSS and CSI are fused together to build a more robust localization model. In this work the authors used a feature-level fusion, to extract features from both data sources. The authors also proved that the multi-modal system was able to surpass both a system based only on RSS and a system based only on CSI. As was presented above, there are works in the interference taxon that use RSS or CSI to infer the breathing rate of one person. As such, it is also possible to envision a system with a similar multi-modal approach as the one in Sánchez-Rodríguez et al. [2020] to unobtrusively sense humans' physiological signals, that could also show performance improvements when compared to a mono-modal approach.

In addition to combining different data streams using one of the mentioned data fusion techniques, it is also possible to use different techniques in a complementary way. For instance, techniques from the image taxon that use facial video feeds for heart rate monitoring are less affected by body movements and, conversely, more affected by rapid head movements. On the other hand, techniques from the reflection and interference taxa are more accurate when the user is stationary or quasi-stationary and are quite affected by rapid body movements, whereas they are unaffected by head movements. Thus, it is possible to envision a system that uses both modalities to address these limitations. One such approach is used in Kefayati et al. [2020], were the authors proposed a system that uses both CSI and a video camera to generate a video stream. The overall aim of this system is to be able to generate video frames from CSI at points in time at which the video camera fails or is unavailable, for instance during an attack or while the camera line of sight is blocked by an object. In this work, the authors also explore a different approach, which only uses the video feed to train a deep-learning model that works with the CSI as input. Complementary techniques such as the ones that are mentioned above can, thus, be valid approaches for unobtrusive sensing.

In Chen et al. [2017], the authors proposed a feature-level fusion method which uses several physiological signals, such as the heart rate and respiration rate, to monitor drivers' stress. This shows the possibility to fuse several physiological signals, and we believe that this approach could also be leveraged in the field of unobtrusive sensing. For instance, there is the possibility of fusing the estimated heart rate and breathing rate from different sensing techniques. Additionally, given the fact that while we are driving our body movements are limited, many of the techniques presented above are quite suited and, as such, this could also be one of the areas that could greatly benefit from the use of unobtrusive techniques. Furthermore, there are also studies that proved that using multi-modal

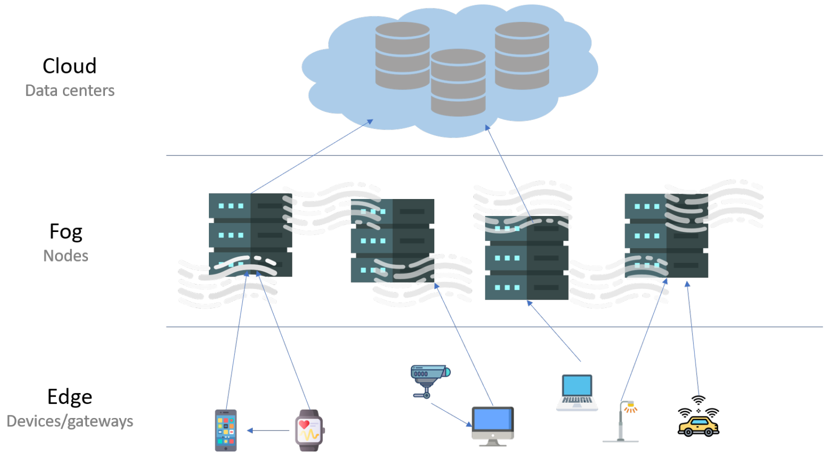


Figure 4.8: Cloud, Fog, and Edge representation.

physiological data can also be leveraged in deep-learning models to detect other emotions, leading to results that are far better than those of single modalities (Yin et al. [2017]).

4.4 Computing Architectures

Distributed systems can use one or more architectural approaches, namely edge, fog or cloud, as can be seen in Figure 4.8. Cloud computing is a model for the provision of remote resources over the Internet, providing high computing power, elastic storage capabilities and high scalability. Fog computing, in turn, offers computing capabilities closer to the network edge, on a smaller scale and with less scalability, but with considerable gains in terms of latency. Lastly, edge computing refers to performing data processing on devices inside the edge network (i.e., next to end user devices), such as sensor nodes that acquire the data or gateways to which sensor devices are connected.

As previously stated, several proposals focus on the development of new sensing techniques, and little considerations are made about optimal architectural designs. That is, most pieces of work are made in experimental setups and the data processing is made offline. However, when considering similar work outside the field of unobtrusive sensing of physiological and emotional states, it is possible to find approaches that take into consideration the underlying distributed computing architecture. As such, in this subsection we will overview the existing designs and address what we believe would be the optimal case for each of the taxonomy nodes.

In Fernández-Caballero et al. [2016], the authors proposed an architecture for the regulation and detection of emotions, through the fusion of behaviour data, emotional data and valence/arousal data. This work was based on the software architecture from Fernández-Caballero et al. [2013]. In order to achieve real time feedback, and since the sensor processing for emotion detection tends to be the more demanding part, the authors propose that the information from each sensor should be processed in a dedicated node. The authors also propose a decision-level fusion of the heterogeneous outputs from the sensor nodes. As such, they proposed the existence of a central node in the system, dedicated to this functionality. This central node is also in charge of communicating with the actuation nodes, which, in turn, interact with the physical devices in the environment. The authors describe their system as a hybrid-distributed system and do not specify the location of their central node, assuming a mixed approach of edge and cloud/fog. We also believe that this is the best approach for this type of system, as it relies on performing pre-processing and dedicated inference in each sensor node (i.e., at the edge), while fusing and dealing with the more complex models in a more robust infrastructure, at fog/cloud level. This approach has also been widely used in other systems such as smart cities, and has proved to be one of the best approaches in terms of system performance and scalability (Du et al. [2018]).

Additionally, most of the works in the image taxon that use video feeds or thermal video feeds could also benefit from performing processing sensor data in a dedicated node. Transferring video feeds, can lead to considerable traffic load and decrease the chance of performing real time inference. Although, it is not trivial, it has been proved that processing video feeds in dedicated nodes is feasible (Zhang et al. [2019]). However, other studies also indicate that the hierarchical architecture <camera/device, private cluster, cloud> is the most common and the only feasible approach for large-scale analyses of video feeds (Hung et al. [2018]). We believe that in the case of unobtrusive sensing, since some systems may be dealing with preventive detection of pathologies, which are related to the subjects' history, the best approach should also be a mixed approach of edge-fog-cloud. In this case, edge nodes should pre-process and extract the more relevant features from the acquired data, while upstream nodes should deal with computing the history, training more complex models, and making inferences. However, in some systems such as those of clinical nature (e.g., Abbas et al. [2011], Villarroel et al. [2014] and Klaessens et al. [2014]), privacy is a must, and for those systems we believe that the correct architecture should be edge-fog, to maintain the personal data in architectural levels that are under direct control of data owners (i.e., users) or primary data processors (i.e., medical staff).

As mentioned before, one of the fields in which techniques for unobtrusive sensing have been used is the field of robotics and human-computer interaction. The systems developed with this purpose try to create the illusion of interacting with other human beings and, thus the response needs to be provided in real time. In Alonso-Martin et al. [2013], the authors proposed a robotic system to detect emotions from facial expressions and speech in human-robot interactions. In this system, the model was created offline, from previously collected data, and then inserted in the robot. Thus, all inferences from the system were directly made on the edge. However, we believe that a more complex robotic system for human interaction should also benefit from a mixed edge-fog/cloud approach. A complex robotic system should be able to perform reactions to user inputs on the fly, but it should also be able to keep track of all the users and the history of each user interactions, and learn from them. As such, the most appropriate approach should be to perform inferences based on existing models and additionally retrain those models in a fog or cloud infrastructure.

Smartphones have also been widely used as a tool for creating unobtrusive sensing solutions, as seen in section 4.2. Most of the works that are based on smartphones use an offline approach, that is, the data is collected by a smartphone and then processed at a later date on a computer. However, in Kwon et al. [2012], the authors also developed a mobile application capable of detecting the user's heart rate using a 20 seconds video feed directly on the smartphone. Other authors also demonstrated that it is possible to run complex models on current smartphones (Cao et al. [2018]). We believe that techniques based on smartphones should also use a mixed edge-cloud architecture, where pre-trained models and simpler processing tasks could run directly on the smartphones, while training of more complex models could be offloaded to a cloud server.

In Patwari et al. [2013], the authors propose the use of a wireless sensor network to monitor the breathing rate and find the location of people. Although, in this work the authors processed the data offline, in their architecture they used a sink node that collected and stored all of the data. In this system, we can envision a fog computing architecture, where a more resource-capable node on the fog would perform the same tasks as the current sink node and, concurrently, process all data and product inferences. Additionally, in Jia et al. [2017] the use of two synchronized geophones could also resort to a fog architecture, as the feed from both devices is needed in order to perform the inference.

Last but not least, it should be noted that all of the reviewed works that used CSI were also made offline. However, in Gu et al. [2018] the authors proposed an architecture where the data would be sent to and processed in a remote server. Additionally, there are works that show that is possible to run complex models directly in the access points (Soltanaghaei et al. [2020]). As such, when considering works based on CSI, we believe that a mixed approach edge-cloud/fog could also be taken.

Based on the considerations made above, in table 4.2 we provide an insight on the possible computational architectures, applicable to the various works in unobtrusive sensing, reviewed in this paper.

Table 4.2: Possible computing architectures applicable to the reviewed works.

	Natural		Artificial	
	Sound	Image	Reflection	Interference
Cloud-Fog-Edge	Yoshitomi et al. [2011], Alonso-Martin et al. [2013]	Yoshitomi et al. [2011], Alonso-Martin et al. [2013], Massaroni et al. [2018], Li et al. [2014], Garbey et al. [2007], Hu et al. [2018], Pereira et al. [2015]	- - - -	- - - -
Cloud-Edge	Jia et al. [2017], Huang et al. [2019a]	- - - -	Wang et al. [2021] Nandakumar et al. [2015]	Liu et al. [2015], Gu et al. [2018], Wang et al. [2017]
Fog-Edge	Jia et al. [2017]	Abbas et al. [2011], Villarroel et al. [2014], Klaessens et al. [2014]	Adib et al. [2015], Zhao et al. [2016], Lee et al. [2015b], Lu et al. [2009]	Liu et al. [2015], Gu et al. [2018], Kaltiokallio et al. [2014], Patwari et al. [2013]
Edge	Jia et al. [2016], Ren et al. [2019]	Kwon et al. [2012], Kwon et al. [2015]	- - - - - - - -	- - - - - - - -

Table 4.3: Open Issues, Challenges and Most Relevant Areas

Open Issues and Challenges	**Predominant Fields**
Noise Reduction	Natural Branch: Sound Taxon Artificial Branch
Multi-Person Monitoring	Natural Branch: Sound Taxon Artificial Branch
Emotional States Detection	Artificial Branch
Machine-Learning Approach	Natural Branch: Image Taxon Artificial Branch: Inference Taxon
Privacy	Natural Branch: Image Taxon Artificial Branch
Open Datasets	Natural Branch: Physiological Data Artificial Branch
Medical Field Exploration	All
Multimodality and Data fusion	All
Standardization	All
HITLCPS Approach	All

4.5 Open Issues and Challenges

There are several open research lines and opportunities to further explore existing unobtrusive sensing techniques and improve them. Table 4.3 will guide us through the discussion of the most relevant open issues and challenges, in what concerns unobtrusive sensing of humans. For this, in the first column we identify the main open issues and challenges in this area, namely noise reduction, multi-person monitoring, emotional states detection, use of machine learning techniques, privacy, access to open data sets, use in the medical field, use of data fusion, and adoption of HITLCPS approaches. For each of these, in the second column of Table 4.3, we identify the predominant fields of application according to the proposed taxonomy.

All the presented techniques are subject to noise, which may lead to significant errors. Sound-based techniques may be affected by static noise and by environment noise, while image-based techniques can suffer from interference caused by lighting and moving artifacts. In the artificial branch of the taxonomy, we observe noise created by movement or by obstructions for both the reflection-based and the interference-based techniques. As we have seen from the review of the state-of-the-art, some approaches already explore these limitations and compare results with and without noise sources (e.g., Li et al. [2014]; Liu et al. [2015]). This noise prone nature of the used sensing mechanisms has impact on the achievable results, and this is one of the reasons why some studies focus on sleep monitoring or on near-static environments (as is the case of neo-natal monitoring). We believe that this is one of the fields that should be further explored, by exploring new techniques to increase system performance in noisy environments, by filtering noise sources, finding new ways to selectively sense the phenomena, or even by adopting a multimodality approach, as previously discussed. This, in turn, would open new opportunities for extending the applicability of existing work to new fields.

Another challenging area for which there is considerable lack of work is the determination or assessment of emotional state, especially in what concerns the use of artificial signals. Most approaches focus on physiological aspects only, and do not address emotional states. Furthermore, several approaches that do address the emotional state of people, extract the required data from physiological data (e.g., Zhao et al. [2016]). This, in turn, shows that several approaches developed for the purpose of detecting physiological state can also be repurposed to address emotional states. Several studies point to a positive correlation between positive emotional states and physical well-being (e.g., Salovey et al. [2000]), while others present results that indicate that mixed emotions with a balance between negative and positive emotional states can be beneficial to the physical well-being (Hershfield et al. [2013]). Nevertheless, information about human psychic state and emotions can be as important as physiological state. Furthermore, we believe that in the future this kind of solutions could lead to the possibility of analyzing people's mental state at a society scale, and as a group indicator. This could be particular important for instance in a pandemic situation, were the emotions and mental states could alert government to the negative effects of certain decisions and restrictions. Thus, we believe that the exploration of unobtrusive techniques to monitor the emotional states of people is one of the most important, promising, and challenging research fields.

Most of the reviewed approaches can only deal with one or two persons at the same time, with Wang et al. [2017] being the exception, as it is able to simultaneously monitor up to five people. Nevertheless, some image-based techniques are applicable to more than one person at the time, such as the techniques described in Kwon et al. [2015] and Li et al. [2014]. Despite this, approaches that address the monitoring of multiple persons typically lead to results that deteriorate as the number of monitored subjects increases. As such, we believe that this is still one of the areas that should be further explored. We believe that techniques that explore tagging/identification and tracking mechanisms inside the sensing environment could contribute to improve the results of the sensing solutions, and lead to more reliable and robust solutions. Furthermore, the identification of people could grant the sensing systems access to historic information of a certain subject's physiological and emotional states. This in turn could be used to take more accurate long-term decisions, or even, for instance project a future outcome for a subject's relative.

As stated before, many techniques are prone to noise and, as such, in most cases the solution is to apply signal filtering. Although, in most situations this is indeed the best approach, the possibility of using machine–learning techniques can also be explored. We believe that approaches such as the use of deep-learning in the case of image-based sensing would greatly enhance the system performance. This approach could also be used in CSI-based techniques, where large amounts of data are produced. Additionally, deep-learning is now accepted as a valid approach for systems that use CSI for a variety purposes, such as activity recognition or occupancy detection (Huang et al. [2019b]; Zou et al. [2018]; Liu et al. [2020]). In Kefayati et al. [2020], it was also proven that it is possible to implement a domain translator to generate video frames from CSI, proving the similarity of these types of data. We believe that future research could benefit from the work done in deep-learning in other applications of CSI, and from the extensive work done in deep-learning with images and videos in order to improve many of the reviewed solutions.

The unobtrusive sensing techniques presented in this paper open up exciting prospects in what concerns better and more affordable health care and e–health systems. However, they also raise several concerns in terms of privacy and security. Questions such as "How do we control who reads our vital signs?", "How can we keep our privacy, when people can see us and track our movements even through walls?", or "How can we make sure that these technologies are not misused?", and many more, must be answered. Unobtrusive sensing techniques create privacy vulnerabilities like no other before. For instance, when dealing with the possibility of tracking movements and vital signs through walls, how can we make sure that this is not used by burglars to rob houses or by governments to keep track of our movements and actions? Furthermore, this could also be used by health insurance companies to keep track of our clinical conditions and deny insurance in high-risk cases. Furthermore, when considering computing architectures many of the presented solutions could leverage the use of distributed edge computing. This in turn could also lead to several security and privacy breaches (Ni et al. [2020]). Thus, assessing data privacy in terms of the different solution architecture options is also important. As we previously discussed, the medical field is one of the fields that would leverage more from the employment of unobtrusive solutions. However, it is also one of the fields where data privacy is of the utmost importance. Other authors, have explored the concerns of privacy in the medical field, and proposed new approaches to deal with attacks, but we believe that new solutions should be proposed with the aim of specifically including these new sensing techniques (Meng et al. [2018]). As such, we firmly believe that, privacy assurance is one of the most important open issues in this field and should be further explored in future research.

One important drawback in this field of research is the lack of open databases/datasets. Although, there are several image and video repositories and datasets, those are not labelled for the recognition of physiological states. Apart from image-based and sound-based emotion detection techniques, to the best of our knowledge there are not any open datasets in this field of research. This creates a significant drawback in this field of research, as this kind of system implementations have high complexity, and the nonexistence of datasets forces researchers in this area to implement their own systems from scratch, and perform trials for data acquisition, which requires considerable effort. Clearly, a collective effort should be made to develop better data acquisition protocols, with well-known and easily replicable factors and constraints. This would allow for independent, yet comparable, data acquisition actions to occur, and would lead to the creation of common datasets, largely available to the research community. Additionally, many of the studies presented in section 4.2 were performed with a small group of subjects. The existence of larger datasets would allow the validation of these techniques in a more reliable and meaningful way.

Although some reviewed approaches address the detection and monitoring of medical conditions, most of them focus on physiological data acquisition. Moreover, although several approaches allow the monitoring of the heart rate signal, they do not compare their results with the results obtained through electrocardiogram, which is the golden standard for HR monitoring. In order to move from the labs to clinical scenarios, we need more evidence that the signal obtained with those techniques is reliable, maintains the information in the time domain, and can be correlated with the clinical metrics. As such, we believe that, as future work, researchers should make an effort to evaluate their solutions against medically approved and certified methods and equipment, as opposed to ad-hoc, and non-approved methods and experimental equipment, such as wearables. This is one of the most pressing research opportunities in this area, since only after tackling this issue we will be able to bring these solutions to real system level, and apply them to improve the medical field. For instance, in Mo et al. [2020] the authors propose the use of robots to monitor the breathing rate and aid in the monitoring of patients with infectious diseases such as Covid-19. The use of techniques such as those depicted in the artificial branch of our taxonomy, that can retrieve these values even through walls, could greatly help in these tasks. Thus, the development of these techniques and the effort to get them certified and approved for their use in the medical field could be an important step in fighting future pandemics.

Additionally, as mentioned in section 4.3 the use of multimodality approaches could bring forth more robust systems. This is still a largely unaddressed field, and, as such, this should also be subject to future work in the area of unobtrusive sensing. In the work that we reviewed in section 4.3, feature-level fusion seems to be the more common approach when creating multimodal approaches. Additionally, some systems use decision-level fusion, where several sub-systems work separately to generate their own outputs, and a decision mechanism weights each output to generate a final classification. This allows the involved systems to mitigate each other's errors (e.g., speech is better suited to detect certain emotions while facial expressions is better suited for others). Although, these approaches could raise other issues, such as dealing with data heterogeneity and data synchronization, we believe that exploring several levels of data fusion can be an interesting opportunity in the field of unobtrusive sensing of humans' emotions and physiological states. The multimodality of sensing techniques could also be used to deal with some other open issues, such as creating systems to detect multiple persons at the same time. Cameras and video feeds have proven to be one of the best solutions to detect, identify and track people (Visakha and Prakash [2018]). As such we believe that in the future, the combination of video feeds and unobtrusive technique could lead to systems that are more accurate and can be used to track multiple persons at the same time.

In addition to data fusion, there is also the option of using several systems to complement each other. For instance, the works in the artificial branch of the taxonomy are less precise when there is significant body movement, but are unaffected by head movements or temperature fluctuations, while works from the image taxon exhibit the opposite behaviour. When considered separately, these techniques may lead to windows of time during which it is not possible to perform inferences. However, an approach that combines both types of techniques may overcome the mentioned problems. As such, we believe that this should also be an issue to address in future work.

Another issue that should be addressed in the field of unobtrusive sensing, should be the issue of standardization. Standardization, helps to bring consensus between industry, the research community, and organizations in general. It also defines benchmarks that set performance minima for these technologies. Additionally, standardization helps to build the trust of costumers in products and systems/services, which, in turn, will help the dissemination of these technologies. There are already some approaches to this issue. For instance, the new 6G design proposals already include the standardization of the cellular signal as a sensing method (Wild et al. [2021]). In this respect, it is proposed that the Radio Frequency sensing capabilities should be natively integrated into the system design of 6G, and that base stations should use the same spectrum for both communication and sensing purposes.

The standardization of this type of techniques is also an important step towards answering the issues we raised about privacy, since privacy enforcing measures could be built into the standard. Furthermore, standardization could also improve the acceptance of these techniques in the medical field. For instance, the use of medically-approved protocols for communications such as HL7 and DICOM (Han and Bae [2016]), could be built into the standards for these new sensing systems, making their integration with existing medical information systems easier, and improving their acceptance by managers of these systems. As such, we believe that standardization is one of the open issues that should be addressed in order to improve the field of unobtrusive sensing.

Lastly, as mentioned previously, systems that take into consideration human intents, actions, emotions and physical state, i.e., HITLCPS systems, maximize the benefits for the users. In order to develop such systems, we need to close the loop and actuate on humans. Although, some approaches already show efforts in this area, as is the case of Fernández-Caballero et al. [2016], where the authors' goal is not only to detect emotions but also to regulate them by automatically changing the environment that surrounds the user. Most of the work in the state-of-the-art still ignores this component.

In Chapter 3 we present a new model for the HITLCPS paradigm, in this model we propose the use of unobtrusive sensing as a tool to gather data for HITLCPSs.

Although most of the solutions reviewed had the goal of developing sensing techniques and were not intended as fully operational and/or readily deployable systems, we believe that implementing a HITLCPS architecture could serve as the next step for many of them. The interactions of the users with these systems and their perception of the system itself is an essential step to better understand and achieve true unobtrusiveness. We believe that in the future this kind of system, using unobtrusive solutions, will be part of several of our society's systems. It is possible to envision that they will be in healthcare systems, security systems or even in our industry, enhancing human beings well-being, and in turn enhancing any system of which they are part of. As such in theory, unobtrusive sensing of human beings' physiological and emotional states could improve every existing system.

In the next chapter, we will present and discuss several implementations of case studies that aim to evaluate the model purposed in Chapter 3. In particular, one of these case studies aims to evaluate the use of unobtrusive sensing in HITLCPSs.

Chapter 5

Implementing the New HITLCPS Model in Several Applications

Contents

In this chapter we will explore the implementations of several case studies. These case studies were developed with the aim of studying the model for HITLCPS proposed in Chapter 3. The case studies and implementations presented in this chapter cover several areas, namely the monitoring of student performance, monitoring of people's behaviour during a pandemic, creation of HITLAI models for recommendations systems, and monitoring of elderly people with chronic diseases. Although none of the case studies covers all the proposed model components, all of them were used to study and propose implementations for one or more key components of the HITLCPS paradigm.

The adaptation to the HITLCPS paradigm is hard for most systems, as such creating implementations and case studies that serve as an example for other system designers is also an important open issue in the HITLCPS field. We believe that the implementations described in this section can be used by others as a starting point to either adapt existing systems to the HITLCPS paradigm or create new ones. Furthermore, most of these case studies were tested in real-world scenarios and the results from those tests are covered in Chapter 6.

5.1 IoT Student Advisor and BEst Lifestyle Analyser (ISABELA)

The aim of ISABELA is to create a HITLCPS system capable of assisting users, by gathering physical data, behavioural data, and social networking data. Initially intended for assisting students in maximizing their academic performance (ISABELA stands for "IoT Student Advisor and Best Lifestyle Analyser"), it can be used for any kind of user in a variety of scenarios. Nevertheless, in our specific case study we concentrated on the student scenario. In this scenario, the aim of the system is to use students' states and behaviours to infer their potential academic performance, in order to detect if students are failing and recommend possible strategies to improve their final outcomes. At the same time the system also aims to detect the students that are doing well and encourage them to continue, or even further improve. The system is also able to gather data from OSNs, which can be used to infer the users' emotional state.

In Figure 5.1 we present a diagram of the ISABELA case study in the context of the HITLCPS paradigm. As can be seen from the figure, with this system we intend to study the implementation of all the phases of a HITLCPS, and several of the key aspects that we believe are needed in order to advance the field of HITLCPS. Namely, social sensors smartphone and IoT data acquisition and actuation through chatbots.

In this section, we start by presenting the overall ISABELA system architecture. Then, we address data acquisition, data processing, and actuation. Last but not least, we provide a detailed description of the OSN-based NLP module, which is central to emotional state inference.

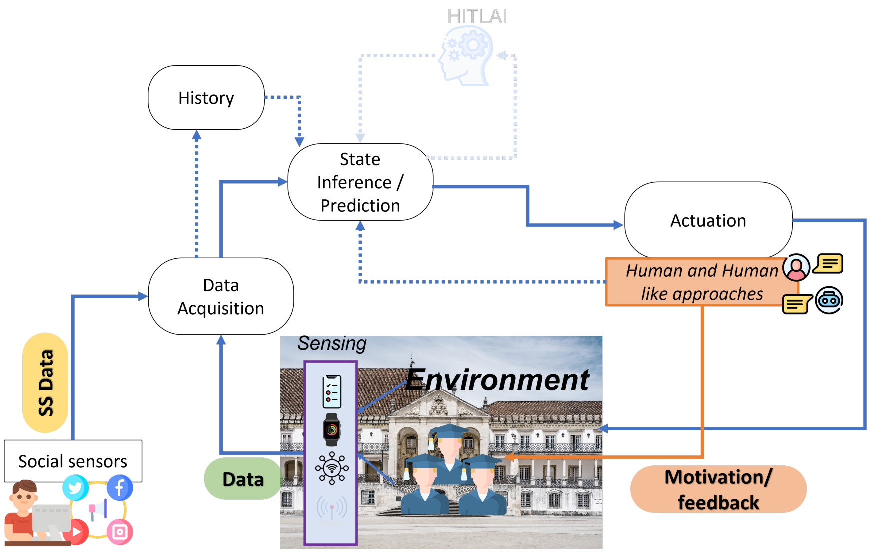


Figure 5.1: ISABELA HITLCPS architecture.

5.1.1 The HITLCPS Architecture

ISABELA was built using a set of components from FIWARE Fazio et al. [2015], with the objectives of reducing development time, allowing for modularity, and maximising interoperability. FIWARE is an European Union (EU) project supported by the FI-PPP (Future Internet Public-Private Partnership) that aims to offer a set of open-source components, open-standards, and an open ecosystem to deal with some challenges of the new IoT-enabled Internet. The project developed the concept of Generic Enabler (GE), designed with well-defined API targeting specific problems in the IoT domain. In the ISABELA system, we used several GEs from the FIWARE project, as can be seen in Figure 5.2.

To deal with the sensing state, the IDAS GE was used to manage all kinds of IoT devices (Telefonica [2022c]). The component supports management and interoperability capabilities, between devices (that communicate using a variety of protocols) and the FIWARE platform. In the ISABELA project we chose the UltraLight agent to connect various IoT devices, contained in ISABELA sensing boxes, with the platform. ISABELA boxes allow to collect temperature, humidity, noise and luminosity data in the user environment. In the student scenario, these boxes were intended to be placed at school and at home and would be important to evaluate the environment's influence in the students' outcome. However, in the study presented, as the Isabela boxes require additional equipment and we did not have direct access to the infrastructure in Ecuador they were not used.

HITLCPS need to be scalable, and dynamic models are needed in order to represent the data of such systems. In order to have such a dynamic ecosystem, the FIWARE project adopts the NGSI9/10 information model, previously developed by OMA (Krčo et al. [2014]). The model is based on entities and attributes. Each entity has its own type and is represented by attributes, using JSON format.

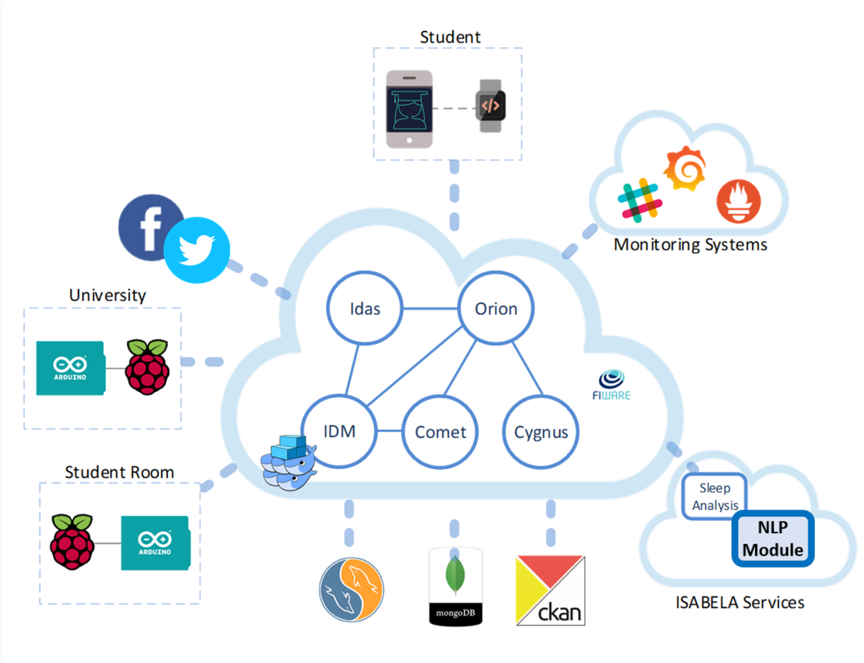


Figure 5.2: ISABELA system architecture.

All the entities in the ISABELA Android application use the NGSI representation and are managed by the ORION GE (Krčo et al. [2014]). Using this API, it is possible to create/delete/retrieve entities, and to update existing ones. To improve the system distribution and scalability, the module acts as a broker, by allowing external systems (consumers) to make data subscriptions to the entities with specific rules and attributes. ORION sends notifications to these systems when subscription rules are fired. Data sent by IoT devices is associated with the entities stored in ORION.

The ORION GE is not fit for storing historic data. Instead, the Cygnus module is the connector "in charge of persisting certain sources of data" in several storages, using the Apache Flume technology and by subscribing to the ORION entities (Telefonica [2022a]). Thus, when a new entity arrives at the Cygnus GE, the listener will put it in a specific channel and forward it to a third-party data storage (e.g., MySQL, MongoDB and CKAN channel). Databases usually do not provide APIs to retrieve the data to applications. Hence, another module of the FIWARE platform handles this issue, namely the Short-Term History or STH-Comet GE, which provides a RESTful API with historic-queries capabilities, and aggregated methods (Telefonica [2022b]). Thus, each time the ISABELA application or the NLP module needs to access historic data, the Comet GE will connect to the database and retrieve the queried data.

Security and privacy are some of the most important requirements of HITLCPS systems. In order to protect communications between GEs, IoT devices, and applications, the IDentity Management (IDM) GE called Keyrock, was used (de Internet de Nueva Generación [2022]). Keyrock allows adding authentication and security to devices, users, and applications, as well as authorization policies. Additionally, from a privacy point-of-view, and in order to meet all requirements from EU privacy laws, such as the General Data Protection Regulation of May 2018 (ENISA [2018]), all data handled by the Android application is anonymized.

Additionally, ISABELA includes two services, namely sleep analysis and NLP module. The former is the component in charge of computing users' sleep hours, based on data collected using smartphone sensors. The latter performs natural language processing, using textual data collected from OSNs to detect expressed sentiments and emotions. The latter module is the focus of section 5.1.4, below.

5.1.2 Data acquisition

In order to capture the indoor location and movements of the users, a third-party API was used, namely the FIND API (Schollz [2018]). Using WI-FI fingerprinting, it was possible to create a model able to retrieve an approximate location of the students inside the buildings. This can be used for instance to infer class attendance, as we also access the calendar stored in the smartphone.

As for the data from the smartphone, it is gathered using one of two ways, namely direct acquisition of sensor values or context acquisition. We use one or more sensors to acquire a virtual sensor value (for example, the already mentioned class attendance) or we get context information from the smartphone's SMSs and calls. The physical sensors are implemented in the smartphone by hardware, and their values can be obtained directly from an API that the Android Software Development Kit (SDK) makes available (Google [2022b]). The physical sensors used by our application are: proximity sensor, light sensor, accelerometer, and gyroscope. All sensors, apart from the accelerometer and the gyroscope, are set for a sampling interval of 5 seconds.

Another capability of the smartphone that can be used to create a virtual sensor is audio recording. In our case study, we were interested in detecting if students remaining for long periods in noisy environments had their academic performance affected. As such, we used the microphone to compute the noise amplitude in the students' vicinity. Apart from the aforementioned indoor location, this study also monitored the outdoors mobility of students. For this purpose, we used the Fused Location Provider API to obtain GPS data (Google [2022d]).

The Activity Recognition API was used in order to capture the activity of students (Google [2022a]). This API automatically detects activities by periodically reading short bursts of sensor data and processing them using machine learning models. The API classifies activity into several possible categories, namely, *running*, *walking*, *on bicycle*, *in vehicle*, *on foot*, *tilting*, and *still*.

Additionally, Google has since 2021 made available a Sleep API (Google [2022e]). This API enables applications to collect data about the users' sleep patterns. It is possible to register for two types of sleep data notification, sleep classification notifications and sleep segment notifications. Sleep classification notifications, are received periodically each 5 minutes approximately, and classify the possibility of the user being at sleep in the last 5 minutes. Sleep segment notifications are received only once a day, and represent a period in which the user was a sleep, with information about the start time, end time and duration of the sleep period. In ISABELA, we collect both of these metrics to retrieve the sleep patterns of students. Since our first tests were performed prior to the release of this API (in 2019), the data from this API was not used in the tests, presented in this thesis. Furthermore, as we discussed in the previous subsection, we created a custom module to perform sleep classification and sleep quality classification. The tests for the created models to perform sleep classification and sleep quality classification are explored in more detail in chapter 6.

5.1.3 Data Processing and Anonymization

The smartphone application also serves as a means to retrieve context information, using not only stored information but also smartphone usage data, such as calendar information, contact information (SMSs/calls), screen lock state, Wi-Fi information, Bluetooth information, alarms information, and app usage information.

In the case of the SMSs/calls information, we only register the number of outgoing, incoming SMSs/calls and missed calls information. Additionally, we keep record of the list of contacted numbers in the form of hashes, to compute the number of distinct contacted persons. This data is hashed in order to maintain users' confidentiality.

Additionally, also to preserve users' confidentiality, all the messages between the user and the chatbot are only stored in a persistence database on the smartphone, which is deleted when the application is uninstalled.

Apart from the hashing process applied to calls and SMS information, we also implement an anonymization mechanism for the storage of sensitive data. Access to the mobile application is protected by an authentication mechanism, namely the one used in the user's Facebook account. From that authentication we receive a hashed access token from the Facebook servers, which we hash once again with a one-way hashing algorithm, using SHA256 (Paar and Pelzl [2009]).

Furthermore, there are still some personalized features, such as GPS location and Wi-Fi Service Set Identifier (SSID)s values. In other to further anonymize the dataset we do not provide data from those features. We replace GPS data with computed values such as travel distance and average pace. Wi-Fi SSIDs values are replaced with *"ssid_<student_id>"*. The same happens for the stored Wi-Fi and Bluetooth scans. The name of the devices/networks are replaced by an id number of each different device/network. This still allows to compute the interactions between different students, while reducing the chance of identifying individuals.

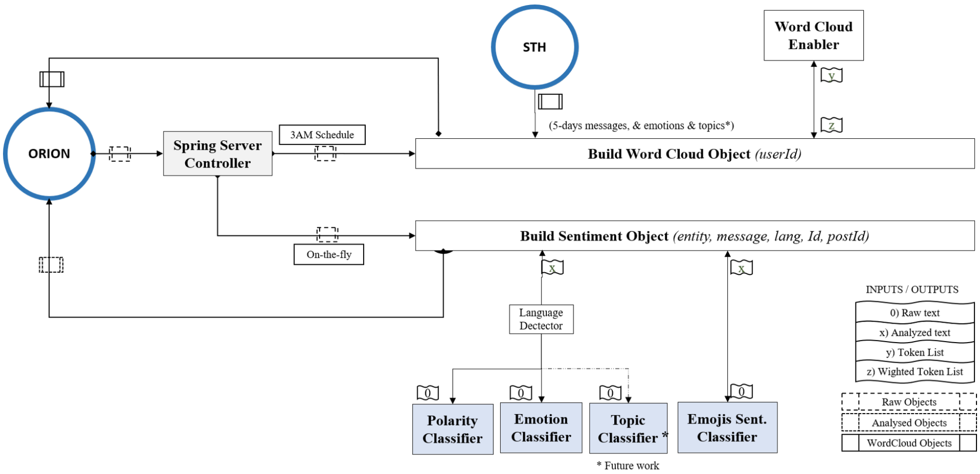


Figure 5.3: OSN sentiment analysis architecture.

5.1.4 OSN-based Natural Language Processing

In current IoT, both sensing and actuation approaches go beyond physical electronic-based devices. Indeed, they also comprise software–based entities like agents and even human beings, for instance, by considering as raw data all human-originated activities from social media. Like the embedded chips in electronic-based sensors, the *"human-based sensors"* also need a mechanism to transform the collected raw data into values which can be interpreted, leading to information and knowledge. NLP techniques can be used for this purpose, thus acting as *"translators"* between what humans input to OSNs and their emotions and context. The NLP module in ISABELA depicted in Figure 5.3 performs this task. The scope of the ISABELA NLP module comprises sentiment and emotion analysis and is designed to include topic modelling in the future. Topic modelling is a data mining technique whose objective is to extract the main related topics from given pieces of text (Hong and Davison [2010]).

We analyse texts produced by ISABELA's users in two popular platforms, namely Facebook and Twitter. On Facebook, we work with the text written by the user in the What's on your mind space. On Twitter, we work with the text written by the user available in the What's happening space, i.e., the user tweets and retweets. With such text entries and applying NLP techniques, we try to produce insights about the users' mood via both sentiment and emotions inferred from the posted texts (Feldman [2013]).

The most straightforward outcome of the sentiment analysis task is: "Does a text express a positive or negative sentiment?". Usually, we assign a polarity value to a text in the $[-1, 1]$ interval, where 1 corresponds to very positive polarity and -1 corresponds to very negative polarity. Emotion analysis is a step further comparing to polarity. Indeed, we try to assign more descriptive scenarios to the contents of the text, classifying it within the scope of emotions such as fear, fulfilment, frustration, boredom, excitation, happiness, etc. (Maia and Santos [2018]).

Figure 5.4 displays a generic architecture for a NLP system. In our case, the in-

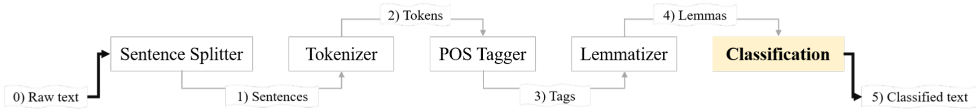


Figure 5.4: A Generic Architecture of a NLP System.

put to the system is a text, and not a speech audio object, that is pre-processed using a variety of linguistic tools such as tokenization, speech tagging, and lemmatization.

Focusing on sentence-level, we assume there is a single opinion in each sentence. This assumption can be extended by splitting the sentence into phrases that contain only one opinion each. This can be something as complex as a predictive classifier to identify sentence boundaries. A typical sentence splitter in the literature is based on a full stop in the text (”.”). Tokenization consists of isolating each component of the text, while POS Tagging consists of finding the grammatical classification of each of those components, e.g., adjectives, punctuation, verb, adverbs. Depending on the tag, the sentiment or emotional load may vary. In general, adjectives attract more attention from a NLP system, because they are considered to be more informative when it comes to emotions. The lemmatization process is responsible for bringing all the components into their original form. For grammatical reasons, documents use different forms of a word, such as organize, organizes, and organizing. Additionally, there are families of related words with similar meanings, such as democracy, democratic, and democratization. Thanks to lemmatization, all conjugated verbs will be turned into their infinitive form, plurals and gender will be eliminated, which reduces the complexity of analysing a text. Finally, the classification unit in Figure 5.4 is where different strategies are used to provide sentiment scores or emotion tags to the sentence.

Studying the sentiment and emotions in a text demands complex design and computations, which is not yet perfect in the literature (Feldman [2013]). The system must deal with complexity such as sarcasm, noisy texts, emojis, emoticons and slang. The classification unit is where the linguistic resources are leveraged to annotate the pre-processed text with an overall mean value for the sentiment and all emotions tags in the text. These NLP results may be attached to whole documents (for document-based sentiment), to individual sentences (for sentence-based sentiment), or specific aspects of entities (for aspect-based sentiment). These annotations are the output of the NLP system, i.e., the classified text. The simplest classification unit is developed by utilising a set of lexicons as linguistic resources.

Lexicons are documents that comprise the sentiment scores, or emotions tags of the entities assumed to represent a written language, i.e., the Corpora/Corpus (Feldman [2013]). As an example of lexicons, we have Sentilex-Flex PT 2 for retrieving the polarities (Silva et al. [2012]), LEED Appendix 1 (Rodrigues et al. [2018]) for retrieving the valence, arousal and dominance values in emoticons/emojis, and ANEW/ANEWPT (Soares et al. [2012]) for retrieving the valence and arousal. Valence expresses the intrinsic *”good”*-ness, neutrality or *”bad”*-ness of a token. Arousal expresses the level of excitement attached to the token. With both valence and arousal scores of a token, proven approaches in the psychology literature, such as the Russel Circumplex model (Posner et al. [2005]), enable to associate its corresponding sentiment and emotion. On the other hand, proven lexicons, such as in LEED Appendix 2 and Linguateca , will enable to associate emotions to emojis and Portuguese tokens, respectively.

Given the scarcity of NLP libraries for the Portuguese language, we developed a sample pipeline in Java programming language based on the Sentilex-Flex-PT2 lexicon for polarities. For both sentiment and emotions in emojis, we used the LEED dictionary. We further used ANEW/ANEWPT for mapping both valence and arousal scores of the sentences, which enables an indirect emotion analysis. Finally, an emotion dictionary by Linguateca enabled us to associate plain emotions with Portuguese tokens.

Contrary to Portuguese, there are abundant NLP implementations for the English language. We used the VADER library, which is both a lexicon and rule-based sentiment analysis tool specifically created for working with messy social media texts. VADER is claimed to offer accuracy over 60% (Hutto and Gilbert [2014]). In VADER, the unit used for measuring the polarity is the valence. For the time being, our solution does not yet integrate an emotion dictionary for English.

5.1.5 Actuation

Finally, the system provides feedback to the users through two different approaches, namely notifications and visual changes in the application. By changing the size and colour of a specific part of the display we let the user know which aspects are more likely to cause poor academic performance. The second approach is through the use of an interactive agent, most commonly known as a chatbot. To implement this approach, a human-computer interaction technology based on natural language conversations was used, namely DialogFlow (Sabharwal and Agrawal [2020]). The ISABELA bot is capable of retrieving data from several APIs available in the ISABELA system, and subsequently provide user recommendations.

In its current state, the chatbot can perform one of four tasks: triggering and showing *"alarm messages"*, performing follow-up response to *"alarm messages"*, requesting information from environmental sensors (ISABELA boxes), and performing daily tasks for the user (e.g., set an Alarm clock, check the cafeteria menu, or set an event on the calendar). The first three types of tasks are intrinsic to the system while the fourth functionality serves only as a mean of encouragement to use the application, by trying to make it more useful.

The *"alarm messages"* functionality is our main actuation focus on the chatbot. As of our first trial we did not have any data a priori, to build a model that was able to detect these alarms. However, we believed that the actuation was an important part of our system, that should be present form day one. As such, we decided to implement *"alarm messages"* based on common-sense recommenda-

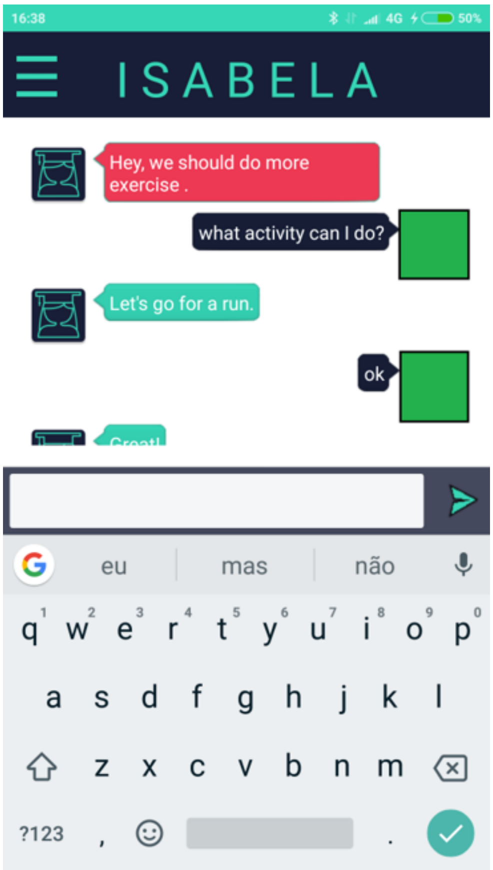


Figure 5.5: Chatbot actuation example for a *"low exercise recommendation"*

tions for human well-being. We had three main focus at the beginning of this trial: the students' daily activity (including their sleep), their whereabouts, and their sociability.

Activity alarms were defined by using the guidelines defined by the World Health Organization for physical Activity (Organization [2011]), while sleep alarms were defined by the guidelines proposed by the American National Sleep Foundation (Foundation [2022]). Location-related alarms were personalized by checking the total duration of classes in the student calendar. In this way, five alarms were implemented: *"too much time still"*, *"Not enough exercise"*, *"Should spend more time at the University"*, *"too much time at home"*, and *"Not enough sleep"*. An example of the chatbot operation can be seen in Figure 5.5. The figure illustrates a situation where the chatbot proactively actuates on the user, triggered by a "Not enough exercise" alarm. The user then asks what type of activity he/she can do to deal with this situation, and the chatbot is able to react with a follow-up proposition of an activity. This alarm, as well as the "too much time still" alarm, is triggered by computing the activity levels from the activity feature, collected through the Google Activity API. Being that the *"too much time still"* alarm is raised if the user is *"still"* more than 12 hours per day. While the *"not enough exercise"* alarm is raised if weekly exercise time is inferior to 8 hours.

In the case of the *"not enough sleep"* alarm, generated when detecting less than

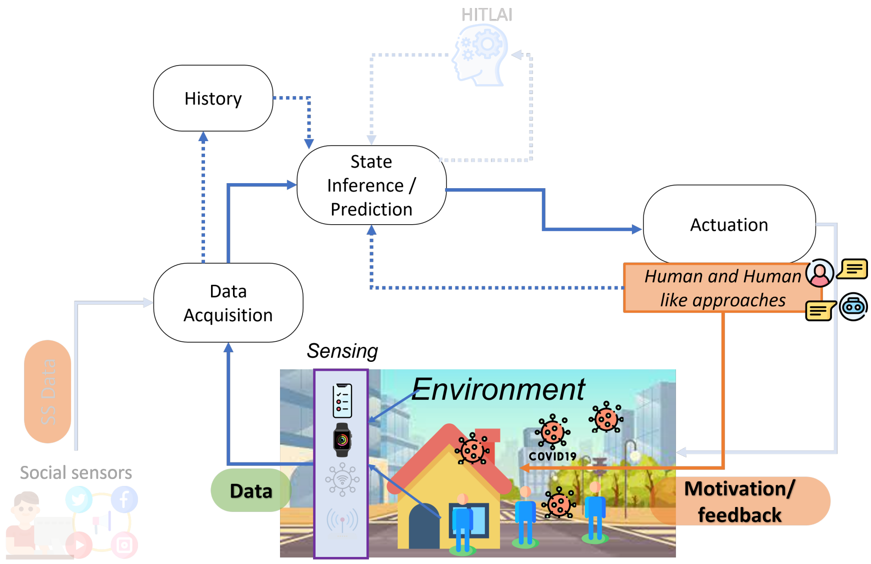


Figure 5.6: Vitoria HITLCPS Architecture.

6 hours of sleep, the chatbot can recommend a time for the user to go to bed considering their next day calendar, with a tolerance of 2 hours from the waking up hour to the first appointed class. In the future we intend to further extend this capability by making the system consider the distance from the student's home to the University and the public transportation schedule. Additionally, we aim to enhance the recommendation model with the automatic sleep detection and sleep quality models.

For the location-based alarms, namely the *"too much time at home"* and the *"not enough time at the University"* alarms, we compute the daily time at each location through the location feature collected from the Wi-Fi scans. We raise an alarm for *"not enough time at the University"* when the time spent at this location is lower than the student's daily schedule minus 3 hours. As for the *"too much time at home"*, the alarm is raised when the student spends more than 2/3 of his/her *"free time"* (i.e., time without events on the calendar) at home, not considering sleeping hours.

5.2 Vitoria- Monitoring during a Pandemic

Although, the Covid-19 pandemic brought forward without doubt many major negative effects, it was also a very relevant case study for HITLCPSs. During the pandemic most national health systems were overloaded, partially due to the fact of people not respecting the measures putted in place or lack of awareness of risk. At the same time we also saw an increase on psychological pathologies and their effects, due to the fact that of medical resources being relocated to deal with the pandemic. As such, during and post–pandemic, it is even more important to have tools capable of monitoring and measuring the emotional states of humans. This highlights the importance of developing HITLCPS.

In this context, we developed the Vitoria system, which is based on our work on

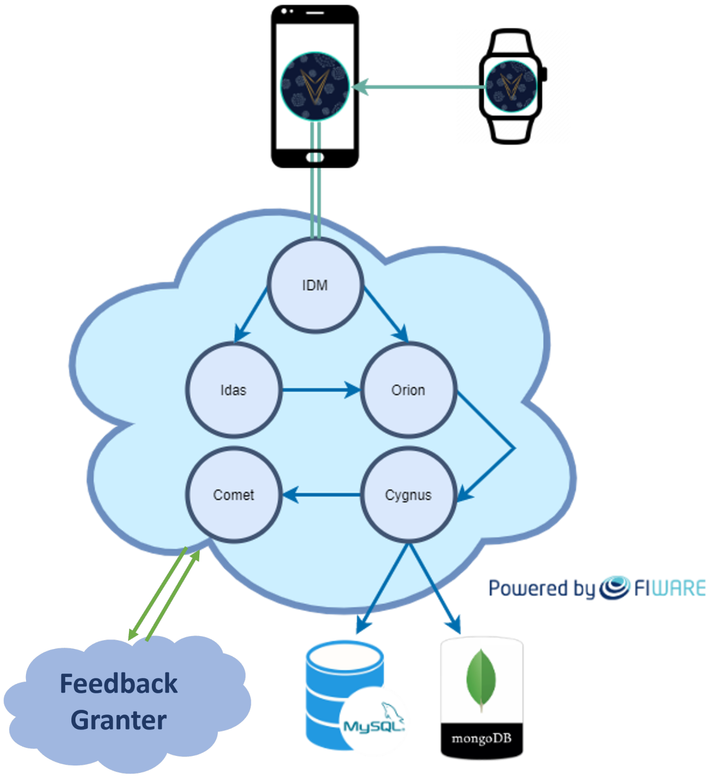


Figure 5.7: Vitoria system Architecture.

the ISABELA case study presented in the previous section. Our previous case study aimed to monitor and profile students, as well as improve their academic outcomes. However, in the Vitoria case study we aim to monitor and profile the general individual, and to determine how his/her behaviour is affected by changes in the pandemic situation, through the layering method presented in the previous section. Understanding these changes can lead to the implementation of more effective measures to fight the pandemic. Vitoria directly translates from *"Win"* in Portuguese, and with this application we aim to create a tool that would help to understand and fight the progression of the current and future pandemics.

In Figure 5.6 we present a diagram of the Vitoria case study in the context of the HITLCPS paradigm. As can be seen from the figure, with this system we also intended to study the implementation of all the phases of a HITLCPS. However, in contrast to the ISABELA case study we did not have direct access to the participants of the case study, as such we did not include IoT sensors in this case study. We also did not address social sensors in order to narrow the scope of the case study. Although, due to the nature of the case study we added several metrics in the data acquisition metrics (e.g., proximity questionnaires, transportation questionnaires), that we will cover in the next section. Additionally, we also aimed to explore the effects of the actuation mechanism in HITLCPSs, by creating a control group to validate the effects of the actuation.

Although the ISABELA and the Vitoria systems target different use cases, they share many architectural similarities, such as the FIWARE-based backend and the Android-based Mobile and Wear applications. As can be seen in Figure 5.7, the system is made up of a smartwatch and smartphone applications that deal with data acquisition and user feedback, and the backend that is able to manage and persistently store the collected data.

The system backend is very similar to the one used in the ISABELA system. As such, most of it components were explained in section 5.1.1 and will not be covered in this section. The main different between both systems is the added custom module, *"Feedback Granter"*, which is responsible for managing the feedback groups (control and active feedback) and generate the personalized feedback for each user. This module is able to communicate with the Comet GE to retrieve historic user data and generate the metrics for the feedback showed to each user. With this case study we wanted to study the effects of the performed actuation in the humans behaviours. As such we choose to divide the users of the system into a control group and an active feedback one, to study these effects. The feedback system is explored in more detail in section 5.2.3.

5.2.1 Passive Data Acquisition

Both the mobile application and the smartwatch application from the Vitoria system serve primarily as means to obtain data from the user. Most of the metrics collected come from the smartphone, while the smartwatch is primarily used to acquire health-related data (i.e., step count, activity, and heart rate).

One of the most important metrics, when considering the pandemic context is user location. Due to privacy concerns, our system does not store any GPS information. Instead, the system is able to detect the discrete location between two available options, *home* or *other*. In order to infer if the user is at home or in other places, when the users first configure their app, they are prompted to scan the available networks and select the name of their home network. The name of their home network is then stored in a local database on the smartphone. Due to privacy concerns this information never leaves the smartphone. The application is then able to periodically scan the networks and compare the scan results with the stored information, in order to infer if the user is at *home* or not.

Apart from mandatory confinement periods, during the pandemic in Portugal there were additional traveling restrictions between municipalities in certain periods of time. In addition, the analysis of the number of cases was carried out at municipality level. As such, user movements between municipalities were also an important metric to obtain. The Vitoria system was able to obtain this metric by first obtaining the GPS information of the user and then querying a reverse geocoding API, namely the OpenStreetMaps API (Maps [2022]). By using this method, the system was able to infer the user district, municipality, and parish. As previously stated, the GPS information was never stored due to privacy concerns, and in the case that no Internet connection was available the system would discard the GPS information. Additionally, storing the name of municipalities or districts could raise privacy concerns as well, even with anonymized user identity. As such, the system anonymized the denominations obtained from the reverse geocoding system as well. Furthermore, in order to further preserve user privacy, the information anonymization was user specific. That is, even for 2 users on the same location, the stored information would be different. This allowed our system to detect if a specific user was in different locations, the duration of said stay and the frequency of those movements, but never the specific location or his/her interactions with other users

In addition, in Portugal, the municipalities were classified in 4 risk levels, ranging from normal to extreme in terms of risk of infection. In order for our system to keep track of this information as well, a back-end service was configured to maintain an updated list of Portuguese municipalities and their respective risk levels. The mobile application was then able to, prior to the anonymization of the geographical information, query this service to obtain the current risk of infection due to geographic location of the user.

Another important metric to consider due to the nature of the pandemic, is user activity levels. Due to sudden change in habits caused by confinement, most people saw their daily routines affected in terms of sleep and physical activity. As happened for the ISABELA application, we also passively collect activity and sleep related data by using both Google's Physical Activity Recognition API (Google [2022a]) and the Sleep Recognition API (Google [2022e]). As mentioned before, the Activity Recognition API is able to automatically detect activities by periodically reading and processing sensor data and outputting a classification whenever the user activity changes. And as we explore in section 5.1, this API is able to classify user activities into: *running*, *walking*, *on bicycle*, *in vehicle*, *on foot*, *tilting*, and *still*. On the other hand, the Sleep Recognition API periodically returns a classification on whether the user is sleeping or not. Additionally, both APIs return a confidence on their classification, which allows the Vitoria system to post-process these classifications.

Another metric that can be used to infer the changes in user behaviour, is the statistics of the used applications. The Android SDK allows us to retrieve information about which applications were used, and for how long they were used. This information can, for instance, be used to infer changes in the user behaviour due to confinement, or to infer the frequency and duration of user interaction with other people through communication apps. This information will be analysed in further detail in sections 6.4.1 and 6.4.2.

Smartphones can also be used to create a virtual sensor that is able to detect ambient noise. In the Vitoria system, we were interested not only in obtaining information about the user but also about their context. As such, our system uses the microphone to compute the noise amplitude in the users' proximity. As with the other metrics, here privacy is also a major concern. As such, no audio is stored nor recorded, and the system microphone is just used to average the ambient noise in short intervals of time.

Another important aspect of people's behaviour in the context of a pandemic is their proximity to other people and the number of people they have contact with. For this, we used the Bluetooth smartphone capabilities to detect nearby devices that could correspond to people in the user vicinity. The Bluetooth API from the Android SDK allows to differentiate the type of devices that are detected with the Bluetooth and Bluetooth Low Energy (BLE) scans (Google [2022c]). These types can be for instance, headphones, smartphones, televisions, cars, wearables, etc. Additionally, all already connected devices are filtered and, as such, we can consider that all detected headphones, smartphones, and wearables correspond to people in the proximity of the user. However, the same person could be using several of these devices. As such, we do not infer this information directly. Instead, we use this data to opportunistically ask the user to answer a questionnaire. This proximity questionnaire is explored in the next section. In addition to the type of device, the strength of the signal is also considered, to infer the user distance to the device.

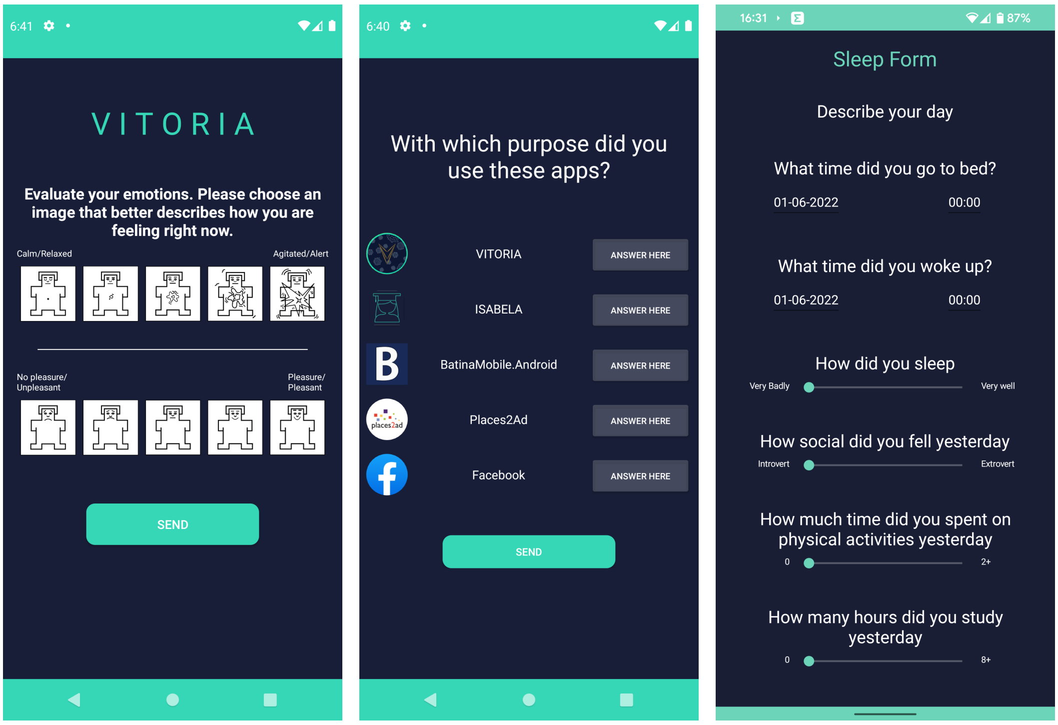
The Vitoria system is also able to scan and store the Wi-Fi information. Usually, places with a large number of people, like supermarkets, shopping centres, gyms and universities have a large number of Wi-Fi APs. As such we believe that this information can also be used to infer when the user is in such places. In order, to maintain the user privacy, the name of the detected Access points is never stored, only the MAC address and the signal strength.

As we previously stated, in addition to the smartphone application, the Vitoria system also comprises a smartwatch application. This application serves as a mean to collect physical and physiological data from the users. The smartwatch application is capable of infer the type of activity that the user is doing, the number of steps taken and the heart rate values of the user. Although activity data is already collected from the smartphone, the smartwatch is closer to the user and can offer a more reliable insight on the user activity levels. Moreover, physiological data can also offer a valuable insight about the user's physical and emotional states, especially the heart rate values, which have been proved to be correlated with stress levels, for instance (Taelman et al. [2009]). Additionally, the heart rate can also be used to post-process sleep recognition data or to infer sleep quality.

In addition to the aforementioned metrics, the Vitoria system is also able to collect raw sensor data such as accelerometer data, gyroscope data, proximity data, and ambient light data. Although, these metrics are not yet used, they can be used to create novel classification systems, or to improve existing ones. For instance, the values from the light sensor can be used to post-process the information from the sleep recognition API. All the mentioned sensors are collected from both the smartphone and smartwatch applications.

5.2.2 Active Data Acquisition

Although smartphone context information and sensors data can be used to infer several aspects of the users' daily life, when considering the users' psychological states these kinds of metrics have yet to be validated. As such, there is a need to use more conventional methods, such as questionnaires. In the Vitoria system, we implemented several questionnaires to infer the user's emotional state, and to complement the already mentioned data. In this section we describe the



(a) (b) (c)

Figure 5.8: Vitoria's Questionnaires: Emotional questionnaire using the SAM scale (a); Application finality form (b); Sleep duration and sleep quality questionnaire (c).

implemented questionnaires.

One of the aspects that we consider important to monitor during the pandemic context is the emotional state of people. Emotional states of humans are difficult to perceive and infer, and no approach that uses sensors has been validated as an adequate means of inferring these states. As such, we implemented a questionnaire-based approach to obtain this type of data. The Vitoria system implements the Self–Assessment Manikin (SAM) scale as a questionnaire that is daily prompted to the user (Bynion and Feldner [2020]). The SAM scale is an image based self-assessment scale to evaluate two-dimensional emotions of the users. The questionnaire can be seen in Figure 5.8a. The two scales correspond to arousal (top) and valence (bottom). This questionnaire was prompted to the user on a daily basis and at random time between 14h-20h. The questionnaire was not released during the morning because, during this period, the user perception can still be affected by the events of the previous day.

As mentioned in the previous section, the used applications and the respective usage duration might hold important information about the user. As such, we implemented another questionnaire to additionally determine the purpose of the applications' use. As can be seen in Figure 5.8b, in this form the user is prompted to select the purpose to which he/she used each of the five most used apps during the last 24 hours. This questionnaire was prompted to the user every day at 14h. The user can select one of four major purposes: communication, *leisure*,

Table 5.1: Number of persons on the surroundings options by type of transportation.

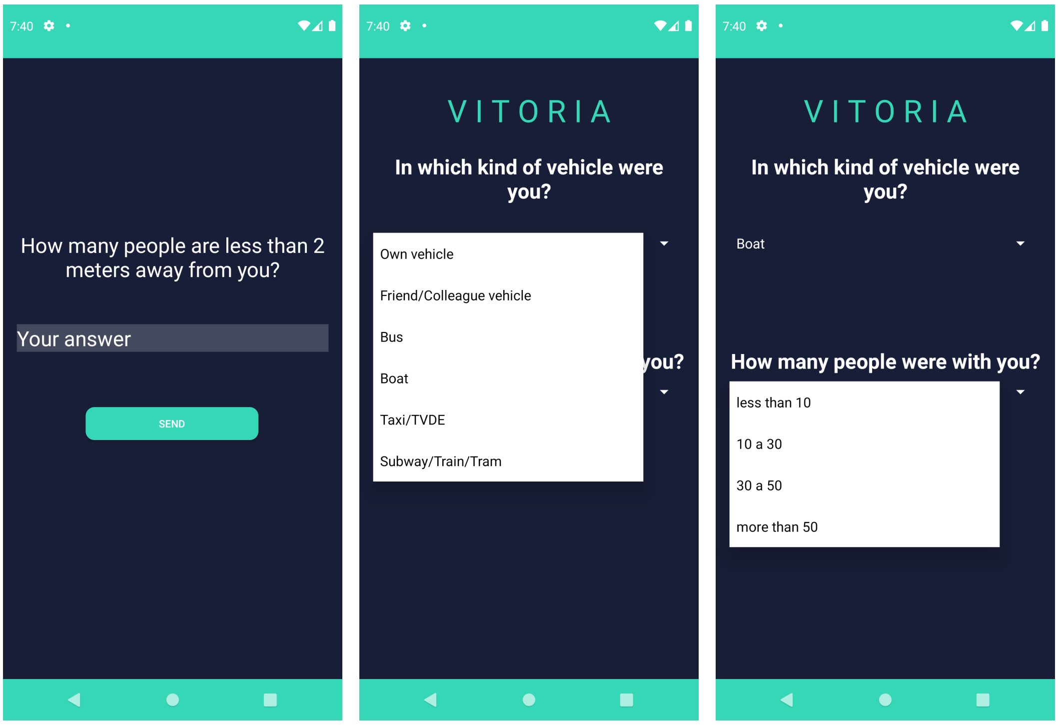
Own car **Friend/Colleague vehicle** **Taxi/TVDE**	**Bus** **Subway/Train/Tram**	**Boat**
0	less than 10	less than 10
1	10 to 20	10 to 30
2	20 to 30	30 to 50
more than 2	more than 30	more than 50

research, and *work*. We believe these four purposes make up most of the purposes for which every application today is used. The data obtained from the user answers to this form will be explored in more detail in the section 6.4.2.

Another aspect that the Vitoria system aims to explore is user sleep habits. As such, the Vitoria system implements a questionnaire that prompts the users to select the time at which they went to bed, at which time did they wake up, and how they classify their sleep, as can be seen in Figure 5.8c. In addition, with the objective of being able to use this questionnaire to validate the data retrieved from the Sleep Recognition API, the system is able to retrieve information about the sleep quality of the users. We intend to use the latter to extend the Sleep Recognition API with the capabilities to detect the sleep quality as well.

Apart from the questionnaires that are prompted to the user on a daily basis, our system is also able to show the user opportunistic questionnaires, based on their current context. One of them, is the *Proximity/Contacts Questionnaire* that can be seen in Figure 5.9a. The number of contacts a person maintains is very important when considering the context of a pandemic. When the application detects more than two devices corresponding to other people, we raise the Proximity Questionnaire. The user is then prompted to answer how many persons are closer than two meters from him/her, in a numeric type answer. To prevent the application from being bothersome to use, we added a cooldown of one hour to this questionnaire.

Additionally, another aspect that can be considered a risk factor for contamination with the Covid-19 virus is the use of public transportation or the use of shared vehicles with restricted space. One of the activities that the Activity Recognition API is able to detect is if the user is in a vehicle. By using this feature, the Vitoria application is able to prompt the user with the *Transport*



(a) (b) (c)

Figure 5.9: Vitoria's Opportunistic Questionnaires: Proximity questionnaire (a); Transport information questionnaire transport selection open (b); Transport information questionnaire with number of people selection open (c).

Questionnaire, which can be seen in Figures 5.9b and 5.9c. In this questionnaire the users should indicate the transport type and the number of people in the vehicle that are close to them. As can be seen in Figure 5.9b, the available options are: *Own vehicle*; *Friend/Colleague vehicle*; *Taxi/TVDE*[1]; *Bus*; *Subway/Train/Tram*; *Boat*. Depending on the type of vehicle, the options to indicate the number of people are different. The full list of options can be seen in Table 5.1. This questionnaire is triggered every time the *"in vehicle"* activity is detected for more than 2 minutes. If more than one transportation is used, or if there are several stops during a trip (e.g., on a bus ride) the system triggers several transportation questionnaires. However, when there are multiple transportation questionnaires, only the last one will be considered, and the remaining ones are discarded.

Although data acquired directly from the user is more reliable and is well–validated in the literature, one of the main issues when collecting this type of data is the user engagement. In order to address this issue, in the Vitoria system every questionnaire is prompted to the user through the Android notification system. The user can then select each notification and will be redirected to the specific page of each questionnaire. These notifications are permanent until the user clicks or dismisses them, since the user is not required to answer them right

[1]Designation for transports from electronic platforms such as Uber or Bolt in Portugal.

Figure 5.10: Vitoria Main screen without feedback.

way. Additionally, the Vitoria system triggers a notification every day at 16 hours to remind the user to answer any unanswered questionnaires. When the user clicks that notification, he/she is redirected to a page where all unanswered questionnaires are listed. The user can then select each questionnaire and answer them. The pending questionnaires have a validity of 24h after which they are deleted and are no longer shown in that page.

5.2.3 Feedback System

Another aspect that the Vitoria system aimed to explore was how the use of active feedback would affect user behaviour. In order to support that, the system implements on screen personalized information feedback. As discussed previously the *Feedback Granter* module was created to deal with feedback generation and distribution in the Vitoria system. This module is able to divide the users into two feedback groups, namely the control group and the active feedback group (with a 50–50% ratio).

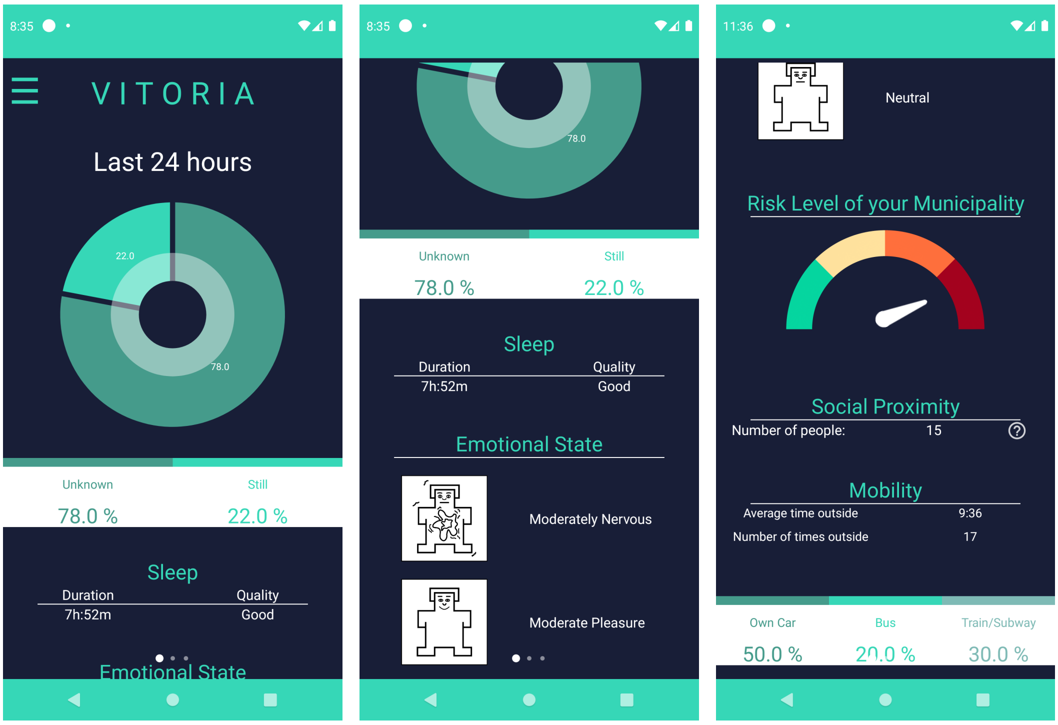
Additionally, the field study presented in this paper was divided into 2 periods: a period without feedback (default measurement), and a period with feedback. In the first period, none of the two user groups receive any feedback. The application layout for this period can be seen in Figure 5.10. In the second

Table 5.2: Feedback differences between groups.

	Control Group	Active Feedback Group
Physical Activity	X	X
Sleep	X	X
Emotional State	X	X
Municipality Risk Level		X
Social Proximity		X
Mobility quantity		X
Mobility duration		X
Mobility type		X
Intervals	last 24 h last 4 days last 8 days	

period both user groups received feedback. However, the feedback differed based on the user group, as can be seen in Figure 5.11. The differences between the groups are presented in Table 5.2. Additionally, users were always presented with the feedback for 3 intervals of time, namely: the last 24 hours, the last 4 days, and the last 8 days. These intervals of time were the same for both feedback groups. Additionally, the feedback information was recalculated every 30 minutes.

As we can see in the table, both groups received the three first types of feedback, namely, sleep information, emotional state information, and physical activity information. The physical activity feedback is generated based on the data provided by the Activity Recognition API, and this information specifies the percentage of time spent by the user in each activity, as can be seen in Figure 5.11a. Sleep feedback is generated from the collected sleep recognition data, and from the sleep report filled daily by the user. This information is shown to the user as the mean time spent sleeping and the mean quality of sleep (i.e., very



(a) (b) (c)

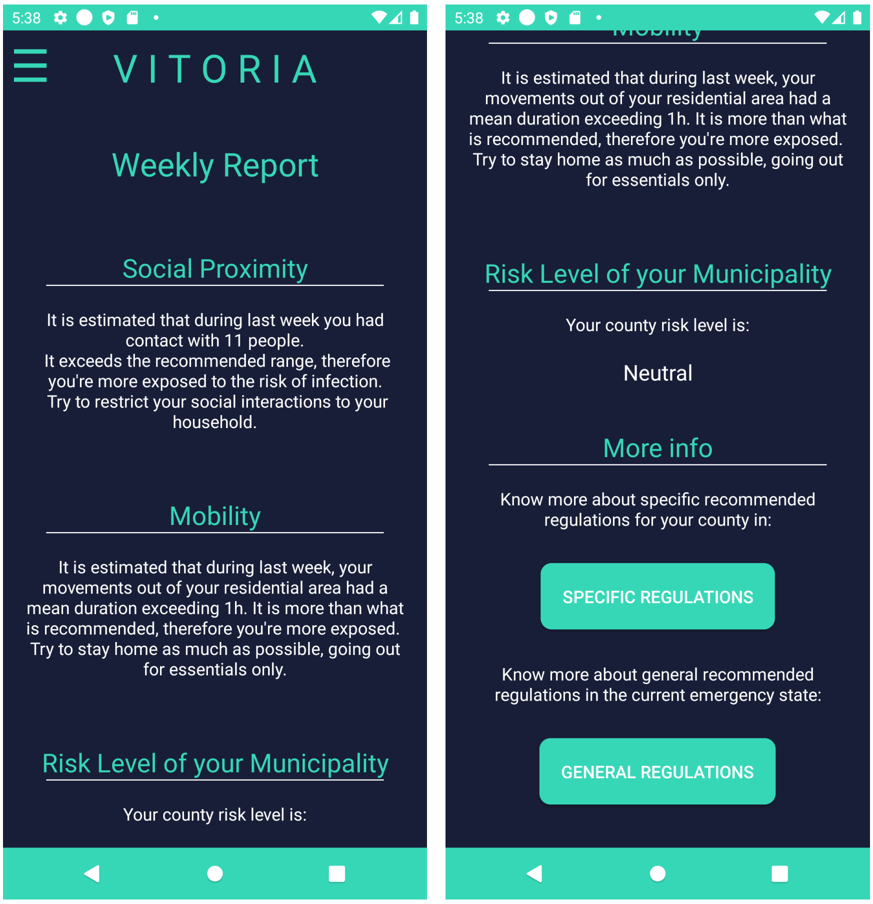
Figure 5.11: Vitoria's Feedback layouts: Control's group Feedback (a) & (b); Specific Active Group Feedback (c).

bad, bad, neutral, good and very good), as can be seen in Figure 5.11a. The Emotional state information is generated from the daily emotional questionnaire. This feedback is presented using the same figures from the SAM questionnaire, as can be seen in Figure 5.11b.

As we can see in Table 5.2, the active feedback group received more information than the control group. This information was directly related to the SARS-Cov-2 risk of infection. For instance, the users on the active feedback group received information about their municipality risk level, based on local infections and incidence rates. As explained before, in section 5.2.1, this was done by comparing the risk level obtained from the reverse geolocation with their discrete location. As can be seen in Figure 5.11c, this information is presented to the user in the form of a gauge divided into 4 sectors that represent the 4 denominations used by the Portuguese authorities (i.e., moderated, high, very high, and extremely high).

Additionally, the users from the active feedback group also received information about the mean number of possible contacts that they had in each specific time interval. This information was generated based on the answers to the *proximity questionnaire*, and was presented to the users in the form of a numeric value. Users also received information about their mobility, namely the mean time the users spent outside their home, and the total number of times that they went outside. In Figure 5.11c we can see an example of this feedback for an 8-day time interval.

As we stated before, some means of transportation can lead to a higher risk of infection due to higher proximity to other people. As such, in addition to the previously mentioned feedback, the application also shows the percentage of time spent on specific means of transportation, as can be seen in Figure 5.11c. This percentage is calculated based on the *transportation questionnaire* that the application prompts the user to fill whenever the *"in vehicle"* activity is detected for more than 2 minutes, by the Activity Recognition API.



(a) (b)

Figure 5.12: Vitoria's Full Weekly Feedback

Furthermore, in addition to the feedback explained so far, the users also receive weekly feedback with the aggregated data for that week. Although, the users already receive feedback for the last 8 days, that information is recalculated every 30 minutes, and, as such, is continuously changing. On other hand, in the weekly report, the information is static until the users receive the next weekly report. The report was set to be sent every Saturday at 21 hours and this feedback can be seen in Table 1, in the annexes.

Starting with the *Control Group* we can see that the feedback given to this group was generalist, with only two of the four items in the weekly feedback items. The last item in Table 1 was common for both groups, with information about the current measures implemented by the Portuguese authorities. The penultimate item of the table was also given to both groups. However, only the group with active feedback received information about their respective municipality risk, while the control group only received the general guideline.

The Active Feedback group received two more items of feedback, namely the feedback about the number of contacts and the mobility outside the home area. As we can see in Table 1, these two items are also divided into positive reinforcement and negative reinforcement. That is, whenever the user meets the recommendation from the health authorities, they receive a positive message; whenever they fail to meet the recommendations, they receive a negative one. In the case of the number of contacts, the recommendation was that users would limit the number of contacts to 10 or fewer people. And in the case of mobility outside their home area, the recommendation was that it should be limited to a minimum, which led us to choose a threshold of one hour.

As explained before, the users received a notification, every Saturday at 21 hours, which redirected them to the weekly feedback page in the mobile application. This information could then be accessed over the course of the following week, until a new notification was received, and the information was updated. This information was presented as shown in Figure 5.12, with the control group only having access to the *"More Info"* part of the layout.

Although one of our objectives was to evaluate the implemented feedback, lack of data also invalidated such analysis, as dividing the dataset into two parts would further reduce the size of the available data. Nonetheless, we believe that the implemented feedback system is functional and would be benefit for the users in the context of a future pandemic. We intend to use this system in future studies and further evaluate how users that are offer active feedback differ from a control group.

5.3 CentroAdapt - Bridging academy and industry

CentroAdapt is a project for the transfer of scientific knowledge and technology in the context of adaptations to climate change, between researchers from the University of Coimbra and entities in the central region of Portugal. Climate change poses a global threat to natural and human ecosystems at an environmental, social and economic level. Implementing concerted and sustainable strategies to mitigate its effects and enhance adaptation to a changing climate is a short-term priority. In this sense, *CentroAdapt* assumes the role of facilitator of information between academia and companies/entities, stimulating the potential definition of agents' needs, and the search for alternatives for future challenges in the face of climate change.

The mission of this project is to establish an open and inclusive innovation platform that incorporates the necessary skills for an efficient and effective transfer of knowledge. This was to be achieved by two main methods: firstly by the creation of networking sessions between researchers and potential problem holders; And secondly by creating a technological platform that would be able to aggregate information about current problems to be solved, as well as research in related fields. This technological platform is also able to use the collected in-

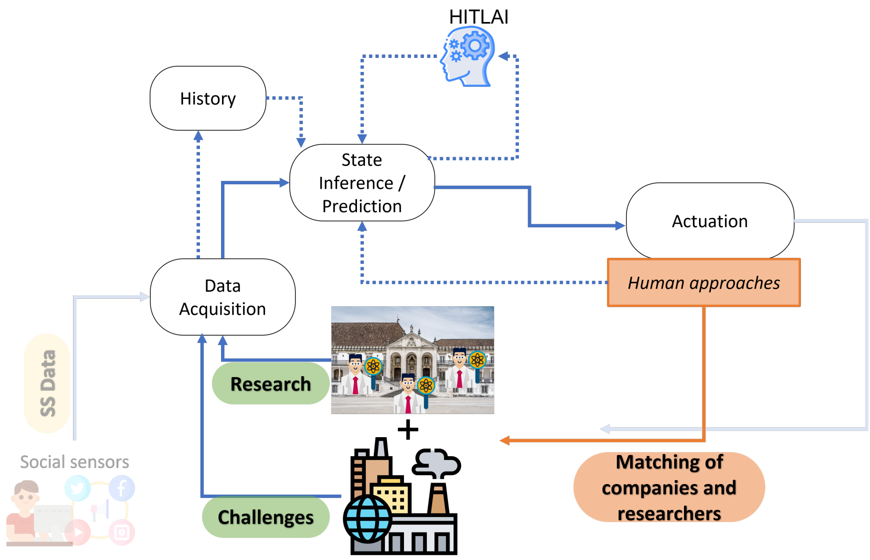


Figure 5.13: *CentroAdapt* HITLCPS architecture.

formation to create a recommendation system and generate recommendations, for potential matches between researchers and challenges in the platform.

The main issue that this technological platform aims to solve is to bridge the gap between research currently being done in academia and real–world challenges that the industry is facing, and finding the *"best person for the job"* for each challenge created in the platform. Although, the *CentroAdapt* project focus was to solve challenges created by the climate changes, the created platform can be used in any other case–study.

In Figure 5.13 we present a diagram of the *CentroAdapt* system in the context of the HITLCPS paradigm. As can be seen from the figure, there are several differences from the previously shown case studies. The main focus of this system, in the context of HITLCPS is to implement and evaluate HITLAI mechanisms. As we have stated previously, having the human in the loop of the decision of machine learning models (i.e., HILTAI), is one of the important concepts to achieve a HITLCPS paradigm. The work developed in the context of the *CentroAdapt* project explores this aspect of the paradigm, by presenting a decision–helping recommendation model that helps the human make better decision and also learns from the human's feedback.

In this section we explore the implementation of the platform, and how data collection is performed, as well as the implementation of the recommendation system.

5.3.1 Platform and data aggregation

As stated before, the web platform is constituted from 3 main components, a web dashboard, a recommendation system and a backend that serves as a mean to aggregate the data and communicate between the other components. The architecture of the system can be seen in Figure 5.14. We can see that the

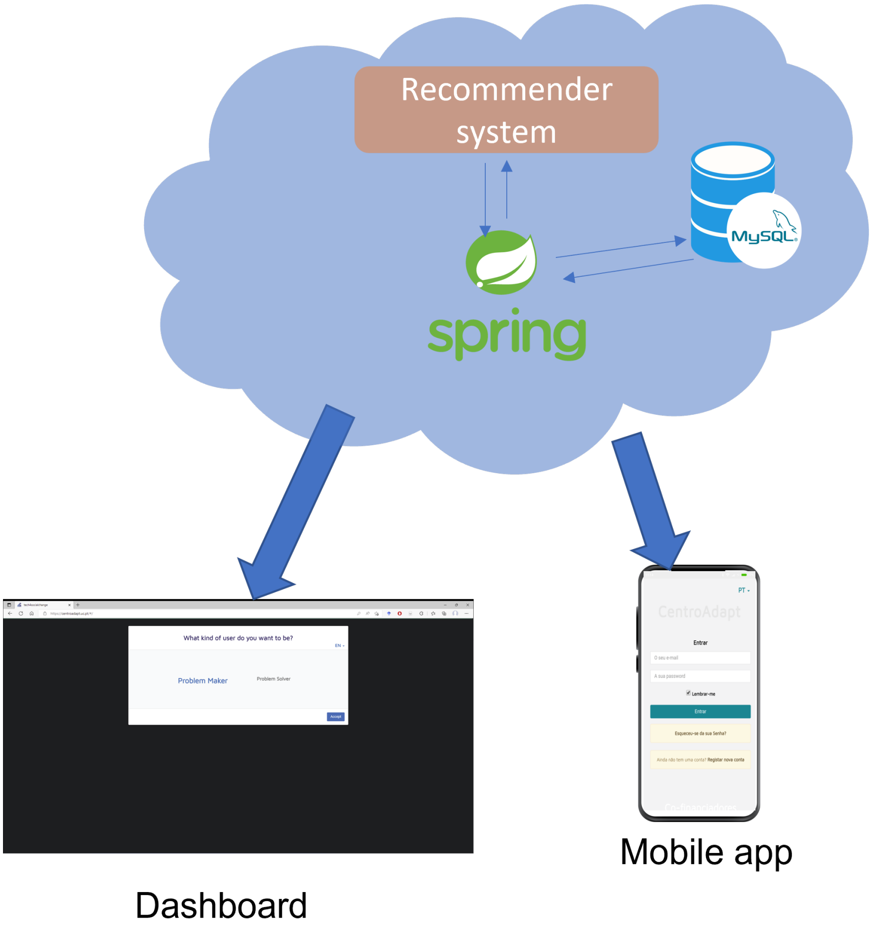


Figure 5.14: Simplified architecture of the *CentroAdapt* platform.

application has both a mobile and a web client application (dashboard) that can be used by both, the administrator, as well as the users. The backend of the application is hosted in a cloud system and is constituted by the web API] backend, built on the springboot framework, the database (MySQL database) and the recommender system which is a custom module built using python and the flask library. The recommendation system will be explored in further detail in the next section.

The web API backend is responsible for implementing the logic of the system and answering any requests from the client applications. It also enables queries to the database to fetch any requested data. Additionally, this component is responsible for scheduling the training tasks of the recommender system every time a new, challenge is created. Once the recommendation system is trained, the web API is also able to request the recommendation list of the top ten *"best matches"* for a specific problem.

As we stated the before, the primary functionality of this system is to enable the aggregation of challenges that the industry faces, and the research being done by the academia in the same system. In order to do this, the web–clients must offer a way for both parties to import this information to the platform. This was done, by implementing three types of users with different permissions in the platform.

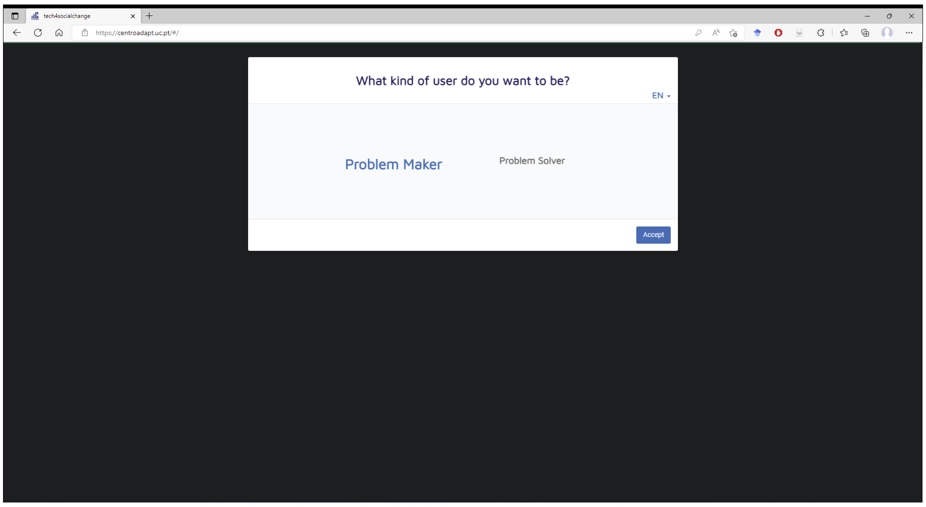


Figure 5.15: Account type selection during first login.

Namely, the *"problem makers"*, the *"problem solvers"* and the administrator of the system. The administrator of the system is a type of account with special permissions, has it can access all information available in the platform. And it can use the recommendation system to obtain recommendations and make matches between *"challenges"* and *"solutions"*. As such, this account is set on boot manually on the database and its information cannot be changed by the web clients.

The other types of account can be initialized by a symple register-login system, using a valid email account and a password. During the first login a user must select which type of account he desires *"problem maker"* or *"problem solver"* as can be seen from Figure 5.15. As we stated previously, the *"problem maker"* account type are ment to target companies that face problems caused by climate change, while *"problem solver"* is the type of account that any research or person of expertise in desired fields should select. The type of account also limits the functionalities available for the user.

Problem makers are able to post challenges to the platform. Users with this profile can also see and edit their own challenges, see their status (resolved or unresolved), or remove them. As can be seen from the Figure 5.16, to add a challenge to the platform a user must insert a title, a description and at least 2 tags/keywords that describe the challenge. The tags can be selected from a list of already existing tags or can be inserted manually, the rest of the information is also manually inserted. All of this information is used by the recommendation system as will be explained in the next section. Additionally, the user also needs to add a phone number (which can also be set to the profile account as a default for all challenges), which can be used by the administrator of the platform to contact the user to obtain more details or to finalize the process of matchmaking.

As stated before, *"Problem solvers"* can be any person with the proved know-how in relevant areas to solve challenges, caused by climate change, nevertheless the platform main target is researchers/professors from universities. Most re-

Figure 5.16: Example of the insertion of a challenge into the platform.

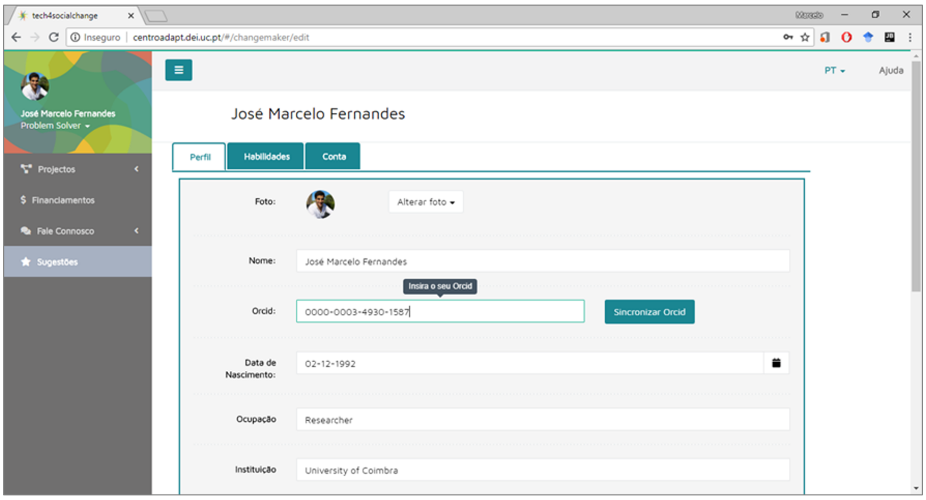


Figure 5.17: Example of the integration with the Orcid platform.

searchers, especially the ones with longer careers, have a long list of projects and reports to showcase their work. Furthermore, there are already several online platforms that keep track of this information (e.g., Google Scholar, Orcid, ResearchGate). We choose to implement an integration with the Orcid platform, by using the API that this platform offers, which allows us to fetch the public profile of any registered researcher by their id. In order to do that a user, only needs to insert their id into the profile section of the platform, as can be seen in Figure 5.17. This information includes, their name, institution, email, current profession, list of projects and list of projects' funding. The user can also filter which information to import from their profile, as can be seen in Figure 5.18. All the retrieve data, can then be used to describe a specific user profile and train the recommendation model.

However, since there is a large of researchers, it would not be viable to ask every researcher to register to the platform, and import their data. As such, the platform also allows the administrator to create of anonymous entries for researchers. That is, it is possible to create a researcher profile without it being

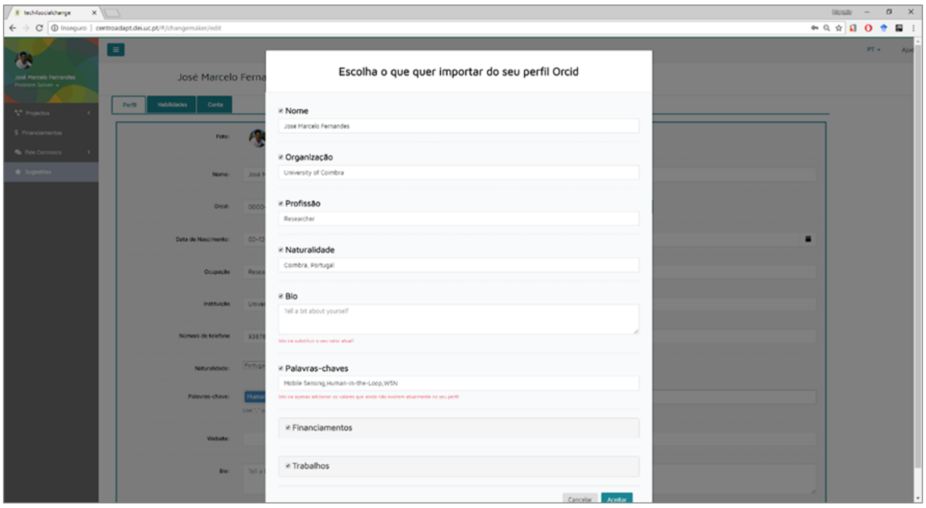


Figure 5.18: Optional filtering of information to be imported from orcid.

bind to a user account. The administrator, is able to register several researchers and import their data from Orcid, by uploading a *.csv* file with the researchers' Orcid ids. Additionally, since the information on Orcid can change over time, the platform also has a service scheduled to update any information imported from Orcid.

The administrator of the platform is intended to be someone that already is responsible for making the connections between the industry and researchers that work at the University of Coimbra. This account type allows the access to all the information available in the platform. Furthermore, this user has access to all create, read, update and delete operations on the platform and database. However, the main goal of this account is to allow a given user access to the recommendation system, this will be explored in the next section.

5.3.2 Recommendation system

When considering recommendation systems to match users with content, there are two main approaches. Namely, content-based filtering and collaborative filtering (Isinkaye et al. [2015]). We can see in Figure 5.19, the example of these two techniques applied to the recommendation of new articles to users. Content-based filtering techniques use the past history of a user to recommend its next *"product"*. While, collaborative filtering techniques compute the similarity between different users to recommend *"products"*, based on user similarity. In our case study, we choose to implement a content-based filtering approach, as it was the most suitable to our business logic.

Considering the specific case, the recommendation model must recommend *challenges*, based on the past history of users (i.e., projects, funding list, description, etc). Since, the entities (i.e., *challenges* and users history) are very different we must first extract the list of keywords that represent each entity and can be used to compute the similarity between the entities.

Another problem that we must consider, is the fact that the platform support

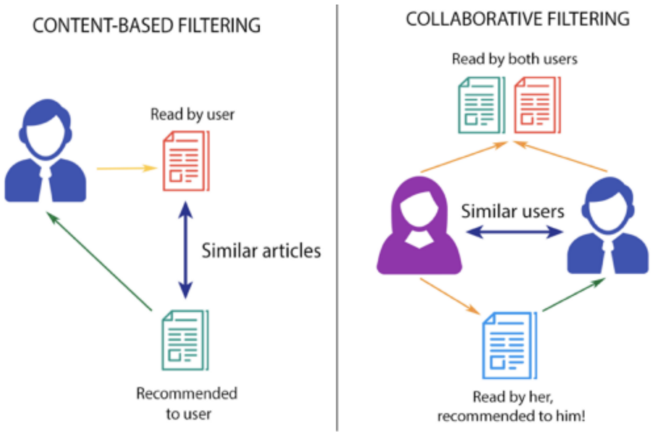


Figure 5.19: Example of recommendation systems for both collaborative and content-based filtering (image from Doshi [2018]).

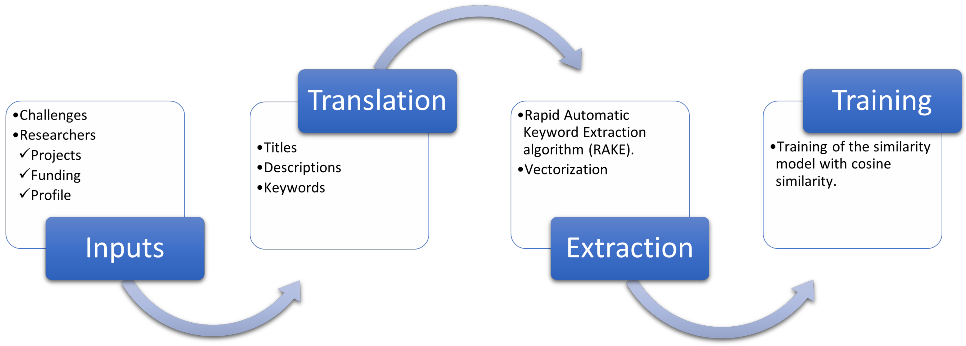


Figure 5.20: Pipeline of the recommender system.

multiple languages on the client applications (i.e., english and portuguese). This would invalidate the match between entities inserted with different languages as the list of keywords would be different. The information from researchers (imported from Orcid) is already mostly in english, however since the *challenges* are inserted manually they can be inserted into any language that the user chooses. Since, the case study is implemented in Portugal, we expected most *"problem makers"* to insert the *challenges* in portuguese. As such we must first convert all data to the same language.

The pipeline for our recommender system can be seen in Figure 5.20. The recommender system was implemented in python, as this language offers a set of libraries that enable the creation of recommender systems. As we can see from the figure, we first extract the inputs from the database. For the *challenges*, we considered the title, the description and the keywords inserted at the time of the challenge creation. Considering the data available to each researcher (user history) we used the projects' title and description, the funding title and description, and the researcher summary/description and keywords. We then translate all the text to a common language, namely to english, by using the Google translate API (PyPy [2020]). This API offers, two main capabilities, namely detecting the language of inputted text and translating inputted text. We first detect the language to ignore any text that is already in english, and

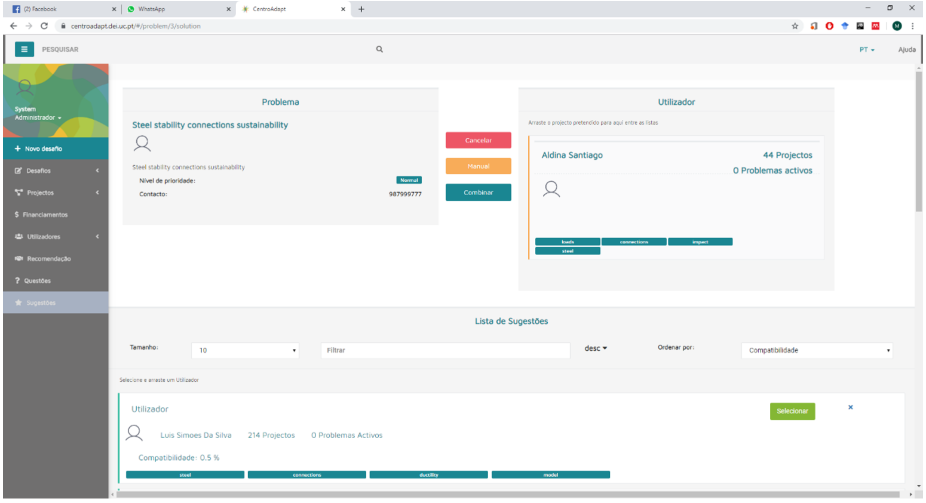


Figure 5.21: Page that allows the administrator to choose the solution to a given problem.

then translate any text that is still on portuguese. Once all the text is on the same language we proceed to extract the keywords for each challenge and for each user history. We use the Rapid Automatic Keyword Extraction (RAKE) algorithm, which tries to determine keywords in the inputted text by analyzing the frequency of word appearance and its co–occurrence with other words in the text. The found keywords are then vectorized by using the *CountVectorizer* of the sklearn library. Finally, the cosine similarity between the matrix containing the *challenges* vectors and the matrix containing the users vectors is computed. Generating a matrix of $N \times M$ (N *challenges* and M users). Recommendations are then generated by selecting a row of the similarity matrix that represents a given problem, sorting it by ascending order and selecting the first 10 columns representing the users which are most similar to the *challenge*.

The model is scheduled to train every half an hour, since the information of all the entities on the platform (*challenges* and user history) can be updated, which can affect recommendations. It is also important that the platform is able to generate recommendation for newly added *challenge*, as such, the model is also trained every time a new *challenge* is inserted. Additionally, the model can also be trained manually, by the administrator of the platform.

Additionally, as we can see from Figure 5.21, the administrator can use the platform to access a list of potential *"problem solvers"*, and their compatibility index (in percentage) to the selected challenge. The administrator, can then lock a match between a *challenge* and the selected *"problem solver"*. The administrator can then contact both parties (using the platform or other means) to start the process of developing a solution for a specific challenge. Even though the recommendations are made automatically, based on the data available to the platform, the match must be validated by the administrator. That is, the platform serves as a decision-helping tool for the administrator and never as a decision–making tool.

By having the human in the decision loop, we are inserting it on the loop of the system, that is the HITLAI. This creates a more reliable system, that is able to better perceive each specific *challenge* and select the best solution. Furthermore, the system itself learns from the *matches* that the administrator selects, that is the feedback of the humans are used to increase the performance of the system. We aim to extend the work developed in this smaller case study to other implementations of the HITLCPS by taking similar approaches in the models used for state inference.

5.4 iFriend - Joining HITLCPS and Unobtrusive sensing

As previously stated, perceiving humans is hard due to unpredictable nature. However, actions and emotions are accompanied by involuntary reactions at the physiological level, namely, subtle changes in the HR and BR signals. As such the implementation of sensing mechanisms able to monitor changes in these signals are a must for the creation of a full HITLCPS.

There is already some work that explore acquiring these signals in the context of HITLCPS, however most of them focus on the use of wearables. Although, wearable technologies have been widely explored and carry several advantages when compared to traditional solutions (e.g., ECG), They are also accompanied by several drawbacks. Namely, the lack of validation for medical use, the fact that many of these devices are not accurate and prone to errors, and the high *dropout* rate of users. The last drawback is especially difficult to overcome, as another open issue from the HITLCPS is the lack of adoption/participation of the humans in systems tasks, being them from a sensing nature or other. This could be even more significant when considering a case–study that targets elderly people or people which are not familiar with the use of technology. As we explored in the proposed model in section 3.2, we believe that one of the open issues of the HITLCPS is the implementation of new sensing techniques that are able to interface with humans in a passive/unobtrusive manner, and that are able to retrieve these physiological signals.

In this section we will explore the case–study developed for the project: *"iFriend – Supervisión inteligente del estado de salud en personas mayores con insuficiencia renal mediante dispositivos inalámbricos"* supported by Fundación CSIC, Interreg Portugal-Espanha. This project was developed in partnership between universities from both Portugal and Spain and a Spanish hospital, as such it included professionals from both the technological field and the medical field. The project aims to create a platform that is able to monitor elderly people that suffer from renal insufficiency. Renal insufficiency, can affect heavily one's life and could even lead to dead when kept unchecked. For that reason, patients with this pathology are recommended to keep a healthy lifestyle, i.e., practice exercise, having a healthy sleep cycle and maintain their medication properly. In this project we aim to create a system that not only is able to track the patients and give detailed information to medical staff, that is responsible for that patient, but also help the patient keep up with the recommendations from its medical staff.

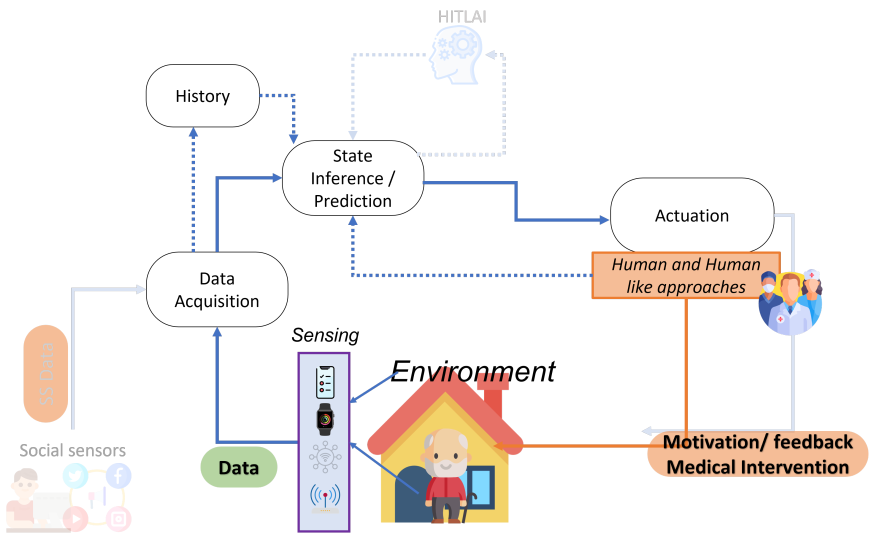


Figure 5.22: iFriend HITLCPS architecture.

In Figure 5.22 we present a diagram of the iFriend case study in the context of the HITLCPS paradigm. As can be seen from the figure, with this system we also intend to study the implementation of all the phases of a HITLCPS. However, we give special focus to the use of unobtrusive sensing as a mean to acquire data from the humans vital signals. Additionally, in this system since it is critical (i.e., it is related to e–Health), the actuation is moved from interactive agent to feedback and actuation directly from the medical team, which can post-process any inference of the system and make decisions.

As we also explored in chapter 4 there are several approaches for unobtrusive sensing. These approaches are very distinct in their sensing nature. We believe that the one that offered the best approach for our case study is the one based on the detection of vital signals by measuring the interference of the patient bodies in CSI of Wi-Fi signals. In the next section, we will give an overview of the general architecture of the system.

5.4.1 The iFriend Architecture

The *"iFriend"* project aims to join a HITLCPS with unobtrusive sensing, as such the architecture of the system must be able to deal with data from different sources and formats. For that reason, we choose an architecture based on a IoT middleware, namely the FIWARE ecosystem, as can be seen from the Figure 5.23. We have explored the implementation of this backend in section 5.1, and except differences in configuration and data formats the system implementation is very similar, as such we will not address this part of the implementation. The full architecture of the system can be seen in the Figure 5.23.

Additionally, in this system we also have data coming from devices dedicated to unobtrusive sensing, namely acquiring HR and BR from the CSI of the Wi-Fi signal. CSI data has a particular format which is represented by a 3D matrix of $T\times\ R\times\ C$, with T being the number of transmitting antennas, R being

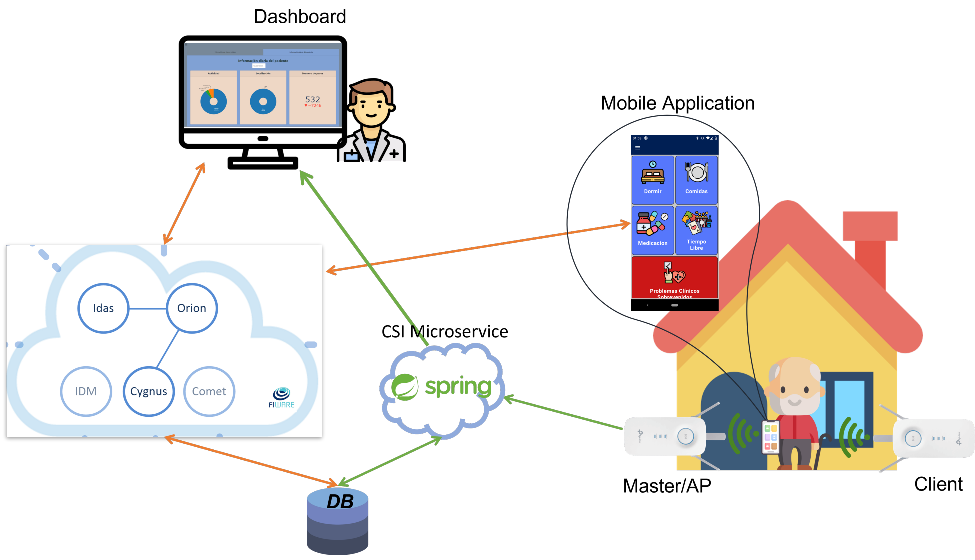


Figure 5.23: iFriend system Architecture.

the number of receiving antennas and C being the number of sub-carriers of the Wi-Fi signal. Additionally, every matrix cell contains a complex number, which represents the amplitude and phase of each component of the signal.

As can be seen from Figure 5.24, the device chosen to implement the CSI collection was the TP-Link RE450 (RE450 [2022]), which was used to implement the client (transmitting device) and the AP (receiving device). These devices were flashed with the OpenWrt operating system (OpenWrt [2022]), which allows us to have finer control of these devices and the Atheros CSI tool (Xie et al. [2015]) was used to transmit and obtain the CSI data. As can be seen from Figure 5.24 the device used 3 transmitting antennas and 3 receiving antennas, and since we were using the 2.4 GHz band at the 20 MHz bandwidth the signal was composed of 56 sub-carriers. As such, every transmitted packet generated a $3 \times 3 \times 56$ matrix of complex components. In addition to the elaborated format of the data, there is also a large volume of it being generated, since the rate of transmission is 20 packets per second. For that reason, a microservice was created to deal with any data created by these devices as can be seen in Figure 5.23 (CSI Microservice). This microservice was implemented with the Spring framework and mainly served as an interface between the end-devices (AP) and the database. The microservice is also able to serve the dashboard with data queried from the database. It is important to note that we chose to extract the raw data from the CSI measurements, instead of any processed metrics, in order to test and evaluate different models for the prediction of HR and BR.

As can be seen from Figure 5.23 in addition to the two devices used to collect the CSI measurements, every patient will be also tracked by a mobile device. The mobile application is able to collect passive data (e.g., activity, location, sleep behaviours), as well as data directly provided by the user. We will cover the mobile application in the next section.

Additionally, we can also see from the Figure 5.23, that this system also has

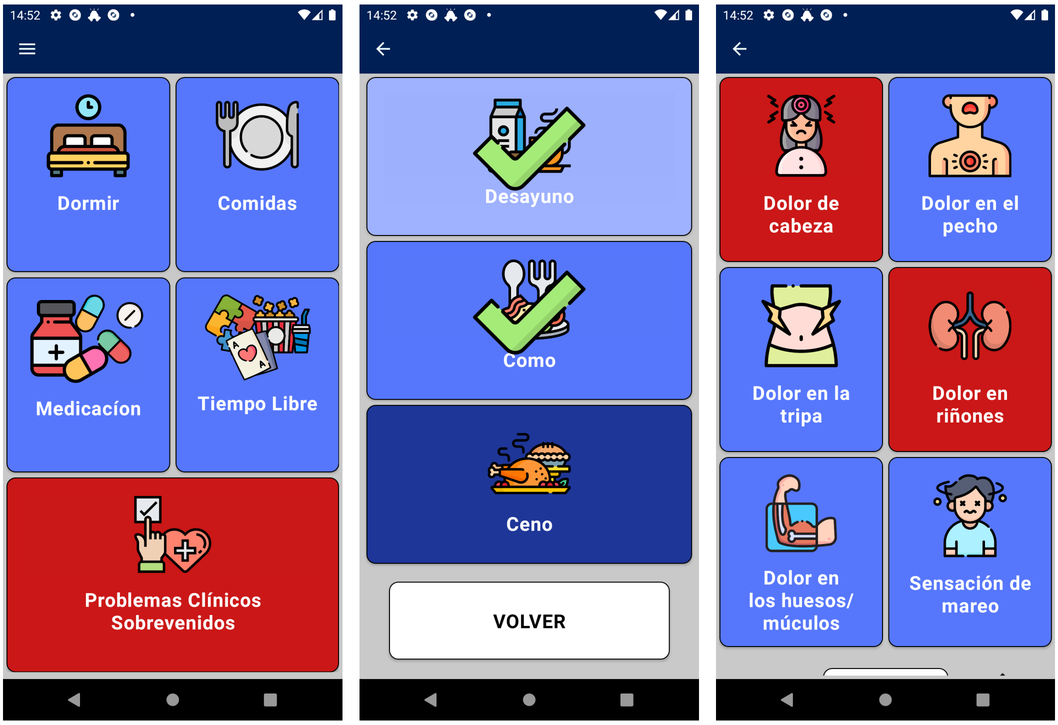
Figure 5.24: Wi-Fi Access Point/Client used in the iFriend system (TP-Link RE450).

a dashboard that is able to give information to the medical staff. Due to the nature of this application it is important that the decisions in the system are made by someone with logic knowledge of the problem. As we stated before, the inclusion of the human in the state inference and actuation process is one of the important issues to address in HITLCPS, especially so in cases that are of sensitive nature such as this case study that deals with a medical pathology. In this dashboard the medical staff is able to view and explore the data collected from both types of devices, i.e., smartphone data and HR/BR estimations from the CSI data. It is possible to receive a feed of real–time data, supported by the subscription/notification capabilities of the ORION GE of the FIWARE ecosystem. And it is also possible to analyse the historic data of each patient.

In addition to the mobile application and the devices used to collect CSI, we also used a smartwatch application and a respiration belt (Vernier [2022b]). These devices were used to obtain the ground–truth for both the measurements of HR and BR. Both of these devices are connected to the smartphone application and all information is sent and store through the FIWARE back–end.

5.4.2 The iFriend Mobile Application

The mobile application design for the *"iFriend"* project followed the same design patterns of the other applications addressed in this section. However, some changes were made in order to improve over past drawbacks and tailored the system to the specific case study. The main change was the move from native Android development to the use of the Xamarin framework (Microsoft [2022]), which is a cross–platform development tool that allows for the creation of mobile applications for both Android and iOS systems that run at native performance. This change was done to overcome one of the challenges that we faced in past studies, which was the difficulty to obtain subjects for the case studies and to maintain their engagement in the study. By including the iOS platform as a target system of our application we are undoubtedly increasing the number of potential participants of our studies. The second drawback is not directly connected to the use of a single platform, but by obtaining more participants we



(a) (b) (c)

Figure 5.25: iFriend's mobile application: Main layout of the application (a); Example of the options available at the *"Meals"* sub-menu (b); *"Clinic problems"* sub-menu (c).

will be able to better deal with dropouts during ongoing studies. Although, the scope of this project will have a smaller group of subjects handpicked to target the particular case–study, the project also aims to present a system that is able to help as many people as possible. As such this was a very important change to our past mobile applications architecture.

Another important aspect that we had to take into account in this project was the User Experience (UX) of the mobile application. As we stated before, HITLCPS should take into account the behaviour of a human and be aware of its context, however, it should also be aware of his/her restrictions/condition. Since we are targeting a case–study with elderly people, several UX design options had to be made to adapt to the difficulty that people of higher age have to adapt to technology. We can see some examples of the layout of the application in Figure 5.25.

One of the UX changes made was the option to use a simplistic layout based on larger images and fewer text. Many smartphones nowadays offer the option to enlarge content to help people with lower sight capabilities, as such we opt to use the same design pattern for the application. Furthermore, as can be seen from Figure 5.25a, the navigation of the application was also designed to be minimalist. Figure 5.25a, shows the main page of the application from which the user can navigate to the available sub–menus. These sub–menus include the main daily routines of our subjects, namely their sleep patterns, their meals behaviours, their medications, and their leisure time.

When the user selects one of these options, he/she is able to navigate to specific options as can be seen in Figure 5.25b. The user can then select the button, that corresponds to the action he/she has done, the application then collects the timestamp of that event and sends it to the back-end of the system. In this particular case–study the medical staff is mostly interested in changes in the behaviour patterns of the user. As such, keeping track of these events and capturing potential changes will help to monitor those changes. Most of the actions covered in the sub–menus are only to be performed once a day (e.g., going to bed, having breakfast, taking morning medication). As it is possible to see from the Figure 5.25b, once a button is clicked in a sub–menu that option is highlighted with a *check mark* and become disable to let the user know that the action was performed.

Another option that is the users are able to select from the main menu, as can be seen from Figure 5.25a, is the option to report clinic problems. Keeping track of symptoms of patients with this clinical condition is very important, as their very prompt to face acute symptoms. Furthermore, it could help the medical staff to give a more timely response, since they will receive the information at almost real-time in the dashboard, and the patient will not necessarily be required to go to the hospital. As it is possible to be seen from Figure 5.25c, the application covers some of the most probable conditions that a patient with renal insufficiency will face. As happened for the behaviours, the application also collects the timestamp of every clinic problem event, as it is not only important to know what problem the user is facing but also the beginning time and duration of said event.

In addition to the data collected actively by interfacing with the user, the application is also able to retrieve data passively, by using the context information and sensors of the smartphone. Namely, it is able to collect the activity, location, connectivity status, step count, accelerometer, gyroscope, proximity sensor values, light sensor values, screen lock status, nearby devices (BLE and Bluetooth), GPS location, application usage metrics and sleep behaviours. Although with this application, we moved from an Android native application to a cross-platform application, the Xamarin framework is able to interact with the native libraries of each system as such it is possible to access low-level APIs such as the sensors' API. Furthermore, the iOS operating system offers several counterparts for the APIs of the Android operating system (Inc. [2022]). We have already covered most of the acquisition process in sections 5.1 and 5.2, as such we will not address it in this section.

In future iterations of the application, we aim to also extend the application with a feedback mechanism that is able to give tailored recommendations to each user based on their past behaviour. And also create a direct feedback mechanism from the medical staff to each patient based on their staff analysis of each situation.

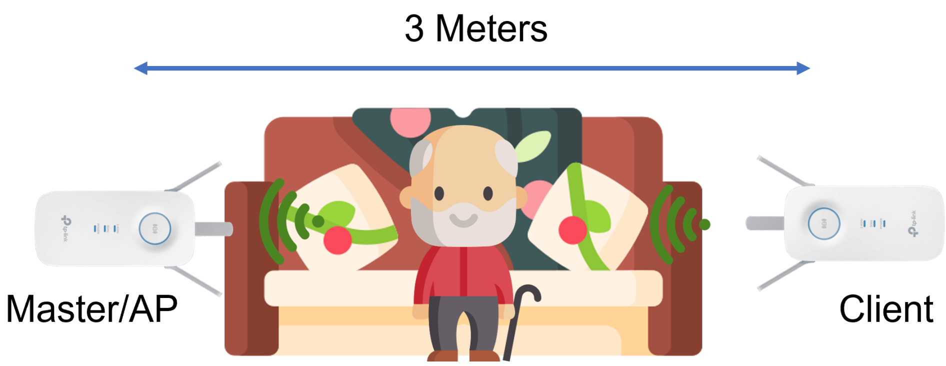


Figure 5.26: Scheme of CSI retrieving using the Wi-Fi Access Point/Client in the iFriend system.

5.4.3 Retrieving HR and BR from CSI signals

As we explored in section 4.2, one of the possible approaches to retrieve the vital signals of humans, namely HR and BR, is by analysing the interference caused on the CSI of the Wi-Fi signal when traversing the human body. The work of Liu et al. [2015], shows promising results while monitoring up to two person during their sleep. However, it also had some limitations, in the *"iFriend"* project we aim to implement a solution that offers better performance and solves some of those limitations.

One of the limitations that we aim to solve was the hardware dependency, which limited the scalability of this type of solution. Although, the solution in Liu et al. [2015] was presented as *"using off-the-shelf Wi-Fi"*, it was implemented by using a CSI tool which only works with the Intel Nic 5300 card (Halperin et al. [2011a]). This would limit the number of systems which would be able to be leveraged as sensing devices. As such, we choose to implement our solution by using the Atheros CSI tool (Xie et al. [2015]) which is, in theory, compatible with any Atheros or Qualcomm device. Furthermore, this library is compatible with both Linux and OpenWrt. OpenWrt is an open-source, and largely compatible operating system for Wi-Fi routers and AP, which allows for the deployment of programmable solutions in these devices. For our case–study we choose to implement our solution by using OpenWrt, since it offers us not only the larger pool of compatible Wi-Fi devices, but also gives control over the transmitting and receiving data. Furthermore, Wi-Fi routers and AP are largely available in most houses, improving the scalability of our solution and making it less expensive.

Additionally, this library allows for the retrieval of more precise information, since it allows for the collection of CSI information not only in the 20MHz bandwidth but also in the 40 MHz, increasing the number of captured sub-carriers from 56 to 114. The library also represents the CSI component values with more precision, since each component is represented with 10 bits for the real part and 10 bits for the imaginary part (while the Intel NIC only offers 8 bits for each component). This allows for more precision when analysing the retrieved data.

Another aspect that we aim to improve is the scope of the sensing activity, and its limitation in more complex situations. In Liu et al. [2015], the authors monitored the HR and BR during sleep. Although, sensing while a person is sleeping can give us important information about one's health, and it is one of the sensing opportunities at which the sensing can be more precise do to the lower level of movement. We believe that this leaves out several sections of a person's daily time that can also be very important to monitor. Furthermore, HR and BR decrease to lower levels while people sleep, as such it is important to validate the performance of this technique during waking time as well.

However, as we explored in section 4.5, one of the challenges of this type of technique is the amount of noise during the sensing process, namely caused by larger movements of the body. As such the sensing must be done in static or quasi-static situations. For the case study of the *"iFriend"* project, we propose the situation depicted in Figure 5.26. As can be seen from the figure, we propose that the sensing will be performed during the time that the subject sits on their couch. The medical staff that advises the project believes that this is where the subjects spend most of their day, either watching tv, reading or on their mobile devices (smartphones/tablets) consuming content from the Internet.

As can be seen from Figure 5.26 the sensing will be performed using two identical devices, namely the TP-link RE450 also shown in Figure 5.24. these are inexpensive devices that can be flashed with a custom OpenWrt firmware, and allow us to develop code to be run on them. One of the devices acts as an AP broadcasting a Wi-Fi network, while the second acts as a client that connects to the Wi-Fi network and sends packets periodically.

Although the case study is yet to be deployed, we performed some preliminary results. We explore these results in chapter 6.

5.5 Chapter Summary

In this chapter, we presented the implementation details of the case studies developed to study our HITLCPS model. We start by exploring the case study named ISABELA which uses an IoT architecture based on the FIWARE backend, smartphones, smartwatches and small IoT devices to monitor students. This system aims to monitor the students' behaviours, lifestyles, and daily routines, and to infer their performance. The system also includes an approach based on a chatbot to closes the loop and give feedback to the students.

During the development of this thesis, the Covid-19 pandemic happened, and it brought several drawbacks to our research work. However, it also constituted an opportunity to employ the HITLCPS paradigm to monitor humans during a never faced before situation. We leveraged the architecture proposed with the ISABELA system to create a new system capable of monitoring people during the Covid-19 and future pandemics, called Vitoria. This application extended the capabilities of the ISABELA system by being able to collect data that can be related to the risk of contagion and with the user perception of that risk. We also implemented mechanisms to evaluate the effectiveness of the feedback given by the system, namely, by creating a module that was able to divide the users into two groups (feedback and control), and manage the feedback that which group received.

Another aspect proposed in our model for the HITLCPS paradigm is the use of HITLAI in the state inference phase of the loop. In this chapter, we presented the "CentroAdapt" platform that is able to generate automatic recommendations for matches between challenges related to climate changes, faced by industry companies, and research done in the academy. HITLAI, is generally implemented in one of two manners, namely, using humans as a mean to evaluate classifications and find misclassifications to help the model learn, or by using models to offer humans more context in a system where the human has the last decision. The "CentroAdapt" platform implements the last, that is, the model generates recommendations, however, it is the responsibility of the humans to make the final decision (match the challenges and the researchers). Additionally, the actuation in this platform is also made by humans, since a human with the role of platform administrator is the one that chooses the best way to approach both researchers and companies.

Lastly, we finish this chapter by presenting the implementation details of the "iFriend" project. The "iFriend" project aimed at monitoring elderly people that suffer from kidney failure, by using smartphones, IoT, and unobtrusive sensing to collect data. In this case study we specifically explored the implementation of unobtrusive sensing as a means to obtain data in a HITLCPS. Additionally, we tackle the challenges of designing a system that takes into account specific human limitations, namely, the limitations inherent to old age.

To further evaluate the implementation of our HITLCPS model several field trials were performed in the real–world, using the systems presented in this chapter. The next chapter covers the implementation details of those trials and presents the results from the tests performed with the data acquired from them.

Chapter 6

Evaluation results

Contents

In this chapter, we will explore several field trials developed using the platforms and applications explained in the last chapter. Furthermore, these trials resulted in the acquisition of several datasets. We will address the results obtained from analysing those datasets and the creation of models for the HITLCPS. Specifically, we present the results from the ISABELA first study, which was developed in both Portugal and Ecuador. We also present an approach to predict students' performance using the datasets from the ISABELA's trials, which were collected in two distinct periods of time. Additionally, we explore the results from the *Vitoria* trials, which were developed in Portugal during 3 and half months, in 2021 during the covid pandemic. And lastly, are presented the preliminary results from tests which aim to analyse the prediction of the Human HR and BR using Wi–Fi CSI, in the "iFriend" project.

6.1 Analysing Students' Behaviour

ISABELA was subject to several tests, involving real users, more specifically, students in their academic context. This section provides some results, organised into three sections:

A) Results of a trial involving 10 Portuguese university students, focused to sentiments analysis;

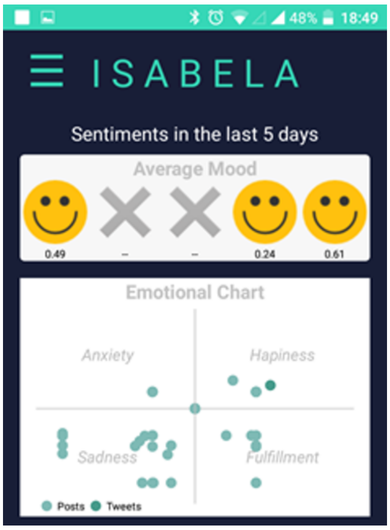
B) Results of a trial whose objective was to correlate users (i.e., students) location with their academic performance; and

C) A trial whose objective was to correlate users sociability with academic performance.

The latter two trials were conducted in Escuela Politécnica Nacional, in Ecuador, from the 12^{th} of May to the 12^{th} of June 2018, with a total of 30 students. It is also worth mentioning that in Ecuador the first term goes from the 9th of April to the 8th of August. Thus, our study was in the middle of the first term, and we were able to gather data during the midterm exams. This is important since to obtain a positive classification the students need to have at least 12 points in the sum of the midterm and the final exams (both exams are graded from 0 to 10 points). As such, we can consider that any grade above 6 is a positive grade, while any grade below 6 denotes poor performance from the student.

6.1.1 OSN Results

As mentioned above, this trial involved 10 university students, whose main language was Portuguese. The trial allowed to validate the ISABELA implementation and assess the effectiveness of Sentiment Analysis classification involving both Portuguese and English languages. On the whole, the obtained classification results were considered correct by the users.



(a)

(b)

Figure 6.1: Sentiment analysis visualization in the ISABELA app.

Figure 6 presents typical data visualizations generated by ISABELA, resulting from the users' Facebook and Tweeter activity. Polarity emojis and emotions scatter plot in Figure 6.1a are entirely processed in ISABELA's Android code, while the weighted lists for the word clouds in Figure 6.1b are processed in the NLP module (that treats both English and Portuguese texts). This hybrid approach (i.e., part of the treatment being done in ISABELA's Android code, while other parts are processed by the NLP module) was adopted to enable immediate display of the Android activity. However, it is possible to have everything processed on either side. Sometimes, users may feel more confident of having all the data processed locally, even if it may consume more energy.

The logic behind the average mood is a computation of mean polarity values over the previous 5 days. Then, each daily polarity value is mapped to one of the three emojis, depending on chosen intervals, which are displayed on the screen. The crossed box image is displayed when no objects were found on a corresponding day. Concerning the Scatter Plot in the Emotional Chart, we retrieve the scores for valence (horizontal axis) and arousal (vertical axis), then normalize them to -4, $+4$, and finally plot the corresponding spots in the Cartesian plane according to Russell's simplified circumplex model (Russell [1980]). Both LEED libraries and Linguateca dictionary are exploited to create the Top Emotions' dataset for Portuguese tokens and emoji text. Finally, Top Expressions is fed by an array, regrouping a weighted list of tokens and emojis in social media.

6.1.2 Location and Performance Correlations

In this section, we present some of the results extracted from Ecuador's dataset, for the students' performance based on their location. We divided the dataset into two groups, based on the student grades in the mid-term exam, group 1

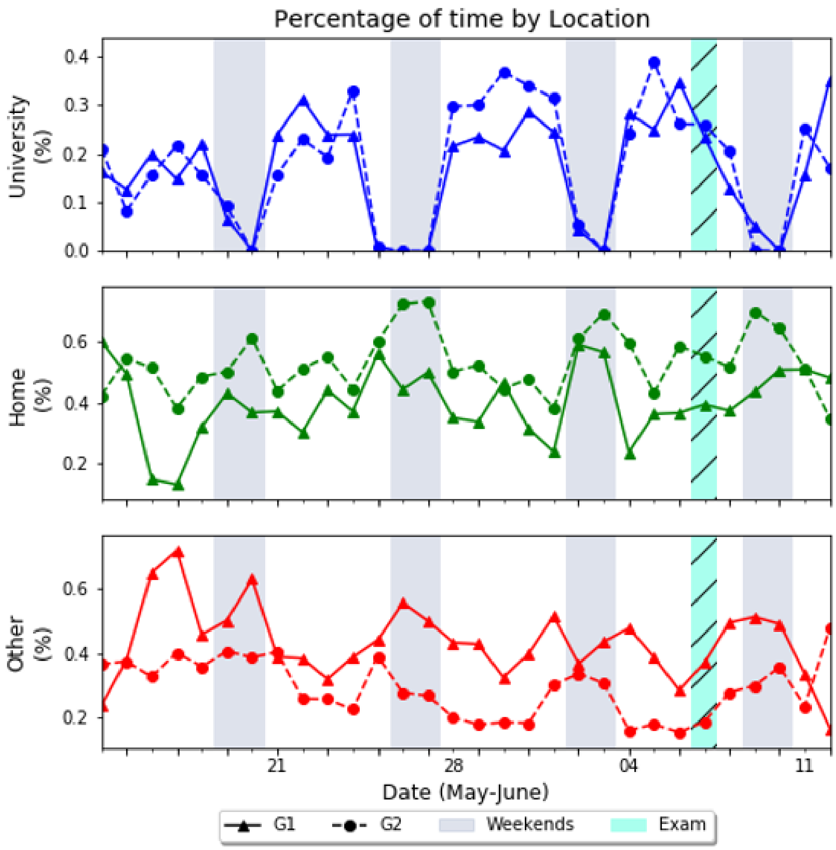


Figure 6.2: Percentage of time by location for students of both groups.

(G1) comprising the students that obtained a grade of 6 or less, and group 2 (G2) comprising the remaining students. G1 had 12 students, and G2 had 18 students. One of the objectives of the trial was to validate the ISABELA implementation in terms of passive data collection via smartphones, concerning location. Another objective was to correlate location and academic performance, with the aim of providing recommendations in future versions of the HITLCPS application.

We determined the students' location based on Wi-Fi. With this approach, we were able to distinguish between three general locations: *university*, *home*, or *other*. In Figure 6.2 we present the results for each of the two groups, over the trial period, concerning the amount of time spent at each location. As expected, the time at university follows a sinusoidal-like form, with high percentages of time during the week and almost null time during weekends. We can see that students from G2 spent more time at the university in the week before the exam, although overall the difference is quite small. It is important to notice that the 25^{th} of May is a national holiday in Ecuador, and, as such, the students do not have classes to attend. It is also worth noticing that the peak time at University for G2 occurs 2 days before the exam, and they then reduce their time at the University, most probably to spend more time studying at home.

On the other hand, when analyzing the time that students spend at other locations, we can see clear differences between the groups. For group G2, the time they do not spend at the University is mostly spent at home, while for students belonging to G1 a greater part of their time is spent in other locations. For both of these locations, we can also see a sinusoidal-like modulation in every week, but in the inverse form, that is, higher values during the weekends and lower values during working days. Students at G2 also decrease their time at other locations during the week, with the course of the semester, with the lowest values occurring in the week of the exam, while students belonging to G1 seem to maintain roughly the same values over the whole period, including the week of the exam.

6.1.3 Sociability and Performance Correlations

In this section, we present some of the results extracted from Ecuador's dataset, for the students' performance based on their sociability. G1 and G2 division is the same as in the previous case. The objectives were similar to the previous case, now concerning sociability-performance correlation.

Several metrics were collected with the aim of monitoring the students' sociability, namely, sent and received SMSs, phone calls, contacted list, and nearby Bluetooth devices.

For the call metrics, we collected data for the number of initiated calls, received calls, and missed calls. However, none of the above metrics showed any significant difference between the groups. We also found that both groups seem to have a lower number of calls at the beginning of each week and tend to increase this number closer to the end of the week. Additionally, both groups made and received less calls during the week of the exam.

We also gathered information not only on the number of devices that surrounded the user on daily basis, but also on the amount of time that the user spent with no devices in his/her proximity. We choose to label those times as *"alone time"*.

As can be seen in Figure 6.3(a), the students in both groups spent higher amounts of *"alone time"* at the weekends, and lower amounts during the week. However, we can see that, in the mid of the week before the exam, students from G1 changed their behaviour and spent less time alone, while students from G2 maintained their normal behaviour. This correlates well with our findings, in a self-reported data questionnaire, that pointed to the fact that students from G1 seem to prefer to study in groups, while students from G2 study alone.

As in the case of the call metrics, the SMS metrics did not reveal significant pattern distinction between the groups. The results for each group can be seen in Figure 6.3(b). Nevertheless, contrary to what happened with the calls, SMSs data revealed some differences between the groups during the exam week. Students from G2 started the week with a higher number of exchanged SMSs and then that number dropped to half the value in the days before the exam. Students in G1 show the opposite behaviour, as they started the week with less exchanged SMSs, and two days before the exam they increased the number of exchanged SMSs.

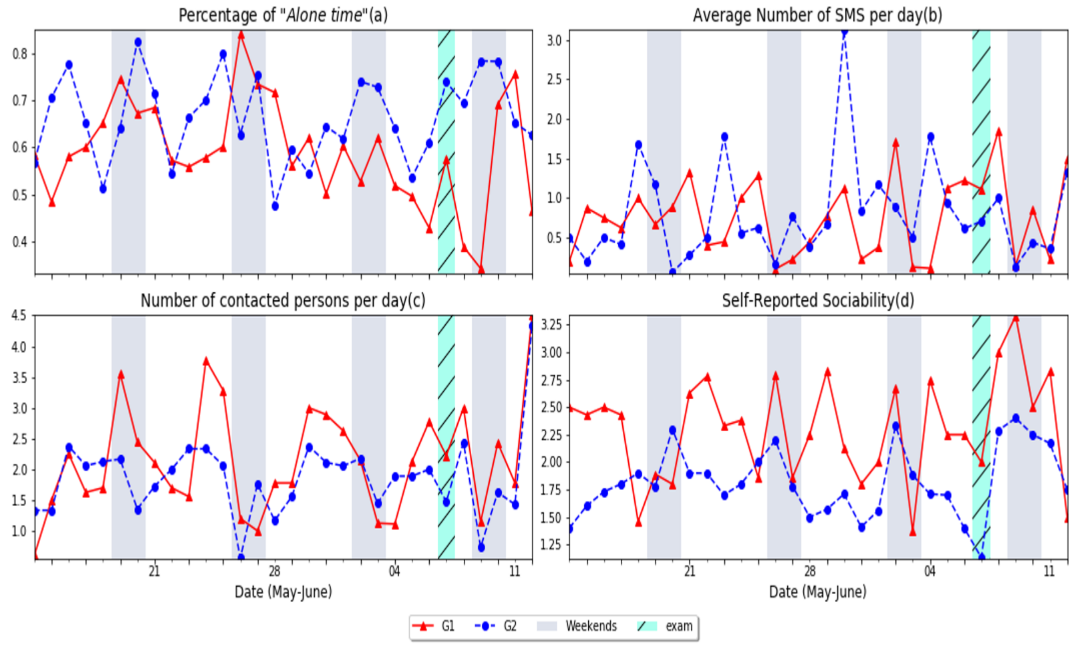


Figure 6.3: Percentage of time alone (a), Average number of SMSs (b), Average number of different destinations (c) and Self-Reported Sociability(d).

As a means to observe the students' sociability, we also kept track of the number of contacted persons per day. In order to do so, we collected each phone number from incoming and outgoing calls and SMSs. All this information was then hashed, to address all privacy and anonymization concerns. The values for the number of contacted people for both groups are shown in Figure 6.3(c). We aggregated the number for both SMSs and calls, and removed the repetitions of the same hash in each day. As we can see in the figure, students from G1 contacted more people that students in G2, including during the week of the exam. Furthermore, G2 students have less variance in their values in contrast to G1 students, which could indicate they have more stable friends/family circle.

Additionally, it was also requested that the users self-reported their sociability level each day. The users filled a form where they would rate how sociable they felt during that day, with the score ranging from 0 to 4. The results of computing the average value of this self-reported data for the day, for each group, can be seen in Figure 6.3(d).

Looking at the sociability levels in each group, some differences seem to emerge. For instance, students in G2 seem to have an increase in their sociability levels at the weekends and decrease their sociability levels during the week. Students in G1 have a more irregular pattern in their sociability levels and have sociability peaks during the week as well. It is also visible that the sociability level of G1 students is almost always higher than the one in G2. Additionally, when considering the week of the exam, we can see that G2 students decrease their sociability level, while students G1 maintain theirs. This could indicate that students in G2 prefer to study alone and/or focus on their objectives, while students in G1 study in groups and/or are more subject to distractions.

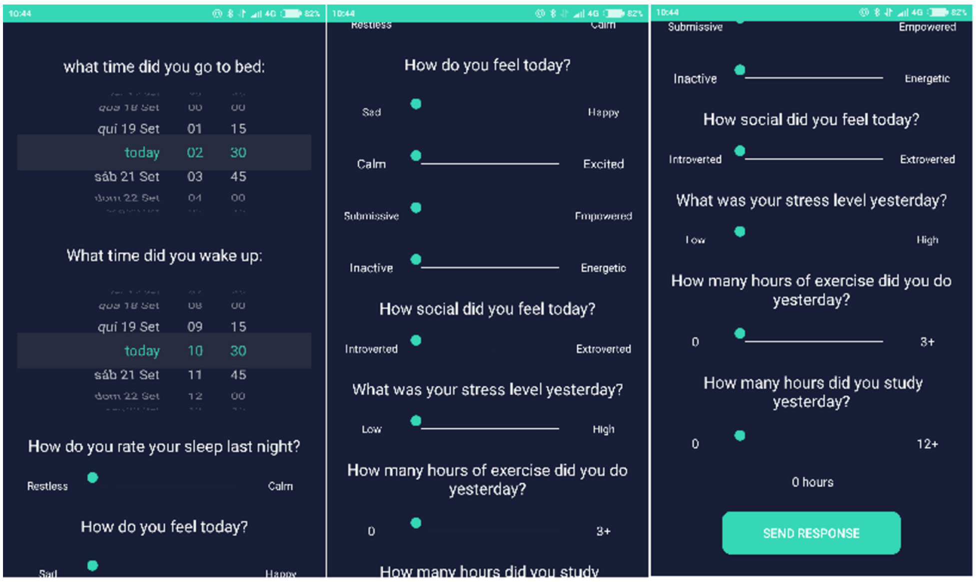


Figure 6.4: Daily form students are asked to fill for ground-truth.

6.1.4 Ground Truth analysis

Our main goal in the ISABELA project was being able to model the students' performance and infer their physical and emotional states. In order to do so, we decided to create a simple form that would allow us to create supervised models. This form was meant to be simple and quickly filled, as many studies already proved that simple questions and short questionnaires improve the response rate (Dillman et al. [1993]). Additionally, for metrics such as valence, arousal and dominance, one of the more reliable ways to extract them is trough self-rating (Reisenzein [2010]).

The form was automatically presented to the student once a day, in the form of a notification, that once clicked redirected them to the form page. The created form can be seen in Figure 6.4, and is composed by eight questions only.

The first and second questions are related to the students sleep patterns. By asking at *"what time did you go to bed?* and at *"what time did you wake up?"*, we were able to label the periods of time in which the students were sleeping. Both questions are answered by selecting a value for the day, hour and minutes. In the minutes section, we used a 15 minutes interval, as many users will not precisely remember the exact minute at which they went to bed or woke up.

The third question, *"How do you rate your sleep?"*, is meant to label the sleeping quality of each night. This question could be answered with one of four values between restless and calm.

The fourth question, *"How do you feel today?"*, allowed the user to rate their feelings. Namely, the degree of happiness (in a continuous scale ranging from sad to happy), which is mapped to valence, the degree of excitation (in a continuous scale ranging from calm to excited) and the degree of energy (in a continuous scale ranging from inactive to energetic), which are both combined and mapped to arousal. This two metrics are the most commonly used metrics to represent sentiments using the Russell's Circumplex (Russell [1980]). And the degree of submission (from submissive to Empowered), which is mapped to dominance. The dominance metric can also be used to improve this sentiment analysis by moving from a 2-Dimensional space to a 3-Dimensional space of sentiment representation (Verma and Tiwary [2017]). We aim to use this data to understand the correlation between the user actual sentiment and the sentiments extracted from our OSN Sentiment Analysis module.

The fifth question, *"How social do you feel today?"*, is used to label the data for the sociability model. Once again, the students have to answer by choosing a value from zero to four, which represent introverted and extroverted, respectively.

The sixth question, *"What was your stress level yesterday?"*, is answered in the same manner as the previous ones. These values allow us to evaluate the stress levels throughout the academic year, and measure which factures contribute to higher stress levels.

As for the last two questions, they are related to students' daily routines. We aim to capture the amount of time that students spend doing physical exercise (i.e., going to the gym, running, playing football) and studying. The physical activity ground truth will be used to evaluate the activity levels measured from the smartphone, while the study ground-truth will be used, in a first instance, to perceive how the number of studying hours influences the academic outcome of students. In future studies we also aim to improve the system in order to automatically detect when the student is studying.

6.2 Sleep and Sociability Classification

We have seen in the last section, that the self-reported sleep patterns and the sociability levels can be correlated with the students' performance. As such, these are metrics that the ISABELA system should monitor, in order to perceive changes in students behaviours and/or performance. In this section we present results for three implemented models, namely the Sleep Detection model, the Sleep Quality model and the Sociability model.

The datasets use to obtain the results presented below were obtained in a study performed in the Escuela Politecnica Nacional, Quito, Ecuador. The study was performed during the months of May to June, 2018.

6.2.1 Sleep Detection

Several works have focused on the monitoring of sleep through smartphone data as a mean to improve well-being and promote better lifestyles (Chen et al. [2013] and Min et al. [2014]) . Concurrently, several research studies show the influence of sleep patterns on the students' academic outcome (Trockel et al. [2000] and Curcio et al. [2006]). As such, modelling the students' sleep patterns is also an important part of our work.

Table 6.1: Sample distribution for the classes of the three, before and after class balancing.

		Original size	***After Balacing***	***Total***
Sleep Detection	*"Not Sleeping"*	245334	161722	323639
	"Sleeping"	78305	161917	
Sleep Quality	*"Restless Sleeping"*	6153	7368	14785
	"Calm Sleeping"	8632	7417	
Sociability	*"Low Sociability"*	197	165	347
	"High Sociability"	150	182	

Based on previous work from Chen et al. [2013] we selected several features from the smartphone namely, light intensity, sound amplitude, phone-lock, and activity. Additionally, we selected other features, namely, the remaining time to the next alarm, the proximity sensor and the day of the week. Nowadays, most people use the smartphone as an alarm clock, and, as such, we believed that the time to the next set alarm is significant towards detecting sleeping periods. We also believe that due to the rigorous scheduling of students' lives, sleep exhibits different patterns for different days of the week. Furthermore, the proximity feature can help to discern between other occasions where the light and sound are similar to sleep environments, such as carrying the phone in ones' pocket or bag.

As stated before, the users need to fill a form on a daily basis, in which they describe at which time they went to bed and woke up. This data is then used to label the dataset between sleeping periods and non-sleeping periods, classifying the sleeping periods as 1 and the non-sleeping periods as 0.

Before normalizing the data, some issues were addressed concerning the sensor values, namely the values of the proximity sensor and the sound amplitude sensor. The proximity sensor on smartphones returns a value in the range between 0 and a max value in cm (e.g., 8 cm). However, all recorded values are either 0 or the max value for the specific sensor. As such, for all proximity sensors we considered a binary value of either 0 or 1 if the value was different from 0. The sound amplitude measured from the smartphone's microphone is between 0 and 32768, i.e., the positive range of an integer variable. Before normalization, we converted the value into decibel (dB)s using the formula provided by the Android documentation.

Table 6.2: Results for accuracy, precision, recall, and f-measure for the several tested models.

	Accuracy	***Precision***	***Recall***	***F-Measure***
Naive Bayes	0.704	0.689	0.750	0.718
Bayes Net	0.795	0.795	0.799	0.797
Logistic	0.755	0.754	0.760	0.757
ZeroR	0.596	0.575	0.758	0.654
J48	0.766	0.764	0.774	0.769
Random Forest	0.906	0.906	0.907	0.906
Random Tree	0.904	0.904	0.905	0.904

$$sound_{dB} = 20 \ \times \ (\ log_{10}(32768) + log_{10}(\frac{value}{32768}) \ ) \ dB$$

The values were then normalized for a maximum value of 90 dBs, which corresponds to the maximum amplitude in dBs. The rest of the features were normalized between 0 and 1. The dataset was then balanced, by using the resample filter of the Weka framework (Witten et al. [2005]), with a bias of 1 towards the minority class. The number of samples in each class can be seen in Table 6.1. As we are dealing with a small set of features, no feature selection was performed. Weka version 3.8 was used for this and other tasks.

After pre-processing the data, we performed a data split of 80% for training and 20% for test. The training data was then used to train the model with a 10-fold Cross Validation. The model was then tested with the remaining dataset. The first tests were run using the Weka Software, in which several algorithms were tested. The results for accuracy, precision, recall, and area under the Receiver Operating Characteristic (ROC) curve are presented in Table 6.2.

As can be seen in Table 6.2, the Random Forest Model worked especially well when compared with the rest. As such, we chose to run tests for different architectures of the model. One of our goals was being able to perform classification within the smartphone itself. As such, we decided to implement the model in JAVA, resorting to the Encog library (Heaton [2015]). The results for those trials are presented in Table 6.3. As we can see in the table, the models increase in accuracy, precision and recall with the increase in the number of decisions trees. However, the increase is not very significant, and the model seems to stagnate above 50 decision trees.

Additionally, as we stated before, other works like the one in Chen et al. [2013] obtained good results while performing sleep classification. In that particular work, the authors implemented an Artificial Neural Network (ANN) model. As such, we decided to test the same approach. Resorting to the Encog library once more, we implemented a Feed Forward ANN with two hidden layers, both with a Log-Sigmoid activation function. In Table 6.4, we can see the results for different architectures. As we can see from the table, once more, increasing the complexity of the model translates into better performance.

Table 6.3: Results for the different Random Forest architectures.

Nº of Trees	***Accuracy***	***Precision***	***Recall***	***F-Measure***
1	0.932	0.933	0.932	0.932
2	0.940	0.940	0.940	0.939
5	0.942	0.943	0.941	0.941
10	0.944	0.945	0.944	0.944
20	0.960	0.945	0.977	0.961
50	0.966	0.948	0.987	0.967

Table 6.4: Results of the different structures of ANN.

ANN Structure	***Accuracy***	***Precision***	***Recall***	***F-Measure***
1-1	0.498	0.501	0.428	0.462
2-2	0.571	0.567	0.620	0.592
3-3	0.769	0.865	0.641	0.736
5-5	0.844	0.863	0.821	0.841
10-10	0.934	0.911	0.963	0.936
15-15	0.944	0.921	0.973	0.946
20-20	0.958	0.941	0.977	0.959

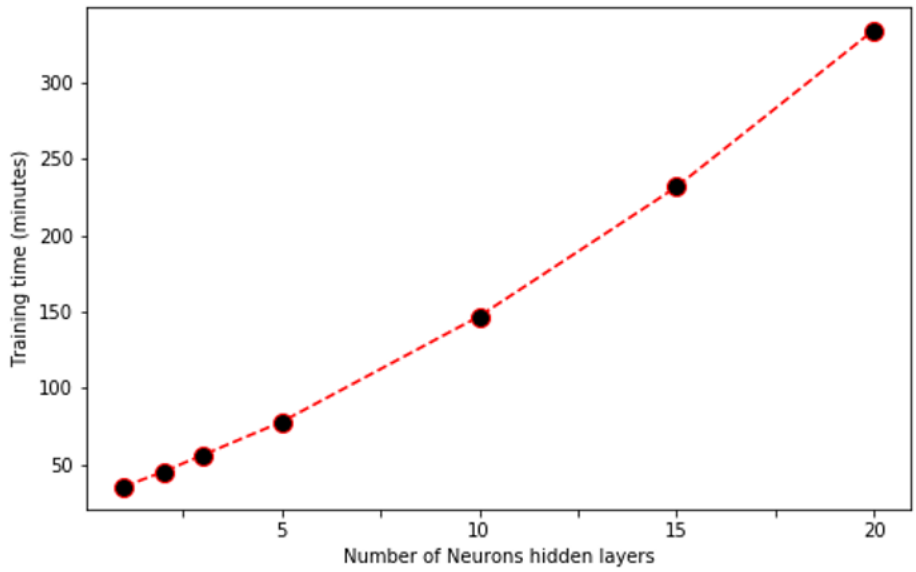


Figure 6.5: ANN average training time as function of number of neurons in each hidden layer.

Our aim is to be able to run and train the model in real-time on the smartphones themselves. This, however, requires Central Processing Unit (CPU) time, which will translate into higher energy consumption. In order to reduce this, we opted for training the models at night, while the smartphone is charging and the users are sleeping. Tests were performed in three Samsung Galaxy sm-J330, that do not support hardware acceleration. As we can see in Figure 6.5, the training time of our ANN in a smartphone largely increases with the number of neurons. Furthermore, the training time of the ANN with 20 neurons its over 5 hours. As such, we decided not to go over 20 neurons in each layer, to reduce the needed time for training.

If we compare both models, we can see that the Random Trees models outperforms the ANN model, showing better accuracy, precision, recall, and F-measure values. Furthermore, the Random Forest model showed lower training times and CPU usage, making a better fit for our proposed goal.

6.2.2 Sleep Quality Classification

As stated before, the ISABELA users had to fill a form every day, answering the question *"how do you rate your sleep last night?"*, with a value from 0 to 4, where 0 corresponds to *"Restless"* and 4 to *"Calm"*. We then used that data to label the dataset for the sleep quality classification model. With the aim of making the model binary, we chose to normalize the values between 0 and 2 as 0 (restless sleeping), and the values within the range of 2.1 and 4 as 1 (calm sleeping).

For this classification we only used the data which corresponded to sleeping periods of time. The sample distribution can be seen from Table 6.1. As we are only considering the samples which correspond to sleeping periods, the dataset is shorter. Furthermore, not all students filled the surveys all days, which led to an even shorter dataset. Additionally, as we are only dealing with the periods of sleeping, we choose to drop time-related features, namely the *"time to the next-alarm"* and the *"day of the week"* features. As happened with the previous model, the dataset was also balanced by resorting to Weka's Resample filter with a bias of 1. As happened for the Sleep Detection model, we also performed a train/test split of 80/20%, and the models were trained and validated using 10-fold cross-validation and tested in the remaining data.

Furthermore, as the Random Forest technique presented better results for the sleep detection model, we choose to implement the same approach for the sleep quality model. The test results for the different structures are visible in Table 6.5.

We can see that with an architecture of 100 random trees the results start to look promising. In future endeavours, we aim to automatically classify the students sleeping quality. The proper functioning of this model, however, will be directly influenced by the ability of the sleeping detection model to correctly detect the sleeping periods. As such, we need to further improve both models in our next trials.

6.2.3 Sociability level Classification

Some studies in the past addressed the topic of automatic detection of human sociability through the evaluation of phones' usage (Eskes et al. [2016b]). From

Table 6.5: Sleep Quality results for the different Random Forest architectures..

Nº of Trees	***Accuracy***	***Precision***	***Recall***	***F-Measure***
1	0.791	0.795	0.793	0.794
2	0.791	0.793	0.796	0.794
5	0.808	0.812	0.809	0.810
10	0.813	0.817	0.813	0.814
20	0.816	0.820	0.816	0.818
50	0.825	0.829	0.825	0.827
100	0.912	0.915	0.910	0.912

the analysis of the literature, we chose to explore the communication component of sociability by profiling the users. The features chosen for this task were: the number of received calls, number of outgoing calls, duration of calls, number of received SMS, number of sent SMS, and number of different destinations for both the calls and SMS metrics.

We already explored this metrics in section 6.1.3, showing how there are different patterns for the two considered groups and how these behavioural patterns can correlate with the students' outcome. However, we can also see from that data that the number of calls and SMS that the students exchange are not very high, with the average number of calls per day of 2.78 calls (σ=1.38), and an average number of SMS per day of 0.78 SMSs (σ=0.56).

The average adult/teenager in America exchanges 94 SMS texts per day. A study points to the fact that the average American adult makes or receives 5 calls per day (Center [2010]). When comparing these results with the data that was obtained, we can see that the number of calls made by the members of our trial was roughly half the expected value, and the number of SMS was lower by a factor of more than 120. Furthermore, the decreased use of SMS to the detriment of other communication methods is also supported by the data presented in Burke [2019].

The students rated their sociability on a daily basis from 0 to 4, using the daily form. We performed the same class division for the case of Sleep Quality, that is, the students' reports were divided into two classes, namely *"low sociability"* and *"high sociability"*. Scores from 0 to 2 fall under the *"low sociability"* class, while the reports from 2.1 to 4 fall under the *"high sociability"* class. The sample distribution can be seen in Table 6.1. Once again, the Weka resample technique, with a bias of 1, was used to balance the classes. The number of samples for this model is more limited, since it corresponds to a daily average of the metrics per student. Additionally, once again, we could not use data for which students did not report sociability values. All the metrics were normalized between 0 and 1. As in the two previous models, the same model selection process was used, with a train/test split of 80/20%. A 10-fold cross-validation was performed over the training set, and the model was then tested with the remaining data.

The results for the Random Forest model created for classifying the sociability

Table 6.6: Results for the sociability Random Forest model.

Nº of Trees	***Accuracy***	***Precision***	***Recall***	***F-Measure***
1	0.681	0.659	0.652	0.653
2	0.692	0.706	0.671	0.671
5	0.694	0.672	0.661	0.662
10	0.695	0.677	0.662	0.663
20	0.692	0.673	0.660	0.661
50	0.723	0.708	0.680	0.694
100	0.728	0.715	0.683	0.698

can be seen in Table 6.6. As we can see from the table, the results are less promising than the ones for the other two presented models. We believe that the main cause of this is the fact that nowadays people tend to use other means of communication, such as instant messaging applications and video call apps, instead of the regular cell phone resources.

This shows that the use of the communication capabilities of smartphones is no longer representative of the users' communication habits. In our future trials we will also collect usage statistics of the smartphone applications.

Furthermore, we believe that all the presented models would benefit from a personalized model for each user. However, we were not able to test this theory due to the lack of sufficient data. We also leave this to be done in future work.

6.3 Automatically Assessing Students' Performance

In section 5.1 we present the ISABELA system, which is capable of collecting both passive and self-reported data from students. That system was used to monitor students during two real–world trials. One in 2018 and another one in 2021. We covered part of the exploration of the data from the first trial (2018) in sections 6.1 and 6.2. However, in that part of our work we do not address the problem of predicting student performance. Furthermore, the previous mention results do not cover any of the data from the trial performed in 2021.

In the past three years, we have been affected by the Covid-19 pandemic. Apart from the obvious human and economic cost, studies have also pointed out the psycho-social effects of the Covid-19 pandemic (Haleem et al. [2020]). The measures adopted in many countries, to limit the spreading of the virus, have led to significant changes in the daily routines of people. At some point in time, almost every affected country had to switch from in-person to online classes, which had substantial impact on the daily lives of students. Some studies analysed the effect of these changes (Afonso [2020]). Evaluating how these changes affect student performance is also an important challenge.

As stated the dataset used for the tests presented in this section is based on two distinct one-month trials executed on different years, namely, a first trial in May-June 2018, and a second trial in March-April 2021. Both trials were executed at the Escuela Politecnica Nacional of Ecuador. During the 2021 trial, the Covid-19 pandemic was ongoing in Ecuador, and the classes were lectured remotely. As such, this dataset also offers an insight on how Covid-19-related changes to the students' daily routines have affect their performance. To the best of our knowledge, we believe that the presented dataset and results are the first to include both data from *non-Covid-19* and *Covid-19* periods. Furthermore, the models presented in this section can predict the students' performance, even periods as different from one another as these.

When considering student's performance, determining which students will do worse or better within a class can be useful information. This can help to reduce dropout rates and take measures to increase the overall performance of the class.

In this section will cover the features as well as some additional information of the datasets used. We explore how the data acquired from both trials can be used to create a model for this purpose. The results for several models created and tests performed are also covered in this section. Additionally, in this section we present a final classification vote mechanism that improves the general performance of our model and can make the model increase its performance over time.

All the tests presented below were performed with Leave–One–Out–Cross–Validation (LOOCV). This particular type of cross-validation uses a number of folds that is equal to the number of samples of the dataset, that is, all but one sample are used as training set and the remaining instance is used as a single-item test set for each run. This allows us to maximize the training set of each cross-validation fold, which also allows the creation of a more generic model.

Additionally, all the code, for the results presented below, was produced using python and the scikit-learn library. The used python version was 3.7.6, and the scikit-learn version was 1.0.2.

6.3.1 Dataset's Features and Information

In this subsection will cover the features as well as some additional information of the datasets used in the models presented below. Such as time of acquisition, demographics, and class labelling.

As we covered in section 5.1, one of the main components of students' lives that the ISABELA system aims to monitor, is their activity levels. The system gathers this data by resorting to Google's Physical Activity Recognition API (Google [2022a]). This API is able to classify user's activity into one of eight possible activities, namely: running, walking, on bicycle, in vehicle, on foot, tilting, still, and unknown. Additionally, since we are not interested in any particular type of exercise, but rather in activity/exercise levels, all the categories that represent exercise were grouped into one category (*"exercise"*). Thus, running, walking, on foot and on bicycle were grouped under the *"exercise"* classification. Moreover, although *tilting* does not represent any specific user activity, and is just an indicator of smartphone tilting, we believe that this classification can also be used to infer the smartphone's usage.

The activity levels of students are one of the metrics that might have changed more due to the pandemic context. Due to sudden change in habits caused by confinement, most people saw their daily routines affected in terms of physical activity. As such, this feature can be important to study and characterise those periods of time.

Another aspect that could be significant to evaluate students' performance is their location. Using the students' location information, it is possible to infer if a student attended classes, spent time at home, or even if it frequently went out at night. As explored before, due to privacy concerns, our system did not store any GPS information. Instead, the system infers the discrete location of the user, namely by categorizing their location into one of three possible categories: *home*, *university*, or *other*. The ISABELA application scans available networks periodically and compares the results of the scans with the stored SSID to infer if a user is at *"home"* or not. All the universities in Ecuador make the *"Eduroam"* network available throughout the campuses. As such the application, searches for this particular network to determine if a user is at the *"university"*. Lastly, if none of the two networks exist in the Wi-Fi scan's result, the application categorizes the student's location as *"other"*.

Due to the lockdowns caused by the Covid-19 pandemic, location is also an important variable during this period of time. Additionally, students in Ecuador attended classes remotely during lockdown periods, eliminating the time spent at their university. Thus, to make both datasets compatible, the *"university"* and *"other"* discrete location categories were grouped together as *"other"* for both datasets.

Although the smartphone's context information and the sensors' data can be used to infer several daily life aspects, there are some aspects that are harder to detect trough passively collected data, especially those pertaining to psychological states, since passive data has yet to be validated as an indicator for these kinds of metrics. As such, there is a need to use more conventional methods, such as questionnaires. In the ISABELA system, we used questionnaires to complement the aforementioned passively collected data.

One of the aspects, that can be relevant when evaluating student's performance is their emotional state. A questionnaire-based approach, based on the SAM scale (Bynion and Feldner [2020]), was implemented in order to obtain this type of data. This questionnaire prompted to the user on a daily basis in order to obtain values of the self-perceived emotional states for each day. The SAM scale is an image based self-assessment scale to evaluate the users' two-dimensional map of emotions. The questionnaire can be seen in Figure 5.8a. From top to bottom, the first scale corresponds to arousal while the second scale corresponds to valence. The user selects, one image from line, corresponding to an integer value ranging from 0 to 4, which indicates their valence and arousal values.

These values can then be converted to a range of -1 to 1, with a step of 0.5, and be subsequently mapped to the 2-dimensional emotional map (Yazdani et al. [2013a]). The questionnaire was prompted to the users at a randomly selected hour between 14–20h. Additionally, the questionnaire was not released during the morning since during this period the user's perception can still be affected by the events of the previous day.

Another type of data that the ISABELA system is able to collect, and that is highly relevant to the students' academic performance is sleep data (Taras and Potts-Datema [2005]). This is done by prompting the user with a questionnaire every morning, which can be seen in Figure 5.8c. In this questionnaire, users can indicate the time at which they went to bed, the time at which they woke up, and how did they classify sleep quality. In addition, the users can also indicate their sociability level for the previous day, how many hours of exercise they had, and how many hours did they study.

The ISABELA system is able to collect other metrics that were already covered in section 5.1. However, we believe that the previously presented data is the most relevant concerning students' performance. Additionally, some functionalities were added for the 2021 trial, while others were removed and, thus, not all collected metrics were presented in both datasets. A list of the selected features for this work are shown in the Table 6.7. As it is possible to see in the table, the 14 features selected are a mix of passively collected data from the smartphone's sensors/context and data that was self-reported by the users. We believe that the use of these two types of data makes the model more robust.

The datasets presented in this section, they were obtained in two distinct periods of time, and in two real–world trials with the ISABELA system. Furthermore, those trials were performed in different years and with different school subjects. As mentioned before the trials, were performed in Ecuador's Escuela Politecnica Nacional, in Quito. The first trial involved 28 students, from the 12^{th} of May to the 12^{th} of June 2018. The second trial involved 33 students, from the 1^{st} of March to the 31^{st} of March 2021.

All the collected data was anonymized and, as such, the dataset does not include any personal information (e.g., age, gender, working status). For this reason, is not possible to determine the age, gender or other information for a specific user. However, from the school subject information, we knew that both trials had students with different ages, genders, and working status. On the other hand, both sets of subjects were enrolled in the computer science degree, which normally includes a larger percentage of male students. Additionally, all students had Ecuadorian nationality. Nonetheless, we believe that the dataset is representative of a larger population.

As previously mentioned, the 2021 dataset was obtained during the lockdown caused by the Covid-19 pandemic. During the lockdowns students were confined to their homes, which could lead to lower exercise levels, did not have as much physical and/or social interactions, did not have the option to do leisure activities outside, and had to attend classes remotely. These changes had the potential to cause negative effects not only on the students' physical well-being, but also

Table 6.7: Dataset's Features

	Feature
Sensors and Context	% of time at *other*
	% of time at *house*
	% of time *still*
	% of time *exercise*
	% of time *in vehicle*
	% of time *unknown*
	% of time *tilting*
Surveys	self-reported arousal levels
	self-reported valence levels
	self-reported sociability levels
	self-reported sleep quality levels
	self-reported sleep hours
	self-reported physical activity hours
	self-reported study hours

on their psychological and social well-being (Haleem et al. [2020]). Additionally, these changes had the potential to affect the performance of students. Because the trial periods can be so different in nature (as is the case with the 2018 and 2021 trials), it is important that any model for automatically predicting students' performance is able to deal with data samples from heterogeneous periods.

The students' final grades in the school subjects, for each trial, were also considered, to be used as classification label. Only the professor of the subject was aware of student's identities. Thus, we only obtained the final grade of the subject, for each corresponding anonymous id of the ISABELA application, hence students' anonymity was maintained. This metric was selected, instead of their average final classification, because not all students were enlisted at the same school subjects. The school subjects that one takes could influence their average

Table 6.8: Final grades metrics in the 2018 and 2021 datasets.

	Min	**Max**	**Median**	**Approved**	**Non-Approved**
2018	9.6	14.4	12.6	20	8
2021	12.14	17.72	14.90	33	0

Table 6.9: Number of students and sample size by dataset and time windows.

<table>
<tr><th rowspan="2"></th><th rowspan="2">Number of Students</th><th colspan="3">Monthly data aggregation</th><th colspan="3">Weekly data aggregation</th></tr>
<tr><th>Below Median</th><th>Above Median</th><th>Total</th><th>Below Median</th><th>Above Median</th><th>Total</th></tr>
<tr><td>2018</td><td>28</td><td>15</td><td>13</td><td>28</td><td>65</td><td>53</td><td>118</td></tr>
<tr><td>2021</td><td>33</td><td>15</td><td>18</td><td>33</td><td>38</td><td>44</td><td>82</td></tr>
<tr><td>Joined</td><td>61</td><td>30</td><td>31</td><td>61</td><td>103</td><td>97</td><td>200</td></tr>
</table>

final classification, and as such this metric would not allow for a direct comparison between students. In Table 6.8, we can see the minimum, maximum and median grade of each dataset, as well as the number of students that obtained a passing grade and the number of students that did not. In these school subjects, a student had to have a minimum grade of 12 points, to be approved.

We can see from the table that the distribution of the students' grades is quite different for both trials. Furthermore, in the 2021 dataset not only all students were approved, but also the median grade is higher than the maximum grade obtained in the 2018 trial. Due to these aspects, the labelling of each dataset was made separately, even when the datasets were jointly considered. The students were divided into two groups to perform a binary classification, namely the ones that obtained a grade *"above the median"* value (True class) and the ones that obtained a grade *"below the median"* value (False class). Even among editions of the same school subject for different years, we can have different levels of difficulty due to the changes in the lecturing or evaluating process. Using the median value allows us to create a model that is able to predict the best and lowest performance students among a class, even considering disparities between school subjects or editions.

We can see the final distribution of the dataset's samples in the two classes in Table 6.9. Additionally, in Table 6.9, we present the distribution of samples considering two data aggregation periods, namely monthly data aggregation and weekly data aggregation. We will explore the implications of using different data aggregation time windows, in the next section.

Table 6.10: Performance of distinct model by time interval, for the 2018 dataset.

	Monthly data aggregation			**Weekly data aggregation**		
	sensitivity	**specificity**	**accuracy**	**sensitivity**	**specificity**	**accuracy**
Decision Tree	0.492	0.567	0.532	0.606	0.705	0.660
Random Forest	0.477	0.607	0.546	0.715	0.794	0.758
SVM	0.385	0.533	0.464	0.736	0.846	0.797
Naive Bayes	0.692	0.533	0.607	0.811	0.646	0.720
K-Near-Neighbours	0.846	0.267	0.536	0.585	0.877	0.746
AdaBoost-SAMME w/ Random Forest	0.569	0.6	0.586	0.751	0.777	0.765
XGBoost	0.769	0.533	0.643	0.755	0.815	0.788

6.3.2 Performance assessing in different periods

As previously stated, the total available dataset is made up of two distinct trials, in two distinct years, namely 2018 and 2021. In both trials students participated for one month each. Furthermore, due to the Covid-19 pandemic the datasets are constituted by data from a period when students had in person classes (i.e., 2018 trial) and data from a period when classes were fully remote (i.e., 2021 trial). As such, the datasets from both trials are intrinsically very different. This is particularly true since we are using both the activity levels and the location of students as features for our models. For these reasons, we decided to start by evaluating the performance of each of the considered models when applied to each individual dataset.

Furthermore, we compare the results obtained by each model considering the data aggregated by student and by different intervals of time, namely the data aggregated by month and by week. Each feature is firstly aggregated by the mean value per day and per user, and then the mean value of each feature is calculated for each month and week respectively. The results for the various models, and for the different aggregation intervals are showed in Tables 6.10 and 6.11, for the 2018 and 2021 datasets, respectively.

Seven different models were tested, namely: Decision Tree (Quinlan [1996]); Random Forest (Breiman [2001]); Support Vector Machine (SVM) (Mammone et al. [2009]); Naive Bayes (Rish et al. [2001]); K-Near-Neighbours (Cover and Hart [1967]); AdaBoost-SAMME (Hastie et al. [2009]) with Random Forest (hereafter referred to as AdaBoost); and XGBoost (Chen and Guestrin [2016]). All the models shown in Table 6.10 and 6.11 were obtained by running the default implementation of scikit-learn, and all models were trained by using LOOCV. Using this method removes any variance that normally is created by the random distribution of the samples by folds. However, each model has a random start state, which can lead to different results. For that reason, every model was run

Table 6.11: Performance of distinct model by time interval, for the 2021 dataset

	Monthly data aggregation			Weekly data aggregation		
	sensitivity	specificity	accuracy	sensitivity	specificity	accuracy
Decision Tree	0.483	0.433	0.461	0.777	0.537	0.666
Random Forest	0.528	0.407	0.473	0.791	0.718	0.757
SVM	0.611	0.600	0.606	0.864	0.711	0.793
Naive Bayes	0.722	0.333	0.545	0.841	0.605	0.732
K-Near-Neighbours	0.444	0.733	0.576	0.864	0.605	0.744
AdaBoost-SAMME w/ Random Forest	0.678	0.533	0.612	0.814	0.745	0.782
XGBoost	0.611	0.467	0.545	0.750	0.684	0.720

10 times and the mean values for sensitivity, specificity and accuracy were calculated. Furthermore, a grid-search was performed for each model and on each dataset, to find the best parameters for each case. The grid search space as well as the best configuration of each model can be seen in the appendix.

As we can see in the tables, in both datasets, the models that use data aggregated by week surpass the models that use the data aggregated by month. This is partly due to the smaller number of samples available to the models, when data is aggregated by month, as we show in Table 6.9. However, as we will see in the next section, even when the sample size increases, the weekly aggregation models still surpass the ones that use monthly aggregation. We believe that the reason behind this is the levelling out of changes in the data when aggregating the data over long periods. For example, if a student skips school one day every other week, by month that would represent 6% of the time, while by week it would represent more than the double (15%) of the time. We believe that a temporal window of a week to aggregate the data, is the best approach since it can better grasp sudden variations in the student's behaviour. As such, for the remaining of the tests, presented in this section, we will focus on this aggregation window.

As mentioned before, both datasets are only slightly imbalanced and, as such we can use accuracy as metric to evaluate the models' performance. We can see that the models which present the best performance in both tables are the SVM and the AdaBoost models. The AdaBoost model offers a performance improvement over the Random Forest model, with only a slight improvement in the 2018 dataset, but a bigger improvement in the 2021 dataset. Furthermore, in both cases, the SVM and AdaBoost models obtained an accuracy higher than 76%.

However, we can also see from the tables that the best two models have distinct behaviours, the AdaBoost model has a more balanced distribution of sensitivity

Table 6.12: Cost-Sensitive Performance of SVM and AdaBoost for both datasets, in the week time window.

		Sensitivity	***Specificity***	***Accuracy***
2018	**AdaBoost**	0.792	0.800	0.797
	SVM	0.774	0.800	0.788
2021	**AdaBoost**	0.795	0.763	0.780
	SVM	0.841	0.711	0.780

and specificity in both datasets, while, the SVM models tends to favour the class with majority of samples in both datasets. For that reason, we also used Cost-Sensitive learning to evaluate if the class imbalance was affecting the model's final performance. For this particular test, we only considered the weekly aggregation interval, since it shows better performance than the monthly aggregation approach. Grid search was used to find the best cost weight for each class in both datasets. Both SVM and AdaBoost were tested with cost-sensitive learning and with cost weight of 1.125 for the minority class of each dataset, namely *Above Median* in the 2018 dataset and *Below Median* in the 2021 dataset. As we can see in the Table 6.12, for the 2018 dataset the SVM model lost some of its accuracy, while the AdaBoost model increased its. The AdaBoost model increased its sensitivity by 4.1% and its specificity by 2.3%, leading to an increase of 3.2% in accuracy. On the other hand, The SVM model increased its sensitivity by 3.8%, but had a decrease in its specificity by 4.6%, which led to a less accurate model. However, this can still be a valid strategy, if the main purpose of the desired model is the correct detection of students that are *above the median* value. In the case of the 2021 dataset, both models lost some of their accuracy. However, only the AdaBoost model had a trade-off between the sensitivity and specificity metrics, while the SVM model lost accuracy in the majority class but did not gain any performance in the minority class. This shows that, in the particular case of AdaBoost, cost-sensitive learning can be used to improve the model performance. Furthermore, applying cost-sensitive learning shows better results in the case of the 2018 dataset. This could be due to the difference in the ratio of the classes being bigger when compared to the 2021 dataset.

Lastly, we also wanted to find out the answer to the question ”can a model trained with one dataset be used to predict the performance of the other dataset?”. This question is particularly relevant due to the difference in students' daily routines caused by the Covid-19 paradigm. For this particular test, once again we only considered the weekly aggregation approach and the models with the best performance, namely AdaBoost, SVM, Cost-Sensitive-AdaBoost (C-S-AB) and Cost-Sensitive-SVM(C-S-SVM). We considered two particular cases, namely: Case *A*, use the 2018 dataset (non-covid dataset) as training set and the2021 dataset (covid dataset) as testing set; and Case *B*, use the 2021 dataset as the training set and the 2018 dataset as testing set.

Table 6.13: Cross testing with datasets.

		Sensitivity	***Specificity***	***Accuracy***
Case A	**AdaBoost**	0.684	0.750	0.720
	SVM	0.711	0.750	0.732
	C-S-AB	0.711	0.682	0.695
	C-S-SVM	0.763	0.841	0.805
Case B	**AdaBoost**	0.642	0.692	0.669
	SVM	0.736	0.692	0.712
	C-S-AB	0.679	0.738	0.712
	C-S-SVM	0.755	0.569	0.653

The results for both cases can be seen in Table 6.13. We can see that case A provides models with better performance. In particular, C-S-SVM led to an accuracy of over 80%, which is higher than the one obtained when using LOOCV to train and test with only the 2018 dataset. This could be due to the fact that the 2018 dataset has more samples and, as such, is able to generate a more generic model. Another reason that possibly contributes to this difference in accuracy, is the changes in lifestyle of students, caused by the Covid-19 pandemic. Several aspects of the students' daily routines were affected during this period, namely the confinement to their household, the decrease in exercise levels, changes in sleep patterns, and more. Because of that some variance between the trial subjects may have been lost and, as a result, the model trained with data from this period turned out to be less generic and representative of other periods. This may lead to its lower performance when evaluating samples prior to the 2021 dataset. However, in order to support these conclusions, tests with bigger datasets from both periods (outside and during Covid times) are needed.

6.3.3 A generalized model for performance assessment

In the previous section we saw, that training with one dataset and testing with the other one can lead to mixed results. The other aspect that we were interested in evaluating was if both datasets, although having been collected in two different periods, could be used to create a more generic model. In fact, by joining both datasets, we increase the number of samples and this may lead to better evaluating the proposed models. After joining the datasets, the final dataset became almost perfectly balanced. As such, in this section we do not test cost-sensitive learning. As stated before, the aim of this work is to create a model able to predict students' performance, considering they are under the same conditions. However, even considering the same school subject, different editions (i.e., same school subject in different years) might be affected by changes in curriculum, teaching method, or even evaluation procedure. As such, class labelling was performed prior to joining the datasets, by computing each dataset

Table 6.14: Performance of distinct models by time interval, for the joined dataset.

	Monthly data aggregation			**Weekly data aggregation**		
	sensitivity	**specificity**	**accuracy**	**sensitivity**	**specificity**	**accuracy**
Decision Tree	0.526	0.490	0.508	0.664	0.692	0.678
Random Forest	0.568	0.54	0.554	0.732	0.717	0.724
SVM	0.710	0.467	0.590	0.825	0.748	0.785
Naive Bayes	0.710	0.567	0.639	0.763	0.621	0.690
K-Near-Neighbours	0.677	0.300	0.492	0.773	0.612	0.690
AdaBoost-SAMME w/ Random Forest	0.619	0.587	0.603	0.826	0.760	0.792
XGBoost	0.645	0.467	0.557	0.794	0.777	0.785

grade median value separately and using them to label each dataset separately. As such, the class distribution is the sum of the students from each dataset in each class, presented in Table 6.9.

The results for the various models being tested, for the joined dataset, considering both time aggregation approaches can be seen in Table 6.14. As in the case of the previous tests, a grid search was performed in order to obtain the optimal parameters for each model. In Table 6.14 we can see that, similarly to what happened with the individual datasets, the models trained with the weekly aggregation approach outperform the ones trained with the monthly aggregation one. Once again, this might be because of the smaller sample size of the dataset with the monthly aggregation approach. However, as we stated previously, this seems to be only part of the reason, since even after doubling the sample size (by joining both datasets), most models see almost no performance increase, and some actually lose performance. This support our thesis that a month might be too big a time window, and some of the changes in students' behaviour might be ignored due to that. However, a bigger dataset is needed to validate this hypothesis.

Considering the results obtained with the weekly aggregation, we can see that similarly, to what happened when working with the individual datasets, the models with the best performance are SVM, AdaBoost and XGBoost. The AdaBoost model is the one that offers a higher accuracy, reaching almost 80% of accuracy. We can also see the confusion matrix for the AdaBoost model in Figure 6.6, where it is apparent that this model is able to predict correctly the class of 160 samples from the dataset. Additionally, the SVM model and the XGBoost model have the same accuracy of 78.5%. However, XGBoost offers a higher specificity, thus is able to detect the students that are *below the median* more accurately.

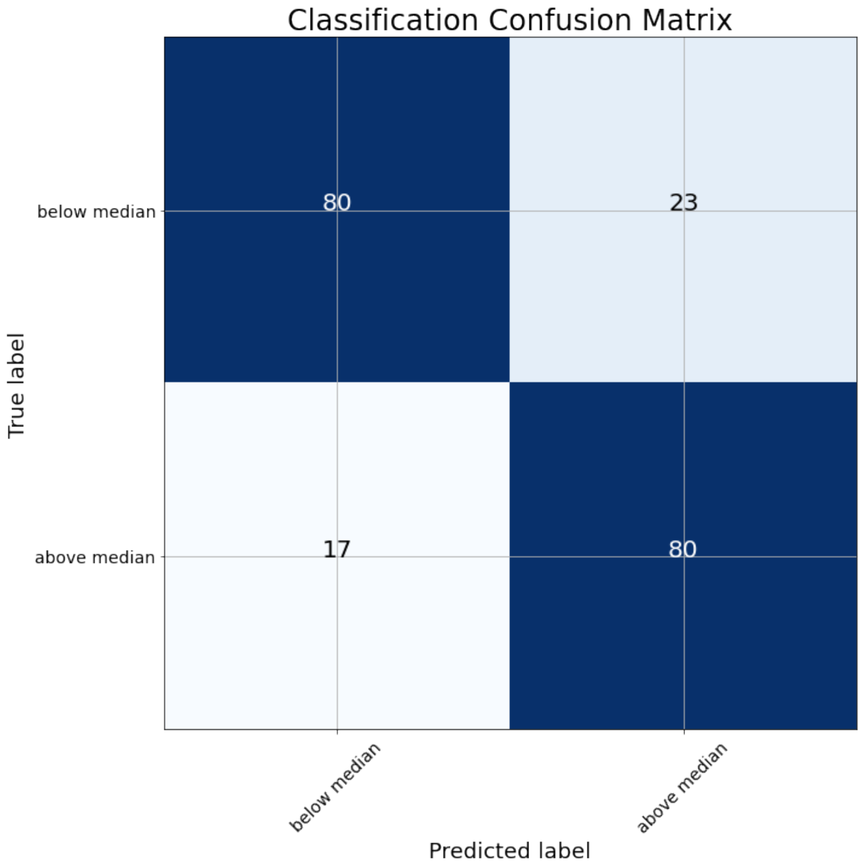


Figure 6.6: Confusion Matrix AdaBoost-SAMME with Random Forest model.

We can see the ROC curves of each model in Figure 6.7. Even though the AdaBoost model offers a higher accuracy, the XGBoost offers a AUC score of 0.856 score, against 0.846 of the AdaBoost model and 0.807 of the SVM model. This means that the XGBoost model is the model that is less prone to false positives. As such, the decision between one of these two models could depend on the particular use-case. That is, if we want to accurately determine both classes, XGBoost is the most suited model. However, if, for a particular case study, we wanted only to determine the students that are the best performers within a subject, AdaBoost would be the best model for that use case.

The performed tests show that the AdaBoost model, the XGBoost model or even the SVM model can be used to accurately predict which students will perform above the median and which students will perform under that threshold. Furthermore, the tests indicate that a model trained with the data from both datasets is able to obtain roughly the same performance as a model trained with each individual dataset. This also proves that is possible to create a model that is able to predict the performance of students with data collected in different periods, in this case before and during the Covid-19 pandemic.

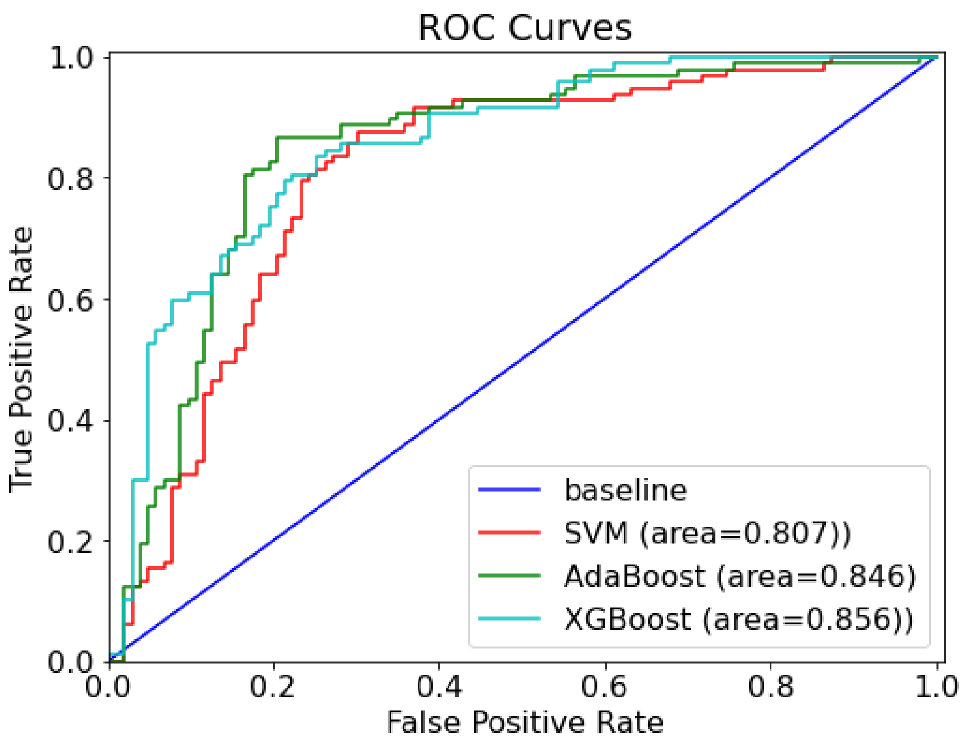


Figure 6.7: ROC curve graph for the best performing models, namely SVM, AdaBoost and XGBoost.

6.3.4 A pipeline for Student's performance classification

As seen above, using a weekly aggregation time window leads to better model performance when compared with a monthly aggregation time window. However, since both datasets were collected for one month each, when using the monthly window we obtain one classification per student. On the other hand, when using the weekly time window, each student has several classifications, that is one per each week of data. However, when considering the real-world context, each student only has one final classification. As such, in this section we propose a method to not only output a final classification per student, but, additionally, to improve the previous models' performance. Below we present the tests for this method using the weekly aggregation time window, the joined dataset, and the three best performing models from last section.

Models to merge classification, have been widely covered in past literatureJahrer et al. [2010]. However, this strategy is mostly used to combine the output of multiple models into one final classification at decision level. This strategy has several possible implementations, namely by computing the median value of the classifications, the majority vote, the product of probabilities, the average of probabilities, the minimum probability, or the maximum probability. Our approach differs from these strategies since it only uses one model to perform the classification of instances. Specifically, we propose an approach that combines all outputs of the model for the same student, to output a final classification per student.

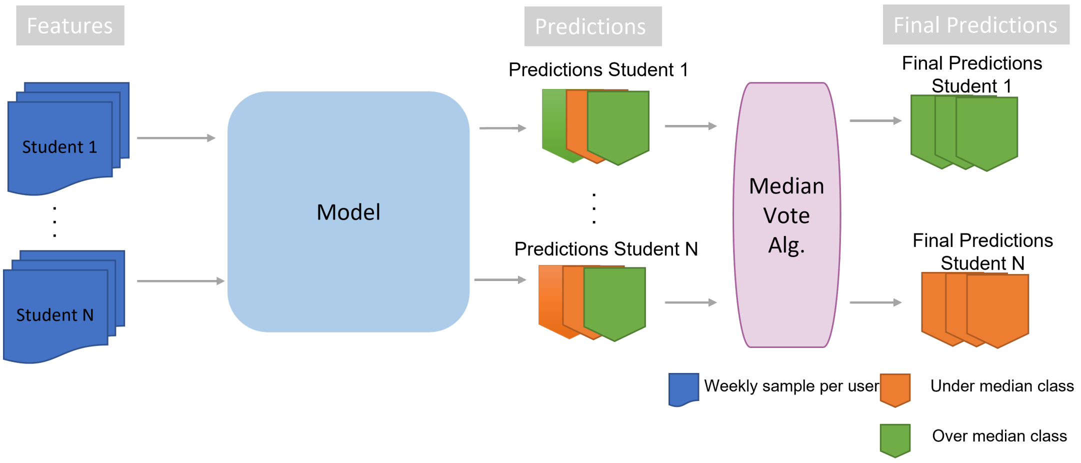


Figure 6.8: Diagram of classification using the proposed median voting algorithm to obtain the final classification.

The pipeline of the proposed mechanism can be seen in Figure 6.8. In the first part of the pipeline, each sample is processed as an independent entry for the model. After obtaining a prediction for each sample, the samples are grouped by student and processed by the median voting algorithm. This algorithm then outputs a final classification, that is the same classification for every sample of the same student. The only requirement for this mechanism to work, is that each sample's student identification must be carried on to the final classification step. Additionally, this mechanism will work with any number of samples per student, equal to or greater than 1.

The median voting algorithm can be seen in Algorithm 6.1. The algorithm requires as inputs the list of predictions of the model and the student identification for each prediction in the same order and with the same size. The algorithm then iterates over the list of predictions and replaces the value of each prediction with the median for the subset of predictions of that prediction's student. When the student has the same number of classifications for each class (e.g., 2 predictions of *above median* and 2 predictions of *below median*), the *above median* class is taken, since we are employing the condition of greater or equal to 0.5. Furthermore, the final predictions should be allocated to a new array, so the median value does not change with every iteration. The algorithm then outputs the final predictions for each sample.

Although simple, this method can improve the previously discussed models' performance. This is due to the fact that even *good* students (that would be classified as *above median*) can have *bad* weeks. A similar thing can happen in the opposite case, that is, a student that performed *below median* can have an *above median* week. However, that *good* week might not be enough to account for the other weeks that negatively affected the performance.

We tested this mechanism for the three models that presented better performance in the previous section, namely SVM, AdaBoost and XGBoost. The results of these models with the voting algorithm can be seen in Table 6.15. We can see that this pipeline effectively increases the performance of all models. In

Algorithm 6.1: Median vote of classifications.

Require:
$P \leftarrow$ Array of Predictions
$S \leftarrow$ Array of Student identifiers for each prediction
$FinalPredictions$

Ensure: $Size(P) == Size(S)$

$P(s) \leftarrow$ subset of Predictions for student s

$FinalPredictions \leftarrow Copy(P)$

$i \leftarrow 0$
while $i < Size(P)$ **do**

$s \leftarrow S[i]$

$p \leftarrow Median(P(s))$
if $p \geq 0.5$ **then**

$FinalPredictions[i] \leftarrow 1$ **else**

$FinalPredictions[i] \leftarrow 0$

$i \leftarrow i + 1$

Table 6.15: Performance of Models after using the Median Voting Algorithm at decision level.

	Weekly data aggregation			By Student		
	sensitivity	**specificity**	**accuracy**	**sensitivity**	**specificity**	**accuracy**
SVM	0.866	0.738	0.800	0.903	0.667	0.787
AdaBoost-SAMME w/ Random Forest	0.938	0.874	0.905	0.903	0.833	0.869
XGBoost	0.897	0.864	0.880	0.935	0.800	0.869

the specific case of XGBoost and AdaBoost, the accuracy increase is in excess of 10%. We can also see that the AdaBoost model is the one that offers the best performance, with an accuracy of 90.5%, followed by XGBoost with 88% accuracy. We can also see that the SVM model was the only one for which one of the evaluated metrics decreased. Specifically, SVM's its specificity decreased by 1%. Additionally, we can see that all models favour the *above median* class (i.e., the sensitivity is higher than the specificity). This is due to the use of the condition of greater or equal to 0.5, in Algorithm 6.1, in the case there is a tie between the number of predictions for both classes for one student. However, depending on the case study (e.g., a higher False Positive Rate is acceptable), this condition can be changed in order to favour the false class.

Using this pipeline, we get the same classification for every instance of data of the same student. As such, this effectively gives us one classification per student. In Table 6.15, we also present the metrics of each model when computed by student. We can see that, once again, AdaBoost and XGBoost offer a better performance, being able to correctly classify 53 of the total 61 students. When comparing these results with the classification using a monthly time window, we can see that the use of this pipeline leads to a big increase in performance. As such, we believe that this approach can be used to create better models to automatically assess students' performance.

Additionally, since this pipeline uses several instances of data from each student, to reach a final classification per student it can in fact improve over time. For that, the model only needs to keep a history of instances of each student. That is, envisioning a model that retrieves data and outputs a classification per student every week, as time progresses the model will have a larger number of instances per student. That means that every week the classification given to the student can be more accurate.

6.4 Analyzing People's Behaviour in a Pandemic Context

As previously stated, a study was developed and conducted in partnership with the Portuguese National Health System, that included the use of online questionnaires and the use of the Vitoria application for smartphone and smartwatch. To evaluate the use of this application, a three and a half months trial was performed from the 1^{st} of February to the 13^{th} of May, in Portugal. Participants were from several locations in Portugal, although predominantly from the district area of Lisbon. The study included 19 participants from all age groups and genders. From those 19 participants, we were only able to obtain data from 14 subjects, as some of them forgot to start the application after booting the smartphone or voluntarily stopped the application. Furthermore, some users did not reply to all the daily questionnaires, presumably because they forgot, which led to fewer samples of data.

During the trial, several Covid-19 related events occurred in Portugal, of which the most relevant ones are listed in Table 2. As we can see in the table, there were four deconfinement phases that gradually removed the restrictions. The trial explored the data in two periods, in co–occurrence with the changes at the national level.

Due to the relatively low number of participants in the trial, we did not evaluate the data of each individual participant. Instead, we only explored the data in terms of the mean value reported by day and by user. Furthermore, as explore before in section 5.2.3, the reduced number of participants also invalidates the analysis of the feedback system.

The low number of users prevents general conclusions about a bigger population. As such, in this paper we aim to present preliminary results of our system, as a proof of concept, that should be further tested with larger samples. Namely, we will explore the usage of different applications and the possible correlation of some of the acquired data with the official numbers reported for the pandemic in Portugal.

6.4.1 Application usage during the pandemic

The types of used applications and the respective usage duration can be used to derive important information about user behaviour. For instance, the preference for applications that are mostly used for leisure as opposed to those that are related to work, could indicate that a user mostly uses his/her phone to relax. In this section we analyse the data retrieved directly from the Android SDK application usage statistics, and the data retrieved from the user responses to the *"application purpose form"*.

Android SDK allows us to retrieve statistics concerning the usage of applications. These statistics come in the form of the total time that an application was used, in milliseconds. In the case of the Vitoria system, we were only interested in the time a user actively used an application. As such, we only retrieved the time an application was in the foreground, that is the amount of time an application was actively on the screen.

In the data collected from the participants it was possible to identify more than 700 different applications. Some of these were applications from the smartphone manufacturers, and others represented Smartphone capabilities like storage of Contacts or Messages. Due to the heterogeneity of the systems, and to the large number of applications on offer nowadays, it was not expected that all users had the same installed applications, and this was in fact exactly what happened. To deal with this issue, in the case of participants that did not have a certain application installed, a zero-minute mean value of usage for that application was considered. Additionally, similar applications were aggregated by categories.

The first aspect that we tried to determine was if the evolution of the pandemic, in Portugal, co–occurred with changes in the use of smartphone applications. We can see in Table 2 that the first phase of deconfinement in Portugal started on the 15^{th} of March. As such, we choose to divide the data into two groups, before the 23^{rd} of March 2021 and after that, since that was the day that marked the middle of the study and was very close to the 1^{st} deconfinement date.

In Figure 6.9 we can see the 25 most used applications in both periods, i.e., before the 23^{rd} of march and after that date. The usage is expressed as the mean value of time spent using an application by day and by user, in minutes. It is also possible to see in the figure that the most used applications are messaging

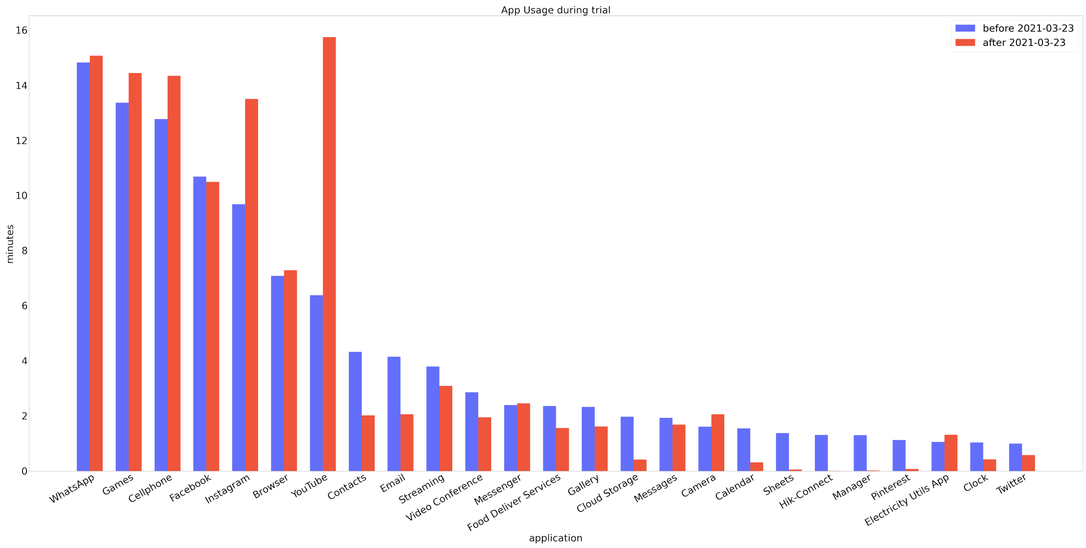


Figure 6.9: Application usage during the 2 halves of the trial.

apps (WhatsApp, Messenger), gaming apps, cell phone apps, social network apps (Facebook, Instagram, Twitter), and streaming apps like Netflix, HBO, Stremio, etc. In the set of most used applications, it is also possible to see some applications that are mostly related to work, such as email applications, cloud storage or video conferencing applications.

Some applications were aggregated into classes of applications. For instance, different games were all aggregated into the Games category. The same happened for applications for Streaming, Email, Contacts, Browser, Cell phone, Cloud Storage, Camera, and Video Conferencing. We choose to do so for three reasons. Firstly, most smartphone manufacturers include different applications for some of those functionalities, such as for cell phone applications and Camera. Secondly, it was not our goal to evaluate the most used application but rather evaluate the type of usage. For instance, it does not matter if a person uses the Netflix application or the Amazon Prime application, since both of them serve the same purpose (i.e., content streaming). The third reason is to eliminate biases and heterogeneity between users. For instance, in the games class we found more than 50 different games, and it would be unreasonable to individually compare their usage, since most users do not have the same games installed.

Concerning the mean total time spent on apps, the times in both periods are very similar, with 178 minutes in the periods before the 23^{rd} of March and 151 minutes after that date, which is in line with the mean usage times in other studies (Andone et al. [2016]). Furthermore, the top 25 applications presented in Figure 6.9 account for approximately 2/3 of the total usage. We can see that although the 7 most used apps seem to all increase in usage in the second period, the mean time spent using applications on the Smartphone is smaller for this period, which indicates that the increase was indeed in these specific applications and not in the use of the smartphone.

The Instagram app, like most social networking applications, can be used to either consume content or publish new content. There was an increase in the use of the Instagram application in the second half of the study. This was accompanied by an increase in the use of the camera application as well. Furthermore, the increase is of similar order, with an increase of 40% in the use of the Instagram app and an increase of 68% for camera applications, which could indicate that, in fact, the use of the Instagram application was more related to publishing content. This could also be related to people leaving their homes more often and finding more ”interesting things” to capture in photos. Other studies have explored the relation between different types of use of this application and depressed states (Frison and Eggermont [2017]). As such, in future studies it would be interesting to monitor the type of use of this and similar applications as they could elicit more information about the users' mental states.

YouTube was the application with the highest increase of all apps, amounting to more than the double. However, due to the low number of participants in the study, it is likely that this was an outlier, i.e., that one person, or a small subset of people, were responsible for this increase. We then verified that the YouTube application was mostly used by two users, and in fact these users had a big increase from an average 30 minutes to 1 hour and 30 minutes in the second period. Although, as we suspected, the increase was indeed affected by the small number of participants, the trend in the increase of YouTube use is still valid, since these users were also the users which presented the highest usage time in the first period.

There was also a decrease in applications related with work tasks, such as email, video conferencing apps (Teams, Skype, Zoom), cloud storage apps, the calendar, and even Sheets. This could point to one of two situations: 1) the use of remote work apps decreased due to the unfavourable work situation (i.e., higher unemployment); 2) people took some time to get used to the remote work situation and, with time, replaced the use of mobile applications with the use of more ”work friendly” ways to use such functionalities, such as using their laptop. It is also important to refer that, with the technological evolution, it is more frequent for people to use social networks as a work tool. Additionally, in Portugal it is very common for some companies to use WhatsApp as a tool for communication within the company. As such, it could be hard to infer if the purpose of the usage of certain application was related to work or not. In the next subsection, we explore the answers of the users to the application purpose questionnaire, which can be used to better infer these aspects.

Additionally, due to privacy reasons, no demographic information from the users was collected, and so we do not have access to information such as age or gender of each participant in the study. It has already been proven in past studies that the use of different applications is highly correlated with users' age and even gender (Andone et al. [2016]). Making future studies with more users, and access to demographic information, may also provide new information related to the use of different types of applications.

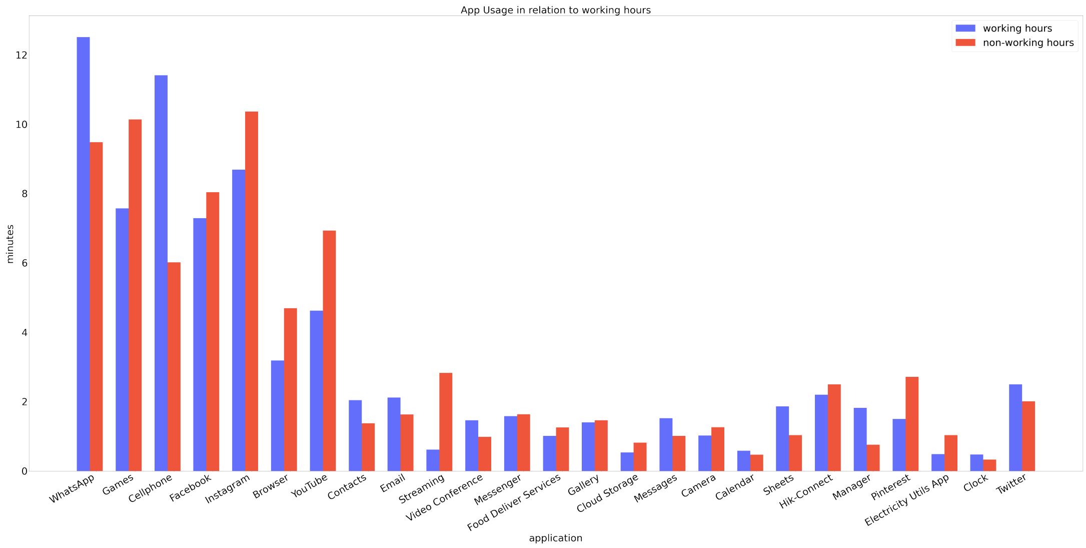


Figure 6.10: Mean application usage during non-working and working hours.

6.4.2 Application finality

Like we have previously stated, an application can have different usage purposes depending on the user. This is also an aspect that we wanted to evaluate in this preliminary study, particularly the use of applications related to work and non-work activities. As we stated before, there are several applications that are highly tied to work activities, such as e-mail and office tools. However, there are other applications that even though their primary use is not for work, can be used as tools for such activities. On the other hand, most applications are not related to work activities and are mostly used for leisure. The use of the various types of applications, their relationship with work/non-work activities, and the time at which they were used during the day could offer some information about the users' daily schedule.

To better evaluate these aspects, we wanted to further inspect how the most used applications relate to working and non-working periods. Since we did not have specific details about the participants' work, or schedules, we considered the most common working schedule in Portugal. Normally in Portugal, the working schedule consists of 40 weekly working hours, with most people working eight daily hours during weekdays. Additionally, most people work from 8 to 18h with a 2-hour break for lunch. As such, this was the arrangement considered when dividing the data in working periods and non-working periods. Additionally, all weekends were considered as nonworking periods.

The mean usage time of each application, in minutes, divided by these two periods can be seen in Figure 6.10. In this figure we present the mean value for the total duration of the study, as we only aimed to evaluate the use of different types of applications in relation to working periods and nonworking periods, and not their evolution.

We can see from the figure that applications used for interpersonal communication, like WhatsApp, Cell phone, Video Conferencing and E-mail tend to

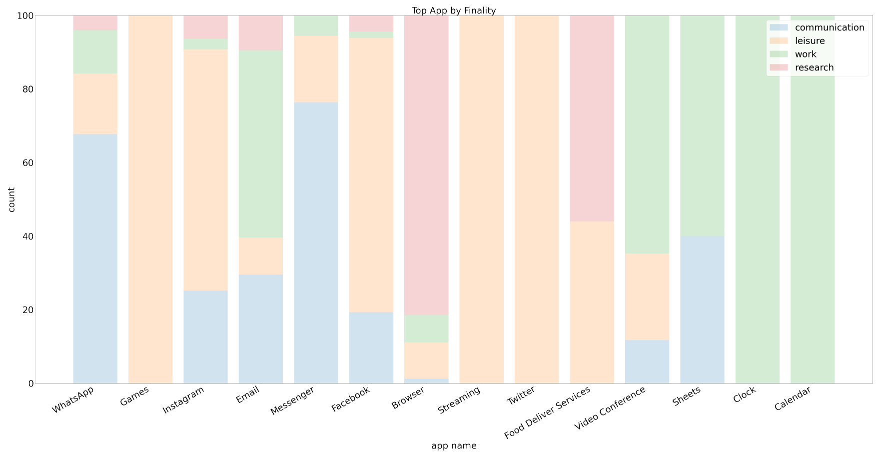


Figure 6.11: Type of use in percentage for most used applications

be more used during working periods. This was expected since most workflows, especially when working remotely, require that people use virtual means of communication with co–workers. However, we can also see significant usage of these applications outside normal working hours. This was also expected, as these applications are also used for personal communications outside work, like WhatsApp and the Cell phone. However, the high usage of Video Conference and Email applications outside normal working hours could point to a shift in the working schedule when working remotely.

Furthermore, as expected, apps related with leisure tended to be used more often outside working hours. This was the case of Games, Streaming applications, YouTube, and social networks. However, as we can also see from Figure 6.10, there was also a considerable usage of Game applications and Social Network applications during working hours. This could point to the participants taking some time for small breaks within their working hours, which can happen more easily due to working from home. Additionally, when analysing the list of distinct games present in the Games category, we verified that many of those games seemed to target very young children. In Portugal, during the duration of this trial study, schools had adopted a remote lecture scheme as well, leading to many parents having to manage their work schedule along with their children's activities. In fact, we believe that some usage of gaming applications was due to participants trying to entertain/distract their children for some time, as some studies have pointed out the rapid increase in the usage of smartphones and other similar devices by small children (Radesky et al. [2015]). However, as previously stated, without more personal information from the participants we cannot draw firmer conclusions about this.

Additionally, as explained in section 5.2, the participants were prompted to answer a daily questionnaire and select the purpose that better described their use of the five most used applications during each day. Since the users were only prompted with the most used applications for that particular day, and these applications vary between users and days, the list of these applications may differ from the previously presented 25 most used applications presented in Figure 6.9. As such, only 14 of the total most used applications are presented in this data. The percentage value of each type of purpose per application can be seen in Figure 6.11. Note, that Figure 6.11 scale carries no relation with the scales of Figures 6.9 and 6.10. That is, a small percentage in one type of purpose may not indicate a small mean usage time for that application and that particular purpose, and vice versa, since the user is only prompted to answer the purpose the of five most used applications in the day, and the total usage time can vary greatly from one day to another, or even from one user to another.

As can be seen in Figure 6.11, some applications were exclusively or almost exclusively used for one purpose. For instance, Games, Twitter and Streaming applications are used exclusively for leisure activities, while others like the Clock or Calendar applications are exclusively used for work. Other applications, like Email, Sheets and Video Conference applications are mostly used for work.

By comparing the answers to this questionnaire with the data shown in Figure 6.10 we can see that some applications used mostly for leisure, such as Games, Twitter and Facebook, were used almost indistinctly during work and non-work periods. This could mean that participants felt the need to take breaks, due to working from home and having more control over their schedules. Applications mostly used for communication, such as WhatsApp and Messenger, presented the same behaviour. However, for this type of applications, some users report using them for work-related activities, which could explain some usage during work periods.

On the other hand, applications that are mostly used for work such as the Clock, Calendar, Sheets, and Email, were used on both working and non-working periods. This could point to a difficulty in *"disconnecting"* from work. Other studies have explored the challenges of working remotely, and pointed out the difficulty in self-managing time (Flores [2019]), which could lead to working more hours and/or later. Additionally, it is not uncommon for employees to contact workers outside working hours, especially in a remote working context, potentially giving some people the impression that they are always at work. In fact, in November 2021, Portugal deemed it illegal for employers or companies to call or message their workers outside normal working hours (Guardian [2021]). However, more information about the participants' working schedule is needed to draw conclusions, and this is something which will be addressed in future work.

6.4.3 Correlation with Covid feeds of Information

The way information is disseminated, and the particular information that is shared with the public in a situation such as the one lived during the Covid-19 pandemic, can directly influence peoples' lives, their psychological states and even their physiological states. As preliminary analysis, we intended to explore the correlation between public information feeds/significant events and the collected data. In these preliminary results we compared the data obtained from

	arousal	valence	sleep quality	sleep hours	number contacts	weekday	risk level	active cases	confirmed cases	t_confirmed cases	n_dead	feeds positiveness
arousal	1.000	-0.560	-0.008	0.022	-0.068	-0.442	0.054	0.146	0.222	0.011	-0.028	0.176
valence	-0.560	1.000	0.057	0.035	-0.069	0.450	-0.080	-0.152	-0.069	0.090	0.094	-0.042
sleep quality	-0.008	0.057	1.000	0.252	-0.081	0.126	-0.316	-0.249	-0.073	0.129	0.153	0.664
sleep hours	0.022	0.035	0.252	1.000	-0.061	0.238	0.074	-0.080	-0.035	0.161	0.065	-0.019
number contacts	-0.068	-0.069	-0.081	-0.061	1.000	-0.120	0.075	0.149	0.077	-0.055	-0.025	-0.002
weekday	-0.442	0.450	0.126	0.238	-0.120	1.000	-0.116	-0.161	-0.186	0.003	0.048	-0.009
risk level	0.054	-0.080	-0.316	0.074	0.075	-0.116	1.000	0.524	0.373	-0.021	-0.192	-0.553
active cases	0.146	-0.152	-0.249	-0.080	0.149	-0.161	0.524	1.000	0.852	-0.363	-0.590	-0.355
confirmed cases	0.222	-0.069	-0.073	-0.035	0.077	-0.186	0.373	0.852	1.000	-0.246	-0.566	-0.125
t_confirmed cases	0.011	0.090	0.129	0.161	-0.055	0.003	-0.021	-0.363	-0.246	1.000	0.843	-0.034
n_dead	-0.028	0.094	0.153	0.065	-0.025	0.048	-0.192	-0.590	-0.566	0.843	1.000	0.078
feeds positiveness	0.176	-0.042	0.664	-0.019	-0.002	-0.009	-0.553	-0.355	-0.125	-0.034	0.078	1.000

Figure 6.12: Time-lagged Cross-Correlation Matrix of collected data and Covid related numbers and events (3 days shift).

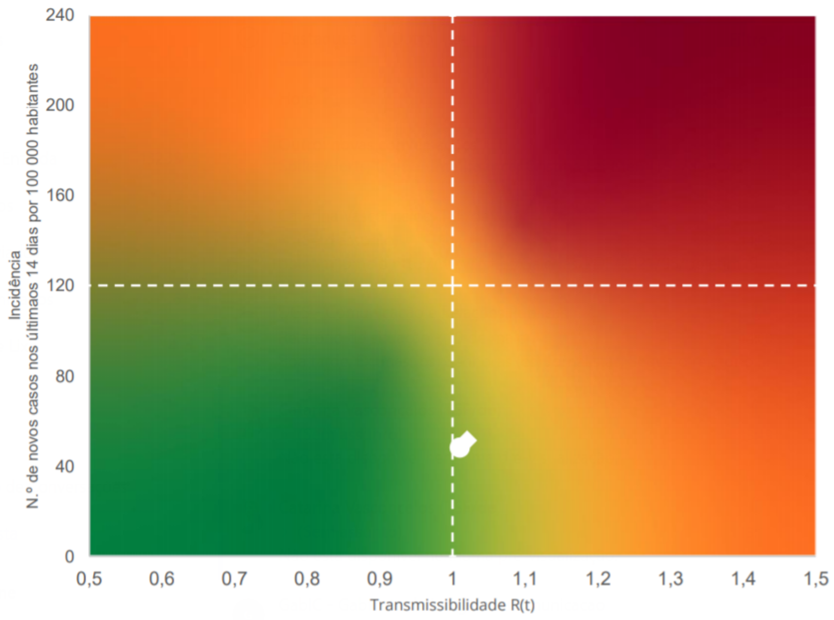


Figure 6.13: Example of Covid Risk Matrix, Extracted from a *Direção Geral de Saude* report.

the mobile questionnaires, namely the sleep questionnaire, the Emotional Questionnaire (SAM scale) and the transport and proximity questionnaires, with the most used metrics by the media to describe the pandemic situation. Furthermore, we compared this data with the general public opinion in relation to the most relevant Covid-related events, presented in Table 2, and with several Covid-related numbers that were disclosed by the competent Portuguese authorities on a daily basis. In addition, the correlation of these variables with weekdays was also considered.

Considering the data collected from the various questionnaires, it is possible to derive five different features. From the emotional questionnaire, it was possible to obtain the daily value of valence and arousal for each participant, which can represent the emotional state of a person. The data collected from the sleep questionnaire, included the daily sleep quality and number of slept hours. And lastly, both the transport questionnaire and proximity questionnaire data were fused to obtain a daily total number of possible contacts by user. For each variable, a time series was generated, considering the mean value by day and by user.

During the pandemic, the Portuguese National Health System maintained a dashboard where daily reports concerning the evolution of pandemic situation and numbers in Portugal were released. From those reports, it was possible to extract a dataset that is now publicly available (covid19pt data [2022]). We selected some features from that dataset that were the most used by the media, namely, the values for the number of active cases, new confirmed cases, total confirmed cases, and number of new death cases for each day. Additionally, a metric not directly present in the dataset but that was one of the most used by the media was the risk level, using a risk matrix that could be calculated from the incidence and the transmissibility of the virus (Rt value), with the top right quadrant/dark red representing the highest level and the bottom left/light green representing the lowest level. An example of this risk matrix can be seen in Figure 6.13. We calculated the risk level values from the incidence and transmissibility values available in the dataset, using the following formula:

$$\begin{cases} 2\ if \text{ incidence} > 120 \text{ AND transmissibility} > 1 \\ 1\ if \text{ incidence} > 120 \text{ OR transmissibility} > 1 \\ 0\ if \text{ incidence} \leq 120 \text{ AND transmissibility} \leq 1 \end{cases}$$

In addition, we made a compilation of the most relevant Covid-related events in Portugal between the months of January and May 2021, which can be seen in Table 2. In this table it is also possible to see the positiveness of each event, i.e., events marked as *"+"*, were perceived as positive by the general public, while the ones marked *"-"* were perceived as negative. We then proceeded to calculate the positiveness index of each date, by considering the +1 value for positive events and -1 for negative ones. The value for each day was then the sum of all the values that preceded that day. This could be seen as an accumulation of events in people's memory, and a relative shift in their perception of the state of the pandemic.

Furthermore, we are aware that the state of mind of people does not change immediately on receiving information, and that their reactions can be influenced by short-term history. Given this, we decided to evaluate the Time-lagged Cross-Correlation, using the Pearson coefficient, between the acquired data and the Covid-related metrics/events. In this particular case we wanted to evaluate the response of people to the current or short-lived feed of information about the pandemic. As such, the data acquired from people was stationary, and all related Covid features suffered a shift in time. We calculated the time-lagged Cross-Correlation with a shift of zero to four days, being that the best results were obtained for a shift of three days. Due to the overflow of new daily information, we believe that it is not important to explore a time-lag greater than four days.

Considering the data presented in Figure 6.12, we can see that, apart from the arousal and valence values, none of the collected data has correlations between them. Furthermore, we can see that the valence and arousal values were strongly and negatively correlated, which indicates that the participants mostly reported emotions that fall within the pleasure and anger sectors of the 2D emotion diagram (Yazdani et al. [2013b]). That is, when the participants felt more positive, they also felt less emotionally energetic, which indicates relaxation, serenity, or contentment. On the other hand, when they felt more energetic, they also felt less positive, which indicates feelings of annoyance and anger. In addition, both valence and arousal had a moderate correlation with the weekdays, namely a negative one for arousal and a positive one for valence. This was to be expected since most people tend to fell less energetic as the week progresses, and normally the weekend is associated with leisure time, when people tend to feel more positive. However, this information also serves as validation for the collected data, since it correlates with the findings of other studies (Helliwell and Wang [2014]).

Considering the correlation between the acquired data, and the time-lagged Covid news/events metrics, the only strong correlation was between the reported sleep quality and the total positiveness of past events with a shift of 3 days. This indicates that people reported that they slept better when the past events related with Covid were more positive. Furthermore, we can see from the coefficients that the number of slept hours and the sleep quality did not have any significant correlation, which indicates that this increase would be mostly due to the participants average perception of the Covid-related events.

Additionally, there was also a moderated negative correlation between the sleep quality and the zone of the risk matrix with 3 days shift. As previously stated, the risk level of the risk matrix was one of the most used metrics by the National Health Organizations and the media. As such, this correlation could indicate that the information broadcasted by the media could affect the way people rated their sleep quality. Note there is also a strong correlation between the risk level and the total positiveness, which could explain this correlation.

On the other hand, the other metrics used by the media, such as the number of new cases and the number of deaths had no significant correlation with the collected data. As previously stated, the risk level matrix was given considerable importance by the Portuguese media. This was due to the fact that the risk level is a more stable metric than the number of cases and number of deaths. Moreover, this metric was directly linked to the deconfinement measures and, thus, had direct implications on peoples' lives, which can explain the differences in terms of the existence of a correlation.

Sleep quality was the only metric from the collected data that presented a significant correlation. This could indicate that self-reported sleep quality might be a good indicator for how well people dealt with the reported Covid situation and events. Furthermore, during the trial, the self-reported emotional data indicates that the participants mostly experienced feelings from the anger and the pleasure sectors of the emotional diagram. However, as previously noted, a larger sample size is needed to draw stronger conclusions about the interplay of these factors. Additionally, it would be interesting to compare the collected data with data collected in a period during which the lockdown and the pandemic are not active. Both of these concerns will be addressed in future studies.

6.5 Analyzing Heart Rate and Breathing Rate with Wi-Fi CSI

As we explored in Chapter 5 one of our case studies aims to evaluate the use of unobtrusive sensing within the context of HITLCPS. In the project *"iFriend"* we aim to monitor elderly people with kidney failure, by using both data collected from the smartphone and by the CSI of the Wi-Fi. As we also covered in the implementation details of the project the first phase of the project will focus on the collection of data, from trial subjects, to be then processed and worked on to generate models able to predict the HR and BR. As such, in this section, we will only present preliminary results of small trials performed in laboratory.

6.5.1 Experimental setup and context

As we previously stated, we implemented the *"iFriend"* prototype by using the Atheros CSI tool (Xie et al. [2015]). This tool allows us to gather CSI, which normally is not available in off-the-shelf devices. The tool is also, in theory, compatible with any Atheros and Qualcomm device and can be compiled for both Linux–based systems and OpenWrt. As such it could be used in the vast majority of AP available in the market, pending the installation of the OpenWrt operating system. We choose to implement the prototype by using two identical devices, namely the TP-Link Re450, which can be seen in Figure 5.24. Since these devices use a different architecture, namely the *mips_24kc*, we performed a cross-compilation of the tool and created a compatible firmware image, which was then installed on both devices.

Additionally, one of the devices was set up as an AP/Master, which was sharing a dedicated Wi-Fi network, and the other one was set up as a client of that network. The network was shared using the 2.4Ghz band and with a bandwidth of 20MHz. The master was also connected to a different network to send the received packets to the server for storage. We choose to use another network, in this case, one connected by the ethernet interface of the device, so no additional interference was generated in the signal by additional transmissions.

A scheme of the experimental setup can be seen in Figure 6.14. As can be seen from the image it resembles the intended scenario in which the prototype will be used (Figure 5.26), that is monitoring elderly people while they are watching tv and sitting on the couch. In this experimental setup, we choose the scenario of someone working on their laptop. The subject was sitting on a chair next to the table with his laptop on it.

Additional, considerations were also made in terms of the positioning of the devices. Namely, the devices were set at a distance of 3 meters from each other

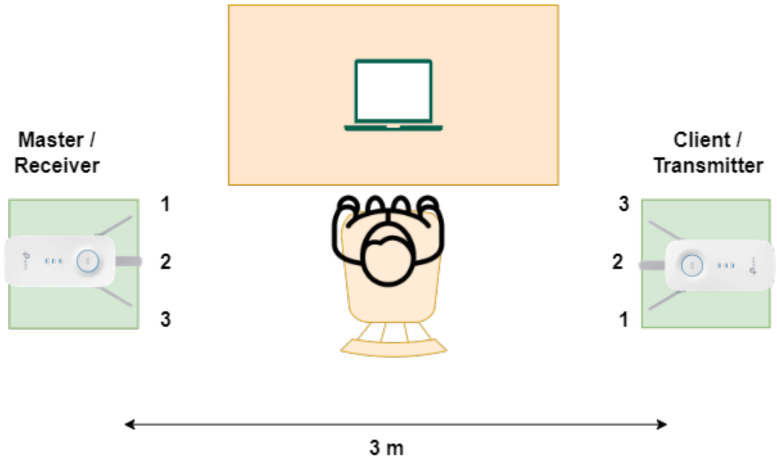


Figure 6.14: Representation of the experimental setup for the collection of Wi-Fi CSI data.

to correspond to the expected scenario for the real–world trials. And they were raised to a height of roughly 85 cm to be at the level of the chest area.

Both of these scenarios are more challenging that the ones found in the reviewed literature, that is monitoring people during their sleep (Liu et al. [2015]). Since there are additional movements from people. One of the challenges of this type of monitoring is the fact that large movements from the human body (e.g., rapidly moving your arms, sitting movement, standing up) generate larger interference in the CSI, which in turn makes it harder for signal processing techniques to retrieve the interference generated by the HR or BR.

In Figure 6.15 we can see the amplitude of a sub-carrier in the CSI. The figure also denotes an interval in time when movements from the subject occur, marked as red. As can be seen from the figure, larger movements cause interference in the signal in both the amplitude and the frequency. Thus, making it harder for the algorithms to estimate the HR and BR.

Lastly, in addition to the previously explained setup, for the collection of the CSI, two additional devices were used for the collection of the ground truth of the HR and BR. For the HR, the Tic Watch S2 smartwatch, which can be seen in Figure 6.16a, was used to collect HR values (Mobvoi [2022]). This device runs the Android Wear OS and as such it is programmable, thus a small application was developed to collect the HR values and send them to the *"iFriend"* server. For the BR values, the Go Direct respiration Belt was used (Vernier [2022a]), as can be seen from Figure 6.16b. This device is attached to the chest with a strap belt and can measure the respiration rate based on the chest's movements. It also offers APIs in some programmable languages and two communication possibilities, by USB connection or by Bluetooth. By trial and error, we found that the Bluetooth interface generated interference with the Wi-Fi signal because of the proximity of the devices, as such, we choose to use the USB connection.

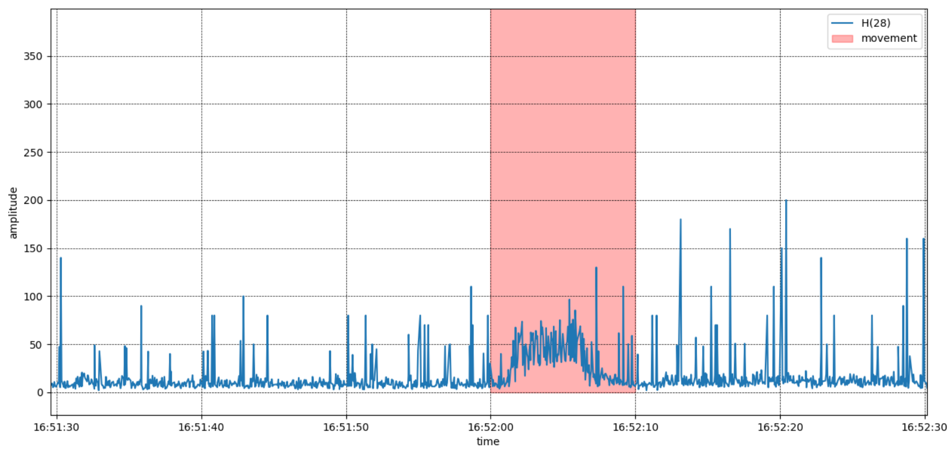


Figure 6.15: Example of interference generated by movement in the CSI signal

(a)

(b)

Figure 6.16: Smartwatch and Respiration belt used for the collection of the ground truth values.

In the next section, we present the preliminary results of some small tests. These tests were performed with one subject, which is a 29 years-old male with no known health conditions. For the duration of the tests, the subject was instructed to perform normal web–surfing tasks on the laptop (e.g., search the web, watch videos, listen to music).

6.5.2 Heart Rate and Breathing Rate estimation

In this section, we will explore the preliminary results obtained by using the CSI to estimate the HR and BR. But firstly, we will explore how CSI can be defined and obtained.

CSI is the channel attribute of each communication link. It describes how much each signal transmission path was weakened, by factors such as power decay over distance, scattering, multipath fading, shadowing fading, and others. CSI is used to adapt the communication system to the current channel conditions, increasing the reliability, in multi-antenna systems also known as Multiple–Input Multiple–Output (MIMO). Channel information can be modelled in the frequency domain as

$$\overrightarrow{Y} = H\overrightarrow{X} + \overrightarrow{N} \tag{6.1}$$

where $\overrightarrow{X}$ and $\overrightarrow{Y}$ represent the transmitted and received signal vectors, respectively. $\overrightarrow{N}$ represents the Gaussian noise vector and the H matrix represents the channel's frequency response. The CSI of all sub-carriers can be estimated by the formula

$$H = \frac{\overrightarrow{Y}}{\overrightarrow{X}} \tag{6.2}$$

MIMO systems have multiple receiving and transmitting antennas, as such for each sub-carrier the CSI matrix has a component for each pair of antennas. The matrix of the n th sub-carrier, with r receiving antennas and t transmitting antennas can be expressed, as follows

$$H_n = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1r} \\ h_{21} & h_{22} & \dots & h_{2r} \\ \vdots & \vdots & & \vdots \\ h_{t1} & h_{t2} & \dots & h_{tr} \end{bmatrix} \tag{6.3}$$

As stated before, in the tests performed we used the 2.4Ghz band and a bandwidth of 20MHz, as such the signal had 56 sub-carriers. Additionally, the devices used had 3 antennas each, as such we have 3 receiving antennas and 3 transmitting antennas. Therefore, for each packet sent we obtained 56 matrices of 3×3 each. Since we only need one stream, we choose to use the pair of antennas $r = 3$ and $t = 1$, which as can be seen from Figure 5.26, is the one that has the more direct path through the human body.

Additionally, each component of the sub-carriers matrix is represented as a complex number, which represents the amplitude and phase of the signal. In our tests, we used only the amplitude information of the signal. Unless indicated otherwise, from this point on when referring to the components of the CSI matrix we are referring to the amplitude.

For these preliminary trials, we used a packet rate of 20 packets per second and the algorithms described in Liu et al. [2015]. For the detection of the HR, we first converted the signal to the frequency domain and then applied a Band-pass filter to remove any frequencies outside the normal HR frequencies during a rest period. That is, any frequency below 0.9 Hz (54 beats per minute) and frequencies over 2 Hz (120 beats per minute). Secondly, the PSD was calculated for each sub-carrier, and the mean PSD of all sub-carriers was also calculated.

Table 6.16: CSI HR estimation error relative to ground-truth, in beats per minute.

	Max Abs Error	***Min Abs Error***	***Mean Abs Error***
1 minute window	17.12	0.08	7.32
2 minutes window	15.57	0.06	5.86
3 minute window	18.05	0.12	6.33
1 minute w/ movements filtered	13.93	0.08	5.65

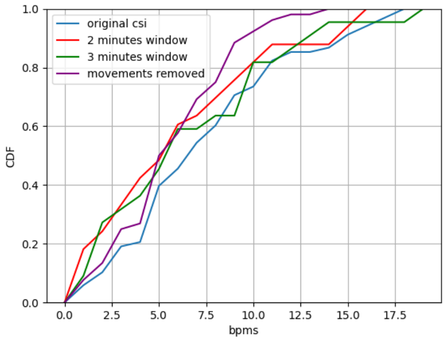
The HR estimation was then found by finding the max frequency component of the mean PSD of all sub-carriers.

A collection of data of one and half hours was performed with the subject using the smartwatch for ground-truth. We also compared the results using different time windows, that is the CSI signal was divided in different windows of time with the same duration and the estimation was made for each individual window. The smartwatch retrieved a value for the heart rate every minute, for larger windows of time the mean value was calculated.

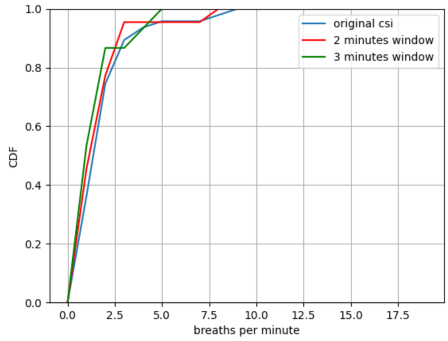
The values for maximum, minimum and mean absolute error of the HR estimations can be seen in Table 6.16. As can be seen from the table, we performed estimation tests for windows of 1 minute, 2 minutes and 3 minutes. Additionally, we present the results for 1 minute windows when removing instances of time that could correspond to the movement of the subject. The largest mean absolute error was for the 1-minute window with 7 beats per minute. It is also possible to see that the minimum absolute error for all the approaches is near 0. However, the maximum absolute error value is larger than 13 beats per minute for all tested approaches. As can be seen from the values, changing the window size results only in small differences showing that this is not the most important factor to consider.

We believe that the larger error values are due to instances in time when the subject was significantly moving. The data was manually annotated and 1 minute windows were removed for any instance that resembled the artifact shown in Figure 6.15. It is possible to see from Table 6.16, that removing these problematic instances in time produces better estimations.

In Figure 6.17a, we present the Cumulative Density Function (CDF) plot for all the above approaches. It's possible to see from the figure, that for almost all approaches 50% of the results are below an error of 5 beats per minute. Furthermore, when removing the instances of time when movements occurred, 90% of estimations are below an error of 10 beats per minute. These results show that there is still a large error when comparing the estimations with the



(a) HR



(b) BR

Figure 6.17: CDF for HR and BR estimations.

Table 6.17: CSI BR estimation error relative to ground-truth, in breaths per minute

	Max Abs Error	***Min Abs Error***	***Mean Abs Error***
1 minute window	8.36	0.09	1.71
2 minutes window	7.82	0.28	1.59
3 minute window	4.32	0.00	1.24

ground-truth from the smartwatch.

Concerning the BR estimation, a similar acquisition of one and half hours was performed with the subject using the GoDirect Respiration belt, shown in Figure 6.16b. The same packet rate of 20 packets per second was also used for this acquisition. The algorithm for the estimation of the BR is very distinct from the one used for the HR, since it is based on the time domain instead of the frequency domain. Firstly, the Hampel filter is used to remove any outliers of the signal. Secondly, a moving average filter is used to smooth the signal. We then find the peaks of the signal and remove any fake peaks. Lastly, the breathing rate for a sub-carrier is estimated from the number of peaks per minute. The final breathing rate estimation is obtained by calculating a weighted mean based on the variance of any sub-carrier, since the sub-carriers with larger variances are more affected by the movements of the body.

The values for maximum, minimum and mean absolute error of the BR estimations can be seen in Table 6.17. As can be seen from the table, we performed estimation tests for windows of 1 minute, 2 minutes and 3 minutes. All approaches had a mean estimation error below 2 breaths per minute. Furthermore, increasing the duration of the time window for estimation lowers the mean absolute error of the estimation. It is also possible to see that for all approaches the minimum error obtained was near 0. However, the maximum absolute error even in the 3-minute window approach is above 4 breaths per minute.

In Figure 6.17b we present the CDF for the BR estimations. It is possible to see from the figure that for all approaches, 80% of the estimations had an error of below 2.5 breaths per minute. Additionally, for the approach of using 2-minute windows, 90% of estimations are below 3 breaths per minute. Although these results show some promise, we believe that there is still room for improvement.

As happened for the HR estimations, we believe that the larger errors are due to instances of time when the subject performed larger movements. In future work, we will improve the algorithms used to filter these instances and perform opportunistic sensing when the subject is quasi-static.

In summary, these results still show significant errors, and we were not able to obtain the same performance as past works in the literature. However, we highlight once again, that these are only preliminary results, we intend to further evaluate these approaches and others with the dataset that will be collected in the future *"iFriend"* project trials.

6.6 Chapter Summary

In this chapter, we focus mainly on presenting the implementation details of the trials realized in our case studies, and presenting the results obtained from analysing and exploring the data obtained in those trials. We cover the results from 3 of the case studies presented in the last chapter, namely, the *ISABELA*, *Vitoria* and *iFriend* case studies.

We start the chapter by presenting the results of two trials of the ISABELA case study, performed during 2018 in Portugal and in Ecuador. In these results, we analysed the use of social sensors (i.e., OSNs) as a source of data, and evaluated differences in behaviour in different groups of students and their correlation with performance. We then present the results for two machine learning models, namely, a model to predict student sociability levels and another model to predict their sleep quality, which were created and evaluated using the data from the same trials.

As stated before, our main objective with the ISABELA system was to monitor students' performance. As such, we needed to create models that were able to predict students' performance, in order to modulate the system. In this chapter, we evaluated several models to predict students' performance using data collected in a trial with the ISABELA system in Ecuador, during 2021. In the context of the same work, we also proposed a pipeline that uses a decision level median voting algorithm to further improve the model's performance.

Additionally, in this chapter, we explore the data collected from a trial performed with the Vitoria system. As we discussed previously, this system aims to monitor humans and their behaviour during the Covid-19 pandemic. The results presented in this chapter mainly focus on the use of different applications and their usage in specific periods of time. We also covered the usage differences of some applications and the purpose for which they are used. In the same results, we present the correlations of the data acquired from the Vitoria system with some metrics used by the media to report the pandemic situation.

Lastly, in this chapter, we explored some results for the *iFriend* project, namely, preliminary results for the estimation of HR and BR, using Wi-Fi's CSI. We started by explaining the experimental setup and the configurations of the acquisition. These tests were performed with one subject, and compared with the ground truth from commercial devices, namely, a smartwatch and a respiration belt. We then explored how larger movements of the body and different time windows can affect these estimations. We presented the results in terms of maximum, minimum and mean absolute error compared to the ground truth, as well as the CDF for each of the metrics and different approaches explored. The results show significant error when compared with the ground-truth and the results found in the state-of-the-art, which shows that some work is still needed in this field.

In the next chapter, we will analyse the work done in this thesis, both theoretical and practical, and address the remaining challenges and future work avenues for the HITLCPS field.

Chapter 7

Conclusions and Future Work

Contents

Advances in several technology fields allow us to develop smarter systems. We now have at our disposal devices that are smaller in size, have more processing power, faster communication capabilities and long-lasting batteries. One such type of device is the smartphone, which is used by the majority of humans and has sensing, processing, and communication capabilities. Furthermore, these devices are interconnected, in what we call the Internet of Things. Additionally, emerging sensing techniques enable the collection of data in manners that were not possible in the past. We can use these devices and all of their capabilities to create systems that better serve humans.

However, most systems still see humans as end-users and not as part of the system. Human actions, intents and emotions can affect the way they interact with a system, impacting its performance, by either improving it or making it perform worse. This points to the fact that there is still a need for proposing and studying new mechanisms for HITLCPS.

In this chapter, we summarize the developed work, highlight the contributions of this thesis, and look into the future challenges of HITLCPS.

7.1 Synthesis

The work described in this thesis focused on creating and studying new models for the HITLCPS paradigm, by using IoT and other sensing techniques. The work was organised and presented as follows.

Chapter 2 introduced the concepts of IoT, CPS and HITLCPS. We reviewed the work concerning the creation of a HITLCPS paradigm, with a special focus on smartphone systems and social sensors. We also explored the state-of-the-art of systems that monitor humans in relevant scenarios, such as the Covid-19 pandemic scenario. Additionally, we also reviewed systems to assess students' performance, as this is one of the fields where the HITLCPS paradigm can be more significant.

Next, in Chapter 3 we focused on the challenges of HITLCPS by addressing the open issues found in Chapter 2. We then proposed a new model for the HITLCPS paradigm, that tries to tackle all of those open issues. This novel model includes humans in every phase of the control loop, and incorporates several design features to better tailor the systems to humans, namely, in the data acquisition phase, by considering humans as both subjects of the acquisition and as a source of data (i.e., social sensors). In this phase, the model also considers unobtrusive sensing solutions as a new source of data for vital signals acquisition. Additionally, humans should be part of the state inference phase as well, which can be achieved through HITLAI mechanisms. Lastly, in the actuation phase, HITLCPS systems should target actuation based on human interaction or human–like interaction, by resorting, for instance, to chatbots.

In Chapter 4 we presented our definition, as well as a taxonomy for the classification of unobtrusive sensing. Unobtrusive sensing comprises several approaches. As such, it can be hard to compare them in a comprehensive manner. To tackle this issue, we proposed a taxonomy to classify these solutions according to their nature, technical approach and multimodality. We also presented the state-of-the-art of these solutions in each of the taxons of the proposed taxonomy. Additionally, we reviewed these solutions in terms of multimodality and computational architecture choices. We finished this chapter by exploring the open issues and challenges that we believe will be more important for the field of unobtrusive sensing and their integration with HITLCPSs.

Advancing to Chapter 5, we presented the implementations made in order to study our HITLCPS model. HITLCPS can be applied to several areas and, as such, we decided to implement several case studies according to the proposed HITLCPS paradigm. One of these areas is the monitoring of students in their academic context. We implemented a case study named ISABELA that uses an IoT architecture based on the FIWARE backend, smartphones, smartwatches and small IoT devices. This system aims to monitor the students' behaviours, lifestyles, and daily routines, and to infer their performance. The system also closes the loop, by giving feedback to the students through a chatbot.

Additionally, in Chapter 5, we covered the implementation details of the Vitoria system. This system leverages the architecture proposed with the ISABELA system to create a new system capable of monitoring people during the Covid-19 and future pandemics. This application also collects data that can be related with the risk of contagion and with the user's perception of that risk. Another objective of this case study was to assess the effectiveness of the actuation phase of the HITLCPS. As such, we implemented the necessary system modules to create two groups of users, one that received feedback and another group that served as control.

Still, in Chapter 5, we also covered the implementation of a case study that aimed to evaluate the use of HITLAI in the HITLCPS paradigm. We presented a platform and a recommendation model developed to automatically create recommendations for matches between challenges faced by industry companies related to climate changes and research done in the academy. Although, the model generates recommendations the responsibility of making the decisions falls on humans (namely, on the administrator of the platform). Furthermore, actuation is also made by humans, since the administrator is the one that chooses the best way to approach both researchers and companies.

The last section of Chapter 5 covers the implementations of the "iFriend" project. This case study aimed at monitoring elderly people that suffer from kidney failure, by merging smartphone data, IoT, and unobtrusive sensing. With this case study, we specifically explored the implementation of unobtrusive sensing in a HITLCPS. The description of this case study also covered the design choices made to adapt to specific human limitations (e.g., limitations typical of elderly individuals).

In Chapter 6, we covered the results of several case studies and trials performed in the context of the projects described in this thesis. We started by presenting the results of two trials of the ISABELA case study, one performed in Portugal with 10 students, and a second study conducted in Ecuador with 30 students, both in 2018. These trials evaluated the use of OSN as a source of data and the correlation between student behaviour and academic performance at the end of the term. Additionally, with the data collected from these trials, we developed and evaluated two machine learning models, namely, a model to predict student sociability levels and another model to predict their sleep quality.

Additionally, a second trial with the ISABELA system was performed in Ecuador, in 2021. This study allowed us to acquire data from a period during which covid restrictions were put in place in Ecuador. Joining both datasets acquired in Ecuador, we evaluated several models to predict students' performance using the collected metrics. The model proved to be able to predict the performance of students even when using data originated in different periods (i.e., with and without pandemic restrictions). In the context of the same work, we also proposed a pipeline that uses a decision level median voting algorithm to further improve the model's performance.

In the same chapter, we also covered the evaluation of the data collected from the study developed with the Vitoria system. As stated before, this system aims at detecting behaviour patterns, risk awareness, and users' emotional states, which can be very important in a context such as the Covid-19 pandemic context. In this study, we covered the use of several applications over different periods of time, specifically, during periods in which the Covid-19 restrictions were more or less restrict active. We also evaluated the use of different types of applications according to their purpose, to study how users divided their time during the period of the trial. Additionally, we compared the correlation of the acquired metrics with the metrics most used by the media to report the pandemic situation, as this could be an indication of risk perception.

Lastly, in Chapter 6, we explored the implementation of preliminary tests for the estimation of HR and BR, using Wi-Fi's CSI. Minute movements of the human body affect the CSI of sub-carriers and, with the use of signal processing techniques, it is possible to extract the HR and BR data. Tests were performed with one subject, and compared with the ground truth from commercial devices, namely a smartwatch and a respiration belt. We evaluated how our results compare with the ground truth in terms of maximum, minimum and mean absolute error.

7.2 Contributions

The research work conducted in the context of this thesis was driven by the objectives described in Chapter 1. With these objectives in mind, this thesis led to the contributions described below:

- **Review of HITLCPS applications and approaches**. To better understand the challenges and propose advances in the HITLCPS paradigm, we reviewed the existing work in this field. We gave special attention to systems based on smartphones, since these devices can be used as an entrypoint into humans' lives. Additionally, we have covered several areas of interest in human–specific applications, such as systems to assess students' performance and systems to track the behaviour of humans during pandemics. We believe that this can be used by other researchers to find the open issues in this field of research and fast track the development in this field of research.

- **Creation of a new model for HITLCPSs**. In order to advance the HITLCPS field, we addressed the open issues that were found during the study of past approaches, and proposed a new model for the creation and implementation of this type of systems. This new model can be used as a reference for future HITLCPSs, which should take humans into account right from the system design phase.
- **Creation of Taxonomy for unobtrusive sensing solutions and review of the state-of-the-art**. Unobtrusive sensing is an emerging and relatively new field of study. Furthermore, due to its nature it is a fairly broad field of study which includes various solutions and approaches. The creation of a taxonomy that can help to better understand, classify, and evaluate these solutions was an important contribution. Furthermore, we reviewed the state-of-the-art of unobtrusive sensing, pointed out open issues and proposed future directions for these solutions.
- **Validation of our HITLCPS model in several case studies**. We developed several case studies to validate our HITLCPS model, and presented approaches to tackle the open issues in this field. We focused mostly on the development of HITLCPS by using smartphone applications. However, some case studies also included web-based applications and unobtrusive sensing techniques. These case studies also covered several fields of interest in human–specific applications, such as students monitoring, human monitoring during the pandemic, and applications in the medical field.
- **Development of real–world trials and analysis/profiling of human behaviour**. Using the applications developed in the implemented case studies, we developed several real–world trials to perform data collection. This data was analysed and different usage patterns were identified and discussed. The analysis of this data also led to the **creation of several machine–learning models to infer states in HITLCPS**. Using the data collected in the developed trials, we created several machine–learning models to infer different states of humans, namely, models to predict sociability, sleep periods, sleep quality, and academic performance of students.
- **Creation of a pipeline for automatic assessment of students' performance**. We proposed a pipeline that uses a decision level median voting algorithm to improve the models' classification performance, by using historic data from the students to further improve predictions. This pipeline showed improvement in all tested models, and can offer a new

way to deal with data from trials with a small number of participants.

In addition to the contributions described above the work conducted in this thesis directly led to several publications in international journals and international conferences. We would like to highlight specially the publications on journals of the first quartile. Additionally, it also led to active cooperation inside and outside the research group which, in turn, contributed to several joint publications.

7.3 Future Work

Although the work presented in this thesis demonstrates that it is possible to create systems that better tackle the HITLCPS paradigm, we believe there is still a lot of room for further work in this field. In Chapter 3 we covered the open issues that we found in the state-of-the-art of the HITLCPS paradigm. We tackled those open issues by proposing a novel HITLCPS model in the same chapter. Additionally, in this thesis we covered the implementation details of several case studies that served as validation for our HITLCPS model. However, we were not able to implement a system that comprised all the proposed approaches to the HITLCPS paradigm, and their interactions. As such, studying the implementation details of these components and how they can interact with each other (e.g., merging data from social sensors and unobtrusive sensing) and with humans should be one of the directions of future work in this field.

Additionally, we are aware of the limitations of our results, due to the small size of the datasets. Humans are highly variable, in what concerns their mind, emotions, and behaviours. Thus, in order to further validate the HITLCPS paradigm, we need to evaluate these systems in larger population datasets. This should also be one of the aims of future work, as broader validation will also ease the adoption of these systems.

However, the smaller datasets are in part due to one of the bigger limitations in the development of HITLCPSs, which is the lack of participation and engagement by humans. Difficulty in recruiting participants for trials and limited participation during the trials was one of the major problems felt during the development of the case studies presented in this thesis. Several approaches were tried, from gift cards, gamification and in the context of ISABELA, even rewards by increasing their academic outcome depending on their participation. All of these approaches were not as effective as expect, since we faced difficulties in finding participants, dropouts during the study, and lack of usage of the systems (i.e., the users stopped answering the questionnaires, and some users turned off the applications). Thus, this is one of the larger limitations when studying and implementing HITLCPS. As such, it is also one of the fields that should be further studied in the future, by searching new mechanisms for incentivization and for maintaining the users' engagement, and evaluating their effectiveness.

Although it was not the main focus of this thesis, in Chapter 4 we also covered the state-of-the-art of unobtrusive sensing. This field of research is vast in approaches and, as such, leads to several future work directions as well. In the last section of Chapter 4, we covered the open issues found in this review. We presented the open issues in relation to the field of the proposed taxonomy we believed they are more relevant to. For instance, the challenge of monitoring several people at the same time has already been tackled in the image-based approaches taxon, while it is still a significant challenge for sound-based approaches and for the techniques that fall under the artificial branch of our taxonomy. Other open issues are relevant to all the reviewed techniques. One such challenge, that should be tackled as future work, is multimodality and data fusion. As we stated, the approaches in this field of research are varied and, as such, they generate heterogeneous data. Evaluating how we can jointly use that data and how we can use distinct approaches together to account for their flaws is also one of the proposed directions for future work.

Additionally, in this thesis we presented the implementation details of a case study that uses unobtrusive sensing to enable HITLCPS to obtain humans' vital signals. However, we were not able to fully evaluate how these two fields of research can interact and leverage each other. We believe that the use of unobtrusive sensing in order to empower HITLCPS data acquisition capabilities is also a relevant topic for future work.

Last but not least, privacy is one of the main concerns when building this type of systems. Human-related data is very sensitive, and humans themselves are increasingly becoming aware of that. In recent years, many countries have put in place several policies and legislation that limit the amount of data, the types of data, and the way systems acquire it. This creates several constraints for HITLCPSs, as we need to maintain a trade-off between usability of the system and privacy. Furthermore, when we add new sensing capabilities, such as unobtrusive sensing, we create new privacy concerns, such as having techniques that can collect human-related data even behind walls and without their knowledge. We believe that addressing these new concerns, by creating and studying new privacy preserving policies, are also one of the important future work directions.

Bibliography

Abbas, A. K., Heimann, K., Jergus, K., Orlikowsky, T., and Leonhardt, S. (2011). Neonatal non-contact respiratory monitoring based on real-time infrared thermography. *Biomedical engineering online*, 10(1):93.

Adib, F., Kabelac, Z., Katabi, D., and Miller, R. C. (2014). 3d tracking via body radio reflections. In *11th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 14)*, pages 317–329.

Adib, F., Mao, H., Kabelac, Z., Katabi, D., and Miller, R. C. (2015). Smart homes that monitor breathing and heart rate. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 837–846. ACM.

Afonso, P. (2020). The impact of the covid-19 pandemic on mental health. *Acta medica portuguesa*, 33(5):356–357.

Almalki, M., Giannicchi, A., et al. (2021). Health apps for combating covid-19: descriptive review and taxonomy. *JMIR mHealth and uHealth*, 9(3):e24322.

Almashaqbeh, G., Hayajneh, T., Vasilakos, A. V., and Mohd, B. J. (2014). Qos-aware health monitoring system using cloud-based wbans. *Journal of medical systems*, 38(10):1–20.

Alonso-Martin, F., Malfaz, M., Sequeira, J., Gorostiza, J. F., and Salichs, M. A. (2013). A multimodal emotion detection system during human–robot interaction. *Sensors*, 13(11):15549–15581.

Alsheikh, M. A., Jiao, Y., Niyato, D., Wang, P., Leong, D., and Han, Z. (2017a). The accuracy-privacy trade-off of mobile crowdsensing. *IEEE Communications Magazine*, 55(6):132–139.

Alsheikh, M. A., Niyato, D., Leong, D., Wang, P., and Han, Z. (2017b). Privacy management and optimal pricing in people-centric sensing. *IEEE Journal on Selected Areas in Communications*, 35(4):906–920.

Andone, I., Błaszkiewicz, K., Eibes, M., Trendafilov, B., Montag, C., and Markowetz, A. (2016). How age and gender affect smartphone usage. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: adjunct*, pages 9–12.

Appelhans, B. M. and Luecken, L. J. (2006). Heart rate variability as an index of regulated emotional responding. *Review of general psychology*, 10(3):229–240.

Armando, N., Almeida, R., Fernandes, J. M., Silva, J. S., and Boavida, F. (2021). End-to-end experimentation of a 5g vertical within the scope of blended learning. *Discover Internet of Things*, 1(1):1–12. Springer.

Armando, N., Fernandes, J., Sinche, S., Raposo, D., Silva, J. S., and Boavida, F. (2019). A unified solution for iot device management. In *2019 22nd International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pages 1–6. IEEE.

Armando, N., Fernandes, J. M., Rodrigues, A., Silva, J. S., and Boavida, F. (2020). Exploring approaches to the management of physical, virtual, and social sensors. In *IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 954–959. IEEE.

Armando, N., Raposo, D., Fernandes, M., Rodrigues, A., Silva, J. S., and Boavida, F. (2017). Wsns in fiware–towards the development of people-centric applications. In *International conference on practical applications of agents and multi-agent systems*, pages 445–456. Springer.

Armando, N., Rodrigues, A., Pereira, V., Sá Silva, J., and Boavida, F. (2018). An outlook on physical and virtual sensors for a socially interactive internet. *Sensors*, 18(8):2578.

Arora, S., Ge, R., Halpern, Y., Mimno, D., Moitra, A., Sontag, D., Wu, Y., and Zhu, M. (2018). Learning Topic Models-Provably and Efficiently. *COMMUNICATIONS OF THE ACM*, 61(4).

Ashton, K. et al. (2009). That 'internet of things' thing. *RFID journal*, 22(7):97–114.

Atalay, A., Atalay, O., Husain, M. D., Fernando, A., and Potluri, P. (2017). Piezofilm yarn sensor-integrated knitted fabric for healthcare applications. *Journal of Industrial Textiles*, 47(4):505–521.

Bachiller, R., Matthys, N., del Cid, J., Joosen, W., Hughes, D., and Van Laerhoven, K. (2015). @ migo: A comprehensive middleware solution for participatory sensing applications. In *2015 IEEE 14th International Symposium on Network Computing and Applications*, pages 1–8. IEEE.

Benedetto, S., Caldato, C., Bazzan, E., Greenwood, D. C., Pensabene, V., and Actis, P. (2018). Assessment of the fitbit charge 2 for monitoring heart rate. *PloS one*, 13(2):e0192691.

Bhas, N. (2013). Smart wearable devices: Fitness healthcare entertainment and enterprise 2013–2018. *Juniper Res., Basingstoke, UK, Tech. Rep.*

Bin, S., Yuan, L., and Xiaoyi, W. (2010). Research on data mining models for the internet of things. In *2010 International Conference on Image Analysis and Signal Processing*, pages 127–132. IEEE.

Bonato, P. (2003). Wearable sensors/systems and their impact on biomedical engineering. *IEEE Engineering in Medicine and Biology Magazine*, 22(3):18–20.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Burke, K. (2019). ”how many text do people send every day?”. https://www.textrequest.com/blog/how-many-texts-people-send-per-day/. Last visited: 2019-09-18.

Buzzi, C., Tucci, M., Ciprandi, R., Brambilla, I., Caimmi, S., Ciprandi, G., and Marseglia, G. L. (2020). The psycho-social effects of covid-19 on italian adolescents' attitudes and behaviors. *Italian journal of pediatrics*, 46(1):1–7.

Bynion, T.-M. and Feldner, M. T. (2020). *Self-Assessment Manikin*, pages 4654–4656. Springer International Publishing, Cham.

Cambria, E. (2016). Affective computing and sentiment analysis. *IEEE intelligent systems*, 31(2):102–107.

Cao, L., Wang, Y., Zhang, B., Jin, Q., and Vasilakos, A. V. (2018). Gchar: An efficient group-based context—aware human activity recognition on smartphone. *Journal of Parallel and Distributed Computing*, 118:67–80.

Caputo, E. L. and Reichert, F. F. (2020). Studies of physical activity and covid-19 during the pandemic: a scoping review. *Journal of Physical Activity and Health*, 17(12):1275–1284.

Carver, C. S. and Scheier, M. F. (1981). The self-attention-induced feedback loop and social facilitation. *Journal of Experimental Social Psychology*, 17(6):545–568.

Center, P. R. (2010). Cell phones and american adults | pew research center. https://www.pewresearch.org/internet/2010/09/02/cell-phones-and-american-adults/. Last visited: 2022-05-19.

Chakladar, D. D. and Chakraborty, S. (2018). Eeg based emotion classification using "correlation based subset selection". *Biologically inspired cognitive architectures*, 24:98–106.

Chee, Y., Han, J., Youn, J., and Park, K. (2005). Air mattress sensor system with balancing tube for unconstrained measurement of respiration and heart beat movements. *Physiological measurement*, 26(4):413.

Cheek, J. M. and Buss, A. H. (1981). Shyness and sociability. *Journal of personality and social psychology*, 41(2):330.

Chen, L.-l., Zhao, Y., Ye, P.-f., Zhang, J., and Zou, J.-z. (2017). Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers. *Expert Systems with Applications*, 85:279–291.

Chen, M., Gonzalez, S., Vasilakos, A., Cao, H., and Leung, V. C. (2011). Body area networks: A survey. *Mobile networks and applications*, 16(2):171–193.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Chen, Z., Lin, M., Chen, F., Lane, N. D., Cardone, G., Wang, R., Li, T., Chen, Y., Choudhury, T., and Campbell, A. T. (2013). Unobtrusive sleep monitoring using smartphones. In *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*, pages 145–152. IEEE.

Clavel, C., Vasilescu, I., Devillers, L., Richard, G., and Ehrette, T. (2008). Fear-type emotion recognition for future audio-based surveillance systems. *Speech Communication*, 50(6):487–503.

Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27.

covid19pt data (2022). dssg-pt/covid19pt-data: dados relativos à pandemia covid-19 em portugal. `https://github.com/dssg-pt/covid19pt-data`. Last visited: 2022-07-04.

Curcio, G., Ferrara, M., and De Gennaro, L. (2006). Sleep loss, learning capacity and academic performance. *Sleep medicine reviews*, 10(5):323–337.

de Internet de Nueva Generación, G. (2022). ging/fiware-idm: Oauth 2.0-based authentication of users and devices, user profile management, single sign-on (sso) and identity federation across multiple administration domains. `https://github.com/ging/fiware-idm`. Last visited: 2022-07-20.

De Silva, L. C., Miyasato, T., and Nakatsu, R. (1997). Facial emotion recognition using multi-modal information. In *Proceedings of ICICS, 1997 International Conference on Information, Communications and Signal Processing. Theme: Trends in Information Systems Engineering and Wireless Multimedia Communications (Cat.*, volume 1, pages 397–401. IEEE.

DiClemente, C. C., Marinilli, A. S., Singh, M., and Bellino, L. E. (2001). The role of feedback in the process of health behavior change. *American journal of health behavior*, 25(3):217–227.

Dillman, D. A., Sinclair, M. D., and Clark, J. R. (1993). Effects of questionnaire length, respondent-friendly design, and a difficult question on response rates for occupant-addressed census mail surveys. *Public opinion quarterly*, 57(3):289–304.

Doshi, N. (2018). Recommendation systems — models and evaluation | towards data science. `https://towardsdatascience.com/recommendation-systems-models-and-evaluation-84944a84fb8e`. Last visited: 2022-07-20.

Du, R., Santi, P., Xiao, M., Vasilakos, A. V., and Fischione, C. (2018). The sensable city: A survey on the deployment and management for smart city monitoring. *IEEE Communications Surveys & Tutorials*, 21(2):1533–1560.

Eisenstadt, M., Liverpool, S., Infanti, E., Ciuvat, R. M., Carlsson, C., et al. (2021). Mobile apps that promote emotion regulation, positive mental health, and well-being in the general population: Systematic review and meta-analysis. *JMIR mental health*, 8(11):e31170.

ENISA (2018). Privacy and data protection in mobile applications — enisa. `https://www.enisa.europa.eu/publications/privacy-and-data-protection-in-mobile-applications`. Last visited: 2022-07-20.

Eskes, P., Spruit, M., Brinkkemper, S., Vorstman, J., and Kas, M. J. (2016a). The sociability score: App-based social profiling from a healthcare perspective. *Computers in Human Behavior*, 59:39–48.

Eskes, P., Spruit, M., Brinkkemper, S., Vorstman, J., and Kas, M. J. (2016b). The sociability score: App-based social profiling from a healthcare perspective. *Computers in Human Behavior*, 59:39–48.

Fazio, M., Celesti, A., Márquez, F. G., Glikson, A., and Villari, M. (2015). Exploiting the fiware cloud platform to develop a remote patient monitoring system. In *2015 IEEE symposium on computers and communication (ISCC)*, pages 264–270. IEEE.

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4):82–89.

Fernandes, J., Raposo, D., Armando, N., Sinche, S., Silva, J. S., Rodrigues, A., Pereira, V., and Boavida, F. (2019a). An integrated approach to human-in-the-loop systems and online social sensing. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 478–483. IEEE.

Fernandes, J., Raposo, D., Armando, N., Sinche, S., Silva, J. S., Rodrigues, A., Pereira, V., Oliveira, H. G., Macedo, L., and Boavida, F. (2020). Isabela–a socially-aware human-in-the-loop advisor system. *Online Social Networks and Media*, 16:100060. Elsevier.

Fernandes, J., Raposo, D., Sinche, S., Armando, N., Silva, J. S., Rodrigues, A., Macedo, L., Oliveira, H. G., and Boavida, F. (2019b). A human-in-the-loop cyber-physical approach for students performance assessment. In *Proceedings of the Fourth International Workshop on Social Sensing*, pages 36–42. ACM New York, NY.

Fernandes, J., Silva, J. S., Rodrigues, A., Boavida, F., Gaspar, R., Godinho, C., and Francisco, R. (2022a). Social sensing and human in the loop profiling during pandemics: the vitoria application.

Fernandes, J., Silva, J. S., Rodrigues, A., Sinche, S., and Boavida, F. (2022b). Automatically assessing students performance with smartphone data.

Fernandes, J. M., Silva, J. S., Rodrigues, A., and Boavida, F. (2022c). A survey of approaches to unobtrusive sensing of humans. *ACM Computing Surveys (CSUR)*, 55(2):1–28. ACM New York, NY.

Fernández-Caballero, A., Castillo, J. C., López, M. T., Serrano-Cuerda, J., and Sokolova, M. V. (2013). Int3-horus framework for multispectrum activity interpretation in intelligent environments. *Expert Systems with Applications*, 40(17):6715–6727.

Fernández-Caballero, A., Martínez-Rodrigo, A., Pastor, J. M., Castillo, J. C., Lozano-Monasor, E., López, M. T., Zangróniz, R., Latorre, J. M., and Fernández-Sotos, A. (2016). Smart environment architecture for emotion detection and regulation. *Journal of biomedical informatics*, 64:55–73.

Fiol-DeRoque, M. A., Serrano-Ripoll, M. J., Jiménez, R., Zamanillo-Campos, R., Yáñez-Juan, A. M., Bennasar-Veny, M., Leiva, A., Gervilla, E., García-Buades, M. E., García-Toro, M., et al. (2021). A mobile phone–based intervention to reduce mental health problems in health care workers during the covid-19 pandemic (psycovidapp): randomized controlled trial. *JMIR mHealth and uHealth*, 9(5):e27039.

Fitbit (2022). Fitbit official site for activity trackers and more. `https://www.fitbit.com/global/eu/home`. Last visited: 2022-07-20.

Flores, M. F. (2019). Understanding the challenges of remote working and it's impact to workers. *International Journal of Business Marketing and Management (IJBMM)*, 4(11):40–44.

Foundation, S. (2022). How much sleep do we really need? | sleep foundation. `https://www.sleepfoundation.org/how-sleep-works/how-much-sleep-do-we-really-need`. Last visited: 2022-07-20.

Francisco, R., Pedro, M., Delvecchio, E., Espada, J. P., Morales, A., Mazzeschi, C., and Orgilés, M. (2020). Psychological symptoms and behavioral changes in children and adolescents during the early phase of covid-19 quarantine in three european countries. *Frontiers in Psychiatry*, page 1329.

Frison, E. and Eggermont, S. (2017). Browsing, posting, and liking on instagram: The reciprocal relationships between different types of instagram use and adolescents' depressed mood. *Cyberpsychology, Behavior, and Social Networking*, 20(10):603–609.

Garbey, M., Sun, N., Merla, A., and Pavlidis, I. (2007). Contact-free measurement of cardiac pulse based on the analysis of thermal imagery. *IEEE transactions on Biomedical Engineering*, 54(8):1418–1426.

Giridhar, P., Wang, S., Abdelzaher, T., Al Amin, T., and Kaplan, L. (2017). Social fusion: Integrating twitter and instagram for event monitoring. In *2017 IEEE International Conference on Autonomic Computing (ICAC)*, pages 1–10. IEEE.

Google, L. (2022a). Activity recognition api | google developers. `https://developers.google.com/location-context/activity-recognition/`. Last visited: 2022-07-20.

Google, L. (2022b). android.hardware | android developers. `https://developer.android.com/reference/android/hardware/package-summary`. Last visited: 2022-07-20.

Google, L. (2022c). Bluetoothclass.device | android developers. `https://developer.android.com/reference/android/bluetooth/BluetoothClass.Device`. Last visited: 2022-07-20.

Google, L. (2022d). Locationrequest | google play services | google developers. `https://developers.google.com/android/reference/com/google/android/gms/location/LocationRequest.html`. Last visited: 2022-07-20.

Google, L. (2022e). Sleep api | google developers. `https://developers.google.com/location-context/sleep`. Last visited: 2022-07-20.

Greenberg, N. (2020). Mental health of health-care workers in the covid-19 era. *Nature Reviews Nephrology*, 16(8):425–426.

Gu, Y., Liu, T., Li, J., Ren, F., Liu, Z., Wang, X., and Li, P. (2018). Emosense: Data-driven emotion sensing via off-the-shelf wifi devices. In *2018 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE.

Guardian, T. (2021). Portugal banned bosses from texting employees after work. could it happen in the us? | life and style. `https://www.theguardian.com/lifeandstyle/2021/nov/15/portugal-boss-texts-work-us-employment`. Last visited: 2022-04-04.

Haleem, A., Javaid, M., and Vaishya, R. (2020). Effects of covid-19 pandemic in daily life. *Current medicine research and practice*, 10(2):78.

Hall, D. L. and Llinas, J. (1997). An introduction to multisensor data fusion. *Proceedings of the IEEE*, 85(1):6–23.

Halperin, D., Hu, W., Sheth, A., and Wetherall, D. (2011a). Tool release: Gathering 802.11 n traces with channel state information. *ACM SIGCOMM computer communication review*, 41(1):53–53.

Halperin, D., Hu, W., Sheth, A., and Wetherall, D. (2011b). Tool release: Gathering 802.11n traces with channel state information. *ACM SIGCOMM CCR*, 41(1):53.

Han, K.-H. and Bae, W.-S. (2016). Proposing and verifying a security-enhanced protocol for iot-based communication for medical devices. *Cluster Computing*, 19(4):2335–2341.

Harari, G. M., Gosling, S. D., Wang, R., Chen, F., Chen, Z., and Campbell, A. T. (2017). Patterns of behavior change in students over an academic term: A preliminary study of activity and sociability behaviors using smartphone sensing methods. *Computers in Human Behavior*, 67:129–138.

Hastie, T., Rosset, S., Zhu, J., and Zou, H. (2009). Multi-class adaboost. *Statistics and its Interface*, 2(3):349–360.

Heaton, J. (2015). Encog: Library of interchangeable machine learning models for java and c. *Journal of Machine Learning Research*, 16:1243–1247.

Helliwell, J. F. and Wang, S. (2014). Weekends and subjective well-being. *Social indicators research*, 116(2):389–407.

Hershfield, H. E., Scheibe, S., Sims, T. L., and Carstensen, L. L. (2013). When feeling bad can be good: Mixed emotions benefit physical health across adulthood. *Social psychological and personality science*, 4(1):54–61.

Hijazi, S. T. and Naqvi, S. (2006). Factors affecting students' performance. *Bangladesh e-journal of Sociology*, 3(1).

Hillyard, P., Luong, A., Abrar, A. S., Patwari, N., Sundar, K., Farney, R., Burch, J., Porucznik, C., and Pollard, S. H. (2018). Experience: Cross-technology radio respiratory monitoring performance study. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, pages 487–496.

Hong, L. and Davison, B. D. (2010). Empirical study of topic modeling in twitter. In *Proceedings of the first workshop on social media analytics*, pages 80–88.

Hu, M., Zhai, G., Li, D., Fan, Y., Duan, H., Zhu, W., and Yang, X. (2018). Combination of near-infrared and thermal imaging techniques for the remote and simultaneous measurements of breathing and heart rates under sleep situation. *PloS one*, 13(1):e0190466.

Huang, K.-Y., Wu, C.-H., Hong, Q.-B., Su, M.-H., and Chen, Y.-H. (2019a). Speech emotion recognition using deep neural network considering verbal and nonverbal speech sounds. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5866–5870. IEEE.

Huang, S., Wang, D., Zhao, R., and Zhang, Q. (2019b). Wiga: A wifi-based contactless activity sequence recognition system based on deep learning. In *2019 15th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN)*, pages 69–74. IEEE.

Hung, C.-C., Ananthanarayanan, G., Bodik, P., Golubchik, L., Yu, M., Bahl, P., and Philipose, M. (2018). Videoedge: Processing camera streams using hierarchical clusters. In *2018 IEEE/ACM Symposium on Edge Computing (SEC)*, pages 115–131. IEEE.

Hutto, C. and Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225.

I. Socionext America "Socionext CMOS 24-GHz Radar Sensor" (Sunnyvale, C. (Last visited: 2018-10-10).

Inc., A. (2022). Featured | apple developer documentation. `https://developer.apple.com/documentation/`. Last visited: 2022-07-20.

innovative XENSIV™ 60 GHz radar chip enables things to see, I. and revolutionizes the Human Machine Interface (Last visited: 2018-10-10).

Isinkaye, F. O., Folajimi, Y. O., and Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian informatics journal*, 16(3):261–273.

Jahrer, M., Töscher, A., and Legenstein, R. (2010). Combining predictions for accurate recommender systems. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 693–702.

Jawbone (2018). Last visited: 2022-07-20.

Jaworski, B. K., Taylor, K., Ramsey, K. M., Heinz, A., Steinmetz, S., Pagano, I., Moraja, G., and Owen, J. E. (2021). Exploring usage of covid coach, a public mental health app designed for the covid-19 pandemic: evaluation of analytics data. *Journal of medical Internet research*, 23(3):e26559.

Jia, Z., Alaziz, M., Chi, X., Howard, R. E., Zhang, Y., Zhang, P., Trappe, W., Sivasubramaniam, A., and An, N. (2016). Hb-phone: a bed-mounted geophone-based heartbeat monitoring system. In *2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pages 1–12. IEEE.

Jia, Z., Bonde, A., Li, S., Xu, C., Wang, J., Zhang, Y., Howard, R. E., and Zhang, P. (2017). Monitoring a person's heart rate and respiratory rate on a shared bed using geophones. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, pages 1–14.

Kaltiokallio, O., Yiğitler, H., Jäntti, R., and Patwari, N. (2014). Non-invasive respiration rate monitoring using a single cots tx-rx pair. In *IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*, pages 59–69. IEEE.

Kefayati, M. H., Pourahmadi, V., and Aghaeinia, H. (2020). Wi2vi: Generating video frames from wifi csi samples. *IEEE Sensors Journal*, 20(19):11463–11473.

Khaitan, S. K. and McCalley, J. D. (2015). Design techniques and applications of cyberphysical systems: A survey. *IEEE Systems Journal*.

Kim, J. and André, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE transactions on pattern analysis and machine intelligence*, 30(12):2067–2083.

Klaessens, J. H., Van Den Born, M., Van Der Veen, A., Sikkens-Van De Kraats, J., van den Dungen, F. A., and Verdaasdonk, R. M. (2014). Development of a baby friendly non-contact method for measuring vital signs: first results of clinical measurements in an open incubator at a neonatal intensive care unit. In *Advanced Biomedical and Clinical Diagnostic Systems XII*, volume 8935, page 89351P. International Society for Optics and Photonics.

Kondylakis, H., Katehakis, D. G., Kouroubali, A., Logothetidis, F., Triantafyllidis, A., Kalamaras, I., Votis, K., Tzovaras, D., et al. (2020). Covid-19 mobile apps: a systematic review of the literature. *Journal of medical Internet research*, 22(12):e23170.

Krčo, S., Pokrić, B., and Carrez, F. (2014). Designing iot architecture (s): A european perspective. In *2014 IEEE world forum on internet of things (WF-IoT)*, pages 79–84. IEEE.

Kwon, S., Kim, H., and Park, K. S. (2012). Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2174–2177. IEEE.

Kwon, S., Kim, J., Lee, D., and Park, K. (2015). Roi analysis for remote photoplethysmography on facial video. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4938–4941. IEEE.

Lane, N. D., Lin, M., Mohammod, M., Yang, X., Lu, H., Cardone, G., Ali, S., Doryab, A., Berke, E., Campbell, A. T., et al. (2014). Bewell: Sensing sleep, physical activities and social interactions to promote wellbeing. *Mobile Networks and Applications*, 19(3):345–359.

Lane, R. D., McRae, K., Reiman, E. M., Chen, K., Ahern, G. L., and Thayer, J. F. (2009). Neural correlates of heart rate variability during emotion. *Neuroimage*, 44(1):213–222.

Ledger, D. and McCaffrey, D. (2014). Inside wearables: How the science of human behavior change offers the secret to long-term engagement. *Endeavour Partners*, 200(93):1.

Lee, C.-H., Yang, H.-C., and Lin, S.-J. (2015a). Incorporating big data and social sensors in a novel early warning system of dengue outbreaks. In *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 1428–1433. IEEE.

Lee, Y.-D. and Chung, W.-Y. (2009). Wireless sensor network based wearable smart shirt for ubiquitous health and activity monitoring. *Sensors and Actuators B: Chemical*, 140(2):390–395.

Lee, Y. S., Pathirana, P. N., Evans, R. J., and Steinfort, C. L. (2015b). Noncontact detection and analysis of respiratory function using microwave doppler radar. *Journal of Sensors*, 2015.

Li, X., Chen, J., Zhao, G., and Pietikainen, M. (2014). Remote heart rate measurement from face videos under realistic situations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4264–4271.

Limone, P. and Toto, G. A. (2021). Psychological and emotional effects of digital technology on children in covid-19 pandemic. *Brain Sciences*, 11(9):1126.

Liu, J., Shen, H., Narman, H. S., Chung, W., and Lin, Z. (2018). A survey of mobile crowdsensing techniques: A critical component for the internet of things. *ACM Transactions on Cyber-Physical Systems*, 2(3):1–26.

Liu, J., Wang, Y., Chen, Y., Yang, J., Chen, X., and Cheng, J. (2015). Tracking vital signs during sleep leveraging off-the-shelf wifi. In *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pages 267–276.

Liu, Y., Wang, T., Jiang, Y., and Chen, B. (2020). Harvesting ambient rf for presence detection through deep learning. *IEEE Transactions on Neural Networks and Learning Systems*.

Lu, G., Yang, F., Tian, Y., Jing, X., and Wang, J. (2009). Contact-free measurement of heart rate variability via a microwave sensor. *Sensors*, 9(12):9572–9581.

Maia, B. and Santos, D. (2018). Language, emotion, and the emotions: The multidisciplinary and linguistic background. *Language and Linguistics compass*, 12(6):e12280.

Mammone, A., Turchi, M., and Cristianini, N. (2009). Support vector machines. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(3):283–289.

Maps, O. S. (2022). Reverse - nominatim 4.0.1. `https://nominatim.org/release-docs/latest/api/Reverse/`. Last visited: 2022-07-20.

Massaroni, C., Lopes, D. S., Lo Presti, D., Schena, E., and Silvestri, S. (2018). Contactless monitoring of breathing patterns and respiratory rate at the pit of the neck: A single camera approach. *Journal of Sensors*, 2018.

McGilloway, S., Cowie, R., Douglas-Cowie, E., Gielen, S., Westerdijk, M., and Stroeve, S. (2000). Approaching automatic recognition of emotion from voice: A rough benchmark. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*.

Meng, W., Choo, K.-K. R., Furnell, S., Vasilakos, A. V., and Probst, C. W. (2018). Towards bayesian-based trust management for insider attacks in healthcare software-defined networks. *IEEE Transactions on Network and Service Management*, 15(2):761–773.

Microsoft (2022). Xamarin | open-source mobile app platform for .net. `https://dotnet.microsoft.com/en-us/apps/xamarin`. Last visited: 2022-07-20.

Min, J.-K., Doryab, A., Wiese, J., Amini, S., Zimmerman, J., and Hong, J. I. (2014). Toss'n'turn: smartphone as sleep and sleep quality detector. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 477–486.

Miyamoto, K., Hashimoto, K., Kasaoka, M., and Kakumu, M. (2020). Wearable sensors corresponding to various applications in medical/healthcare field. In

2020 27th International Workshop on Active-Matrix Flatpanel Displays and Devices (AM-FPD), pages 115–118. IEEE.

Mo, H., Ding, S., Yang, S., Zheng, X., and Vasilakos, A. (2020). The role of edge robotics as-a-service in monitoring covid-19 infection.

Mobvoi (2022). Ticwatch smartwatch and smart products. `https://www.mobvoi.com/eu/products/ticwatchs2`. Last visited: 2022-07-04.

Mohanty, S. P., Choppali, U., and Kougianos, E. (2016). Everything you wanted to know about smart cities: The internet of things is the backbone. *IEEE Consumer Electronics Magazine*, 5(3):60–70.

Monarch, R. M. (2021). *Human-in-the-Loop Machine Learning: Active learning and annotation for human-centered AI.* Simon and Schuster.

Nandakumar, R., Gollakota, S., and Watson, N. (2015). Contactless sleep apnea detection on smartphones. In *Proceedings of the 13th annual international conference on mobile systems, applications, and services*, pages 45–57.

Nepal, S., Wang, W., Vojdanovski, V., Huckins, J. F., daSilva, A., Meyer, M., and Campbell, A. (2022). Covid student study: A year in the life of college students during the covid-19 pandemic through the lens of mobile phone sensing. In *CHI Conference on Human Factors in Computing Systems*, pages 1–19.

Ni, J., Zhang, K., and Vasilakos, A. V. (2020). Security and privacy for mobile edge caching: challenges and solutions. *IEEE Wireless Communications*.

Nunes, D. S., Zhang, P., and Silva, J. S. (2015). A survey on human-in-the-loop applications towards an internet of all. *IEEE Communications Surveys & Tutorials*, 17(2):944–965.

on Smart Systems Integration, E. T. E. T. P. (2013). ”eposs response to the self assessment exercise launched by the european commission for renewed recognition as european technology platform”.

OpenWrt (2022). [openwrt wiki] welcome to the openwrt project. `https://openwrt.org/`. Last visited: 2022-07-20.

Organization, W. H. (2011). Physical activity recommendations who. `https://www.who.int/dietphysicalactivity/physical-activity-recommendations-18-64years.pdf`. Last visited: 2019-09-19.

Osmanbegovic, E. and Suljic, M. (2012). Data mining approach for predicting student performance. *Economic Review: Journal of Economics and Business*, 10(1):3–12.

Ozdalga, E., Ozdalga, A., and Ahuja, N. (2012). The smartphone in medicine: a review of current and potential use among physicians and students. *Journal of medical Internet research*, 14(5):e128.

O'Connor, A. J. and Jahan, F. (2014). Under surveillance and overwrought: American muslims' emotional and behavioral responses to government surveillance. *Journal of Muslim Mental Health*, 8(1).

Paar, C. and Pelzl, J. (2009). *Understanding cryptography: a textbook for students and practitioners*. Springer Science & Business Media.

Pareek, P. and Thakkar, A. (2021). A survey on video-based human action recognition: recent updates, datasets, challenges, and applications. *Artificial Intelligence Review*, 54(3):2259–2322.

Park, N. and Lee, H. (2012). Social Implications of Smartphone Use: Korean College Students' Smartphone Use and Psychological Well-Being. *Cyberpsychology, Behavior, and Social Networking*, 15(9):491–497.

Parkka, J., Ermes, M., Korpipaa, P., Mantyjarvi, J., Peltola, J., and Korhonen, I. (2006). Activity classification using realistic data from wearable sensors. *IEEE Transactions on information technology in biomedicine*, 10(1):119–128.

Patwari, N., Brewer, L., Tate, Q., Kaltiokallio, O., and Bocca, M. (2013). Breathfinding: A wireless network that monitors and locates breathing in a home. *IEEE Journal of Selected Topics in Signal Processing*, 8(1):30–42.

Pereira, C. B., Yu, X., Czaplik, M., Rossaint, R., Blazek, V., and Leonhardt, S. (2015). Remote monitoring of breathing dynamics using infrared thermography. *Biomedical optics express*, 6(11):4378–4394.

Perera, C., Zaslavsky, A., Christen, P., and Georgakopoulos, D. (2014). Context aware computing for the internet of things: A survey. *IEEE Communications Surveys and Tutorials*.

Pérez-Fuentes, M. D. C., Molero Jurado, M. d. M., Oropesa Ruiz, N. F., Martos Martínez, Á., Simón Márquez, M. d. M., Herrera-Peco, I., and Gázquez Linares, J. J. (2020). Questionnaire on perception of threat from covid-19. *Journal of Clinical Medicine*, 9(4):1196.

Pierce, A. (2017). Walabot diy can see into walls. *Tech Directions*, 76(5):8.

Piwek, L., Ellis, D. A., Andrews, S., and Joinson, A. (2016). The rise of consumer health wearables: promises and barriers. *PLoS medicine*, 13(2):e1001953.

Posner, J., Russell, J. A., and Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology*, 17(3):715.

Psomakelis, E., Aisopos, F., Litke, A., Tserpes, K., Kardara, M., and Campo, P. M. (2016). Big iot and social networking data for smart cities: Algorithmic improvements on big data analysis in the context of radical city applications. *arXiv preprint arXiv:1607.00509*.

PyPy (2020). googletrans · pypi. `https://pypi.org/project/googletrans/`. Last visited: 2022-07-20.

Quinlan, J. R. (1996). Learning decision tree classifiers. *ACM Computing Surveys (CSUR)*, 28(1):71–72.

Radesky, J. S., Schumacher, J., and Zuckerman, B. (2015). Mobile and interactive media use by young children: the good, the bad, and the unknown. *Pediatrics*, 135(1):1–3.

RE450, T.-L. (2022). Re450 | ac1750 wi-fi range extender | tp-link portugal. `https://www.tp-link.com/pt/home-networking/range-extender/re450/`. Last visited: 2022-07-20.

Reisenzein, R. (2010). Broadening the scope of affect detection research. *IEEE Transactions on Affective Computing*, 1(1):42–45.

Ren, Y., Wang, C., Chen, Y., Yang, J., and Li, H. (2019). Noninvasive fine-grained sleep monitoring leveraging smartphones. *IEEE Internet of Things Journal*, 6(5):8248–8261.

Ren, Y., Wang, C., Yang, J., and Chen, Y. (2015). Fine-grained sleep monitoring: Hearing your breathing with smartphones. In *2015 IEEE Conference on Computer Communications (INFOCOM)*, pages 1194–1202. IEEE.

Rish, I. et al. (2001). An empirical study of the naive bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, volume 3, pages 41–46.

Rivadeneira, J. E., Silva, J. S., Colomo-Palacios, R., Rodrigues, A., Fernandes, J. M., and Boavida, F. (2021). A privacy-aware framework integration into a human-in-the-loop iot system. In *IEEE INFOCOM 2021-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 1–6. IEEE.

Rodrigues, D., Prada, M., Gaspar, R., Garrido, M. V., and Lopes, D. (2018). Lisbon emoji and emoticon database (leed): Norms for emoji and emoticons in seven evaluative dimensions. *Behavior research methods*, 50(1):392–405.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161.

Sabharwal, N. and Agrawal, A. (2020). Introduction to google dialogflow. In *Cognitive virtual assistants using Google Dialogflow*, pages 13–54. Springer.

Salovey, P., Rothman, A. J., Detweiler, J. B., and Steward, W. T. (2000). Emotional states and physical health. *American psychologist*, 55(1):110.

Sanchez, O. T., Fernandes, J. M., Rodrigues, A., Silva, J. S., Boavida, F., Rivadeneira, J. E., de Lemos, A. V., and Raposo, D. (2022). Green bear - a lorawan-based human-in-the-loop case-study for sustainable cities. *Pervasive and Mobile Computing*, page 101701.

Sánchez-Rodríguez, D., Quintana-Suárez, M. A., Alonso-González, I., Ley-Bosch, C., and Sánchez-Medina, J. J. (2020). Fusion of channel state information and received signal strength for indoor localization using a single access point. *Remote Sensing*, 12(12):1995.

Sandoval-Palis, I., Naranjo, D., Vidal, J., and Gilar-Corbi, R. (2020). Early dropout prediction model: A case study of university leveling course students. *Sustainability*, 12(22):9314.

Schollz, Z. (2018). schollz/find3: High-precision indoor positioning framework, version 3. `https://github.com/schollz/find3`. Last visited: 2022-07-20.

Scully, C. G., Lee, J., Meyer, J., Gorbach, A. M., Granquist-Fraser, D., Mendelson, Y., and Chon, K. H. (2011). Physiological parameter monitoring from optical recordings with a mobile phone. *IEEE Transactions on Biomedical Engineering*, 59(2):303–306.

Shiffman, S., Stone, A. A., and Hufford, M. R. (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol.*, 4:1–32.

Silva, M. J., Carvalho, P., and Sarmento, L. (2012). Building a sentiment lexicon for social judgement mining. In *International Conference on Computational Processing of the Portuguese Language*, pages 218–228. Springer.

Sinche, S., Hidalgo, P., Fernandes, J. M., Raposo, D., Silva, J. S., Rodrigues, A., Armando, N., and Boavida, F. (2020). Analysis of student academic performance using human-in-the-loop cyber-physical systems. In *Telecom*, volume 1, pages 18–31. MDPI.

Sinche, S., Polo, O., Raposo, D., Femandes, M., Boavida, F., Rodrigues, A., Pereira, V., and Silva, J. S. (2018). Assessing redundancy models for iot reliability. In *2018 IEEE 19th International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)*, pages 14–15. IEEE.

sm 24, G. (Last visited: 2018-10-10).

Soares, A. P., Comesaña, M., Pinheiro, A. P., Simões, A., and Frade, C. S. (2012). The adaptation of the affective norms for english words (anew) for european portuguese. *Behavior research methods*, 44(1):256–269.

Soltanaghaei, E., Sharma, R. A., Wang, Z., Chittilappilly, A., Luong, A., Giler, E., Hall, K., Elias, S., and Rowe, A. (2020). Robust and practical wifi human sensing using on-device learning with a domain adaptive model. In *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, pages 150–159.

Sousa Nunes, D. S., Zhang, P., and Sa Silva, J. (2015). A Survey on human-in-The-loop applications towards an internet of all. *IEEE Communications Surveys and Tutorials.*

Statista (2022). Iot connected devices worldwide 2019-2030. https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/. Last visited: 2022-05-19.

StayWayCovid (2022). Stayaway covid. https://stayawaycovid.pt/landing-page/. Last visited: 2022-05-19.

Strelan, P., Osborn, A., and Palmer, E. (2020). The flipped classroom: A meta-analysis of effects on student performance across disciplines and education levels. *Educational Research Review*, 30:100314.

Taelman, J., Vandeput, S., Spaepen, A., and Van Huffel, S. (2009). Influence of mental stress on heart rate and heart rate variability. In *4th European conference of the international federation for medical and biological engineering*, pages 1366–1369. Springer.

Talevi, D., Socci, V., Carai, M., Carnaghi, G., Faleri, S., Trebbi, E., di Bernardo, A., Capelli, F., and Pacitti, F. (2020). Mental health outcomes of the covid-19 pandemic. *Rivista di psichiatria*, 55(3):137–144.

Tapia, E. M., Intille, S. S., Haskell, W., Larson, K., Wright, J., King, A., and Friedman, R. (2007). Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. In *2007 11th IEEE international symposium on wearable computers*, pages 37–40. IEEE.

Taras, H. and Potts-Datema, W. (2005). Sleep and student performance at school. *Journal of school health*, 75(7):248–254.

Telefonica, S. (2022a). telefonicaid/fiware-cygnus: A connector in charge of persisting context data sources into other third-party databases and storage systems, creating a historical view of the context. https://github.com/telefonicaid/fiware-cygnus. Last visited: 2022-07-20.

Telefonica, S. (2022b). telefonicaid/fiware-sth-comet: A component of the fiware ecosystem in charge of managing historical and aggregated time series context information. https://github.com/telefonicaid/fiware-sth-comet. Last visited: 2022-07-20.

Telefonica, S. (2022c). telefonicaid/iotagent-ul: Iot agent for a ultralight 2.0 based protocol (with http, mqtt and amqp transports). https://github.com/telefonicaid/iotagent-ul. Last visited: 2022-07-20.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Trockel, M. T., Barnes, M. D., and Egget, D. L. (2000). Health-related variables and academic performance among first-year college students: Implications for sleep and other behaviors. *Journal of American college health*, 49(3):125–131.

Vatsikas, S., Kalogridis, G., Lewis, T., and Sooriyabandara, M. (2017). The experience of using the ies cities citizen-centric iot platform. *IEEE Communications Magazine*, 55(2):40–47.

Verma, G. K. and Tiwary, U. S. (2017). Affect representation and recognition in 3d continuous valence–arousal–dominance space. *Multimedia Tools and Applications*, 76(2):2159–2183.

Vernier (2022a). Go direct® respiration belt. `https://www.vernier.com/product/go-direct-respiration-belt/`. Last visited: 2022-07-04.

Vernier (2022b). Go direct® respiration belt - vernier. `https://www.vernier.com/product/go-direct-respiration-belt/`. Last visited: 2022-07-20.

Villarroel, M., Guazzi, A., Jorge, J., Davis, S., Watkinson, P., Green, G., Shenvi, A., McCormick, K., and Tarassenko, L. (2014). Continuous non-contact vital sign monitoring in neonatal intensive care unit. *Healthcare technology letters*, 1(3):87–91.

Visakha, K. and Prakash, S. S. (2018). Detection and tracking of human beings in a video using haar classifier. In *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, pages 1–4. IEEE.

Wang, A., Nguyen, D., Sridhar, A. R., and Gollakota, S. (2021). Using smart speakers to contactlessly monitor heart rhythms. *Communications Biology*, 4(1):1–12.

Wang, D., Szymanski, B. K., Abdelzaher, T., Ji, H., and Kaplan, L. (2019). The age of social sensing. *Computer*, 52(1):36–45.

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., and Campbell, A. T. (2014). Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*, pages 3–14.

Wang, R., Harari, G., Hao, P., Zhou, X., and Campbell, A. T. (2015). Smartgpa: how smartphones can assess and predict academic performance of college students. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pages 295–306.

Wang, T., Zhang, D., Zheng, Y., Gu, T., Zhou, X., and Dorizzi, B. (2018). C-fmcw based contactless respiration detection using acoustic signal. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4):1–20.

Wang, X., Yang, C., and Mao, S. (2017). Tensorbeat: Tensor decomposition for monitoring multiperson breathing beats with commodity wifi. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 9(1):1–27.

Watanabe, K., Watanabe, T., Watanabe, H., Ando, H., Ishikawa, T., and Kobayashi, K. (2005). Noninvasive measurement of heartbeat, respiration, snoring and body movements of a subject in bed via a pneumatic method. *IEEE transactions on biomedical engineering*, 52(12):2100–2107.

Wild, T., Braun, V., and Viswanathan, H. (2021). Joint design of communication and sensing for beyond 5g and 6g systems. *IEEE Access*, 9:30845–30857.

Witten, I. H., Frank, E., Hall, M. A., Pal, C. J., and DATA, M. (2005). Practical machine learning tools and techniques. In *Data Mining*, volume 2.

Xie, Y., Li, Z., and Li, M. (2015). Precise power delay profiling with commodity wifi. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, MobiCom '15, page 53–64, New York, NY, USA. ACM.

Yang, S., Adeel, U., and McCann, J. (2015). Backpressure meets taxes: Faithful data collection in stochastic mobile phone sensing systems. In *2015 IEEE Conference on Computer Communications (INFOCOM)*, pages 1490–1498. IEEE.

Yazdani, A., Skodras, E., Fakotakis, N., and Ebrahimi, T. (2013a). Multimedia content analysis for emotional characterization of music video clips. *EURASIP Journal on Image and Video Processing*, 2013(1):1–10.

Yazdani, A., Skodras, E., Fakotakis, N., and Ebrahimi, T. (2013b). Multimedia content analysis for emotional characterization of music video clips. *EURASIP Journal on Image and Video Processing*, 2013.

Yin, Z., Zhao, M., Wang, Y., Yang, J., and Zhang, J. (2017). Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. *Computer methods and programs in biomedicine*, 140:93–110.

Yoshitomi, Y., Asada, T., Shimada, K., and Tabuse, M. (2011). Facial expression recognition of a speaker using vowel judgment and thermal image processing. *Artificial Life and Robotics*, 16(3):318–323.

Zenonos, A., Khan, A., Kalogridis, G., Vatsikas, S., Lewis, T., and Sooriyabandara, M. (2016). Healthyoffice: Mood recognition at work using smartphones and wearable sensors. In *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, pages 1–6. IEEE.

Zephyr (2022). Zephyr™ performance systems | performance monitoring technology. `https://www.zephyranywhere.com/`. Last visited: 2022-07-20.

Zhang, Q., Sun, H., Wu, X., and Zhong, H. (2019). Edge video analytics for public safety: A review. *Proceedings of the IEEE*, 107(8):1675–1696.

Zhao, M., Adib, F., and Katabi, D. (2016). Emotion recognition using wireless signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, pages 95–108.

Zou, H., Zhou, Y., Yang, J., and Spanos, C. J. (2018). Towards occupant activity driven smart buildings via wifi-enabled iot devices and deep learning. *Energy and Buildings*, 177:12–22.

Appendixes

A Vitoria system weekly feedback

Table 1: Weekly feedback for the Actuation groups

<table>
<tr><th rowspan="2">Control Group</th><th colspan="2">Active Feedback Group</th></tr>
<tr><th>Positive Reinforcement</th><th>Negative Reinforcement</th></tr>
<tr><td>- - - - -</td><td>If social proximity below 10 people in the last week:
”We estimate that in the last week you had contact with
<Number of Contacts> people, what is within
what is recommended.
Congratulations! Keep doing what you’re doing,
limiting social contacts to your household.”</td><td>If social proximity exceeds 10 people in the last week:
”We estimate that in the last week you had contact with
<Number of Contacts> people.
Because it is a greater number than what is recommended,
you are more exposed to the risk of infection by the coronavirus.
Limit your contacts to members of your household as much as possible.”</td></tr>
<tr><td>- - - - -</td><td>If the user had low mobility
(less than 1 hour outside your home area on average):
”We estimate that in the last week your trips outside
your area of residence had an average duration of less than 1h,
which is within what is recommended.
Congratulations! Please continue to stay at home except for essential travel.”</td><td>If the user had high mobility
(1h or more outside the home area on average):
”We estimate that in the last week your trips outside
your area of residence had an average duration of more than 1 hour.
Because it is a greater number than what is recommended,
you are more exposed to the risk of infection by the coronavirus.
Try to stay at home as much as possible, except for essential trips.”</td></tr>
<tr><td>”Learn more about the specific measures
recommended for the municipality
where you live at:
https://covid19estamoson.gov.pt/”</td><td colspan="2">”The risk level in the municipality where you live is
<Municipality Risk Level>. Learn more about the specific measures recommended
for the municipality where you live,
at: https://covid19estamoson.gov.pt”</td></tr>
<tr><td colspan="3">”Learn more about the general measures recommended to follow in the current state of emergency:
<Link for current measures>”</td></tr>
</table>

B Covid-19 related events

Table 2: Covid-19 related events in Portugal from the 8^{th} of January to the 23^{rd} of May 2021

Date	**Event**	**Perception**
08-01-2021	The number of daily cases of infections by Covid-19 surpasses the 10000	-
12-01-2021	The Pandemic numbers increase in all age groups	-
14-01-2021	The Government announces new measures to control the proliferation of the pandemic	-
18-01-2021	The new measures come into effect	-
21-01-2021	New containment measures and the government closes educational institutions	-
28-01-2021	The barrier of 300 new deaths from Covid-19 was surpassed and the record for the highest number of new cases ever is registered with 16432 new cases	-
09-02-2021	Epidemic shows decreasing trend	+
12-02-2021	Government announces that *"the current level of confinement will be maintained during the month of March"*	-
22-02-2021	The Minister of Health highlights the decreasing trend, but remembers that *"nothing is certain"*	+
01-03-2021	35% of people over 80 years old have already been vaccinated	+
03-03-2021	A year of pandemic in Portugal is marked	-
08-03-2021	Study reveals perceptions and concerns of the Portuguese over a year of pandemic	-
12-03-2021	Government unveils gradual reopening plan until May 3rd	+
13-03-2021	Study sets red lines for intervention in Covid-19 epidemic	-
15-03-2021	Portugal suspends administration of Astrazeneca vaccine; The 1st phase of deconfinement	+
22-03-2021	Portugal resumes administration of Astrazeneca vaccine	+
26-03-2021	More than 1 million Portuguese were vaccinated with the first dose of the Covid-19 vaccine	+
05-04-2021	The 2nd phase of deconfinement begins (e.g., face-to-face teaching in the 2nd and 3rd cycles; reopening of restaurants' outside spaces)	+
13-04-2021	Portugal registers an increase in incidence between 0 and 9 years old	-
19-04-2021	The 3rd phase of deconfinement begins (e.g., face-to-face teaching in high-school and higher education; reopening of restaurants)	+
23-04-2021	Opening the schedule of vaccination for users over 65 years of age	+
26-04-2021	Registered a day without deaths by Covid-19 in Portugal	+
03-05-2021	The 4th phase of deconfinement begins (e.g., no commercial time restrictions; outdoor events w/ low capacity)	+
11-05-2021	Sporting becomes national football champion, in a match played at Sporting's stadium in Lisbon, people take to the streets to celebrate	-
23-05-2021	SC Braga becomes wins the national cup of football, in a match played at the Jamor stadium in Lisbon, people take to the streets to celebrate	-

C Grid–Search used in the Students' Performance prediction models

Table 3: Grid search parameters for the K-Near-Neighbors model, with selected configurations highlighted for each dataset.

	algorithm	***weights***	***p***	***n_neighbors***
2018	***auto*** ball_tree kd_tree brute	***uniform*** distance	***1*** 2	1 ***2*** 3 4 5 6 7
2021	***auto*** ball_tree kd_tree brute	uniform ***distance***	1 ***2***	1 2 3 ***4*** 5 6 7
Joined	***auto*** ball_tree kd_tree brute	uniform ***distance***	***1*** 2	1 2 3 4 5 ***6*** 7

Table 4: Grid search parameters for the XGBoost model, with selected configurations highlighted for each dataset.

	booster	***n_estimators***	***learning_rate***
2018	***gbtree*** linear dart	***20*** 50 75 100 200	0.1 0.5 ***0.6*** 0.7 0.8 1
2021	***gbtree*** linear dart	20 ***50*** 75 100 200	0.1 0.5 ***0.6*** 0.7 0.8 1
Joined	***gbtree*** linear dart	20 50 75 100 ***200***	0.1 0.5 0.6 0.7 ***0.8*** 1

Table 5: Grid search parameters for the SVM model, with selected components highlighted for each dataset.

	Kernel	***C***	***Gamma***
2018	poly linear sigmoid ***rbf***	***1*** 10 100 1000	***auto*** scale
2021	poly linear sigmoid ***rbf***	1 ***10*** 100 1000	***auto*** scale
Joined	poly linear sigmoid ***rbf***	***1*** 10 100 1000	***auto*** scale

Table 6: Grid search parameters for the Naibe Bayes model, with selected components highlighted for each dataset.

	var_smothing
2018	1e-5
	1e-6
	1e-7
	1e-8
	1e-9
	1e-10
	<u>***1e-11***</u>
	1e-12
	1e-15
2021	1e-5
	1e-6
	1e-7
	1e-8
	1e-9
	1e-10
	<u>***1e-11***</u>
	1e-12
	1e-15
Joined	1e-5
	1e-6
	1e-7
	1e-8
	1e-9
	1e-10
	<u>***1e-11***</u>
	1e-12
	1e-15

Table 7: Grid search parameters for the Random Forest model, with selected configurations highlighted for each dataset.

	bootstrap	***n_estimators***	***max_features***	***max_depth***	***min_samples_leaf***	***min_samples_split***
2018	***True*** False	10 ***50*** 100 200	log2 sqrt ***None***	10 20 30 40 50 60 70 80 90 100 ***None***	1 ***2*** 4	2 5 ***10***
2021	***True*** False	10 50 ***100*** 200	log2 ***sqrt*** None	10 20 30 40 50 ***60*** 70 80 90 100 None	1 ***2*** 4	2 5 ***10***
Joined	True ***False***	10 50 ***100*** 200	log2 ***sqrt*** None	10 20 30 40 50 60 70 80 ***90*** 100 None	***1*** 2 4	2 5 ***10***

Table 8: Grid search parameters for the Decision Tree model, with selected configurations highlighted for each dataset.

	criterion	***max_depth***	***max_features***	***min_samples_leaf***	***min_samples_split***
2018	***<u>gini</u>*** entropy	5 8 10 20 30 40 50 60 70 80 ***<u>90</u>*** 100 None	5 6 7 8 9 10 11 12 13 ***<u>sqrt</u>*** log2 None	***<u>1</u>*** 2 4	2 ***<u>5</u>*** 10
2021	***<u>gini</u>*** entropy	***<u>5</u>*** 8 10 20 30 40 50 60 70 80 90 100 None	5 6 7 8 9 10 11 12 13 ***<u>sqrt</u>*** log2 None	***<u>1</u>*** 2 4	2 ***<u>5</u>*** 10
Joined	***<u>gini</u>*** entropy	5 8 10 20 30 40 50 60 70 80 90 100 ***<u>None</u>***	5 6 7 8 ***<u>9</u>*** 10 11 12 13 sqrt log2 None	1 2 ***<u>4</u>***	2 5 ***<u>10</u>***