NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTMENT OF PHYSICS

HUMAN-COMPUTER INTERACTION AND BIOSIGNALS: EVALUATING HUMAN LEARNING PROCESSES

RUI PEDRO SOUSA VARANDAS

Master in Biomedical Engineering

DOCTORATE IN BIOMEDICAL ENGINEERING: NOVA INSTRUMENTATION FOR HEALTH

NOVA University Lisbon
March, 2024

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTMENT OF PHYSICS

HUMAN-COMPUTER INTERACTION AND BIOSIGNALS: EVALUATING HUMAN LEARNING PROCESSES

RUI PEDRO SOUSA VARANDAS

Master in Biomedical Engineering

Adviser: Hugo Filipe Silveira Gamboa
Full Professor, NOVA University Lisbon

Co-adviser: Hugo Humberto Plácido da Silva
Senior Researcher, IT - Instituto de Telecomunicações, Lisbon, Invited Assistant Professor, Instituto Superior Técnico, Universidade de Lisboa, and Head of Research, Instituto Técnico (IST), Lisbon and Chief Innovation Officer, PLUX - Wireless Biosignals, S.A., Lisbon

Examination Committee

Chair: Orlando Manuel Neves Duarte Teodoro
Full Professor, NOVA University Lisbon

Rapporteurs: Hui Liu
Researcher, University of Bremen

Maria Micaela Leal da Fonseca
Assistant Professor, Universidade Lusófona

Adviser: Hugo Filipe Silveira Gamboa
Full Professor, NOVA University Lisbon

Members: Orlando Manuel Neves Duarte Teodoro
Full Professor, NOVA University Lisbon

Valentina Borissovna Vassilenko
Assistant Professor, NOVA University Lisbon

DOCTORATE IN BIOMEDICAL ENGINEERING: NOVA INSTRUMENTATION FOR HEALTH

NOVA University Lisbon
March, 2024

Human-Computer Interaction and Biosignals: Evaluating Human Learning Processes

This document was created with the (pdf/Xe/Lua)LaTeX processor and the NOVAthesis template (v6.10.17) [1].

Acknowledgements

This PhD thesis marks the conclusion of a long and challenging period of work that would not be possible without the contribution of several people.

Firstly, I would like to thank my advisors, Professor Hugo Gamboa and Professor Hugo Silva. Professor Hugo Gamboa for his guidance, invaluable support, and high standards of excellence throughout the duration of this project. His responsiveness during moments of uncertainty, adept steering of my focus, and the plethora of insightful suggestions and opinions have significantly enriched the quality of my work. To Professor Hugo Silva I would like thank his invaluable suggestions and new insights that helped to enrich my work and the challenges to improve its quality. I am particularly thankful for the opportunities provided and the continuous inspiration to delve deeper into the realms of learning and education. Their mentorship has not only shaped the trajectory of this endeavour but has also instilled in me a profound curiosity for ongoing research and knowledge acquisition.

Thank to PLUX Wireless Biosignals, to the Portuguese National Funding Agency FCT for the financial support, and a word of appreciation also goes to Instituto de Telecomunicações (IT). I would like to especially thank PLUX Wireless Biosignals for the privilege of working within its offices and generously equipping me with the essential tools for the development of this research. I am genuinely thankful to the entire PLUX Wireless Biosignals team for their consistent accessibility during critical moments. A special acknowledgment goes to the Software Team, especially to Gonçalo Telo, Daniel Osório, and Guilherme Ramos, whose unwavering collaboration and expertise played an indispensable role in bringing this thesis to fruition. I extend a special thanks to Patrícia Gamboa and Katrin for their invaluable contributions in expanding upon my ideas and providing fresh insights in various sections of this work. The support and resources provided by PLUX Wireless Biosignals have significantly contributed to the success of this project, and I am grateful for the invaluable opportunity to collaborate with such a dedicated and innovative team.

Thanks to Inês Silveira and Pedro Veiga, both of whom I had the privilege of supporting during their master's theses. Their insights and contributions proved to be invaluable,

enriching my own understanding and approach to our shared academic endeavours.

I extend my appreciation to the Cognitive Science Department of ETH Zurich, namely, Dr. Sandra Andraszewicz, Anita Maria Schärer, Dr. Marcus Cheetham, and Professor Christop Hölscher. Their generosity in providing me the opportunity to work and conduct research at ETH Zurich has been a significant privilege. I am particularly grateful for their expertise and assistance that were crucial in expanding my research endeavours. I am truly thankful for the enriching experience of being part of such a distinguished department.

Thanks also to the Libphys team - Cátia, David, João, Luís, among many others, that helped anytime I needed and supported expanding and guiding my ideas especially in the early stages of this work.

I express my heartfelt gratitude to my friends, who deserve special acknowledgment for providing with motivation and sharing enjoyable moments throughout this period. Their support and camaraderie played a pivotal role in maintaining a positive attitude, making the journey more meaningful and enjoyable. I am thankful for the encouragement and laughter we shared, creating a supportive environment that significantly contributed to the overall success and fulfilment of this endeavour.

A special thanks to Filipa whose patience and support was pivotal in the last year of this thesis. Thank you for all the moments.

Lastly, my deepest gratitude goes to my family, especially my parents, brother, and sister, without whom this achievement would never be possible. Thank you for pushing me to be better, providing endless support, and being a pillar of strength in times of need. Your unwavering belief in me has been the driving force behind my success.

”

“A teacher affects eternity: he can never tell where his influence stops.”

— Henry Adam, The Education of Henry Adam

Abstract

In recent years, the widespread adoption of e-learning has led to a surge in solutions for delivering and accessing educational content online. However, the challenge of monitoring learners' cognitive states in real-time becomes pronounced in distance-learning scenarios where tutors are absent. Suboptimal cognitive states can impede the learning process, prompting the need for innovative approaches to incorporate cognitive monitoring in e-learning environments.

In this sense, this work developed 3 different studies: 1) attention classification using biosignals and machine learning; 2) cognitive fatigue detection using functional near-infrared spectroscopy (fNIRS); 3) complex learning states (Neutral, Interest/Flow, Surprise, Boredom, Distraction, Confusion, Eureka, and Frustration) classification using both a myriad of biosignals and Human-Computer Interaction (HCI) metrics. In each of those, different cognitive tasks to monitor the distinct states were applied: 1) N-Back, Mental Subtraction, and a programming language lesson to detect attention; 2) Corsi-Block task, a concentration task, and a lesson about the electrocardiogram (ECG) that helped to elicit cognitive fatigue; 3) a modified version of the ECG lesson combined with a self-reporting stage to classify the complex learning states reported by the participants. In the three studies, six different biosignals were monitored - fNIRS, electroencephalogram (EEG), respiratory inductance plethysmography (RIP), ECG, electrodermal activity (EDA), and an accelerometer - as well as HCI metrics, namely, mouse-tracking.

Attention and cognitive fatigue detection using machine learning was possible with some restrictions. Personalized user-tuned classifiers were required to properly detect those cognitive states due to individual differences between participants. Furthermore, different biosignals and combinations were revealed to be suited differently for each participant. Analogously, the classification of the complex learning states detection in the third study was studied under various conditions, where the user-tuned classification of the gathered similar states was revealed to be the best approach. In this case, the mouse-tracking features were undoubtedly the best-performing features, reaching an F1-Score of 0.87 in this task.

This work and the presented results demonstrate that cognitive and learning monitoring is possible with minimal intrusion in e-learning contexts, allowing the development of solutions that integrate them and adjust to users in real-time.

Keywords: Biosignals, Human-Computer Interaction, Learning, Cognitive Monitoring, Machine Learning

Resumo

Nos últimos anos, a ampla adoção do ensino eletrónico (e-learning) tem gerado um aumento do número de soluções para a divulgação e acesso de conteúdo educacional online. No entanto, o desafio de monitorizar em tempo real os estados cognitivos dos alunos torna-se evidente em cenários de ensino à distância, onde os tutores estão ausentes. Estados cognitivos subótimos podem prejudicar o processo de aprendizagem, evidenciando a necessidade de abordagens inovadoras para incorporar monitorização cognitiva em ambientes de e-learning.

Desta forma, neste trabalho desenvolvemos três estudos distintos: 1) classificação de atenção utilizando biossinais e aprendizagem automática; 2) deteção de fadiga cognitiva usando espectroscopia funcional de infravermelho próximo (fNIRS); 3) classificação de estados complexos de aprendizagem (Neutro, Interesse, Surpresa, Aborrecimento, Distração, Confusão, Eureca e Frustração) utilizando uma variedade de biossinais e métricas de Interação Humano-Computador (HCI). Em cada um deles, aplicámos diferentes tarefas cognitivas para monitorizar os estados distintos: 1) Tarefa N-Back, Subtração Mental e uma lição sobre uma linguagem de programação para detetar atenção; 2) tarefa Corsi-Block, uma tarefa de concentração e uma lição sobre o sinal de eletrocardiograma (ECG) que ajudou a elicitar fadiga cognitiva; 3) uma versão modificada da lição de ECG combinada com uma etapa de auto-análise para classificar os estados complexos de aprendizagem relatados pelos participantes. Nos três estudos, monitorizámos seis biossinais diferentes - fNIRS, eletroencefalograma (EEG), pletismografia de impedância respiratória (RIP), ECG, atividade eletrodérmica (EDA) e um acelerometria - além de métricas de HCI, nomeadamente, movimentos do rato.

A deteção de atenção e fadiga cognitiva utilizando aprendizagem automática foi possível com algumas restrições. Verificámos que classificadores personalizados ajustados ao utilizador eram necessários para detetar adequadamente esses estados cognitivos devido às diferenças individuais entre os participantes. Além disso, diferentes biossinais e combinações mostraram ser mais adequados de maneiras distintas para cada participante. Analogamente, estudámos a classificação de estados complexos de aprendizagem no terceiro estudo sob várias condições, onde a classificação ajustada ao utilizador da agregação

de estados semelhantes mostrou ser a melhor abordagem. Nesse caso, as características do movimento do rato foram indiscutivelmente as mais eficazes, alcançando um F1-Score de 0.87 nessa tarefa.

Este trabalho e os resultados apresentados demonstram que a monitorização cognitiva e de aprendizagem é possível com um mínimo de intrusão em contextos de e-learning, permitindo o desenvolvimento de soluções que os integrem e se ajustem aos utilizadores em tempo real.

Palavras-chave: Biossinais, Interação Humano-Computador, Aprendizagem, Monitorização Cognitiva, Aprendizagem Automática

Contents

List of Figures xvii

List of Tables xix

Acronyms xxi

1 Introduction 1
- 1.1 Motivation . . . 1
- 1.2 Relevance . . . 2
- 1.3 Research Questions / Objectives . . . 3
- 1.4 Thesis Outline . . . 4

2 Theoretical Concepts 7
- 2.1 Cognitive Processes of Learning . . . 7
  - 2.1.1 Attention . . . 8
  - 2.1.2 Memory . . . 9
  - 2.1.3 Cognition and Emotion . . . 10
- 2.2 Paradigms of Learning Theories . . . 12
- 2.3 E-Learning . . . 13
- 2.4 Physiological Signals . . . 15
  - 2.4.1 Brain Monitoring Techniques . . . 15
  - 2.4.2 Techniques Comparison . . . 24
  - 2.4.3 Indirect Cognitive State-Related Biosignals . . . 25
  - 2.4.4 Electrodermal Activity . . . 28
  - 2.4.5 Respiratory Inductive Plethysmography . . . 29
  - 2.4.6 Accelerometry . . . 30
- 2.5 Human Computer Interaction - Mouse Tracking . . . 31
- 2.6 Machine Learning . . . 33
  - 2.6.1 Features Extraction and Selection . . . 34
  - 2.6.2 Features Normalisation and Scaling . . . 35

2.7 Performance Evaluation . . . 35
2.8 Summary . . . 36

3 State of the Art **39**
3.1 Cognitive State Monitoring . . . 39
3.2 Human-Computer Interaction Behaviour Tracking . . . 44
3.2.1 System Logs Tracking . . . 46
3.2.2 Eye-Tracking . . . 48
3.2.3 Camera-Based Tracking . . . 51
3.2.4 Other Methods and Combinations . . . 53
3.3 e-Learning . . . 54
3.3.1 Intelligent Tutoring Systems . . . 56
3.3.2 Recommendation System Assisted E-Learning . . . 57
3.4 Summary . . . 57

4 Methods **59**
4.1 Materials and Methods . . . 60
4.1.1 Biosignals Acquisition Materials . . . 60
4.1.2 Behaviour Acquisition Materials . . . 61
4.1.3 Data Synchronization . . . 62
4.2 Attention Classification Based on Biosignals . . . 63
4.2.1 Cognitive Tasks . . . 63
4.2.2 Programming Language Lesson . . . 65
4.2.3 Data Acquisition Procedure . . . 66
4.2.4 Data Pre-Processing, Segmentation, and Labelling . . . 67
4.3 Cognitive Fatigue Detection Based on Biosignals . . . 67
4.3.1 Cognitive Tasks . . . 68
4.3.2 Introduction to the ECG Lesson . . . 70
4.3.3 Data Acquisition Procedure . . . 71
4.3.4 Data Pre-Processing, Segmentation, and Labelling . . . 71
4.4 Multimodal Data Sources Classification for Learning State Monitoring . . . 72
4.4.1 Introduction to the ECG Lesson - Modified Version . . . 73
4.4.2 Self-Reporting . . . 73
4.4.3 Data Acquisition Procedure . . . 75
4.4.4 Data Pre-Processing, Segmentation, and Labelling . . . 76
4.5 Data Processing . . . 77
4.5.1 Electroencephalography Data Processing . . . 77
4.5.2 Functional Near Infrared Spectroscopy Data Processing . . . 78
4.5.3 Electrocardiography Data Processing . . . 79
4.5.4 Electrodermal Activity Data Processing . . . 81
4.5.5 Respiratory Inductive Plethysmography Data Processing . . . 81

4.5.6 Accelerometer Data Processing . . . 82
4.5.7 Mouse-Tracking Data Processing . . . 82
4.5.8 Agnostic Features Extraction . . . 83
4.6 Summary . . . 85

5 Results **87**
5.1 Attention Classification Based on Biosignals . . . 87
5.2 Cognitive Fatigue Detection Based on Biosignals . . . 90
5.3 Multimodal Data Sources Classification for Learning State Monitoring . . . 93
5.4 Summary . . . 102

6 Discussion **105**
6.1 Attention Classification Based on Biosignals . . . 105
6.2 Cognitive Fatigue Detection Based on Biosignals . . . 107
6.3 Multimodal Data Sources Classification for Cognitive State Monitoring . . . 108
6.4 Summary . . . 111

7 Conclusions **113**
7.1 Main Results . . . 113
7.2 Contributions . . . 116
7.3 List of Publications . . . 116
7.4 Applications . . . 118
7.5 Future Work . . . 118
7.6 Future Implications . . . 119

Bibliography **121**

Annexes

I Annex 1 - Questionnaires **147**

II Results All Sources **151**

List of Figures

1.1 Number of publications about e-learning by year. 2
1.2 Thesis structure. 5

2.1 Learning Process as described by Dias [23]. 7
2.2 Illustration of how single action potentials contribute little to the full electroencephalography (EEG) signal due to their low amplitude. 18
2.3 Excitatory postsynaptic potentials demonstrating how the summation of action potential contributes to the amplitude of the measured signal. 18
2.4 Common brain waves found in the normal EEG. 19
2.5 Absorption coefficients of the brain and its components. 21
2.6 Illustration of the functioning of the functional Near Infra-Red Spectroscopy (fNIRS). 22
2.7 Illustration of the electrocardiography (ECG) signal and its various waves and components. 27
2.8 Standard measurement technique of the ECG consisting of the application of 12-leads to cover all of the heart's information in three directions. 27
2.9 Illustration of the typical electrodermal activity (EDA) signal. 29
2.10 Illustration of how the signal from the Respiratory Inductive Plethysmography (RIP) signal is generated. 30
2.11 Examples of an Accelerometer (ACC) signal showing its three axis. 31
2.12 Examples of binary forced-choices. 32

3.1 Venn diagram depicting the topics discussed in this state of the art. 39

4.1 Illustration of the biosignals acquisition system. 60
4.2 Physiological sensors positioning used during data acquisition in this work. 61
4.3 Latent's communication diagram. 62
4.4 Synchronizing biosignals and mouse-tracking databases into one. 63
4.5 Schematic of the procedure of the Attention Classification Based on Biosignals study. 64

4.6 Schematic of the N-Back task. 65
4.7 Schematic of the mental subtraction task. 65
4.8 Start of the Programming Language Lesson. 66
4.9 Schematic of the procedure followed during the Cognitive Fatigue Detection Based on Biosignals. 68
4.10 Interface showing the nine blocks of the Corsi block task. 69
4.11 Interface showing the matrix of the concentration task. 69
4.12 Example of a portion of the Introduction to the ECG lesson. 70
4.13 Example of a question and respective answer that is shown after clicking "Show Answer". 71
4.14 Schematic of the procedure followed during the Multimodal Data Sources Classification for Learning State Monitoring. 73
4.15 First section of the modified version of the Introduction to the ECG lesson. . 74
4.16 Interface of the self-reporting stage. 75
4.17 Welch plot of an EEG signal. 78
4.18 Transformation of the EEG signal into a figure from where the Hu moments are calculated. 78
4.19 Example of an ECG signal with the R peaks marked with red dots, detected by the Pan-Tompkins algorithm. 79
4.20 Example Poincaré plot with SD1 and SD2 identified. 80
4.21 Example of the calculus of the magnitude signal based on the three axis of the accelerometer. 83

5.1 Confusion matrices of each participant. 92
5.2 Transition matrix of the learning states obtained by the self-reports. 93
5.3 Transition graph of the learning states obtained by the self-reports. 94
5.4 Results of the statistical analysis considering the Binary classes 95
5.5 Results of the statistical analysis considering the Cascade classes 96
5.6 Results of the statistical analysis considering the Gathered classes 97
5.7 Results of the statistical analysis considering the Original classes 98
5.8 F1-Score results for all data sources combined relative to the time-window, type of normalization, and label type. 99
5.9 F1-Score results for all data sources relative to the time-window (10-30-seconds), given the raw features, and Gathered classes. 100
5.10 F1-Score results for all data sources relative to the time-window (40-90-seconds), given the raw features, and Gathered classes. 101

I.1 Initial Questionnaire. 147
I.2 Final Questionnaire. 149

List of Tables

2.1 List of terms around e-learning. 14

3.1 Characteristics of the the three mentioned cognitive state monitoring methods. 40

4.1 Cut-off frequencies in Hz of the second-order band-pass Butterworth finite impulse filters applied for each acquired biosignal. 67
4.2 List of variables tested regarding the used classifiers, normalisation techniques, time windows for segmentation, and class types for classification. 77

5.1 Initial number of samples for the considered classes for each participant. . . 89
5.2 Individual accuracy to predict attention based on each sensor. 89
5.3 Individual accuracy to predict attention based on combined biosignals. . . . 90
5.4 Time each participant spent on each individual task. 90
5.5 Classification results for each individual for the task of detecting cognitive fatigue. 92
5.6 Number of samples for each segmentation procedure. 94
5.7 Best 10 F1-Score results achieved using the features from all data-sources. . 97
5.8 Best 10 F1-Score results achieved using the features from all the studied combinations of data-sources considering Gathered labels, Multi-Classifier (MC) classifier, and raw features. 102

II.1 Results of classification considering the combinations of all data sources taking into account the Binary paradigm. 151
II.2 Results of classification considering the combinations of all data sources taking into account the Cascade paradigm. 156
II.3 Results of classification considering the combinations of all data sources taking into account the Gathered paradigm. 161
II.4 Results of classification considering the combinations of all data sources taking into account the Multiclass paradigm. 167
II.5 Results of classification considering the combinations of all data sources taking into account the Leave One User Out (LOUO) paradigm. 172

Acronyms

AB Adaboost *(pp. 76, 97, 98, 115, 151–177)*
ACC Accelerometer *(pp. xvii, 15, 25, 31, 57, 60, 61, 67, 82, 83, 85, 88–90, 102, 114, 120)*
ADHD Attention Deficit and Hyperactivity Disorder *(p. 118)*
AI Artificial Intelligence *(pp. 2, 15, 56, 57, 120)*
ANS Autonomic Nervous System *(pp. 25, 27, 28, 79)*
ATP adenosine triphosphate *(pp. 16, 20)*
AUC-ROC Area Under the Receiver Operating Characteristic Curve *(pp. 36, 107)*

BCI Brain-Computer Interface *(pp. 41, 42)*

CW cognitive workload *(pp. 1, 41)*
cwfNIRS continuous wave functional Near Infra-Red Spectroscopy *(pp. 21–24)*

DOT near-infrared diffuse optical tomography *(p. 24)*
DT Decision Trees *(pp. 76, 88, 91, 97, 151–177)*

ECG electrocardiography *(pp. xvii, 15, 25–27, 44, 57, 59–61, 67, 68, 70, 72–74, 79, 83, 85, 88–90, 95, 106, 107, 109, 111, 113, 114)*
EDA electrodermal activity *(pp. xvii, 15, 25, 28, 29, 43–45, 57, 60, 61, 67, 75, 81, 85, 88–90, 95, 106, 113, 114, 120)*
EEG electroencephalography *(pp. xvii, 4, 15–20, 24, 25, 40–45, 53, 54, 57, 60, 61, 67, 75, 77, 78, 85, 88–90, 95, 102, 106, 110, 111, 113–115, 120)*

FD-fNIRS frequency domain functional Near Infra-Red Spectroscopy *(pp. 22–24)*
fMRI functional Magnetic Resonance Imaging *(pp. 16, 20, 32, 41)*
FN False Negative *(p. 36)*
fNIRS functional Near Infra-Red Spectroscopy *(pp. xvii, 4, 15, 16, 20–25, 40–43, 57, 59–61, 66, 67, 71, 72, 75, 78, 79, 88–91, 93, 95, 102, 106–109, 111, 113, 114, 118, 120)*
FP False Positive *(p. 36)*

GNB Gaussian Naive Bayes *(pp. 76, 97, 151–177)*
GP Gaussian Processes *(pp. 76, 115, 151–177)*
GVR Galvanic Skin Response *(p. 29)*

Hb deoxygenated haemoglobin *(pp. 16, 20–24, 41, 72, 78, 79, 85, 91)*
HCI Human-Computer Interaction *(pp. 3, 4, 7, 31, 32, 36, 44, 46, 53, 57, 61, 63, 66–68, 70, 75, 107, 111, 113, 114, 116)*
HR Heart Rate *(pp. 43, 44)*
HRV Heart Rate Variability *(pp. 45, 53)*

ITS Intelligent Tutoring System *(p. 56)*

LDA Linear Discriminant Analysis *(pp. 76, 151–177)*
LMS Learning Management System *(pp. 3, 14, 15, 44)*
LOUO Leave One User Out *(pp. xix, 96, 110, 172)*
LSTM Long-Short Term Memory Neural Network *(p. 33)*

MC Multi-Classifier *(pp. xix, 76, 97, 98, 102, 110, 111, 114, 151–177)*
MEG magnetoencephalography *(pp. 16, 41)*
ML Machine Learning *(pp. 4, 7, 33–37, 41–43, 46, 48, 52, 53, 96, 108, 110, 113–116, 118, 120)*
MLP Multi-Layer Perceptron *(pp. 76, 97, 98, 151–177)*
MOOC Massive Open Online Course *(p. 3)*

O_2Hb oxygenated haemoglobin *(pp. 16, 20–24, 41, 72, 78, 79, 85, 90, 91)*

PET positron emission tomography *(pp. 16, 41)*
PFC prefrontal cortex *(pp. 41, 93)*
PGR Psycho-Galvanic Reflex *(p. 29)*
PLE Personalised Learning Environments *(p. 15)*

RF Random Forest *(pp. 76, 87, 88, 91, 151–177)*
RFECV Recursive Feature Elimination with Cross-Validation *(p. 91)*
RIP Respiratory Inductive Plethysmography *(pp. xvii, 15, 25, 30, 60, 61, 67, 81, 82, 85, 89, 90, 102, 106, 109, 110, 113, 114)*
RMS Root Mean Square *(pp. 78, 80–82, 84)*
RS Recommendation System *(p. 15)*

SCR Skin Conductance Response *(p. 28)*

SNR signal to noise ratio *(p. 23)*
SSR Sympathetic Skin Response *(p. 29)*

TD-fNIRS time domain functional Near Infra-Red Spectroscopy *(pp. 22–24)*
THb total haemoglobin *(pp. 72, 78, 85, 91)*
TN True Negative *(p. 36)*
TP True Positive *(p. 36)*

UX User Experience *(pp. 44–46, 54)*

1 Introduction

1.1 Motivation

Biomedical engineering consists of applying engineering principles to medical applications, focusing on designing and developing medical devices and tools to aid healthcare. As new challenges arise, the field of biomedical engineering also evolves, and this evolution is reflected in the training of new professionals who must continually learn and adapt to keep pace with advancements in the field.

Learning involves various brain processes, such as interest, attention, and emotional responses (see Chapter 2). Specifically, interest guides attention, which can lead to more effective learning [2, 3]. Emotions also play a relevant role, and learners' emotions can be categorized as positive activating (e.g., enjoyment), negative activating (e.g., anxiety), and negative deactivating (e.g., boredom) [4, 5]. However, these processes are typically assessed using surveys in meta-analyses, which may not fully capture the dynamic changes in students' mental and emotional states during learning sessions [6].

E-learning has become a popular method for delivering and accessing educational and training resources despite its limitations in accommodating the dynamic cognitive changes occuring during the learning process. In traditional face-to-face classrooms, learners and instructors can interact in real-time, allowing for immediate feedback, which may not be possible in e-learning settings, which might result in lower learner engagement, decreased motivation, and hindered learning. Additionally, e-learning often provides rigid, pre-designed content that may not be suited for every type of learner and may not accommodate individual learning paces and needs. However, as reviewed in [7], some e-learning platforms provide personalized experiences such as personalized learning paths, navigation, recommendations, assessments, and feedback. Nonetheless, these approaches do not consider the instantaneous state of the learner, which can result in subpar learning outcomes. Therefore, since e-learning lacks the ability to monitor and adapt to learners' state in real-time, it may significantly impact their learning process as they may fail to keep learners engaged [8]. Namely, these platforms should consider the attention, cognitive workload (CW), and the emotional state during learning to create more comfortable and

successful learning atmospheres [8–11].

The integration of cognitive and emotional monitoring during e-learning sessions can serve as support to enhance the effectiveness of learning. This might be particularly useful for biomedical engineering students, who focus heavily on the study of biosignals. By monitoring their own biosignals, they can learn about biosignals while interacting with them in real-time.

1.2 Relevance

The evolution of education has been powered by the evolution of industry and Ciolacu et al. (2018) introduced the concept of Education 4.0 in 2018 as virtual courses including Blended Learning with the integration of various Artificial Intelligence (AI)-powered features, including personalized learning and adaptive technologies [12]. One of the differentiating aspects of education coming from the last years was e-learning, which refers to the deployment of educational materials in digital format, most commonly through the internet in recent years [13].

During those years, e-learning has experienced significant growth, as illustrated by the number of publications over the time in Figure 1.1, primarily due to providing flexibility, accessibility and allowing the deployment of teaching and learning materials online [14–16]. This flexibility empowers learners to self-regulate their pace and content of learning [15]. Consequently, e-learning has become a powerful method for disseminating educational materials via the internet, enabling experts to share information and expertise, thus granting learners extensive access to valuable resources [16].

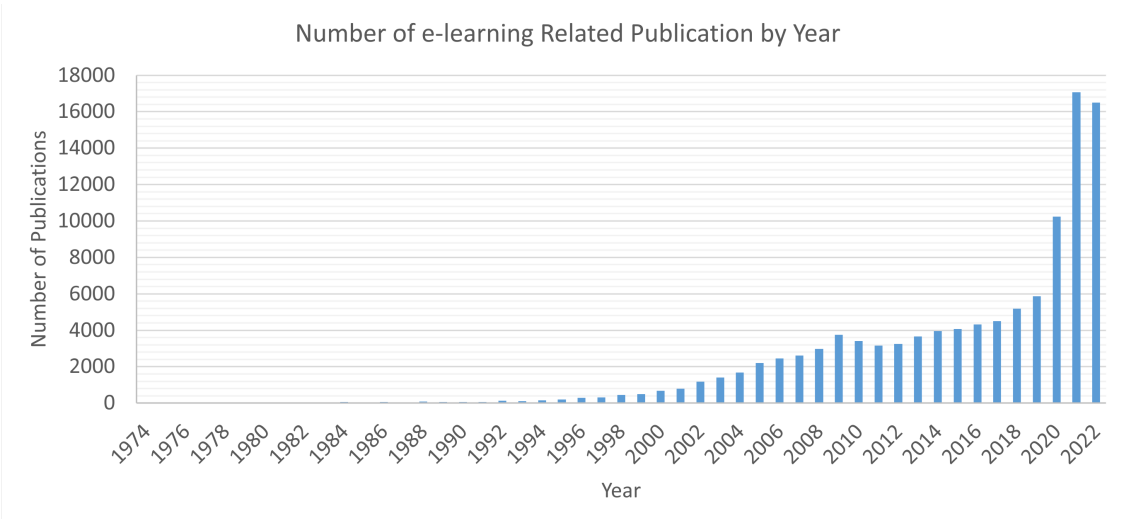

Figure 1.1: Number of publications about e-learning by year. The used query was ["computer-based-learning"OR "internet-based-learning"OR "web-based-learning"OR "e-learning"OR "m-learning"OR "distance learning"] in https://app.dimensions.ai/ [Accessed in 06-07-2023].

Learning Management System (LMS)'s are commonly utilized to centralize educational resources, employing Massive Open Online Course (MOOC)'s that incorporate various tools such as content creation, forums, chats, and gamification elements. Additionally, LMS's can be designed to enhance user engagement by facilitating personalized learning, in contrast to the traditional static approach applied to all students' profiles. Exploring the integration of recommendation systems with LMS's, which consider user preferences and history, holds the potential to expand the knowledge accessible to users and suggest relevant courses for each individual [17, 18].

Although e-learning solutions have made significant progress, most of them still lack the capability to understand and adapt their approaches to individual learners' cognitive states in real-time, which may lead to inattention and disengagement resulting in poor learning performance [19]. To enhance engagement and learning outcomes during e-learning sessions, virtual tutors need to address this challenge and be able to adapt to learners' cognitive states dynamically.

1.3 Research Questions / Objectives

The limited monitoring of cognitive states during e-learning can present challenges in effectively guiding learners to maintain engagement and an optimal cognitive state for learning. Although significant technological advancements have been made, such as the widespread use of biosignals monitoring through devices like smartwatches, their application in monitoring cognitive states during learning or e-learning is still relatively uncommon. Similarly, while the study of behaviour monitoring, such as real-time mouse tracking data, has been explored, its utilization within the educational context remains largely unexplored.

Biosignals have previously been used to measure cognitive states in different contexts with success, however, these are commonly limited to discrete variables, such as, cognitive load, attention, and cognitive fatigue (see for example [19, 20]). Since learning depends on all those variables, one research question arose: is it possible to monitor and influence cognitive state with the objective of optimise the learning process?

This thesis will then address the identification of biosignals and behavioural metrics that might be able to monitor the cognitive states of learners during e-learning lessons. A system composed of cognitive state monitoring combined with content adaptation in real-time would then be able to increase learning by keeping learners engaged and in an optimal learning state. Thus, the questions that guide this thesis are:

- **RQ1 – Is it possible to use biosignals to detect specific cognitive states**, namely, attention and cognitive fatigue? And complex learning states?
- **RQ2 – Can Human-Computer Interaction (HCI) data be used to infer about cognitive states related to learning**, namely, attention and cognitive fatigue? And complex learning states?

- **RQ3 – How does learning occur over time?** Namely, what states are involved during learning and how do they transition between them?
- **RQ4 – Which is the best data-source to automatically assess learning states**, assuming learning can be assessed using more than one?
- **RQ5 – Does the Machine Learning classification algorithm influence the classification performance?** Given different algorithms perform differently, what is the best in the context of learning assessment using the different data-sources?
- **RQ6 – Does the segmentation of the time series influence the classification performance** considering how different biosignals and behaviours may have different periods (e.g., heart rate and respiration rate) or may be aperiodic (e.g., HCI data)?

For the first research questions, a thorough state-of-the-art review will be presented where the most commonly used biosignals will be identified in the context of cognitive states monitoring. Moreover, given the biosignals are measured in specific sites of the human body, those will also be identified for each of the biosignals. The fact that there are biosignals that may require more than one channel (e.g., EEG and fNIRS) may introduce additional difficulty to use them outside of the laboratory and in natural and social scenarios. Thus, in this work the number of channels will be kept as low as possible to optimize the ecological validity of the proposed solution.

An analogous procedure will be followed in order to identify the best behavioural metrics that can be used without the requirement of specific lab-equipment (e.g., eye-tracker devices). The goal is to identify metrics that can be monitored in every e-learning environment, namely when interacting with a computer.

The developed data acquisition protocol will serve as the basis to test the validity of the proposed hypothesis, i.e., that it is possible to monitor learners' cognitive state during e-learning lessons using biosignals and/or behavioural metrics. Common cognitive tasks used in cognitive psychology will be identified, which might be used as the ground work to measure cognitive states related to learning. A data acquisition protocol will be defined and used to acquire data to obtain the identified biosignals and behavioural metrics in specific controlled scenarios.

Finally, a Machine Learning (ML) pipeline will be built to test if it is possible to assess cognitive state during learning automatically using the biosignals and behavioural data acquired during the data acquisition protocol developed during this thesis.

1.4 Thesis Outline

This thesis is structured as shown in Figure 1.2. The structure is divided in 3 main areas: basis and concepts, the developed studies and the conclusions.

The rest of this work is organised as follows: Chapter 2 presents the theoretical concepts that help to support this thesis. In Chapter 3, a literature review presenting

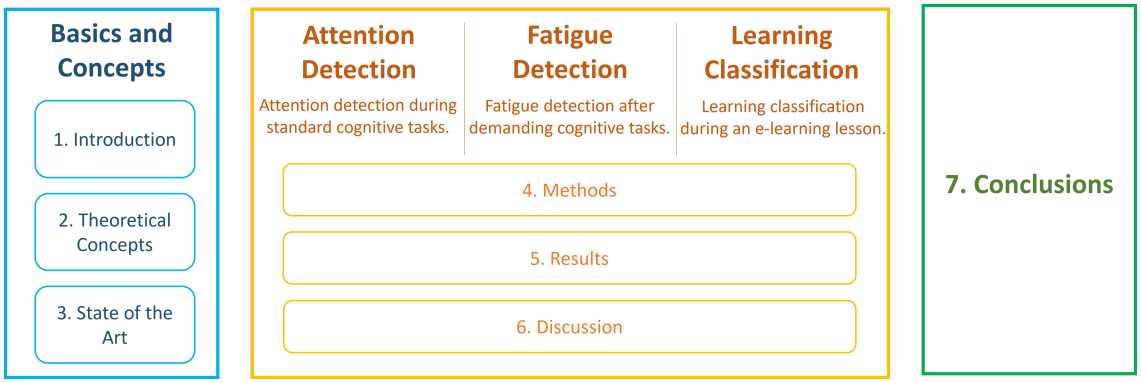

Figure 1.2: Thesis structure.

works using behaviour and cognitive monitoring focusing on learning assessment will be given. Chapter 4 will describe each of the developed data acquisition procedure and the corresponding studies. The data processing methodology employed in this work will also be presented. Chapter 5 will contain the results achieved in each of the performed studies, while Chapter 6 will discuss them. The conclusions of this dissertation project are outlined in Chapter 7.

2

Theoretical Concepts

This chapter presents the theoretical concepts that underpin this thesis. It begins by examining learning from a cognitive psychology perspective and discussing the main theories that have been developed to define it. The following section delves into the concepts associated with e-learning, followed by an explanation of the used biosignals. The next section focuses on HCI with particular emphasis on mouse tracking, and the chapter concludes with a description of ML.

2.1 Cognitive Processes of Learning

Learning is a hard concept to define given it depends on the numerous theories developed by different researchers over time [21, 22]. In simple terms, one might argue that learning refers to a change and, depending on the underlying theory, a change of understanding, a change of behaviour or simply the acquisition of new knowledge, and it might be considered a process or a result [23]. More broadly, learning can be defined as the process of how living beings acquire new knowledge, develop new skills and change their behaviour [23].

The full learning process as described by Dias (2018) is schematized in Fig. 2.1. In this model, the basic pillar for learning is motivation. Without motivation, the remaining steps of the process are not triggered and, thus, learning does not occur. However, motivation might relate to the knowledge acquired itself or simply to get good grades, but it might also be negative, for example, to avoid punishment or to avoid having bad grades.

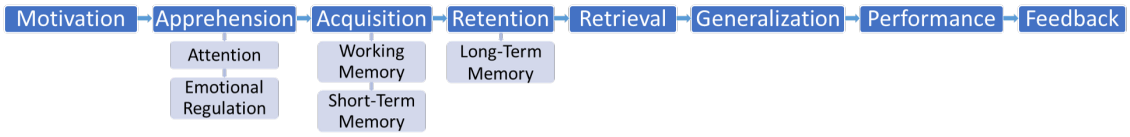

Figure 2.1: Learning Process as described by Dias [23]. Each step depends directly on the previous one, which means that if one of the steps fails, the next steps do not occur. The vertical concepts are directly associated to the specific learning step.

After motivation to start the learning process, the second step is apprehension. For this stage to occur, besides motivation as a precursor, attention plays a key roll. Attention was defined by William James in 1890 as "[...] taking possession of the mind in clear and vivid form of one out of what seems several simultaneous objects or trains of thought", as cited in [24].

However, attention is highly influenced by emotion. For example, sometimes inattention is used as a defence mechanism for individuals to excuse failing when they feel insecure and prone to errors, even though it happens unconsciously. Thus, emotional regulation abilities are important to increase and keep attention, meaning that learning environments should feel safe and should value conquests more than to point out mistakes and errors [23].

After apprehension, there should be an effective acquisition process. In this case, acquisition refers to the process of encoding new information, knowledge or motor skills into working memory (a system that holds the information currently being processed [4]) or short-term memory (a system that holds limited bits of information for short periods of time [4]). Then, the encoding should promote retention into long-term memory (the system responsible for the storage of information, facts, and knowledge over an extended period of time, ranging for a few hours until a lifetime [4]). The better the encoding and storing of information, the better the efficacy of its retrieval. In this regard, the more value assigned to a given content, the higher the probability of effectively retrieving it. Thus, when memory fails, it might be due to poor retention (source) or due to a problem in retrieving information (posterior). Memory is also a process that connects bits and pieces of information and knowledge to generate new ideas, helping to make daily decisions. Thus, if the previous processes were effective, the learning process continues into generalization [23].

Besides these core processes, learning also involves performance, which refers to the ability to show what was learnt. However, if the performance is poor, any of the previous steps may have failed or the person may simply lack the will to answer questions or to explain the acquired new content (information, facts, or knowledge) [23].

Lastly, learning involves feedback, consisting of the evaluation of performance and it is essential to start new learning or to compensate any lacking of the previous steps [23].

This model helps to identify learning difficulties by enabling the evaluation of each individual step.

Next, the key concepts referred here, namely, attention, memory, problem-solving, and emotion related to cognition, will be explored from a cognitive psychology point-of-view and based on the book of Eysenck and Keane [4].

2.1.1 Attention

Attention is often described as the selectivity of processing. In order to consciously perceive something, one must actively direct their attention toward it (e.g., we look in

order to see; we listen in order to hear, etc.).

There is a distinction between active (top-down) and passive (bottom-up) attention. The former is the controlled attention driven by the subject and is goal-oriented, while the second is the attention driven by external stimuli, such as, a loud noise. Passive attention requires less processing effort because it does not depend on the decision of what stimulus is more relevant at any given time.

According to Eisenck and Keane (2000) attentional research has two main areas: focused attention, in which the subject should respond to one of the multiple stimuli presented at a given time; and divided attention, in which the subject should respond to multiple stimuli at a given time [4], however, more recently, Dias (2018) identified 4 different types of attention: focused attention that consists on focusing on one specific task or stimulus excluding all others; alternated attention where attention is alternated between different stimuli; sustained attention, where focus is kept on one continuous and repetitive long task or stimulus; and selective attention, in which the individual consciously focuses on one task or stimulus by choosing where to focus [23]. However, studies on this topic suffer from three limitations:

- We can pay attention to the external or internal (thoughts and memories) environment, but experiments only take into account the external stimulus.
- Attention is highly influenced by motivational states, that are not taken into account in experimental settings, in which attention is determined by the experimental instructions and not the motivational state.
- Experimental settings are usually carried on taking into account 2D stimuli that require arbitrary responses, which is not what happens in the real world, where we decide what actions are more suitable to deal with three-dimensional people and objects.

Research on attention often exhibits bias based on the specific task or stimuli used, as it tends to overlook individual motivation, internal factors, and the inability of experimental settings to encompass all variables present in the real-world beyond the controlled laboratory environment.

2.1.2 Memory

Memory comprises three stages: encoding, storage, and retrieval which, in Dias' model, are involved in the learning steps of acquisition, retention, and retrieval [4, 23].

Although working memory and short-term memory seem to be simple systems to process and retain few elements for a short amount of time respectively, long-term memory is a more complex system that comprises multiple variants.

Special attention should be given to working memory in this context. Working memory plays a crucial role in executive functions, encompassing tasks like inhibition, goal shifting,

and goal updating. Interestingly, individuals with higher working memory capacity often demonstrate superior attentional control and are less prone to goal neglect compared to those with lower capacity. Additionally, working memory capacity is intricately linked to fluid intelligence – the capacity to apply non-verbal reasoning to novel problems. This interconnection complicates the study of isolated effects of working memory, as variations in performance may also be influenced by different levels of fluid intelligence.

Long-term memory encompasses both episodic memory and semantic memory. Episodic memory involves the encoding and retrieval of specific events or episodes, including details of where and when they occurred (e.g., recalling what one had for breakfast). In contrast, semantic memory stores our general knowledge about the world, consisting of abstract representations. The key distinction between these two memory types is that episodic memory pertains to personal, past experiences, whereas semantic memory is impersonal and can encompass abstract information unrelated to specific past events.

Long-term memory can further be categorized into explicit and implicit memory. Explicit memory involves conscious recollection of past experiences or facts, while implicit memory impacts task performance without requiring conscious recall. Implicit learning also encompasses the acquisition of complex information without a complete verbal expression of the newly acquired knowledge.

Long-term memory can be categorized into two distinct systems: declarative and procedural knowledge. Declarative knowledge, often described as "knowing that", encompasses factual information, such as recognizing a country's capital or recalling morning events. This type of knowledge is considered explicit and is associated with heightened activity in the right hippocampus, with potential contributions from the frontal lobes and diencephalon. On the other hand, procedural knowledge, referred to as "knowing how", relates to the ability to perform tasks such as riding a bicycle or playing the piano. It is viewed as implicit knowledge and is associated with specific brain regions. Sensori-motor skills learning involves the basal ganglia and cerebellum, perceptual skills learning is connected to the right parietal cortex and the left inferior occipito-temporal cortex, visual perceptual priming tasks implicate bilateral occipito-temporal areas, and conceptual priming is linked to the left frontal neocortex.

In contrast with typical memory studies, everyday memory takes into account three factors/assumptions: 1) it is purposeful; 2) it is personal, thus, influenced by the individuality and personality; 3) it is influenced by situational demands. Thus, accuracy is not always the main goal. For instance, this is evident in cases of repression, where individuals tend to selectively forget distressing and traumatic events.

2.1.3 Cognition and Emotion

Thinking can occur in two main ways: goal directed, such as arithmetical problems and puzzles; and non-goal directed, such as, introspection and dreaming. Furthermore, however thinking may be a conscious process, we only tend to remember its products

and not its processes. Then, thinking tasks may require different levels of knowledge (for example, puzzles typically require problem-solving skills rather than specific knowledge). According to Eysenck and Keane (2000), experts exhibit enhanced memory regarding the problem space, they employ different problem solving strategies, and they become experts through extensive practice. As a result, distinct behaviours may arise based on varying levels of expertise, with experts approaching problems differently compared to novices.

On another level, emotion was largely ignored in cognitive studies due to the difficulty it introduced, notwithstanding the importance it has in cognition. There are four concepts related to emotions: affect - broad concept used to cover emotions, moods and preferences; emotion - used to refer to brief and intense experiences; mood and state - low-intensity and prolonged experiences.

Many authors have focused on the proposition of models of emotional aspects related to cognition. For example, Zajong cited in [4] argued that affective judgements are made faster than cognitive processes take to occur and corroborated in an experiment showing that the difference of emotions provoked by a stimuli was greater on a 4 ms setting than in a 1 s setting, where there was no significant difference. Thus, two conclusions were taken: 1. affective processing can sometimes occur faster than cognitive processing and 2. the initial affective processing of a stimulus may be different from the later cognitive processing.

On the other hand, Lazarus, also cited in [4], argued that cognitive appraisal plays a crucial role in emotional experience. Cognitive appraisal can be divided in primary appraisal - an environmental situation is classified as positive, negative or irrelevant -, secondary appraisal - taking into account the available resources to cope with the situation - and re-appraisal - monitoring of the situation and the coping strategy in order to modify primary and secondary appraisals if necessary. Various experiences were conducted, where participants had to watch anxiety-evoking films. It was found that most subjects entered a phase of denial or intellectualisation proving that cognition had an important role in emotional response. However, these findings were not easy to replicate by others. The strongest critic against Lazarus' theory is that it ignores or de-emphasizes social context in which emotion is normally experienced.

Multi-level theories have been proposed to explain the intervening processes that lead stimuli to the emotional response. For example, LeDoux studied anxiety and stated that the process of emotional response is localised in the amygdala and the relation between amygdala, thalamus and cortex. Specifically, it was argued that there are two different emotion circuits in anxiety, a slow-acting thalamus-to-cortex-to-amygdala circuit involving detailed analysis of sensory information and a fast-acting thalamus-amygdala circuit based on simple stimulus features (intensity) that bypasses the cortex [4]. While the second process is essential for survival as a mechanism that works in life-threatening situations, the first process allows us to respond to different situations in an appropriate way.

Another theory proposed by Power and Dalgleish known as Schematic Propositional Associative and Analogical Representational Systems divides these process into four

systems, namely, the Analogical System that is involved in basic sensory processing of environmental stimuli, the Propositional System which is an emotion-free system containing information about the world and the self, the Schematic System that constructs an internal model of the situation based on the combination of knowledge in the propositional system and the individual's current goals where there is an emotional response if the current goals are being thwarted, and, finally, the Associative System (Dalgleish (1998), cited in [4]) that constructs an associative rule if an event is consistently handled in the same manner on a Schematic level, resulting in involuntarily triggering the corresponding emotion every time the same event is encountered in the future [4].

Discrete emotions theories suggest humans have a set of innate basic emotions independent of culture that may interact with each other, although there is not a consensus. Namely, different model include different basic emotions such as anger, disgust, fear, happiness, sadness, surprise, and contempt, among others [25].

In contrast, dimensional models consider that there is an underlying system to all affective states and thus, that all emotions are connected [26]. For example, the circumplex model is based on subjective feelings that can be arranged given three dimensions: valence, arousal, and dominance of feelings, i.e., positive or negative, intensity, and the extent of control felt in a given situation, respectively [26]. Furthermore, the flow theory also considers two dimensions, in this case, the perceived difficulty of a task and the knowledge or skills of an individual [27]. In this case, flow occurs when the level of challenge is similar to the level of knowledge or skill, if task difficulty is too high one gets frustrated, and if the difficulty is too low one gets bored [27].

In summary, the study of emotions in cognition is a multifaceted field with diverse theories and models that contribute to our understanding of how emotions influence and interact with cognitive processes. These insights have practical applications in various domains, including psychology, neuroscience, and education, and continue to be an area of active research and exploration.

2.2 Paradigms of Learning Theories

Given there are multiple views of what learning is and how it occurs, several paradigms of learning theories were proposed by different authors, hence the difficulty to find a common definition for the term "learning".

In this sense, behaviourism was one of the first theories developed by authors like Pavlov, John B. Watson, Edward L. Thorndike, and Burrhus F. Skninner [23]. It assumes that learners are mostly passive and only respond to outside stimuli. In this case, learners' behaviour is changed mostly through reinforcement, being it positive or negative and, thus, learning is viewed and assumed as a change in a given learner's behaviour. In this theory, the internal state of learners does not play a role in learning, being influenced solely by the principles of contiguity and reinforcement.

Cognitivism, another paradigm of learning theories developed by authors such as David P. Ausubel, Bernard Weiner, and Charles M. Reigeluth, focuses on the internal mental processes that occur during learning [23]. Here, thinking, memory, knowledge and problem-solving are studied and taken into account when considering learning. Learning is thus defined as the changes in the mental schemes of learners, that are viewed as rational beings that actively participate during learning and their actions or behaviour are a consequence of their mental processes.

In contrast, humanism was developed by figures such as Abraham H. Maslow and Carl R. Rogers. This approach places a primary focus on human development, emphasizing the affective and subjective world, personal freedom, choice, motivations, and feelings [23].

Socioconstructivism, popularized by authors such as Jean Piaget and Lev Vygostky, defines learning as being determined by mutual interactions between an individual and their surroundings. The basic principle is that people are not born intelligent, but they are also not passive under the influence of their surroundings, i.e., they respond to external stimuli, acting upon them to build and organize their own knowledge with increasing complexity [23]. In this case, inspired by evolution in biology, knowledge is not viewed as a means to reach an absolute truth but as a matter of adaptation to the surroundings.

Finally, connectivism found proponents in George Siemens and Stephen Downes and acknowledges knowledge as being distributed in a network. In this case, learning consists on the ability to build and use those networks, i.e., the process of creating new connections and the amplification of this network [23]. Thus, the experience of learning is defined as the moment in which knowledge is actively acquired in order to complete a necessary task or to solve a problem. In this case, the nodes of a network include any source, such as, people, organizations or websites. Thus, in the words of Dias (2018), "know-where"(to find information) complements the "know-how"and the "know-what". Learning is a continuous process, as learners seek knowledge when and where it might be necessary.

This thesis takes from various of these theories to identify learning. Namely, behaviour will be used to infer about the learning state of individuals (behaviourism), while their internal state will be monitored to search for patterns that describe them (cognitivism). However, learners will be able to access the learning contents while being asked to recall and perform on open answers questions about them, so, they will be able to use their "know-where"too (connectivism). These aspects will be made clearer in Chapter 4.

2.3 E-Learning

E-learning, also known as online-learning, distance learning, and web-based education, consists of the utilization of information and communication technologies to deploy and access learning materials in electronic format allowing the collaboration and exchange of information remotely [16]. It may involve the deployment of learning materials over

the internet and might be accessed via computer (computer-based learning) or smartphone (mobile learning or m-learning) and presented in browser, in virtual reality or in augmented reality formats [13, 28].

E-learning can also be categorized based on the mode of access to classes. If it requires the presence of a tutor and mandates simultaneous participation, it is termed synchronous e-learning. Conversely, if the course materials are developed in advance and accessed at different times by participants, it falls under asynchronous e-learning [29]. Additionally, there is a concept known as Blended Learning, which combines traditional face-to-face instruction with online or digital learning (e-learning). In this approach, the strengths of both methodologies are leveraged to enhance the overall educational experience [30]. A summary of the list of terms around e-learning and their descriptions is presented in Table 2.1.

Table 2.1: List of terms around e-learning.

Term	Description
E-learning	Utilization of information and communication technologies to deploy and access learning materials in electronic format.
Online-learning	Deployment and access of learning materials in online formats, including various digital platforms and tools, such as mobile applications, interactive simulations, and virtual classrooms.
Web-based education	Education deployed over the web, involving websites, webinars, and other online resources delivered through a web browser.
Distance learning	Educational process where instructors and learners are physically separated, often geographically.
Computer-based learning	Integration of computers and digital technologies into the educational process, applicable to both distance and in-person learning.
Mobile learning or m-learning	Analogous to computer-based learning but using smartphones or other mobile devices instead of computers.
Synchronous learning	Learning that requires the simultaneous participation of tutors and learners synchronously.
Asynchronous learning	Educational approach where learners engage with instructional materials and complete activities at different times, without the need for real-time interaction.
Blended learning	A mixture of synchronous, asynchronous, or in-person learning to enhance the qualities of all forms of learning.
Personalised Learning	Educational approach that tailors instruction to the individual needs, preferences, and pace of each learner, using adaptive technologies, varied instructional strategies, and flexible content delivery methods.

The contents of e-learning are usually centralized in LMS comprising of course management, student management, feedback management, course materials management,

examination, and assessment of assignments and activities [17]. However, these types of system are typically unable to provide personalized suggestions, which might hinder the learning process as the leaners might reach an impasse [17].

Solutions to this major concern could be Personalised Learning Environments (PLE), dynamic learning environments that can adapt to learners' interests, facilitating student-centred learning. These can be viewed as LMS coupled with data mining to perform personalized recommendations [17].

Analogously, there are Recommendation System (RS) that use AI to find relevant information suited for the evolving learners. The integration of LMS with RS, that suggest the content given the preferences and history of each learner [17, 31], is a possibility not only to increase the range of knowledge that the learner may access, but also to recommend classes or courses that the learner may need at each stage [32]. The interconnection of recommendation systems with LMS would pose an advantage as learners would automatically find new activities within classes that would better suit their interests, expand their knowledge, and learn at their own pace. However, these systems work at the preferences level, not taking into account the real-time cognitive state of learners.

In this context, pedagogical agents emerge as software agents integrated into learning environments, specifically in e-learning solutions, which are grounded in AI. These agents autonomously manage their behaviour and internal states, possess the capacity to detect changes, and engage in goal-oriented actions. They are capable of activation as needed and can interact with other agents or humans, exhibiting a degree of social capability [33]. Pedagogical agents have been widely applied in e-learning due to their ability of providing cognitive assistance as well as posing as a social layer in an otherwise static learning environment [33].

2.4 Physiological Signals

Cognitive monitoring can be achieved through multiple mediums. This work will use physiological signals and behavioural metrics. This section will present the utilized physiological signals and explain how they are related to the cognitive state. The signals are: fNIRS, EEG, ECG, EDA, RIP, and ACC.

This section will be divided into the signals that directly monitor brain signals (fNIRS and EEG) from those signals that allow to infer about the cognitive state indirectly (ECG, EDA, RIP, and ACC).

2.4.1 Brain Monitoring Techniques

This section will present the techniques that rely on the neurological processes of the brain, which generate electrical signals by the neurophysiological processes occurring at the neurons level, followed by a vascular response [34]. Brain monitoring techniques can be invasive or non-invasive. Invasive methods are typically considered the methods

that require surgery to implant sensors, such as the case of electrocorticography [35]. Non-invasive methods are the ones that do not require surgery, such as the case of medical imaging methods (e.g., functional Magnetic Resonance Imaging (fMRI) and positron emission tomography (PET)) [36] and neurophysiological measures (e.g, EEG, fNIRS, and magnetoencephalography (MEG)) [37]. In this work, emphasis was placed on the methods that provide enhanced ecological validity, i.e., the methods that can naturally be used outside the laboratory, considering that e-learning can be accessed anywhere and anytime.

Neurophysiological measures are the techniques that use the electrical currents of the brain to produce the signal or image, while haemodynamic activity refers to the vascular response to metabolic demands, usually oxygen deficiency, that occurs after the neurological activity in order to restore the oxygenation and energy required to carry the normal processes of the brain. Techniques that rely on the haemodynamic activity are, for example, fMRI and fNIRS [37].

Non invasive imaging techniques are important to assess about the functioning of the brain without the need of skull removal or surgical procedure [38]. This is particularly important for research applications in which the main focus is to understand brain functioning during specific activities or environments, relying on the aware state of the subjects to perform the activities or react to stimulus in the environment.

Specifically, the study of the neurological processes during learning tasks relies on the awareness, concentration and attentional state of subjects. In simple terms, the learning process may include sensing, in order to perceive the new information, for example, by seeing or hearing it; memory, to recall old experiences and interconnect known contents with the new; thoughts, to understand how different information relates; and meta-analyses of beliefs, to understand if the new information still fits with the cemented old beliefs [4]. These require the activation and interconnection of multiple brain regions in order to execute all the processes simultaneously and thus, learn [39]. These activations can be detected non-invasively by the imaging techniques enumerated before.

Regardless of the type of technique, all activity occurring in the brain, such as, generation and propagation of action potentials, binding of vesicles in the pre-synaptic junction, release of neurotransmitters, among other processes, require the consumption of energy in the form of adenosine triphosphate (ATP) [34]. One way to produce ATP is the aerobic pathway, where glycolysis occurs, consisting of the degradation of glucose molecules with the production of carbon dioxide and other toxins with the mediation of oxygen molecules (O_2) [39]. During the onset of activation, there is a decrease of the concentration of oxygenated haemoglobin (O_2Hb) and an increase of the concentration of deoxygenated haemoglobin (Hb) in the area, due to the energy production process. These poses needs, namely, oxygen supply and production of toxins, that are chemically signalled, leading to an increased blood flow to the region of interest. The increased blood flow is accompanied by an increase of the concentration of O_2Hb and decrease of concentration of Hb after a period of 1 to 2 seconds [34].

2.4.1.1 Electroencephalography

The brain is constituted by, essentially, two types of cells: glia cells and neurons. Neurons are responsible for the transfer and processing of information in the form of action potentials or release of chemical substances to other neurons in the brain [39]. These may have various shapes and are constituted by dendrites, axons and the soma. Dendrites are the structures that receive the information passing to the soma, which is the body cell where the cell nucleus is, that may summate the inputs from various dendrites to form an action potential that flows trough the axon to further pass it to the next cell [39].

Information may be transmitted by chemical or electrical signals in the brain. The chemical pathway consists on the release of neurotransmitters to the postsynaptic gap, acting on the proteins of the membrane of the receptor neuron, exciting it, inhibiting it or altering its sensitivity to other stimulus. This pathway, transmits information in one way only, from one neuron to the next.

Electrical signals may be generated by neurotransmitters or by other electrical signals. Cell membrane of neurons have voltage-gated channels that respond to the gradient of voltage between the interior and exterior of cells, by opening if the difference is enough. Specifically, cation channels allow the passage of Na^+, Ca^{2+} and K^+, while anion channels allow the passage of Cl^- and other anions. The electric potential inside the cells is typically lower than the electric potential outside the cells, though this equilibrium is maintained by active transport of the respective ions to and from the cell, hence involving energy consumption.

Neurotransmitters act by opening cation or anion channels, instantaneously altering the polarity of the cell. If the cell becomes more negative, it is considered an inhibitory stimulus and if it becomes less negative, it is considered an excitatory stimulus, because action potentials are only formed if the tension rises above a threshold. These action potentials are created and propagated in the axon due to the high number of voltage-gated channels of this component of the cell.

These processes generate currents in the brain that, in high number and in synchrony, have a constructive effect and can be measured outside of the head.

EEG consists on the measurement of electrical currents that are generated in the brain by the placement of electrodes in direct contact with the skin on the surface of the head. This technique captures both intensity and the pattern of activity that depend on the excitation level of the brain and synchrony of the firing neurons [39].

The measured signals are mainly the result of the postsynaptic potential and not the action potentials, because action potentials vanish too rapidly in order to have constructive effects, while postsynaptic potentials have a duration of tens or hundreds of ms (instead of 1 ms of action potentials), giving the chance for the summation of potentials, as illustrated in Fig. 2.2 [40]. Thus, the measured signals are secondary, which means they correspond to the extracellular currents (primary currents would correspond to intracellular currents).

Synchrony is necessary because the currents generated by individual action potentials

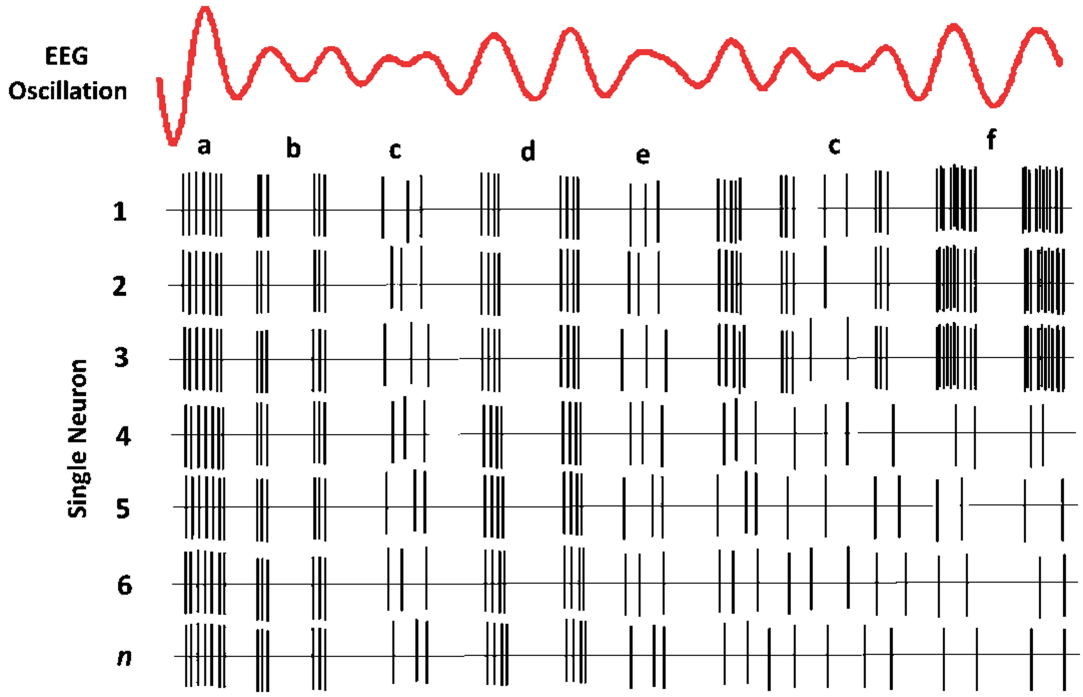

Figure 2.2: Illustration of how single action potentials contribute little to the full EEG signal due to their low amplitude. Excitatory postsynaptic potentials result from the summation of numerous action potentials. For example, a), d) and f) have higher amplitude than b), c) and d) given the first have more synchronous firing than the latter. Adapted from [39].

are not enough to be indirectly measured outside of the brain, as shown in Fig. 2.3. Furthermore, if neurons fire irregularly or in asynchrony, currents may be annulled, decreasing the intensity of the measured signals. Thus, the intensity of the measured signals is inversely correlated with the degree of mental activity [39].

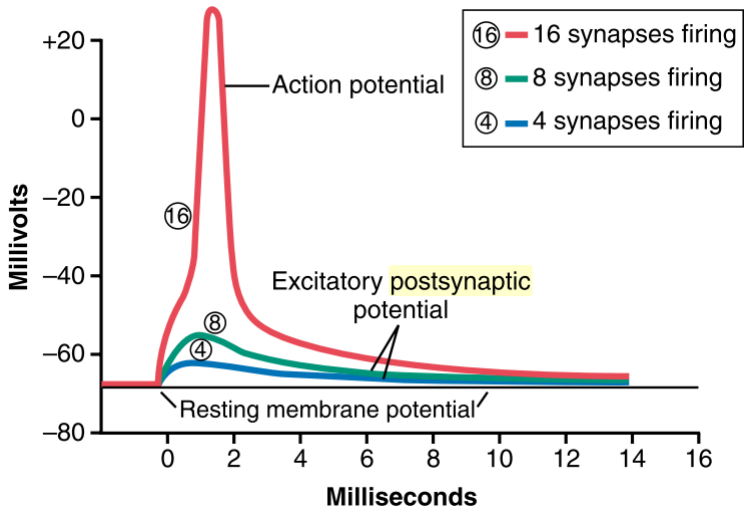

Figure 2.3: Excitatory postsynaptic potentials demonstrating how the summation of action potential contributes to the amplitude of the measured signal. From [39].

EEG is characterised by intensities between 0 and 200 mV and frequencies from 0 to above 50 Hz, which are divided in alpha, beta, theta and delta waves as demonstrated in Fig. 2.4 [39]:

Alpha Waves - characterised by frequencies 8-13 Hz that appear in wake and quiet adults during resting state of cerebration and are more intense in the occipital area.

Beta Waves - characterised by frequencies of 14-80 Hz that appear in directed attention to a specific mental activity. It usually involves lower intensity due to lower synchrony caused by the activation of more neurons. These are better detected specifically in the frontal and parietal regions of the brain.

Theta Waves - characterised by frequencies of 4-7 Hz and occur naturally in the parietal and temporal regions in children. They also appear in adults during emotional stress, specifically, during disappointment and frustration. Theta waves may also occur in brain disorders, usually during degenerative states.

Delta Waves - characterised by frequencies lower than 4 Hz and high intensity. They occur in very deep sleep, during childhood and organic brain disease.

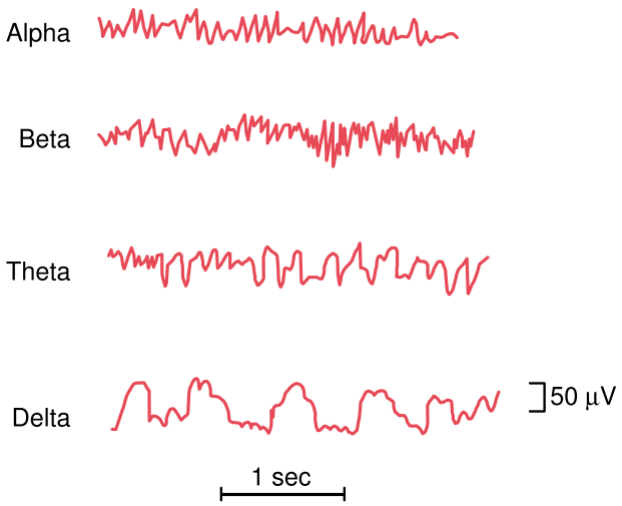

Figure 2.4: Common brain waves found in the normal EEG. From [39].

Thus, contrary to intensity, the frequency of the measured signals are directly correlated with the degree of mental activation.

Typically, EEG recordings involve a high number of sensors following the 10-20 system, but can also use fewer number of sensors [41]. The advantage of using more sensors is that they provide more information than a system with less sensors, but the setup of the electrodes takes longer and more computational power is necessary to process all information. Thus, a trade-off is usually considered and non-medical applications typically use a reduced number of sensors (less than 50 and typically around 30), while medical applications use more than 100 sensors [41].

Once the measured signals reflect macroscopic local field potentials of intracranial acquisitions, temporal resolution of those potentials is preserved in the order of milliseconds, but spatial resolution is low (in the order of 6-9 cm [40]). Spatial resolution is also affected by the volume conduction, which has the effect of smearing the signals when passing through the surrounding tissue and the skull [40].

Nonetheless there are considerations regarding the acquired signals that need to be made that involve complex mathematical formality, primarily to estimate source localisation, as exposed in [42]. This technique has been widely applied due to its many advantages over other techniques. EEG being a widely studied technique, there is extensive information available about all of its aspects - instrumentation, signal acquisition, electrodes placement and signal processing [41, 43]. Besides, instrumentation is typically small and portable, being usable in outdoor settings, thus studies are not restricted to laboratory [44], though results may be affected by the constrained behaviour of users due to the unnatural sensors on their head.

One of the drawbacks of this technique is the fact that intensity may be influenced by the magnitude, synchrony of sources and the area of the activated surface, but it is not possible to identify what is the underlying cause for the variation of the signals, thus, it is an ambiguous measure [40].

Furthermore, the acquired signal is the result of the difference potential between two electrodes and, in referential montages, one of them should be electrically null, but such point does not exist in the human body. Thus, frequency, amplitude and phase of the acquired signals may be ambiguous, because they depend on the positioning of the reference electrode [40]. The other modality, bipolar montage, refers to relative measurements of pairs of scalp electrodes and, thus, is always relative [45].

Finally, EEG recordings are commonly degraded by artefacts. Such artefacts may arise from muscular activity of the scalp, eye blinks, swallowing and general eye and tongue movements [45].

2.4.1.2 Functional Near Infrared Spectroscopy

Metabolic processes may require the consumption and supply of oxygen and glucose to produce energy in the form of ATP. As aforementioned, those processes involve the response of increased blood perfusion to the activated tissues due to the activation of brain areas, which is called neurovascular coupling [38].

When activated, the metabolic activity of the neurons rises, increasing the consumption of energy and the production of toxins, resulting in a decrease of the concentration of O_2Hb and increase of the concentration of Hb. The toxins signal the body to increase blood flow to the region of interest, augmenting local oxygenation, which is maintained higher than the metabolic needs, resulting in an increased concentration of O_2Hb and a decreased concentration of Hb locally [38].

Haemodynamic based techniques, namely, fNIRS and fMRI, acquire signals based

on the concentration of the two different chromophores of haemoglobin, O_2Hb and Hb (they are referred as chromophores because the two forms exhibit different colours when exposed to light) [46].

fNIRS is an optical technique that considers the differences of the optic properties of the two chromophores of haemoglobin, namely the absorption coefficients, in order to generate the output signal [47]. As illustrated in Fig. 2.5, those differences happen in most of the utilised spectrum.

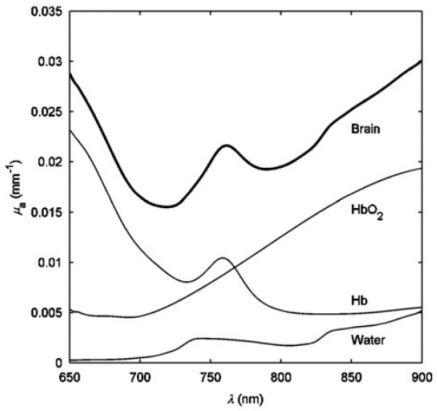

Figure 2.5: Absorption coefficients of the brain and its components, namely, oxygenated haemoglobin (HbO_2 in the figure), deoxygenated haemoglobin (Hb in the figure), and biological tissues, which is comparable to water given the wavelength. The brain is a result of the addition of all tissues. The plot also shows that the absorption coefficients of the chromophores invert around the 775 nm, hence the importance of using LEDs with different wavelengths. From [48].

This technique is made possible due to the near transparency of the surrounding tissues to the light in the spectrum between 650 nm and 1000 nm. The dominant interaction of light with the tissues in that spectrum is scattering, allowing for the photons to reach blood vessels, but also introducing noise in the system, especially considering the multiple types of tissues the light must travel to reach them: skin, skull, superficial blood vessels and cerebrospinal fluid [49]. However, it is the scattering process that allows to detect the radiation non-invasively, that causes the redirection of photons outside of the tissue, as illustrated in Fig. 2.6 [50]. According to Harrison and Hartley (2019), depth specificity is approximately half of the distance between pairs source-detector that should be 30-50 mm in adults and 20-30 mm in infants [51].

The traditional cwfNIRS consists of the coupling of source-detector pairs, which continuously illuminate the surface of the skin and then measure the outcoming intensity of the corresponding wavelength, as illustrated in Fig. 2.6 [48]. Given the different absorption coefficients, the measured light intensity of each specific wavelength informs about the relative concentration of each chromophore. Specifically, red and infrared light sources are used simultaneously, hence, according to Fig. 2.5, on the onset of activation, when the concentration of Hb is higher than the concentration of O_2Hb, the measured red

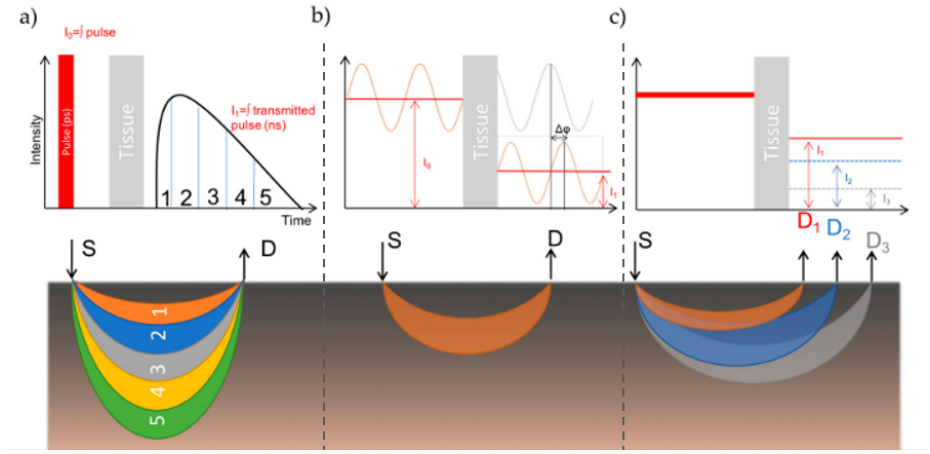

Figure 2.6: Illustration of the functioning of the fNIRS. The physical assumptions regarding each type of fNIRS are illustrated: a) shows TD-fNIRS, b) shows FD-fNIRS and c) shows cwfNIRS. Adapted from [50].

light intensity is lower than the infrared light intensity. On the other hand, after the latent period of one to two seconds required for the perfusion of blood to happen, the relative concentrations invert, and the same happens with the intensity of the two different light beams.

Nonetheless the heterogeneity of tissues through which the light scatters, it is usually considered the simplification that those tissues are optically homogeneous and so, the modified Beer-Lambert law, shown in Eq. 2.1, is applicable, allowing to calculate the relative changes of the concentration of each chromophore [48].

$$\Delta A = \ln(\frac{I_{det}}{I_0}) = L\Delta\mu_a \tag{2.1}$$

This law states that the variation of the fraction of intensity of the detected light (I_{det}) that goes through a material over a baseline measurement (I_0 is a prior measurement) depends on the variation of its absorption coefficient ($\Delta\mu_a$) in the tissues and the mean path length of the detected photons (L) [48, 52]. According to [48], $\Delta\mu_a$ corresponds to the weighted sum of the change in concentration of each chromophore, as expressed in Eq. 2.2.

$$\Delta\mu_a = \alpha_{O_2Hb} \times \Delta c_{O_2Hb} + \alpha_{Hb} \times \Delta c_{Hb} \tag{2.2}$$

This may be used to derive the expressions related to the variation of the concentration of each chromophore, as depicted in Eq. 2.3 and Eq. 2.4, where $\alpha_{XHb}^{\lambda_X}$ is the absorption coefficient of the photon with wavelength λ_X for the chromophore XHb.

$$\Delta c_{O_2Hb} = \frac{\alpha_{Hb}^{\lambda_1}\frac{\Delta A^{\lambda_2}}{L^{\lambda_2}} - \alpha_{Hb}^{\lambda_2}\frac{\Delta A^{\lambda_1}}{L^{\lambda_1}}}{\alpha_{Hb}^{\lambda_1}\alpha_{O_2Hb}^{\lambda_2} - \alpha_{Hb}^{\lambda_2}\alpha_{O_2Hb}^{\lambda_1}} \tag{2.3}$$

$$\Delta c_{Hb} = \frac{\alpha^{\lambda_1}_{O_2Hb} \frac{\Delta A^{\lambda_2}}{L^{\lambda_2}} - \alpha^{\lambda_2}_{O_2Hb} \frac{\Delta A^{\lambda_1}}{L^{\lambda_1}}}{\alpha^{\lambda_1}_{O_2Hb} \alpha^{\lambda_2}_{Hb} - \alpha^{\lambda_2}_{O_2Hb} \alpha^{\lambda_1}_{Hb}} \tag{2.4}$$

Thus, this technique allows to calculate the relative changes of concentration of each chromophore rather than their absolute values. Furthermore, since it is not possible to measure the path length of photons by this technique, it is usual to use tabulated values or modelling results, which may compromise the accuracy of the calculation [50].

To measure the absolute values, there are the techniques:

- Frequency domain functional Near Infra-Red Spectroscopy - radiation is emitted using intensity modulated to a specific frequency, allowing the determination of the phase shift, in addition to the attenuation, when the scattered light is detected outside the head, as illustrated in Fig. 2.6b. Thus, the path length of photons can be assumed and the exact degree of light scattering calculated, hence, the concentration of the chromophores are accurately determined [53].

- Time domain functional Near Infra-Red Spectroscopy - photons are emitted in short pulses (typically in the picosecond range) and are attenuated and broadened in time (typically to nanoseconds) by the scattering and absorption effects of the tissues. Thus, using a probe to collect the scattered photons, the measured signal is the distribution of time of flights (see Fig. 2.6a) [50, 54]. Through different mathematical modelling (e.g., finite element methods, Monte Carlo methods, inverse models or semi-empirical approaches), researchers were able to accurately calculate the optical parameters that define tissue-photon interaction and, thus, calculate the absolute concentration of the haemoglobin chromophores using the Beer-Lambert law [54].

Though the possibility to calculate the absolute concentrations of Hb and O_2Hb, these techniques are more expensive relative to the cwfNIRS and the instrumentation is larger, heavier and more complex [55].

Overall, fNIRS systems present a time sensitivity of 0.1-1 second, spatial resolution of less than 1 cm and depth resolution dependent on the source-detector distance (the longer the distance, the higher the sensitivity and higher the attenuation, thus the worse signal to noise ratio (SNR)) [49]. Furthermore, in the case of cwfNIRS, commercially available devices allow to acquire data in natural positions, due to the small size, enabling off-laboratory acquisitions [46]. Also, once the commercially available devices are portable and silent, they may be used in social situations. However, these may be hindered by the fact that the devices are visible, which might influence the natural behaviour of tested subjects [49].

Further drawbacks of fNIRS technique are: impossibility to access non-cortical structures of the brain and no-depth information, sensitivity to motion artefacts, delay between onset of cerebral activation and its detection, and influence of superficial tissues on the acquired signals.

Near-infrared diffuse optical tomography (DOT) is a fNIRS technique that attempts to gain depth sensitivity by coupling multiple fNIRS channels with different source-detector distances in a high density grid of optodes, with lost of whole head mapping [56]. Another technique to overcome one of the drawbacks of typical fNIRS techniques consists on the coupling of short distance source-detector pairs, that measure systemic haemodynamic fluctuations of superficial layers, allowing to remove this noise from the signals measured with the longer distanced source-detector pairs [49].

2.4.2 Techniques Comparison

The main advantage of EEG relative to fNIRS is the superior temporal resolution, which is able to capture brain activity in the millisecond range. Furthermore, it measures direct brain activity, while fNIRS measures the haemodynamic response to that activity. Another advantage is the fact that being a widely used technique in multiple applications, there are numerous solutions and information for instrumentation and signal processing, which does not occur for fNIRS.

Its main disadvantage is the poor spatial resolution, which hinders the source localisation methods and forces the use of models for both sources and the head to estimate it. Thus, as an imaging technique, EEG is highly inaccurate. Another disadvantage is the number of electrodes necessary for acquiring data and the process required to place them, as it involves depositing gel on the head in order to increase conductivity between electrode-skin or placing adhesives on the head. Additionally, depending on the number of electrodes, the process to place them may be tedious, due to the precision involved to respect the standard 10-20 system. The fact that the intensity, frequency and phase depend on the placement of the reference electrode may also affect the comparison of results between different acquisitions. Moreover, since the electrodes are placed on the skin, artefacts from muscular and other organs activity often occur. Finally, EEG is prone to electromagnetic noise while fNIRS, being an optical sensor, is not.

One advantage of fNIRS relative to EEG is, even though a high number of channels may be used, optodes are used instead of electrodes, which need to be in direct contact with the skin but do not require the usage of gel or other types of adhesives. fNIRS also allows to do acquisitions in natural settings off-laboratory, though being susceptible to movement artefacts.

The lack of depth sensitivity of the fNIRS technique is the main disadvantage, though EEG is also limited in that aspect. Furthermore, the model of the head and of the underlying physical process (photon scattering) is also needed to compute the underlying quantities that this technique measures, namely, the concentration of Hb and O_2Hb. Although TD-fNIRS and FD-fNIRS allow for the calculation of absolute aspects of optical properties of tissues and to calculate the absolute concentrations of the two chromophores of haemoglobin, the most wide spread technology of cwfNIRS only allows to calculate the relative concentrations, which depend on the baseline level and, thus, the comparison

between acquisitions may be ambiguous. Also, source-detector distance determines the results, namely the depth of the measured signals, which hinders the comparison of results from different devices. Artefacts from superficial blood flow and optical properties of the multiple mediums between the optodes and the blood vessels in the brain (for example, skin colour, hair colour, and skull thickness) can also alter the measured results.

Nonetheless, measurements using only one source-detector pair have been made and results show that, for specific applications, it is feasible because validation results show positive correlation with standard devices with higher number of sensors, thus, the portability is increased and acquisitions in natural settings are facilitated [57]. This highly simplifies the optodes placement, though it requires the identification of the correct placement given the specific activity. According to [39, 57], dorsolateral prefrontal cortex should be monitored during learning tasks, because it is responsible for carrying thought processes and is involved in the storage and transmission of working memory, which is used to combine new thoughts as they form in the brain.

Though the correct placement for fNIRS optodes have been identified, it measures indirect metabolic activity, contrary to the EEG, which measures direct activity of the brain. Thus, though most studies using EEG involve the employment of a large number of channels, theoretically, it should be possible to use only one pair of electrodes. However, due to the volume conduction effect, the signals measured in one electrode would be influenced by signals originated in other regions of the brain. The solution would be to use a full cap with high number of electrodes, which would greatly increase the complexity of electrode placement and signal analysis, but would allow to locate the sources of activation and to extract more information.

Finally, both fNIRS and EEG deal with models for the head, dealing with photons scattering and electricity conduction, respectively, and so the outcomes depend on the quality of those models.

2.4.3 Indirect Cognitive State-Related Biosignals

In this subsection, we will delve into the biosignals that may indirectly infer about the cognitive state of individuals, namely, ECG, EDA, RIP, and ACC.

These biosignals originate from different physiological processes, such as, heart function, respiration, sweating, and involuntary movements, which are regulated by the Autonomic Nervous System (ANS) [58]. Thus, the biosignals allow us to indirectly infer about the cognitive state by allowing us to infer about the state of the ANS.

ANS is composed by the sympathetic nervous system and the parasympathetic nervous system [58]. Specifically, the sympathetic nervous system's action increases when stress levels increase, when an individual is in danger or if an individual is physically active, which leads to increased heart rate, increased breathing ability, and increased sweating. On the opposite, the parasympathetic nervous system's action increases when stress levels are low and in a rest and digest state, leading to decreased heart rate, lower respiration

ability, and decreased sweating [58, 59].

2.4.3.1 Electrocardiography

The heart pumps oxygen-rich blood from the left atrium to the left ventricle to the aorta and then into the body. Analogously, oxygen-poor blood is pumped from the right atrium to the right ventricle to the lungs. What causes the heart to function as an efficient pump is the spread of the action potential that causes regular and repetitive myocardial contractions. The sinoatrial node, also known as the heart's pacemaker, is located on the posterior right atrium and is responsible for initiating electrical activity. After initiation of an action potential, it spreads to the atrioventricular node via both atria. The atrioventricular node delays the spread of action potential to the ventricles due to its slow conduction. The action potentials reach the left and right bundle branches and the Purkinje fibers coming from the atrioventricular node. The spread of action potentials through the Purkinje fibers is very fast, enabling the activation of ventricular cells simultaneously and hence an effective pumping of blood out of the ventricles.

The cardiac cycle includes the diastole, which consists of the refilling of the heart chambers with blood, and the systole, which corresponds to the contraction and emptying of the heart chambers from blood. The excitation of a heart muscle cell during the cardiac cycle leads to an electric dipole because the cell becomes more positive compared to neighbouring cells, leading to the generation of an electric vector. A sum vector originates through the spreading of the excitation to the neighbouring cells. During the spread and the regression of excitation, the shifts of the charges can be measured from the skin surface. Each part of the heart has different timing for the excitation and spread of the action potentials, which affects the components of the ECG signal.

Figure 2.7 illustrates the cardiac cycle with its multiple components in a typical ECG signal. The first component is the P-wave, which represents the contraction and spread of excitation of the atria. The PQ-interval corresponds to the spread of excitation from the atria to the ventricles. During the QRS complex, the ventricles contract and spread the excitation. In the ST-interval, the ventricles are fully excited and the excitation begins to regress. The T-wave represents the regression in the ventricles and the TP-complex represents the unexcited atria and ventricles. The RR interval is typically used as a measure of the interval between consecutive cardiac cycles.

The standard measurement technique of the ECG is the application of 12-leads to cover all of the heart's information in three directions as illustrated in Fig. 2.8. According to Einthoven, the first three leads are the bipolar leads, representing the frontal plane with both arms and left leg, as illustrated in the Fig. 2.8A. The first lead (I) is the measurement from right arm (-) to left arm (+). The second lead (II) measures from right arm (-) to left leg (+), and the third lead (III) measures from left arm (-) to left leg (+). The augmented unipolar limb leads (aVR, aVL, and aVF) also represent the frontal plane, as shown in Fig. 2.8B. The unipolar chest leads (V1-V6) represent the horizontal plane with electrodes

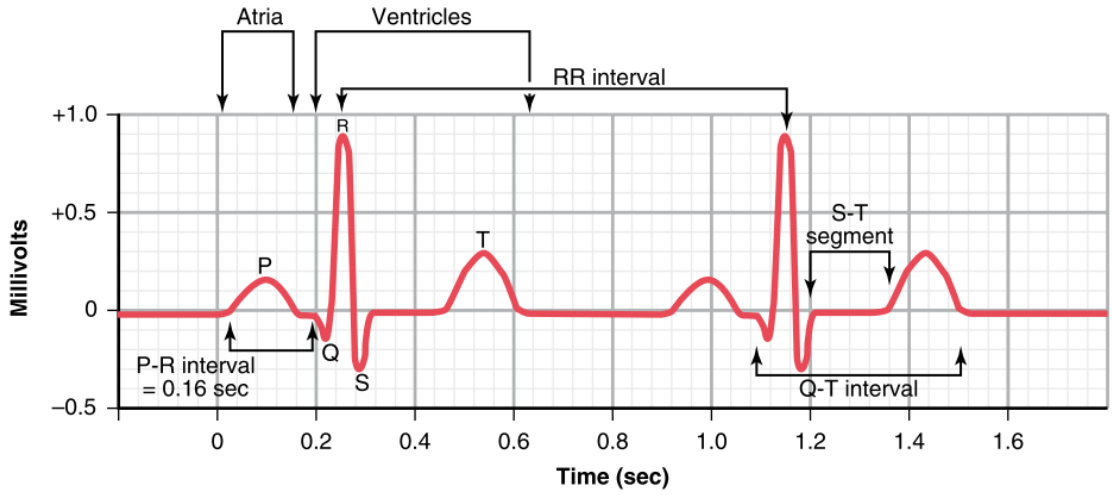

Figure 2.7: Illustration of the ECG signal and its various waves and components. P-wave (atrial excitation), PQ-interval (atria to ventricles excitation), QRS complex (ventricular contraction and excitation), ST-interval (full ventricular excitation, regression begins), T-wave (ventricular regression), TP-complex (unexcited atria and ventricles), and RR interval (consecutive cardiac cycle measure). From [39].

positioned in six locations on the chest.

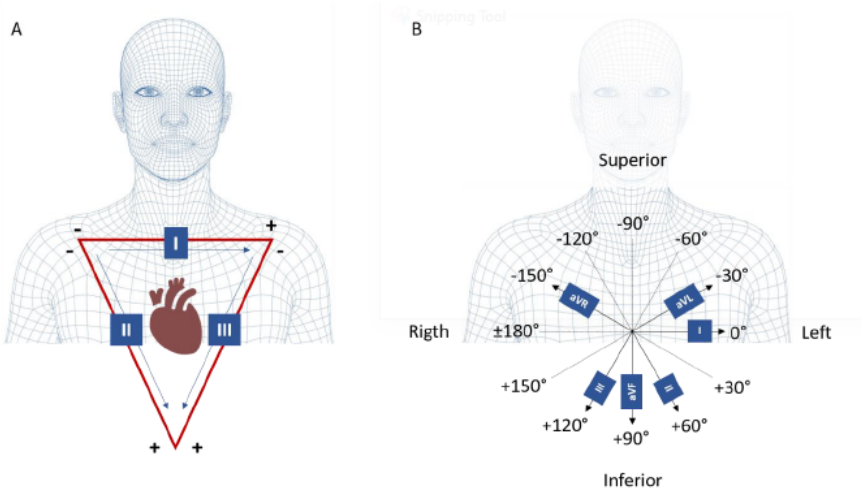

Figure 2.8: Standard measurement technique of the ECG consisting of the application of 12-leads to cover all of the heart's information in three directions. According to Einthoven, the first three leads are the bipolar leads, representing the frontal plane with both arms and left leg. The augmented unipolar limb leads (aVR, aVL, and aVF) also represent the frontal plane. The unipolar chest leads (V1-V6) represent the horizontal plane with electrodes positioned in six locations on the chest. From [60].

ECG can be related to the cognitive state by considering the impact that the ANS has on it. Namely, the sympathetic nervous system increases the heart rate, the conduction velocity, and contractility, while the parasympathetic reduces heart rate, conduction

velocity, and atria contractility [58]. Thus, by measuring heart rate and heart rate variability, it is possible to infer about the cognitive state of an individual, given the correlation between the two aforementioned systems.

2.4.4 Electrodermal Activity

EDA refers to the changes of the electrical conductance of the skin due to the sweat glands activity in the skin and it corresponds to an indirect measure of the ANS activity, as previously mentioned.

The skin is the body's first physical barrier against external agents besides being the largest organ of the human body. It is responsible for numerous functions, such as, protection against microorganisms, dehydration by avoiding loss of fluids, ultraviolet light, and mechanical damage, and is also responsible for the sensation of pain, temperature, touch, and pressure. It also has the function of mobility by allowing smooth movement of the body, endocrine activity being involved in the production of vitamin D, exocrine activity by secreting sweat among other fluids, and finally, temperature regulation by conserving or releasing heat, helping to maintain the body's water and homeostatic balance [61].

The skin is composed of three layers: epidermis, dermis, and hypodermis or subcutaneous tissue. Being the outermost layer of the skin, the epidermis serves as a waterproof barrier and contributes to the skin tone while the dermis, the layer right beneath the epidermis contains connective tissue, hair follicles, blood vessels, lymphatic vessels, and sweat glands. Finally, the deepest layer, the hypodermis, is composed of connective tissue and fat [61].

Given the mentioned functions of the skin, sweat is an integral part of it as a way to regulate temperature and the body's homeostasis. Sweat glands are divided in eccrine and apocrine, that differ in terms of embryology, function, and distribution. Eccrine sweat glands have a thermoregulatory function by evaporative heat loss. In this case, when the temperature of the body rises, these glands release water onto the skin surface, which in turn evaporates by an endothermal reaction, cooling the surrounding skin and blood beneath it. These glands also participate in ion and nitrogenous waste excretion and they can produce between 500 mL to 750 mL in a day in response to emotional or thermal stimuli [62]. Apocrine sweat glands on the other hand, start to function during puberty due to the stimulation of sex hormones and are present in specific locations, such as, the groin and axillary regions [62].

In 1888, Féré studied the electrical properties of the skin, namely, that its conductivity increased when an external stimuli was applied and the next year, in 1889, Tarchanoff found that the electrical changes also happened even without the application of an external current on the skin [63]. In fact, the larger the amount of sweat, the higher the conductivity of the skin. The number of activated sweat glands and the quantity of produced sweat is controlled by the ANS. The measured signal is called EDA, Skin

Conductance Response (SCR), Sympathetic Skin Response (SSR), Galvanic Skin Response (GVR), or Psycho-Galvanic Reflex (PGR).

The measurement sites for this kind of signal are typically the palms of the hands or feet, given the high density of eccrine sweat glands besides these being more responsive to psychological stimuli than to thermal stimuli [63]. Thus, the setup commonly consists of two electrodes placed on the palms or fingers to measure the current flow between them to determine the changes on the electrical conductance.

The EDA signal is composed of two components: a phasic component and a tonic component (shown in Fig. 2.9). The tonic component corresponds to slow changes of high amplitude and is low during sleep and rises during mental work, while the phasic component may relate directly to the presentation of stimulus or it may occur to non-specific stimuli. These relate to attention and correlate with novelty, intensity, and significance [63].

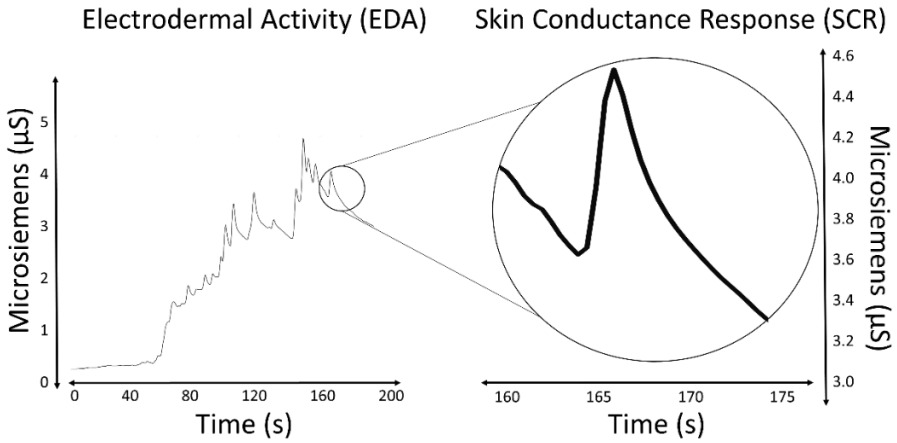

Figure 2.9: Illustration of the typical EDA signal. The left hand side shows the full signal while the right hand figure shows a zoomed in portion showing one part of the phasic component of the signal. From [64].

2.4.5 Respiratory Inductive Plethysmography

Respiration is a vital function that involves inhaling and exhaling providing oxygen and expelling carbon dioxide and other toxic gases. This involves the movement of the rib cage, as its volume increases during inhalation and decreases during exhalation, meaning that its cross section varies accordingly.

Respiratory Inductive Plethysmography (RIP) corresponds to an elastic band with an electric conductor inside and a constant current passing through the conductor. In this case, the magnetic flux passing through the cross section of the elastic band is given by the expression of Eq. 2.5, where Φ is the magnetic flux (measured in Wb or T·m^2), A is the cross section of the band (measured in m^2), influenced by the thorax, B is the magnitude of the magnetic field (measured in T or in Wb/m^2), and θ corresponds to the angle between the magnetic field vector and the cross section.

$$\Phi = B \cdot A \cdot cos(\theta) \tag{2.5}$$

Given Faraday's Law, the induced electromotive force opposes the variation of the flux, thus, mathematically, it is described as seen in Eq. 2.6.

$$E = -\frac{d\Phi}{dt} \tag{2.6}$$

This means that, given the current and the magnetic field are constant, the electromotive force changes with the change of the area of the cross section of the elastic band. Thus, when the rib cage expands the electromotive force increases and when it contracts the electromotive force decreases, in absolute terms, as shown in Fig. 2.10. This means that, in cases where the band is stationary (e.g., holding breath, apnoea), the measured signal is 0, as there is no variation of the magnetic flux.

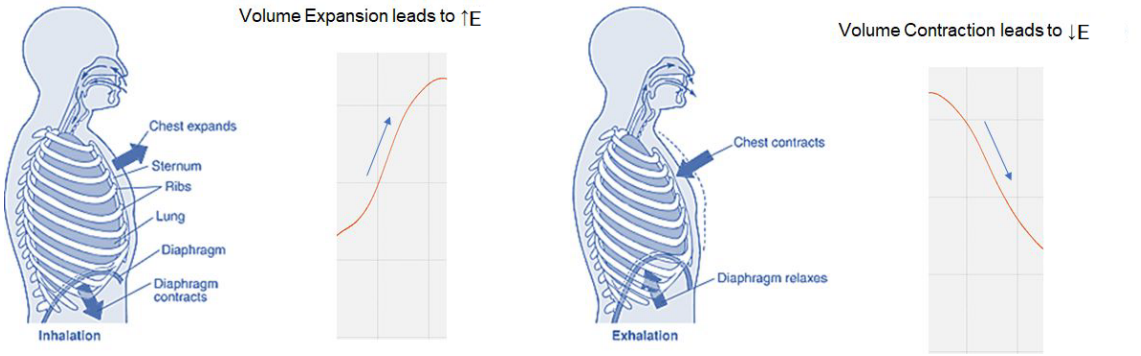

Figure 2.10: Illustration of how the signal from the RIP signal is generated. When the thoracic cage expands, the cross section of a positioned RIP elastic band increases and, thus, signal also increases. During exhalation, the thoracic cage contracts and the cross section decreases leading to a decreasing signal. Adapted from [65].

With this device, it is then possible to extract the respiratory rate and to infer about the deepness of respiration by measuring the time between peaks and the amplitude of those peaks, respectively.

2.4.6 Accelerometry

Accelerometers are devices that measure acceleration in relation to free-fall, which refers to the acceleration that people and objects experience relative to an inertial frame. These measure static or dynamic acceleration, such as, gravity or human movement, respectively, and is measured in m/s^2 or in g. Thus, at rest, the measured value is 1 g or approximately 9,8 m/s^2, while a dynamic accelerometer would measure 0 g or 0 m/s^2.

The functionality of an accelerometer hinges on the relative motion of a proof mass, often referred to as a seismic mass. This mass is affixed to a mechanical suspension system and moves concerning a reference inertial frame. The displacement of the mass within the system is quantified using an electrical mechanism. This mechanism typically

involves the use of piezoelectric materials that alter their electric potential in response to volume changes, piezoresistive materials that modify their resistance properties in a manner analogous to piezoelectric materials, or capacitive components. In the case of capacitive components, variations in the distance between two plates lead to changes in capacitive properties. This alteration enables the accelerometer to gauge the proper acceleration experienced by the proof mass [66].

Given its functioning, if attached to a person, it can measure the acceleration of the body part that person moves. For example, if attached to the head of a person, it would be possible to measure its acceleration and even the relative position, if we consider the signal of the 3 axis xx, yy, and zz. Given that head movements and pose can be used to estimate mental states such as fatigue and drowsiness (e.g., [67]), it is possible to use this sensor as a complement to the remaining to get more information about the mental states of subjects. Fig. 2.11 shows an example of the accelerometer signal where it is possible to see the differences of the signals of the three axis in three different poses when learning in e-learning settings.

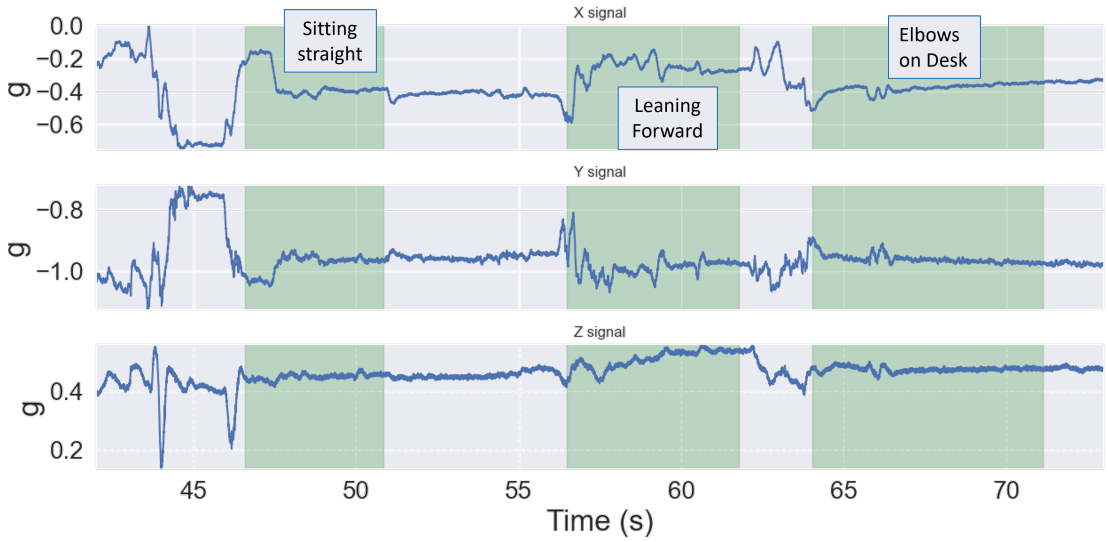

Figure 2.11: Examples of an ACC signal showing its three axis. The green coloured portions represent the signals during the poses of sitting straight, leaning forward while sitting, and elbows on desk in order. The signals of the three axis differ and the changes of poses are also obvious by the increased energy of the represented signals.

2.5 Human Computer Interaction - Mouse Tracking

HCI studies the implications and influences between computer interfaces or intelligent agents and people [68]. On one level, HCI includes the investigation and design of methods and techniques to enhance usefulness of computers, while on another, the main focus are intelligent technologies that should adapt the natural interaction methods between

computers and humans [68].

HCI can involve different modalities of data streams, such as, sound (e.g., listening and speaking), reading and writing, visual (e.g., face detection and recognition, facial expression recognition), and others (e.g., biosignals) [68]. According to Cepeda et al. (2019) web behaviour metrics, used in HCI applications, involve mouse tracking, keyboard tracking, geolocation, screenshots, audio, snapshots, video, and touch events that can be used to evaluate the interaction between users and webpages, as is the case of most e-learning courses [63].

In cognitive science, researchers have employed self-report measures, objective behavioural or psychophysiological measures such as eye-tracking, and neuroimaging methods such as fMRI. However, mouse tracking has been added to these methods in recent years [69]. Using mouse tracking to infer cognitive processes relies on the premise that cognitive activities influence real-time motor activations and subsequent cursor movements [69].

One paradigm where mouse-tracking has been used the most is the binary forced-choice, illustrated in Fig. 2.12. There, participants are required to choose between two options represented as buttons on a computer screen while mouse movements are recorded. In this case, the task requires that mouse starts always from the same place on the screen but the timing of starting may vary. For example, it might be imposed a time limit to choose one of the two options or the options may not be shown before mouse movements start [70]. In this case, the behavioural procedure is studied using known metrics: deviation from and towards the chosen option, velocity and acceleration of the mouse movement, entropy, and number of zero-crossings in the axis separating the two available options. Although this paradigm might be useful in specific cases, it restricts movement and does not represent the natural environment of web navigation or e-learning.

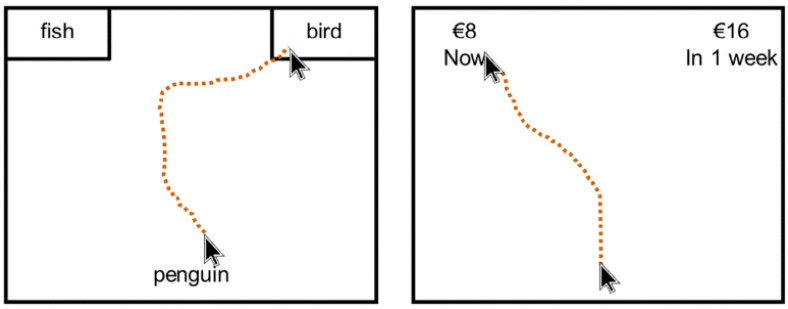

Figure 2.12: Examples of binary forced-choices. The dashed orange line represent the mouse movements performed since the starting point until the choice. Adapted from [69].

In some instances, mouse tracking is applied in less structured, real-world scenarios. Various authors adopt diverse approaches to handle such data, which may not yield features as easily interpretable as those in binary forced-choice tasks, where intentions are clear. In these cases, data processing often involves extracting specific metrics like peak velocity and acceleration, transforming data into various types of images and use

computer-vision methods, or directly use Deep Learning algorithms that handle time series, namely, Long-Short Term Memory Neural Network (LSTM) [71–74].

Mouse tracking, due to its versatility in analysing user interface usability, optimizing the overall user experience, studying design aspects, and inferring cognitive processes, holds significant potential for monitoring users' cognitive states within the context of e-learning. Its widespread availability further enhances its applicability in this domain.

2.6 Machine Learning

ML refers to the science of programming computers so they can learn from data, i.e., it gives the computers the ability to learn without being explicitly programmed [75]. This is especially useful in cases where complexity of programming is so high, that it would require too long sets of rules and where it would be too hard to maintain. In such cases, ML techniques would automatically learn the rules and facilitate programming [75].

ML can be divided based on whether the training requires human supervision, whether it can learn iteratively, or whether they are instance-based or model-based.

Regarding the supervision of the training, it can be divided into supervised learning, unsupervised learning, semisupervised learning or reinforcement learning [66, 75].

Supervised Learning - this type of ML involves a training set that consists of training samples and the corresponding labels, that can be represented as $D = \{(x_i, y_i)\}_{i=1}^{N}$, where x_i corresponds to a sample, y_i to its corresponding label, and N to the number of data points of the training set. Analogously, there is usually a test set that is used to assess the performance of the ML models in which the labels are not used or do not exist. Given the trained model, when a point from the testing set is given, the model will assign a label from the training set $y_i, i \in \{1, ..., N\}$. On this level, a sample x can be regarded as a single point in the most simple case, but it can also represent an array (e.g., time series, features) or a matrix (e.g., image). Moreover, the label or output y may be continuous, in which case the problem is called regression, or discrete, in which case it is a classification problem.

Unsupervised Learning - in this case, the training data is unlabelled, meaning that the training data is represented simply as $D = \{(x_i)\}_{i=1}^{N}$, where x_i corresponds to a data sample and N to the number of samples. In this case, the objectives might be to group the input data into different clusters, to detect anomalies in data, for visualization and dimensionality reduction, and association rule learning.

Semisupervised Learning - algorithms that can handle partially labelled data, where it is mostly unlabelled with some examples of labelled samples. In this case, most algorithms combine unsupervised with supervised learning algorithms. For example, an algorithm can firstly work by finding patterns in the data and then use

the examples that are labelled to label the remaining data according to the similar samples.

Reinforcement Learning - in this case, the model or agent, observes an environment and selects actions to perform. In return it receives rewards if the action is positive or penalties if it is negative. The agent then learns the best strategy to get the most rewards over time. The strategy defines what actions the agent performs in each situation.

In terms of iterative learning, a given model might be able to learn incrementally, called online learning, or it might be unable to do so, which is called batch learning. In online learning the system can learn on the fly by feeding it new input data instances sequentially or in mini-batches. In this case, each learning step occurs fast and is computationally cheap and it is useful in cases where the system needs to adapt continuously, for example to learn the behaviour of stock market changes over time. On the other hand, if the system cannot learn incrementally, it is called batch learning. In this case, the system is trained using all available data and is typically done offline to be later launched into production, running without any more learning. Here, if more data needs to be introduced into the system, it needs to be retrained, the old model needs to be stopped and replaced with the new one in production [75].

Finally, ML systems can be divided based on whether they are instance-based or model-based. In instance-based systems, the models learn and adapt solely based on previous examples, which is useful in cases where there is no information about the problem at hand. However, if the problem is somewhat known, it might be interesting to use specific models to learn from examples. For example, if one knew that the solution for a given problem was a linear regression, then it would be more efficient and possibly more accurate, to directly train a linear regression model instead of using a general ML model. The latter case is called model-based learning and is useful in cases where the solution to the problem is known to some extent [75].

2.6.1 Features Extraction and Selection

In the context of ML it is usual to represent the input data as series of features, i.e., characteristics of the data. This is the step of features extraction and the goal is to extract the best characteristics that are able to separate the data into the different classes. For example, regarding the biosignals, it could be interesting to extract features such as maximum, minimum, and average heart rate to separate individuals by their age. In case the extracted features are not relevant, they might increase the complexity of the problem by increasing the number of dimensions and, thus, of the number of regions of interest in the data. This problem is called the curse of dimensionality, and it might also happen when there are redundant features that carry similar information, since it also increases the number of dimensions and complexity, without carrying new information.

In this sense, feature selection might be an important step to avoid the curse of dimensionality. Feature selection consists on the application of techniques or algorithms to remove the features that do not contribute or worsen the performance of the ML algorithms. This can be done by removing or agglomerating highly correlated features that carry similar information for the given problem or removing features that do not contribute to an accurate classification when training.

2.6.2 Features Normalisation and Scaling

Another commonly applied step that might be essential depending on the ML models used, is the scaling or normalization of the extracted and/or selected features. This serves to modify the weight of each feature. For example, if the range of a given feature is in the order of thousand and another is in the order of units, then the first would dominate the results given its weight. In this case, it would be important to scale the features correctly. There are many algorithms to achieve this, however, the most widely applied are simply the min-max scaling and the z-score standardization:

Min-Max Scaling - consists on the application of the Eq. 2.7, that scales the distribution to a range of 0 to 1. In this case, the distribution of values is kept, but all features will weight the same to the algorithm at hand.

$$X_{mod} = \frac{X - min(X)}{max(X) - min(X)} \tag{2.7}$$

Z-Score Standardization - consists on the application of the Eq. 2.8, where μ is the average value and σ is its standard deviation. This modifies the distribution of values so that every feature has mean 0 and standard deviation of 1, which might result in features with different weights depending on the original distribution.

$$X_{mod} = \frac{X - \mu}{\sigma} \tag{2.8}$$

Scaling or normalization can be applied before or after features selection, and the appropriate algorithm can be chosen given the ML model at hand or it can be empirically chosen by evaluating the results and choosing the methods that optimizes them.

The fact that ML depends on these steps—feature extraction, scaling/normalization, and feature selection—represents a disadvantage because an imperfect combination of methods might result in suboptimal outcomes.

2.7 Performance Evaluation

ML models' performances can be assessed using a set of commonly used metrics, namely, the accuracy score, precision, recall, and F1-score. The corresponding equations are in Equations 2.9-2.12, where TP, TN, FP, and FN mean True Positives, True Negatives, False Positives, and False Negatives, respectively, and represent:

True Positive (TP) - samples predicted as the positive class that correspond to that class;

True Negative (TN) - samples predicted as the negative class that correspond to that class;

False Positive (FP) - samples predicted as the positive class that correspond to the negative class;

False Negative (FN) - samples predicted as the negative that correspond to the positive class.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.9}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2.10}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2.11}$$

$$\text{F1-Score} = \frac{2 \cdot (\text{Precision} \cdot \text{Recall})}{(\text{Precision} + \text{Recall})} \tag{2.12}$$

Additionally, the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) might also be reported in cases of binary classification. This curve shows the dependence between True Positive Rate and the False Positive Rate allowing to analyse the models' performance at various threshold levels.

Moreover, it is also common to analyse the confusion matrices that show how the predictions compare with the expected outcomes of classification.

2.8 Summary

In this chapter, the fundamental concepts related to learning and various theories that elucidate this intricate process were presented. The significance of e-learning was also explored, highlighting the possibilities to adapt and integrate emerging technologies into educational systems. Furthermore, the role of biosignals and HCI, namely mouse-tracking, was discussed in the context of identifying cognitive states during the learning journey. Lastly, a foundational understanding of ML was provided.

The complexity inherent in defining learning was demonstrated, as well as how multiple interconnected systems contribute to this phenomenon. While these systems are intricately linked and challenging to study in isolation, they collectively offer insights into understanding the learning process as a whole. By considering these factors during the design of e-learning solutions, we open the door to the potential improvement of the learning experience. Moreover, the integration of cognitive monitoring through biosignals and mouse-tracking enables real-time adaptation of e-learning content, thereby optimizing

the learning process. ML plays an important role in this context by autonomously discerning patterns within the data that signify cognitive state monitoring.

This chapter underscores the multifaceted nature of learning and its dynamic potential when harnessed with advanced technologies and thoughtful design considerations.

3

State of the Art

Cognitive and behavioural monitoring hold extensive relevance across various domains, with numerous studies investigating them in diverse settings. Within this chapter, a concise narrative review of the literature pertaining to the monitoring of cognitive states and behaviour tracking in the context of e-learning is presented. Figure 3.1 shows how these topics overlap in the study of e-learning, namely, when considering users' states evaluation in e-learning contexts.

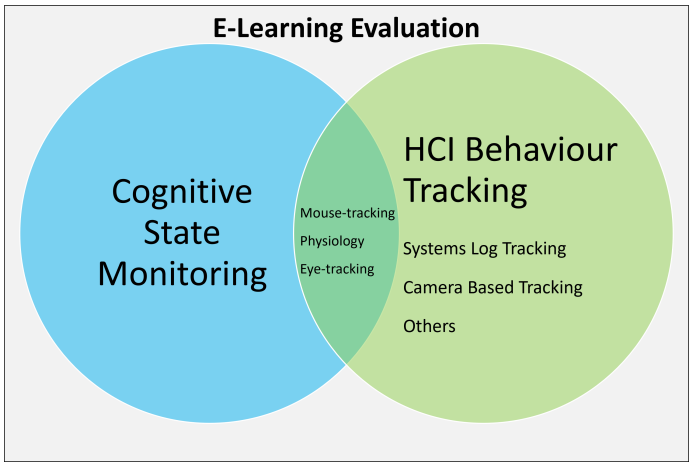

Figure 3.1: Venn diagram depicting the topics discussed in this state of the art.

3.1 Cognitive State Monitoring

Cognitive state monitoring has been extensively studied, and its assessment is usually performed given one of three methods: subjective methods–the self-assessment is typically achieved via questionnaire, being subjective to each person; behavioural methods–study the relation between specific behaviours and cognitive states; physiological-based methods–use physiological sensors to monitor the person and then, by processing the acquiring biosignals, relate them with the different cognitive states [20].

Subjective methods employ questionnaires such as Situational Fatigue Scale [76], State-Trait Inventory for Cognitive Fatigue (STI-CF) [77], or Subjective Exercise Experiences Scale

(SEES) [78] (in the context of cognitive fatigue), which can only be applied at the end of the given tasks. Thus, although useful to assess the real personal perception of cognitive state, these are not applicable unobtrusively, because there needs to be an interruption to answer them. Moreover, such tests are bound in time - the answers are respective to a given instant.

Behavioural methods, on the other hand, can be employed in a time-continuous fashion. In these methods, personal behaviour is tracked using characteristics of the specific tasks at hand, meaning that it is possible to monitor cognitive state over time. However, there is a cost as different tasks involve different types of behaviour, e.g., the behaviour during driving and during computer-based tasks are not comparable, because one involves dealing with the steering wheel and pedals while watching the road, while the other involves interacting with the mouse and keyboard while focusing on a screen. Thus, typically, behavioural methods are specific to different tasks. Such methods include eye-tracking [79] and human-computer interaction features [80].

Finally, physiological-based methods can be continuously applicable and are agnostic to the type of activity or behaviour. In this case, wearable sensors that include EEG and fNIRS can be applied to monitor the mental state of subjects [81, 82]. Then, specific information can be drawn from the generated physiological signals to classify the mental states of the monitored individuals. The main disadvantage is that this monitoring involves using external instruments (the sensors), which might interfere with the natural behaviour of the monitored subjects.

A summary of the pros and cons of each of the cognitive monitoring methods is shown in Table 3.1. Subjective methods are inherently the only subjective, as they require answers to specific questionnaires, deemed as external devices. Behavioural and physiological-based methods are both continuous, but only the first is specific to each individual task or way of interaction. Finally, physiological-based methods also require external devices, namely physiological sensors that may be wearables.

Table 3.1: Characteristics of the the three mentioned cognitive state monitoring methods.

Method	Subjective	Continuous	Specific	External Devices
Subjective	✓	✗	✗	✓*
Behavioural	✗	✓	✓	✗
Physiological	✗	✓	✗	✓

* External device being questionnaires.

Cognitive state includes numerous aspects, namely, attention and cognitive load. While Chapter 2 presents the fundamental concepts mentioned here in more depth, attention can be defined as the process of taking into account a determined number of stimuli while ignoring all others. On the other hand, cognitive load refers to the load on working memory that a given task demands. The fundamental idea of the cognitive load theory, is that learners, in particular, have a limited cognitive capacity to handle information and,

thus, it should not be overloaded for learning to be effective [83]. Cognitive load is typically divided into three: intrinsic - related to the complexity of the task -, extraneous - related to how the task is presented -, and germane - originated by processing and constructing schema to handle new information [84].

Because extraneous cognitive load is a critical aspect of the learning process that can be influenced by external materials, it is important for instructional designers to prioritize reducing this type of cognitive load in the first place. Doing so will increase learners' capacity to focus on intrinsic and germane load [84].

Cognitive load has been increasingly measured using fNIRS and EEG. Commonly, fNIRS is placed to monitor the prefrontal cortex (PFC) as it encompasses executive functions, i.e., high-order cognitive processes [85, 86]. Herff et al. (2013) used fNIRS to monitor participants during various psychological tasks, reaching accuracy results between 62–71% for the discrimination of tasks and baseline [47]. According to the study of Herff et al. (2014) of the following year, cognitive load induced by a memory task, N-Back, and fNIRS-based evaluation of PFC constitute a realistic method for passive Brain-Computer Interface (BCI) [87]. In turn, Harley et al. (2020) obtained an accuracy of 78% using only N-Back tasks [88]. Mazher et al. (2017) on the other hand, used EEG and the results suggested that α waves were the most appropriate brain waves for measuring cognitive load in learning-based cognitive tasks and for discriminating different learning states [89].

In the cognitive monitoring aspect using physiological sensors, BCIs are systems that provide a mean for individuals to communicate with a computer using only their thoughts [90]. This means that, typically, sensors capable of detecting shifts in brain functioning are a requirement. Sensors such as low-channel EEG and low-channel fNIRS can be applied in these contexts (e.g., [91, 92]) due to their ecological validity when compared to other techniques, e.g., MEG, fMRI, and PET, as presented in Chapter 2 [93, 94]. Active BCIs yield the ability to actively control equipment, e.g., communication of patients with mobility problems and the interaction with computers [92]. Passive BCIs, on the other hand, are used to monitor the individual's state, which may serve, for example, as the basis for improving an interface with which the user interacts [95]. Moreover, the detection of a mental state shift is usually automatic by applying artificial intelligence methods, e.g., ML algorithms [92].

Related to cognitive load, CW can be defined as the relationship between mental processing capabilities and the performed task demands [96], being inversely related to performance [89, 96]. While cognitive load is usually divided into the three aforementioned parts, CW is typically a subjective measure of how each person perceives the difficulty of a given task. From a neurophysiological point of view, CW is directly associated with the variation of the concentration of O_2Hb and inversely with the concentration of Hb [97, 98].

Besides cognitive load and CW, attention is crucial for learning performance [8, 9]. Attention can be assessed using passive BCIs by applying appropriate sensors. For example, Ko et. al. (2017) applied a 32-channel EEG to assess sustained attention in the classroom

and demonstrated that there was a relationship between spectral dynamics and the performance in a specific sustained attention task [99]. In another study, Hu et al. (2018) applied a similar EEG sensor with 32 channels and performed classification using ML, with the algorithm k-nearest neighbours, to detect self-reported attention levels in a simulated distance learning environment, achieving an accuracy of around 80% [100]. Fahimi et al. (2018) used a single prefrontal channel EEG to measure attention during a cognitive test (Stroop test) and a neurophysiological test (repeatable battery for the assessment of neurophysiological status—RBANS), concluding that, specifically, the alpha–gamma ratio and theta–beta ratio are related to the individual's response time [101]. The work of [8] aimed at assessing students' attention levels based on features extracted from gamma, beta, alpha, and theta waves of EEG signals and reached an average accuracy rate of around 89%. Instead of BCI, Zhang et al. (2017) took a different approach and assessed attention in the classroom using embedded sensors, namely accelerometers and gyroscopes, as well as cameras, for the analysis of the head motion, pen motion, and visual focus monitoring [102]. The assessment was made with a rule-based approach (rules defined by the authors) and a data-driven approach (applying ML models), reaching around 60% and 80% accuracy, respectively [102]. Still in the classroom, Zaletelj and Košir (2017) used 2D and 3D models of the Kinect One sensor to detect attention based on image-related features, reaching a classification accuracy of around 75% [103]. Abate et al. (2021) developed a system for attention monitoring with integrated feedback for synchronous distance-learning based on image-related features. The system was applied during distance-learning lessons and also in an engineering company. The aim was to provide the information to the supervisors as feedback to decide how to better adapt the content to the overall users' state [104]. Though there is a high prevalence of attention monitoring works related to learning, other applications include driver vigilance [105–107], construction workers [108], and other attention-dependent workers (e.g., nuclear power plant operator [109]).

Cognitive fatigue, which can negatively impact learning, has been subject of assessment using fNIRS in multiple domains, such as image interpretation in medical settings [82], physical performance [110], operational settings [111], during simulated driving [81], and during serious games [80]. However, there is still a lack of evaluation of cognitive fatigue during learning.

Nihashi et al. (2019) studied the relation between the concentration of oxygenated haemoglobin with cognitive fatigue in radiologists using fNIRS sensors, where they found that, in most cases, the concentration of this chromophore was the lowest when the fatigue was the highest [82].

Analogously, Dehais et al. studied the possibility of using a passive BCI involving the fNIRS combined with an EEG to monitor pilots' engagement features for cognitive fatigue detection during simulated and real flights [112]. Considering the fNIRS sensor alone, the accuracy score for the cognitive fatigue detection was around 81.5% in the simulation case and 83.2% for the real flight case. However, the individual differences were evident, as individual accuracy scores ranged from around 70%–90%. The combination of EEG-fNIRS

revealed better results, indicating that additional data sources may be beneficial for the detection of cognitive fatigue.

In the context of cognitive evaluation, the fNIRS Pioneer™, a wearable fNIRS sensor, has been validated and compared to the NINScan device during standard cognitive tasks, reporting a positive correlation of the results by showing an increase in relative concentration of oxygenated haemoglobin (HbO_2) and a decrease in the performance of the user with the increase of task difficulty [57].

Trakoolwilaiwan et al. (2017) used LABNIRS with 34 channels to monitor motor execution during predefined tasks. In this work, the authors classified three classes: rest, right-hand movement, and left-hand movement. The accuracy scores to classify these classes were around 86–93% using ML algorithms. The lowest results (86%) were obtained using ML trained with custom features extraction, while the highest results (93%), were obtained with automatically extracted features using a convolutional neural network, demonstrating the importance of using appropriate features to classify fNIRS signals [113].

Again, inter-subject variability has been studied and is reportedly relevant for a proper analysis of fNIRS. On a more general level, age was demonstrated to influence the cerebral haemodynamic [114]. Namely, older people revealed fewer changes in blood oxygenation when considering the ones with a normal pattern of an increase in the concentration of oxygenated haemoglobin and a decrease in the concentration of the deoxygenated haemoglobin during task execution.

On the individual level, Quaresima et al. (2005) studied the relative changes of the concentration of the chromophores of haemoglobin relative to the execution of a verbal fluency task, where only half of the population studied showed significant changes [115]. Zohdi et al. (2021) also applied the fNIRS during a verbal fluency task, but added systemic physiological activity measurements, namely, Heart Rate (HR), EDA, SpO_2, mean arterial pressure, and $P_{ET}CO_2$ to study the differences in cerebral haemodynamics [116]. Again, the expected pattern was only visible in half of the studied population.

Holper et al. (2011) applied the fNIRS sensors in 11 subjects while performing two tasks: motor imagery and motor execution. On the group level, they demonstrated that it was possible to discriminate between tasks and also between task complexity. Nevertheless, they also verified that the expected pattern of the chromophores was subject dependent, which might contribute to the inter-subject variability expressed in other works [117].

In the specific context of learning, other sensors have been employed to infer about the cognitive state of learners. For example, Kim (2018) used EDA to assess engagement of students in real-time by applying a Fuzzy inference algorithm and sending the results into an app on the teacher's phone in the classroom, although it lacked validation [118].

Gupta et al. (2023) on the other hand, used two datasets containing EEG data to detect attentive and inattentive states using ML algorithms, where features importance was calculated, validated with clustering algorithms, and finally classified using different classifiers [19]. Although they were able to achieve accuracy results above 91% when

testing, the ground-truth of the state of each learner was inferred based on the waves of the EEG signal, which were then used as features for classification, and may have had a positive impact on the results, resulting in optimistic accuracy scores.

Giannakos et al. (2020) conducted a study where they used a wearable device that measured EDA, temperature, HR, and blood volume pressure data to monitor students during learning activities in in-person classes [119]. In this work, the authors showed the feasibility of using such a system to predict the outcomes of standard questionnaires, however, the instantaneous cognitive state of learners was not taken into account.

Furthermore, recent literature has highlighted the significance of emotions in e-learning environments [11]. Some studies advise research on emotions in e-learning to focus on negative activating emotions (e.g. frustration) and on positive activating emotions (e.g. enjoyment) [88, 120]. While most studies on this topic rely on surveys, [10] adds that EDA and ECG might be the future means of assessing emotions.

3.2 Human-Computer Interaction Behaviour Tracking

HCI behaviour tracking methods have been applied in numerous fields to monitor and possibly improve users' experience in specific circumstances. An important area closely related to e-learning platforms that employs behaviour tracking, is User Experience (UX), defined as *user's perceptions and responses that result from the use and/or anticipated use of a system, product or service* according to ISO 9241-210:2019 [121].

UX is relevant in e-learning to understand if the states of the users during content interaction is caused by the content itself or by the platform that displays it. Hussain et al. (2018) state that poorly designed mobile apps may cause user's disappointment, frustration and dissatisfaction, all negative states that may result in denial of use or even rejection [122]. For example, in the mentioned work, the authors studied the user experience during utilization of Amazon's Kindle mobile app using questionnaires, achieving very positive results among the studied sample [122].

In previous years, UX assessment was exclusively made using questionnaires [123]. Recently, in a 2017 study, the authors studied the most used methods to assess UX, where expert evaluation, user testing, and data mining techniques revealed to be the most efficient [124]. However, they also concluded that questionnaires were still the main method to assess UX design [124]. In 2019, the authors of a literature review that reviewed over 500 papers revealed that AttrakDiff, UEQ, and meCUE were the most utilised questionnaires in this context, although methods such as NASA TLX, PANAS or structured interviews were combined with those questionnaires in 61.5% of the papers [125]. Though less frequent, physiological measures have also been employed as UX assessment, which might reveal users' feeling during the process of experimenting the handled platform [126].

In terms of products where UX metrics evaluation is most frequent, the educational domain leads, where e-learning systems, LMS, and mobile applications for education are included, followed by the health domain, the entertainment domain, the travel and

guiding domains, the business and finance domain, the commercial domain, the social and communication domain, the industrial domain and, finally, the culinary domain [126].

The most used metrics in the context of UX assessment are (1) issue-based, i.e., what kinds of issues did the users find during the use of the product, through interviews, think aloud method, and questionnaires, (2) emotion and stress metric measured with resource to physiological signals (e.g., EDA, Heart Rate Variability (HRV), respiration, and EEG), video-based emotion recognition and eye-tracking, (3) interaction metrics, namely, performance metrics and behaviour metrics, and, finally, self-reported metric that report usability, enjoyment, aesthetic, among others [126].

In [127], for example, the authors propose a new framework that use psychophysiological signals to monitor users' state while interacting with the systems, allowing for a more objective view of their experience, bridging the gap between UX and behaviour tracking. In addition, frameworks that apply big data algorithms, and deep learning algorithms have been proposed for behaviour tracking [128–130].

Tupikovskaja-Omovie and Tyler (2022) studied the different behaviours of experienced and inexperienced consumers while shopping online on smartphones using eye-tracking and noted several differences among the two groups, such as the time and number of steps to reach a desired result [131]. Still in the context of e-commerce, [132] discussed that high complexity websites lead to high cognitive loads, which is detrimental to cognitive function. In this case, the complexity was assessed using eye-tracking to study users' search behaviour to measure subjective perception and objective search efficiency. Similarly, Deepalakshmi and Amudha (2021) studied the possibility of detecting eye-gaze focus in fluid frames to capture where the user was paying attention to on each video frame, with high applicability in real-world settings [129]. Moreover, in [133] the authors studied the usability of systems' recommendations in websites that employed recommendation systems by implicitly evaluating users' interactions using mouse tracking features, revealing that even horizontal vs. vertical recommendations could reflect on the effectiveness of capturing users' attention.

Behaviour tracking has also been highly used in psychotherapy for several years in the form of journaling [134]. Armstrong et al. (2021) proposed the development of a journaling app that would allow therapists to track their patients' behaviour prior to each appointment, optimising appointment times [134]. On another level, Alevood et al. (2019) used a smartphone combined with wearable sensors to preventively track the behaviour of people with depressive, anxiety, or psychotic disorders by monitoring their sleeping behaviour [135]. Behaviour tracking was also applied on the level of diagnosis. For example, Jiang et al. applied behaviour tracking by monitoring facial features using computer vision techniques for schizophrenia and psychosis assessment, which would otherwise depend on subjective views, such as self-reporting, and clinicians' observations [136].

Cepeda (2019) developed work showing how behaviour tracking using mouse-tracking

data and biosignals data were able to assess personality traits of users while interacting with questionnaires [63]. With this intent, the author developed Latent, a Chrome extension that allows the acquisition of multiple HCI variables and deeply analysed mouse-tracking features that could be used in questionnaires and also in open-ended problems [137–139].

In the workplace, behaviour tracking has also been revealed to be relevant. Atta-ur-Rahman et al. (2019) developed a Fuzzy Rule-Based System that took network, web, and local machine logs for suspicious activity detection. The output of the system was then combined with user feedback as a Gaussian Radial Basis Function Neural Network's inputs to finish classification and prediction of users' tendency to attempt suspicious actions [140]. On an increased-productivity view, Callara et al. (2019) proposed a behaviour tracking system using ML to predict when determined applications launch or when users' sessions start with the aim to anticipate those actions in remote access environments, thus reducing the time employees take to start their work-related activities [141].

Specifically applied in learning tools, learning experience design refers to the design of human-centred approaches to develop learning experiences that provide pleasing and effective digital learning tools that are easy-to-use and efficiently guide learners towards their learning goals [142]. Analogously, [143] the authors made a scoping review about usability, a term closely associated with UX, in the context of blended learning programs, revealing its importance in systems specifically designed for learning.

In the following subsections, the most utilized HCI behaviour tracking methods in e-learning, specifically, system logs tracking, eye-tracking, camera-based tracking, and other less common methods and combinations are reviewed.

3.2.1 System Logs Tracking

System logs are the logs that e-learning platforms may save that include mouse tracking, the number of accesses, the time spent in pages, the clicked links, among others. Such metrics allow to track users HCI behaviour while interacting with the platform seamlessly.

For example, Baptista et al. (2008) proposed a tracking tool to facilitate the development of e-learning contents [144]. This tool had the ability to track learners actions, allowing teachers to verify the relation between learners' grades and the time they spent with each learning object. Based on this information, teachers would be able to adjust their teaching to each individual learner.

Hardy et al. (2008) also developed a tracking tool, but included a data visualisation feature to ease the process of analysing the spatial and temporal routes taken by the students [145]. In this case, the comparison between final marks and the considered metrics (number of page hits, number of distinct sessions, average session length, the total time spent, and number of hits within a particular section of the online material) showed no correlation, revealing that there might be other better suited metrics to gain insights about learning behaviours.

May et al. (2008) investigated ways to exploit collected tracking data of learners' activities to assist learners and teachers during and after their communication [146]. In this case, the authors assessed each learner reading the messages the teacher posted in the forum, namely, if they read it entirely, partially or not at all. This approach might also be interesting to assess if course content is fully read and to understand if it should be improved to enhance learners' engagement and interest.

Alexandra Louise Webb and Sunhea Choi (2014), for example, studied the impact of an e-learning platform during radiology study using the number of accesses of learners [147]. In this simple case where system logs consisted of the number of accesses, the authors noted an accentuated increase of accesses close to test periods.

Baharudin et al. (2017) studied whether tracked behaviour correlated with different learning styles to build an e-learning platform that takes them into account [148]. The authors showed that some learning styles (in their model, the theorists and reflectors) accessed the learning materials significantly more times than other (activists and pragmatists), though the correlation was very low. Thus, they demonstrated the possibility of discriminating different learning styles based on the number of accesses to the learning materials.

Tellakat et al. (2019) used the number of mouse clicks to quantify the students' approaches to studying, problem solving, and information search over the entire semester [149]. They found that learning as measured through out-of-class clicks is a reliable and predictive measure of class outcomes, because the number of clicks positively correlated with academic performance.

Krisztina Sára Lukics and Ágnes Lukács (2021) developed a task to assess learning online, allowing to gain understanding of the learning process over time, instead of only in a post-hoc manner, as it was often done [150]. With this task, authors performed data acquisition by measuring the participants' reaction time, revealing that it decreased and accuracy increased with time, indicating the occurrence of learning.

Charmaine Lloyd (2021) used blogging to improve students' learning in the context of medical microbiology [151]. Their objective was to lead students to connect foundational knowledge and laboratory techniques learned in class with relevant field topics. This approach allows students to broaden their views on the studied topics and promotes active and independent learning. In terms of tracking, blogging might help teachers to understand learners' rationale during the course of the semester and help them overcome possible difficulties.

Al-Hawari et al. (2021) described the development of an e-learning platform to track students' attendance and to allow instructors to access students tracking details, namely, who accessed a file or page of each course, when a file or page was opened and viewed, and how many times did a user access a file or page [152].

Analogous to self-report questionnaires used in cognitive monitoring, system logs are primarily employed for retrospective analysis, as the data they capture is generally not utilized in real time. This data is valuable for evaluating students' engagement and

reactions during learning activities.

3.2.2 Eye-Tracking

Eye-tracking is one of the most utilised methods to track learners' and even teachers' behaviours in e-learning platforms. Typically, this technique requires specific cameras that capture visible light and, in some cases, infrared light, in order to capture eye-related metrics, such as eye blinking, saccades, fixations, and eye movements. Through calibration techniques, some devices also allow to capture eye-gaze on a screen, allowing researchers to study points/areas of interest during learning.

In the specific case of infants 5-7 months old, Shukla et al. (2011) studied the implementation of a real-time performance measuring system capable of understanding the degree of learning by measuring anticipatory eye movement [153]. This type of eye movement occurs specifically when the individual anticipates the occurrence of an event after watching it happen, i.e., after learning about the expected event.

In another study, de Mooij et al. (2020) studied the implication of different stimuli in verbal working memory and inhibitory control in the performance of children aged 8-11 years old using eye-tracking during the execution of addition and multiplication problems [154]. One of the main aspects the authors intended to study was whether the visibility of time pressure with an animation, which should increase cognitive load, would affect performance and if it depended on the level of working memory and inhibitory control of each participant. The particular aspect authors studied with eye-tracking was whether learners attended to or actively inhibited their attention to irrelevant/distracting stimuli by analysing their fixations and eye movements. Although the employed verbal working memory task was a positive predictor of arithmetic performance, individual differences of inhibitory control only predicted performance when the time pressure was not visible. Eye-tracking was relevant as it showed that participants fixated on different stimuli depending on whether there was time pressure, the level of inhibitory control, and their age. An important aspect was that eye-tracking showed more fixations on the question when time pressure was visible, showing that participants had increased difficulty remembering the question.

Li et al. (2020) applied ML models to eye-tracking data in order to classify and predict the different difficulty levels of solving spatial visualisation problems [155]. In this case, all participants were first-year engineering students. The authors concluded that the proposed approach accurately classified the difficulty level of the presented problems as well as accurately predicted the difficulty of the upcoming questions.

Jónsdóttir et al. (2021) intended to model the effect of language use and time pressure on English as first language and English as second language students by measuring eye movement data in a self-directed learning environment [156]. The results showed that students of English as a second language had higher fixation rates and longer completion times. Moreover, high time pressure resulted in short completion times with the cost of low

correct rates, high fixation rates (associated with higher difficulty), and high gaze entropy.

In an emotion/cognitive monitoring-focused approach, Calvi et al. (2008) presented an e-learning environment that used eye-tracking for emotion recognition and to study the attention given to the various regions of interest while learning [157]. Specifically, teachers could define the regions of interest and the duration for which learners should focus on each of them. Then, when changing pages, the system informs the learner about the effectiveness of their attention, i.e., if they should have paid more attention to any of the regions of interest and to what degree. The authors reached around 80% accuracy to detect all eye-tacking measures (e.g., increased fixation or blinks) and then manually related each metric to the participants' state. Finally, those metrics and states could be used to recommend easier content or to recommend taking breaks if a high workload, confusion, or fatigue was detected.

Tsianos et al. (2009) used eye-tracking to validate their proposed learners' model, which accounted for individual differences in learning styles, namely, in the level of *imagers* (learners that learn better with images, such as graphics and diagrams), *verbalizers* (learners that learn better with reading text), and *intermediates* [158]. Eye-tracking data seemed to validate the existence of those different learning styles by revealing the diverse behaviours: *imagers* spent more time looking at figures, *verbalizers* invested more time reading text, and *intermediates* looked at both content types in similar degrees.

Qin et al. (2009) employed eye-tracking to propose an e-learning platform to adapt according to the eye movements of its users [159]. In this example, they showed that it was possible to develop different artworks by modifying the Mona Lisa painting in real-time as a proof-of-concept based on eye gaze.

Kai Li and Yurie Iribe (2011) proposed using eye-tracking technology to detect and show what the teacher is looking at in the slides and to evaluate its effects on learners' attention compared to only using enhanced content (e.g., circles and underlines of important information) [160]. Although the students responded positively to the eye gaze feature because it made them more aware of where they should focus during the teacher's explanations, the perceived learning outcomes did not change significantly.

Porta et al. (2012) also explored the possibility of assessing the learners' emotional state during an e-learning experience using eye-tracking data [161]. The authors found a relationship between mental processes and pupil size. For example, when reaching a conclusion, the pupil size typpically increases (e.g., they found the correct answer). Then, this information could be used to adjust the content according to learners' states.

As a continuation, Cantoni et al. (2012) studied new correlations between eye behaviours and emotional and cognitive states [162]. In this study they found that eye fixations and gaze revealed relevant information about the learners' understanding. An illustrative example is that if the fixation in a text leads to a fixation in the part of an accompanying figure and that part of the figure is not related to the text the learner was reading, it is safe to assume that the students might not be understanding the question/text.

Ivanović et al. (2017) also studied the possibility of combining eye-tracking technologies

with an e-learning system PROgramming Tutoring System - Protus [163]. In this case, eye-tracking was designed to identify the learning styles of learners in order to improve the user interface according to those styles. Moreover, the authors posed the possibility of identifying learners' mental states to improve their attention and to notify the teachers where the problematic areas may be.

Analogously, Nugrahaningsih et al. (2021) aimed to assess learners' learning styles using eye-tracking data [164]. Namely, they wanted to distinguish between visual and verbal learners and, as a secondary objective, to distinguish more specific types, namely the active/reflective, sensible/intuitive, and sequential/global bipolar styles. The authors found that verbal learners spent more time looking at the text while visual learners spent more time looking at figures and diagrams, thus, it was possible to distinguish them, as other authors proved earlier. However, the group in between those learning styles, i.e., that did not entirely belong to any of them, did not show a specific gaze behaviour. Moreover, the relation between gaze and learning style was only found when the slides had the picture on the left and the text description on the right, thus, the display is relevant for this assessment. Still, the metrics and protocol employed in the study were not effective in distinguishing the more specific learning types. The learning styles were assessed using the Index of Learning Styles (ILS) questionnaire.

Kao et al. (2013) explored a method to study learners' navigation focus and visual tracks with a combination of gaze-tracking with the contents of web page blocks on an e-learning platform [165]. The authors used the Hough transformation to identify the lines in the figures and establish the boundaries of the objects to comprehend when and where the gaze was fixed, as well as to determine which learning objects it was directed towards.

Charoenpit and Ohkura (2015) explored the building of a prototype e-learning system and analysed the relation between learners' emotions (i.e., boredom and interest) and eye-tracking measures [166]. In this study, both boredom and interest were assessed through questionnaires. Then, the authors calculated the correlation between those and the eye-tracking measures. The authors concluded that the fixation duration ratio, the number of fixation ratio, and the pupil diameter could be used as indexes of interest because they positively correlated to it and, on the contrary, these metrics were inversely correlated to boredom. Intuitively, during boredom conditions, the authors found that there were more fixations outside of the learning areas.

Liu et al. (2019) explored the relation between pupil diameter and blink frequency to positive and negative emotions during online learning [167]. Contrary to the reviewed literature, the authors concluded that pupil diameter is indirectly proportional to emotional state, i.e., better emotions lead to smaller pupil size. Furthermore, blink frequency were directly proportional to emotional state, i.e., better emotions lead to higher blink frequency. In the context of learning, the authors concluded that in easy settings, there was an increase in blink frequency and a decrease in pupil diameter, while the opposite pattern was observed in difficult circumstances.

Joe Louis Paul et al. (2019) studied the possibility of employing eye-tracking to understand if learning materials are tuned to the learners' capabilities [168]. In this study, gaze-tracking allowed to recognize when and to each parts of the slides the learners paid attention to. With this, the authors improved on one particular slide (as no one understood the slide, they explained before how to read that kind of diagram), leading to higher attention to that same slide. This is an example of the importance of the application of eye gaze tracking in e-learning to adapt the content and facilitate learning.

Van Der Sluis and Van Den Broek (2022) employed eye-tracking capabilities to forecast subjective understandability and interest. Their discoveries indicate that these capabilities can offer insights into the depth of subjective evaluations, rather than relying solely on traditional binary assessments, enabling the improvement of the customization of personalized information systems [169].

Eye-tracking can subsequently be employed in e-learning environments not only to autonomously enhance learning materials but also to provide instructors with insights on improving learning materials and delivering teaching content more effectively. Moreover, specific measures can be used to assess the cognitive and mental state of learners in this context. The main disadvantage of these systems is that they require specific devices and setups, hindering their application outside the laboratory or outside the classroom and, thus, lack ecological validity [169].

3.2.3 Camera-Based Tracking

Contrary to eye-tracking, web-cam devices are ubiquitously available and even embedded in most laptops today. Thus, they can be used in e-learning settings to monitor learners' faces and other attributes that help to capture their experience.

For example, Nosu and Kurokawa (2006) aimed to discriminate emotions from facial features in real-time to support e-learning [170]. In this case, facial expressions consisted of the position of the mouth, of the eyebrow corners, and of the chin. The authors presented a model of 8 emotions specific to the e-learning context. Emotion identification was performed by comparing a distance measure between the measured facial features and the features they considered to represent each emotion. With this approach, they reached an accuracy of 69% in discriminating the different emotions.

Robal et al. (2018) studied the effectiveness of existing frameworks for detecting the face in webcam video [171]. In this case, the authors considered that if there was no face in the field of view or if the face was not directed to the computer screen, then the learner was not paying attention. The authors used an eye-tracker device to serve as the best-score, i.e., the maximum performance the camera-based approach could reach, and concluded that the performance of the camera-based method, although lower, was still acceptable. Thus, this method can be applied to detect distractions in e-learning conditions. However, it should be noted that data acquisition was performed during 50 well-defined tasks and not directly during unstructured e-learning courses.

Khurshid and Scharcanski (2020) made an adaptive face tracking algorithm capable of detecting facial features, namely yawning [172]. The developed algorithm achieved better performance than the literature and, although, it was tested during driving scenarios, it should be generalisable to detect interest, fatigue and other states in learning and e-learning contexts.

Ramaha et al. (2021) explored the possibility of using a webcam-based face direction tracking to detect the learners' attention towards e-learning content [173]. With this intent, the authors developed an acquisition protocol that consisted of taking snapshots and saving results of face orientation relative to the camera (left, right, forward, up, down) while solving an exam of 3 questions. The authors concluded that there was high agreement between the system's output and the label given to each photo (between 8/10 and 10/10 agreement).

Khan et al. (2022) studied the development of an e-learning platform using webcam/camera to capture and analyse the gaze data to quantify the attention level of learners during remote learning sessions [174]. For this, the authors developed a ML model to classify student attention with accuracy score of around 90%, where the ground-truth was given by a feedback button the participants had access to in order to inform if they were paying attention or not in each instant.

Franz Pernkopf (2008) took a different approach and developed an algorithm to track faces continuously in meeting scenarios using particle filters and demonstrated it was better than the state-of-the-art of that time [175]. The algorithm can be used to ascertain whether learners are within the camera's field of view and track their presence in both traditional classrooms and e-learning environments.

Bandung et al. (2017) proposed a video conference system with object tracking used for distance learning [176]. In this case, object tracking is used in order for the cameras to move automatically if the object of interest is about to leave the field of view. Analogously, Zhao et al. (2020) explored a similar approach where they built an object tracker robust to occlusions and other unexpected events to be applied in e-learning scenarios [177].

Abate et al. (2021) designed a camera-based monitoring system aimed at assessing attention levels during synchronous distance-learning sessions. Employing image-related features, this system was deployed in both distance-learning contexts and within an engineering company environment. Its core purpose centred on providing supervisors with real-time feedback derived from camera monitoring, facilitating informed decisions for optimizing content to better suit the overall engagement and attention of users [178].

Sutanto et al. (2021) combined face detection with gesture recognition to perform real-time and stable lecturer localisation [179]. These measures were used to build a motorised system to automatically adjust camera position for online learning so that the teacher could move freely without the need to manually adjust it. Moreover, manual commands by hand gesture could be used to give any other command, such as zoom in and zoom out.

Thus, camera and webcam-based systems have been extensively used in e-learning

to track learners' cognitive and emotional state and to assist on synchronous e-learning classes by easing the burden on teachers that typically need to control the camera when teaching. These systems can be valuable because most laptops today incorporate webcams. However, using the learners' image poses a privacy issue that should be taken into account.

3.2.4 Other Methods and Combinations

In addition to the methods mentioned earlier, some studies utilized biosignals-based methods and a combination of the above-mentioned approaches.

Van Drunen et al. (2009) developed a system to study how mouse clicks and movements, HCI variables, related with physiological data, namely, eye fixation, HRV, and skin conductance to monitor attention and workload. They showed that in one of the performed tasks, mouse data correlated with biosignals, revealing the feasibility of employing it in the context of cognitive monitoring [180].

Mailhot et al. (2018), for example, assessed the feasibility and acceptability of using a wireless EEG device to measure the affective and cognitive states of patients and healthcare professionals during e-health or e-learning interventions [181]. The employed device had integrated algorithms to detect anxiety, attention, engagement, enjoyment, interest, and relaxation with a degree of intensity between 0-100. The results showed that the method was feasible and that the participants showed high acceptance of using the device. However, the discomfort level increased with the increase of time of use.

Taub et al. (2022) studied system logs of click-stream data in combination with the achievement goals questionnaire-revised to investigate learners self-regulated learning behaviours over a semester-long blended learning physics course [182]. For this, they used ML, namely, a combination of hierarchical clustering, process mining, and sequence mining to analyse the collected data. With the complementation of learners' tracked behaviour and perceptions that can give a bigger picture of their experience, researchers had access to improved information on learners' behaviours within e-learning environments.

Wan et al. (2022) employed a combination of gaze behaviour, focus group discussions, and surveys to assess the effectiveness of an interactive Digital Pathology Repository (iDPR) tool to improve clinical pathology learning [183]. Results showed that after reviewing the iDPR website, knowledge scores increased.

Porter-Szucs and DeCicco (2022) used a combination of surveys, attendance tracking, and performance metrics to assess the experience of learners in three modalities, namely, face-to-face classes, synchronous online, and asynchronous online [184]. The distribution of learners was based on the pre-survey to assess the preferences of each student regarding the available learning modalities. The authors found that what mattered to the students and what enhanced their success in the course was the appropriate match between their preferences and the modality of attendance. Moreover, motivation over time was maintained for all (synchronous and asynchronous online learners) except the face-to-face students who needed to switch to synchronous learning during the pandemic because

their preference ceased to match their learning modality.

Li and Iribe (2011) used the Wiimote with MotionPlus to capture and process the swings of people learning tennis in order to improve the distance learning and teaching process of the sport [160]. For this, the authors acquired a dataset composed of data of people of different sizes performing the different swings to understand if it was possible to standardise the data and recognise the type of attempted swing. They verified that it was possible to remotely recognise the type of swing and to analyse its flaws based on the collected data and, thus, distance-learning was possible using the mentioned approach.

Notaro and Diamond (2018) described the development of a system that integrated EEG, eye-tracking, and mouse-tracking data to support online learning [185]. They used Duolingo to show that the system could integrate all data but failed to show how it related to the learning experience.

Zardari et al. (2021) employed a four-methodology user experience evaluation to identify usability problems in an e-learning platform [186]. Namely, the authors used heuristic evaluation, usability testing, user experience questionnaire and eye tracking. In this case, three usability experts employing a set of well-known heuristics carried out the heuristics evaluation. Usability testing consisted of the users performing specific tasks on the platform to assess its effectiveness and efficiency. As for the user experience questionnaire, the authors asked the users to fill out the User Experience Questionnaire after the participants finished all tasks. Eye-tracking was also employed while participants completed the usability tasks. The proposed UX evaluation methodology resulted in increased participants' satisfaction and on the overall usability of the e-learning platform.

In conclusion, different approaches and combinations of approaches might enhance the learning experience by complementing the flaws of the most commonly used methods. Namely, the combination of questionnaires with other methods allows to overcome the most noticeable disadvantage of the questionnaires, which is the fact that they do not take into account the variations during the evaluation time, but can only capture the view the individuals have of the whole process in the end.

3.3 e-Learning

In its infancy, e-learning was seen as less costly than presential learning, more effective in data recording and tracking, and more effective in delivering education materials, resulting in higher course completion rates [187]. However, even then, the lack of social components of the typical classroom and the inaccessible access to individual tailored educational content (e.g., teachers' explanations) were seen as relevant drawbacks [187].

Booth et al. (2009) proposed a framework to enhance the e-learning experience by considering ten aspects: applicability, attractiveness, usability, offline working, asynchronous engagement, learner interaction, peer support, moderated learning, formal support, and assessment [188]. This framework expands on previous experiences by proposing further support and increased communication among peers, overcoming the previously

mentioned limitation of low social interaction.

Koch et al. (2010) studied the impact of a web-based course delivered to nursing students on their learning outcomes [189]. The authors concluded that adherence was high compared with other works, and the students enjoyed various positive points of the experience, namely, the higher level of interactivity relative to in-person classes. Students who spoke English as a second language revealed that online learning allowed them to take the course at their own pace, thus achieving better understanding. However, overall, blended learning was preferred over online only.

Oeffner et al. (2011) studied the impact of blended learning in the context of teaching genetics courses to students of different areas and concluded that the time spent in online learning varied among fields of study [190]. Thus, students' behaviour may help to cluster them into different types, allowing for targeted teaching. Moreover, the degree of acceptance of such courses increased over the years, and both students and teachers revealed a decreased workload due to the effective sharing of information across communication forums.

Cook et al. (2018) assessed the openness for online and simulated learning platforms specifically by physicians [191]. In this specific context, physicians seem to be most receptive to short, high-relevance, and patient-focused activities, besides innovations that automate administrative tasks. Moreover, the authors concluded that age did not impact the interest, willingness, and capability to use online or simulated-based learning.

In an example of an application where Sinnayah et al. (2021) studied the possibility of employing H5P to build course contents in a blended learning scenario, the authors concluded that, although the workload of students increased as they spent more time solving problems in the platform, they believe they have learnt better by repetition [192]. Moreover, the faculty members improved their proficiency with technology, which in pedagogical terms might be positive, but in terms of workload, might pose higher loads with little reward or recognition, as stated by [184].

Co et al. (2021) evaluated the effectiveness of distance learning on anatomy and surgical training [193]. They concluded that distance learning was non-inferior, if not superior, to in-person learning in this context. The benefits include easier monitoring of learning progress, improved participation, increased flexibility of schedule, efficiency, and facilitation of academic collaboration. However, they identified disadvantages to tackle: teaching anatomy should be supported by 3D models, there are technological problems, high implementation costs, lack of students' engagement, and reduced concentration.

Shah et al. (2022) studied the impact of age in using remote learning consisting of a series of health-related seminars [194]. In that work, they could perform better mental tracking in remote participants and hypothesised that this was due to less social anxiety and fewer peer distractions. However, they also noted that, despite the high adherence program, people who did not complete it were statistically significantly older than the ones who completed it.

Analogously, Taub et al. (2022) examined the feasibility (e.g., completion rate), acceptability (e.g., satisfaction), and participant-reported impact (e.g., memory concerns, behaviour change, goal attainment) of a self-guided, e-learning adaptation of a validated, facilitator-guided, in-person memory intervention for older adults [195]. The intervention led to reduced age-related memory concerns, showing that self-guided interventions in e-learning are feasible, overcoming the problems that in-person, clinician-led intervention have (e.g., schedule-dependent, difficulty of access for people with low mobility).

Evidently, e-learning is widely perceived as a valuable tool with a wide range of uses, extending beyond the realm of education and into our daily lives. Nonetheless, these applications and platforms consider personal preferences, age, and technological aptitude, as some individuals may lack motivation to engage with these systems, or it might result in an increased workload.

3.3.1 Intelligent Tutoring Systems

Artificial Intelligence has been applied in e-learning to improve the learning process for distance-learning applications. According to Zawacki-Richter et al. (2019), there are three applications of such systems: personal tutors, intelligent support for collaborative learning, and intelligent virtual reality [196].

In the case of personal tutors, Intelligent Tutoring Systems (ITSs) simulate personalized tutoring for individual learners. These systems make decisions about the learning path of individuals and the content they select based on learners' models to maximize engagement in dialogue to simulate one-to-one tutoring [196]. These systems contain an expert module that aggregates the learning object to teach, a learner diagnose module to monitor the evolution of learners' knowledge, skills, learning styles and other personal attributes, an instructional module that focuses on different teaching methods and procedures to deliver personalized learning content, and finally, a user interface module to enable the interaction between the system and its users [197]. Notwithstanding the advantages of personalized accompaniment of ITSs that can mitigate the lack of teachers where one-to-one tutoring is impossible (e.g., large scale around the globe courses), these systems lack the social and collaborative aspect of learning, which may be detrimental to this process as a whole.

In this sense, AI in education may support the formation of groups based on their models, may be used to facilitate the interaction of online groups, or can be used to summarize discussions used by human tutors to guide students towards the aims and objectives of a course [196]. Even using ITSs for personal accompaniment and AI powered collaborative tools in e-learning, it still lacks the on-site materials, for example, for practical classes (e.g., laboratory materials). On that account, intelligent virtual reality is used to mimic the conditions of real learning environments to engage and guide students in such contexts [196]. These systems can replicate the environments and also mimic in-person peers, tutors and other types of learning facilitators.

By combining these three components, these systems may imitate a real-world learning

scenario or classroom to optimize engagement, although, as pointed out in [197], distance-learning platforms and systems cannot replace their real physical counterparts, being more effective as a complement in blended-learning systems instead of serving as replacements.

3.3.2 Recommendation System Assisted E-Learning

On another level, AI can also be used for recommendation systems applied to e-learning systems, where the learning materials are recommended to each individual. In such cases, the system may adapt to the learners' individual preferences given their history and learning styles, which require the fulfilment of questionnaires so they can indicate those characteristics a priori. Then, when learners utilize the systems for long periods, their knowledge about the learners' interests increases, and their adaptations should improve accordingly [198]. In this case, the adaptation occurs on the level of course recommendation instead of adapting the contents in real time.

Finally, considering the instantaneous state of learners, it should be possible to adapt the learning contents in real-time. Thus, the learning state may be inferred using cognitive monitoring, as explored in 3.1, using HCI behavioural metrics, as explored in 3.2, or using combinations. Contrary to the other methods, these methods should be able to recommend or make changes to the learning objects in real time to enhance the learners' cognitive states.

3.4 Summary

In this chapter the advancements in cognitive state monitoring, behaviour tracking in the context of HCI, and e-learning were presented.

Specifically, cognitive state monitoring has been applied in many cases, such as cognitive fatigue, attention detection, and the emotions scanning. The most used sensors in this context are the fNIRS and EEG. Nonetheless, sensors such as ECG and EDA are also proposed to infer about some of the mentioned aspects of learning.

On the other hand, behaviour tracking in HCI has been employed using many other approaches. Systems-log tracking, eye-tracking, and camera-based tracking were identified as the most used techniques, although combinations of physiological sensors and HCI metrics to track learners' behaviours during learning tasks were also found. In this context, sensors such as ACC and other accessible devices, such as webcams, that can be embedded into everyday items can be used seamlessly to monitor the cognitive state of learners based on their behaviours.

This chapter also presented how e-learning has had an impact in different contexts and how it has been accepted over the years by different types of learners and, finally, presented how AI has been implemented into e-learning systems to assist in the optimization of the learning process.

4

Methods

Before, this work aimed to assess how cognitive processes related to learning can be monitored, thus, a series of data acquisition procedures to elicit and monitor them were developed. An incremental model, leveraging the findings from each procedure to enhance the subsequent ones was adopted.

On the first procedure, two standard cognitive tests, n-back and mental subtraction involving short-term memory and logical reasoning were used. In the end, a scientific computing learning subject (based on Python programming language) using Jupyter Notebooks and a Jupyter extension to support it were developed.

On the second developed procedure, the followed approach was modified. First of all, the first procedure involved little mouse interaction, hindering the application of mouse tracking analysis. Thus, the cognitive tasks were replaced by a digital version of the Corsi block task and a concentration task presented by Gamboa (2008) [199]. The learning task was replaced and instead of a scientific computing subject, the participants partook an introduction to the ECG signal digital lesson also developed in a Jupyter Notebook. In this case, all tasks involved mouse movements, although the cognitive states were estimated based on the tasks.

Finally, a third study was performed where the focus was on the lesson about the ECG signal, similar to the second study, however, the participants reported about the evolution of their own cognitive states over time.

The baseline periods of the different sutdies were selected considering the different recommendations Herold et. al (2018) that presented a set of recommendations that fNIRS studies should follow [97]. Furthermore, the questionnaires applied during these studies are shown in Annex I.

This chapter will present the three mentioned exploratory studies describing in detail the data collection procedures, data pre-processing steps, and the classification methodology.

4.1 Materials and Methods

Biosignals data and behavioural metrics were acquired, both described in Chapter 2, to build a common database allowing us to study how these data can be used for the identification of different learning states. In this sense, different data acquisition systems were used together with timestamps saved in the different databases to synchronize the biosignals and the behaviour metrics.

4.1.1 Biosignals Acquisition Materials

For the acquisition of biosignals data, two biosignalsplux devices developed by PLUX Wireless Biosignals[1] were used. All the physiological sensors, namely, EEG, fNIRS, ECG, EDA, RIP, and ACC, were connected to these devices that acquired the corresponding physiological signals and sent them via Bluetooth to a computer running OpenSignals, a data acquisition software also developed by PLUX Wireless Biosignals. This application connects to the acquisition devices, saving and showing the signals in real-time. Besides the aforementioned physiological sensors, a push-button that the participants could press whenever they felt the need to stop or pause the acquisition procedure was also provided, allowing us to process the data accordingly. A schematic of the acquisition system is shown in Figure 4.1.

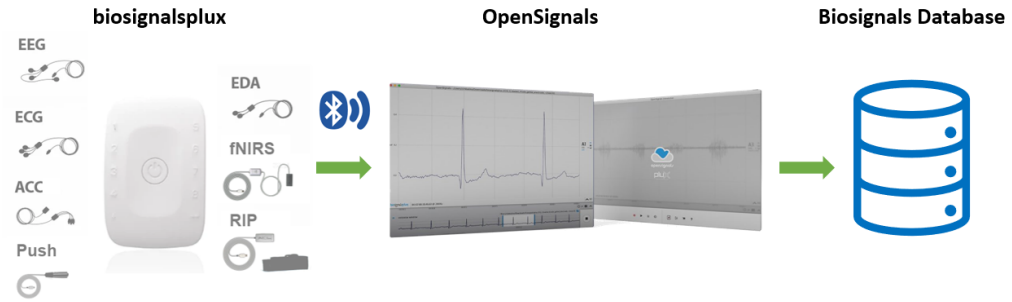

Figure 4.1: Illustration of the biosignals acquisition system. The sensors are connected to the biosignalsplux device that sends the collected data via Bluetooth to OpenSignals that stores it into a local database (.txt and .h5 files).

Each of the mentioned sensors was consistently placed at the same locations for all studies and participants, described next and shown in Figure 4.2:

- **EEG** - two channels positioned on the AF7 and AF8 of the 10-20 system [200].
- **fNIRS** - two channels to measure the dorsolateral pre-frontal cortex close to the AF7 and AF8 positions of the 10-20 system.
- **ECG** - three electrodes placed in the abdominal region to measure an approximation of the Lead I of the Einthoven triangle.

[1]PLUX Wireless Biosignals website (Accessed in 13/11/2023): https://www.pluxbiosignals.com/

- **EDA** - two electrodes placed on the hypothenar eminence of the non-dominant hand, at a distance of approximately two centimetres.

- **RIP** - elastic band positioned around the upper abdominal region.

- **ACC** - placed on the right side of the head, where the xx axis is aligned with the sagittal plane, the yy axis is aligned with the longitudinal plane, and the zz axis is aligned with the frontal plane.

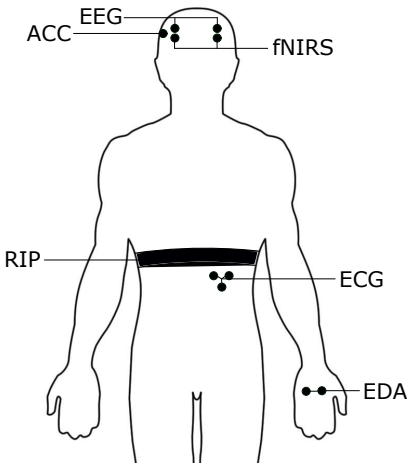

Figure 4.2: Physiological sensors positioning used during data acquisition in this work. EEG and fNIRS are placed around the AF7 and AF8 positions of the 10-20 system, the ACC is placed on the right side of the head, the RIP is placed on the upper abdominal region, the ECG is placed on the abdomen to measure the approximated Lead I of the Einthoven triangle, and the EDA is placed on the palm of the non-dominant hand. From [201].

This setup allowed to capture all intended biosignals with minimal intrusiveness. Despite this, the quantity of sensors might have potentially caused discomfort during task performance. To address this, a push button was provided to participants, allowing them to communicate any discomfort. This enabled us to pause data acquisition promptly without any loss of data.

4.1.2 Behaviour Acquisition Materials

Behavioural metrics consisted of HCI measures acquired using Latent, a Chrome extension that performs locally by monitoring web browser pages, sending the data into a local server that stores it in a local MongoDB database, as illustrated in Figure 4.3 [139].

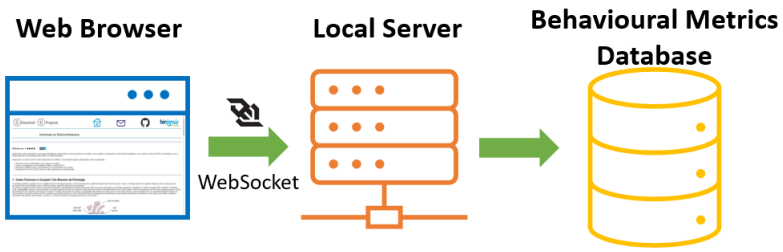

Figure 4.3: Latent's communication diagram. The system uses websockets to communicate the data acquired in the web browser to a local server and then to a local database. Adapted from [139].

In this case, the first two studies included an extended set of measures, while the final one included only a few of them. The full set is composed of:

- **Audio** - audio recording during the performance of tasks that could be used to detect unexpected events related to distracting factors, such as, knocking on doors.
- **Snapshots** - pictures taken using the laptop's webcam to capture facial expressions of the participant.
- **Screenshots** - pictures of the screen captured periodically.
- **Keypresses** - events corresponding to the keyboard keys pressed by the participant while interacting with the computer.
- **Mouse-Tracking** - events corresponding to mouse movements, clicks, and scrolls while interacting with the computer.

From this list, only screenshots, keypresses, and mouse-tracking were acquired on the third study. Besides the mentioned acquired data, each had a corresponding timestamp corresponding to the time of acquisition, as well as the identification of where the event happened in the case of keypresses and mouse-tracking, i.e., in which browser separator and element ID. Other specific information included the key pressed in keypresses events and the type of event in the mouse-tracking data.

Although all this data was acquired, the analysis was focused on mouse-tracking data and used the snapshots and screenshots as helpers when needed (e.g., for data segmentation and labelling).

4.1.3 Data Synchronization

Since different data sources were used and two different databases were built, one containing the biosignals data and the other containing the behavioural metrics, there was a need to synchronize the data to analyse it appropriately. In this sense, the timestamps

corresponding to each sample were considered to build a unique database from the first two, as shown in Figure 4.4.

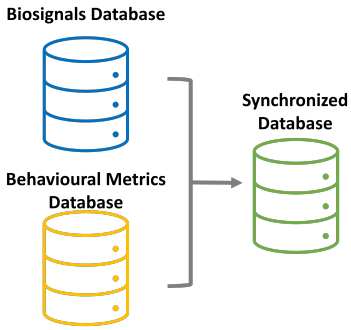

Figure 4.4: Synchronizing biosignals and mouse-tracking databases into one.

This step was only applied on the last study and the proper in depth data analysis will be explained in Section 4.4, since the mouse-tracking data in particular, needs to be resampled to generate a constant sampling rate, otherwise, its analysis could not be performed aligned with the biosignals data. The resampling step will generate mouse-tracking data with the same sampling rate as the acquired biosignals, allowing to build a unique database containing the biosignals and mouse-tracking data.

4.2 Attention Classification Based on Biosignals

The first study performed in this work aimed to classify different attention states during the execution of well known cognitive tasks. For this, an acquisition protocol involving the execution of the N-Back task and a mental subtraction task separated by baseline periods was developed using PsychoPy and tested and validated in other works [87, 202, 203]. Besides, a programming language lesson, in this case, a Python tutorial was developed in a Jupyter Notebook that ran in Google Chrome so that HCI data using Latent could be acquired. The procedure is shown in Figure 4.5.

First, participants were instructed to focus on an "X"centred on the laptop screen and relax for 60 seconds, serving as a baseline period during which biosignals were acquired in the absence of any activity. Therefore, there would be no attention during this period. Then, they would perform the N-Back task 4 times with increasing difficulty and then another baseline period of 60 seconds. After this, they performed the mental subtraction task again followed by a similar baseline period. Finally, they would fill a demographics questionnaire, followed by the Python tutorial and a final questionnaire about their feedback on what could be improved.

4.2.1 Cognitive Tasks

The initial cognitive task used in this study was the N-Back task. This task involves recalling the letter presented N steps back and comparing it to the current one. In this

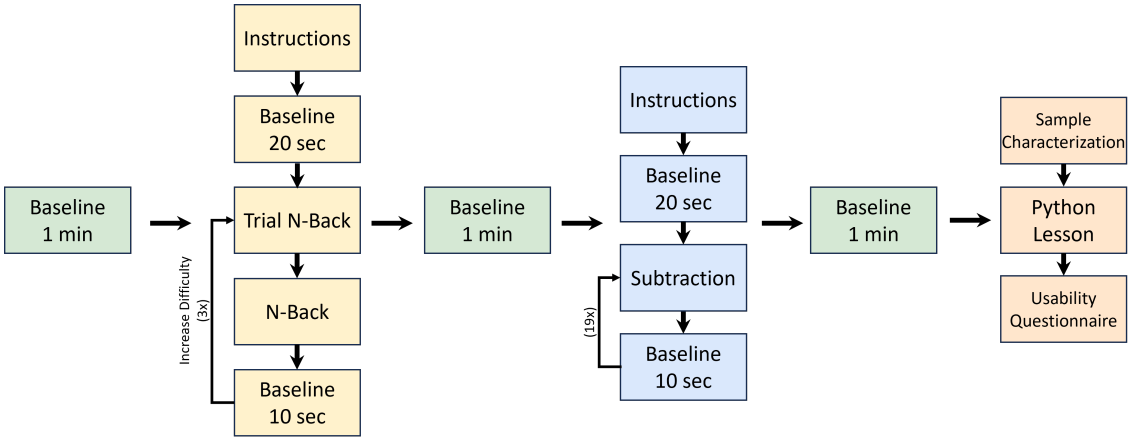

Figure 4.5: Schematic of the procedure of the Attention Classification Based on Biosignals study. It starts with a baseline period, followed by an N-Back task repeated four times with increasing difficulty. Then there is a mental subtraction task that repeats 20 times and end with a programming language lesson, in this case a Python lesson. In between each task there is a 60-second baseline period.

study, a single letter from the set $\{A, B, C, D, E\}$ was displayed at a time. Participants were first presented with instructions, followed by a 20-second baseline period. Subsequently, a trial comprising 20 consecutive random letters was administered, providing feedback on the accuracy of responses to ensure participants understood the task. Following this familiarization phase, the task commenced with 40 rounds without feedback. Finally, a concluding 10-second baseline period was provided.

This procedure involves four difficulty levels and is schematized in Figure 4.6:

- 0-Back: Participants are instructed to press 'Y' if the displayed letter is "A", and 'N' otherwise.
- 1-Back: Participants should press 'Y' if the current letter matches the one shown immediately prior, and 'N' otherwise.
- 2-Back: Participants are tasked with pressing 'Y' if the current letter matches the one displayed two steps back, and 'N' otherwise.
- 3-Back: In this level, participants compare the current letter with the one shown three steps back.

The difficulty increases progressively due to the greater demand on short-term memory with the growing number of letters to recall.

The second cognitive task comprises 20 cycles, each lasting 10 seconds, wherein participants engage in consecutive mental subtraction of two numbers. This involves subtracting the second number from the result of the preceding operation. For instance, a calculation such as "330 – 2" is displayed for five seconds, followed by a prompt saying

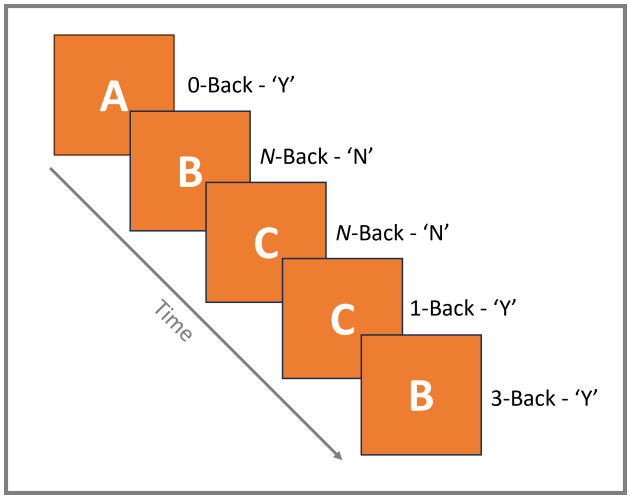

Figure 4.6: Schematic of the N-Back task. The cases in which the participant should press 'Y' are identified given the level (in this case, 0-Back, 1-Back, and 3-Back), while the others are always 'N' stimulus, regardless of the difficulty level.

"Go!"and a cross appearing at the screen's centre. Participants are then tasked with mentally computing "330 – 2 = 328" and subsequently "328 – 2 = 326", continuing this pattern by subtracting "2" from the result until a new message reading "Stop!"is presented. Each of these calculation periods was alternated with 10-second baseline intervals. The procedure is schematized in Figure 4.7

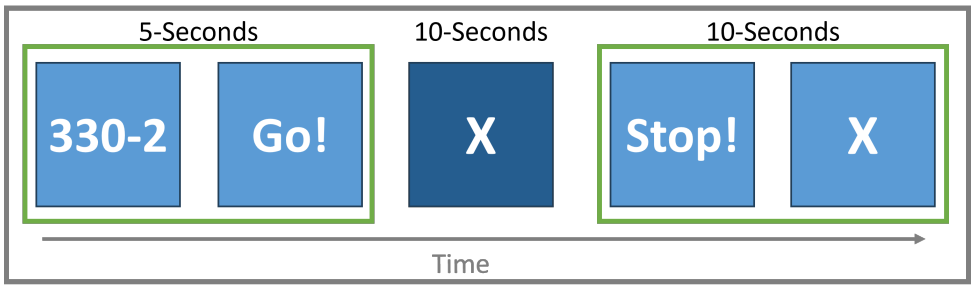

Figure 4.7: Schematic of the mental subtraction task. The task starts with the presentation of the first operation and after the "Go!"cue, the "X"is shown during which the participant should keep subtracting the second number from the results of the consecutive preceding operations. After 10 seconds, the "Stop!"is shown and a 10-second baseline follows. This is repeated 20 times with varying numbers.

At the conclusion of each task, a 60-second baseline period is implemented, allowing participants to rest without any additional activities or interference.

4.2.2 Programming Language Lesson

A programming language lesson in the format of a Jupyter Notebook involving practical examples was developed to mimic an e-learning lesson. The lesson consisted of several Python programming exercises in incremental difficulty to allow participants to learn

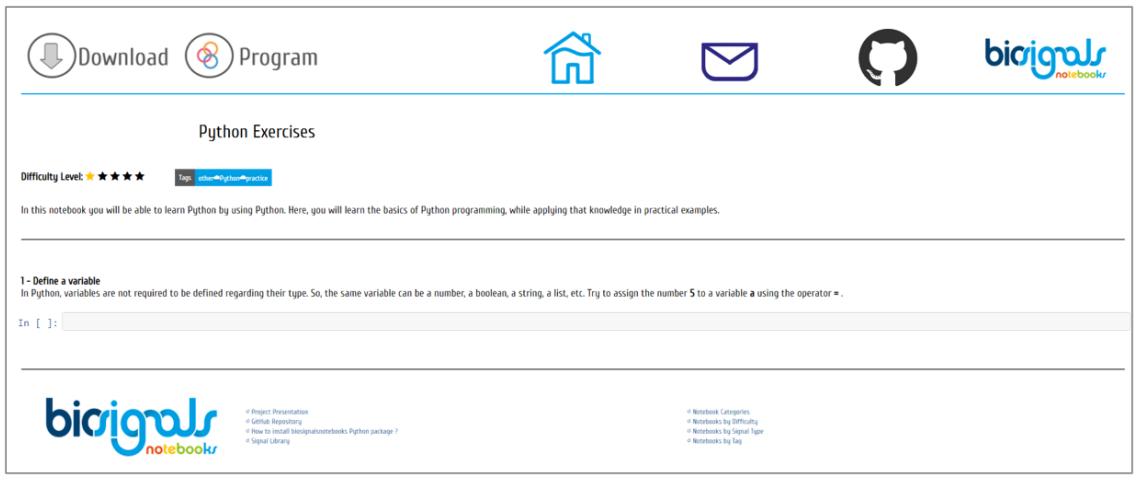

Figure 4.8: Start of the Programming Language Lesson.

about Python while solving problems by using. The start of the lesson is shown in Figure 4.8.

The lesson development involved creating a Jupyter extension to facilitate incremental learning. Specifically, the extension was designed to restrict interaction with text portions of the Jupyter Notebook when activated. It also concealed all questions except the initial one initially. Participants were required to provide a correct response to the first question; a hidden cell then evaluated their answer and code. Only upon a correct response would the subsequent question become visible. The testing cells remained hidden throughout the lesson execution when the extension was active. As for assessing answers, the extension provided feedback. A correct answer prompted a green message, revealing the next question, while an incorrect answer triggered a red message displaying the Python error, if any.

During this lesson, all mentioned biosignals and HCI metrics were acquired using the biosignalsplux device, OpenSignals, and Latent software programs.

4.2.3 Data Acquisition Procedure

Eight volunteers (4 females), aged 20 to 27 (M = 22.9, SD = 2.1), participated in this study. These individuals were students recruited from the NOVA School of Science and Technology. All participants were Portuguese, proficient in English, right-handed, and reported no history of psychological or neurological disorders nor were they taking any medication other than contraceptive pills. Prior to participation, written informed consent was obtained in accordance with the approved protocol by the Ethics Committee of the NOVA University of Lisbon. The study procedures were conducted in a dedicated room where all equipment underwent thorough disinfection before and after each session.

The biosignals were acquired using a sampling rate of 1000 Hz and a resolution of 16 bits, except for fNIRS that was acquired with a resolution of 24 bits.

4.2.4 Data Pre-Processing, Segmentation, and Labelling

This study focused on the processing and classification of biosignals data during the performance of the cognitive tasks. Namely, the main objective was to distinguish between task execution and baseline, assuming that attention was only present during the execution of cognitive tasks.

Biosignals can be prone to motion artefacts or other noise sources and should be pre-processed to minimize their impact. In this sense, a second-order band-pass Butterworth finite impulse filter was applied to each acquired biosignal with cut-off frequencies adapted to each of them, as indicated in Table 4.1.

Table 4.1: Cut-off frequencies in Hz of the second-order band-pass Butterworth finite impulse filters applied for each acquired biosignal.

Biosignal	**Low-Pass** (Hz)	**High-Pass** (Hz)
fNIRS	0.05	1
EEG	1	30
ECG	5	15
EDA	0.016	35
RIP	0.01	1
ACC	0.01	10

After preprocessing, the signals were segmented based on tasks and baseline periods. Then, a non-overlapping windowing technique was employed, dividing the signals into 10-second segments. Segments with potential irregular sampling were filtered out using specific criteria: segments with over 20% of data points having a period five times greater than the 1×10^{-3} s sampling period, or any data point with a time difference greater than 0.2 s compared to adjacent points were not further considered. Additionally, the latter half of segments from each task (n-back and mental subtraction) was excluded to mitigate potential fatigue effects on the results.

Each extracted segment was labelled as either 'task' or 'baseline.' If a segment corresponded to a cognitive task, it was labelled as 'task,' while segments during baseline periods were labelled as 'baseline'. Given the base assumption, the segments labelled with 'task' were the segments that represent external attention states.

4.3 Cognitive Fatigue Detection Based on Biosignals

In the second study performed during this work, the focus was shifted to the detection of cognitive fatigue induced by different cognitive tasks. With this intent, a second data acquisition protocol was developed to elicit cognitive fatigue that also allowed us to acquire HCI data during the execution of all tasks. For this, a new data acquisition protocol was developed that involved the execution of a digital Corsi block task, a concentration task, and a digital lesson about the introduction of the ECG signal, also developed in

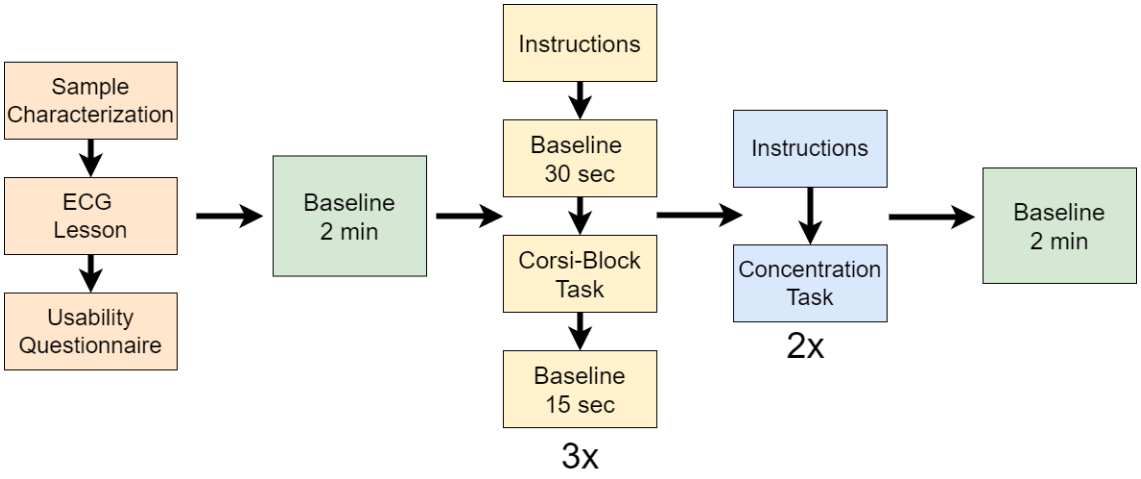

Figure 4.9: Schematic of the procedure followed during the Cognitive Fatigue Detection Based on Biosignals. The procedure starts with the lesson about the ECG, followed by a 2-minutes baseline period. Then, the participants completed the Corsi block task three times, a concentration task two times, and, finally, another 2-minutes baseline period. From [205].

Jupyter Notebook [199, 204]. In this case, the cognitive tasks were developed using HTML, JavaScript, and JQuery, and thus also ran in Google Chrome, allowing us to monitor HCI metric using Latent. The full procedure is shown in Figure 4.9.

In this study, participants began by completing a demographics questionnaire, proceeded to the ECG lesson, and then filled out a usability questionnaire. Subsequently, a 2-minute baseline period was implemented to capture biosignals during a resting state. Following this, the digital Corsi block task was conducted three times, followed by a concentration task performed twice, inducing significant cognitive fatigue. Finally, another 2-minute baseline period allowed for monitoring biosignals during a fatigued state.

4.3.1 Cognitive Tasks

The first cognitive task of this study was the Corsi block task. This task consists of the presentation of nine squares in fixed positions as standardized by Kessels et al. (2000) [204], as shown in Figure 4.10. First, the instructions are shown to the participant and a 30-seconds baseline period follows. Then, the blocks are shown and blink in order, starting with two blocks blinking. The participants are asked to click on the same blocks following the blinking pattern and if they do it correctly the number of blinking blocks increases. This is repeated until the participants make a mistake or if all patterns are correctly followed. Then, there is a 15-second baseline period and the procedure is followed two other times.

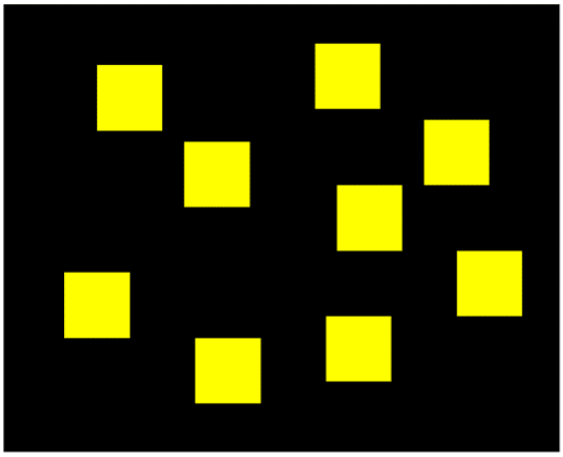

Figure 4.10: Interface showing the nine blocks of the Corsi block task. From [205].

In this case, the increasing number of blinking blocks leads to increased cognitive workload due to the increasing number of items in short-term memory.

The second cognitive task, the concentration task, was presented and developed by Gamboa (2008) [199] and involves the presentation of a matrix of 20×40 containing numbers from 1 to 9 generated using the following algorithm: the consecutive numbers are generated randomly with a probability of 0.8 and should be the complement to 10 ($n_{i+1} = 10 - n_i$) with a probability of 0.2. An example of a matrix is shown in Figure 4.11.

Figure 4.11: Interface showing the matrix of the concentration task. The first line is blurred because it was already blocked and the red numbers correspond to pairs of selected numbers that sum to 10. The blue numbers correspond to the numbers being hovered. From [205].

The task began with instructions for participants to follow a matrix line by line, simulating a reading task using a mouse cursor. They were instructed to focus on consecutive number pairs and click on pairs that summed up to 10. In the interface (Figure 4.11), the hovered pair appeared in blue, and any selected pairs were highlighted in red. As participants progressed to a new line, the previous line became blocked and

blurred, preventing corrections and emphasizing the need for sustained attention during the task.

Given the level of attention or cognitive workload involved during these tasks, it was assumed there was high cognitive fatigue present during the 2-minutes baseline after the concentration test compared with the initial period.

4.3.2 Introduction to the ECG Lesson

To mimic a real e-learning lesson, the previous programming language lesson was altered to a lesson with an introduction to electrocardiography, in particular the ECG signal that should be better suited for biomedical engineering students. This lesson was also developed and presented in a Jupyter Notebook in Google Chrome, thus enabling the use of Latent to acquire the mentioned HCI data. In this case, the lesson consisted of a page with multiple paragraphs of text and images to support them. The information was blurred at the beginning and got visible only when the mouse cursor hovered over the paragraph or the associated figures. Given Latent saves the ID's of the elements with which the mouse cursor interacts with, it is then possible to infer about what paragraph each participant was reading at any given time. An illustration of this is shown in Figure 4.12.

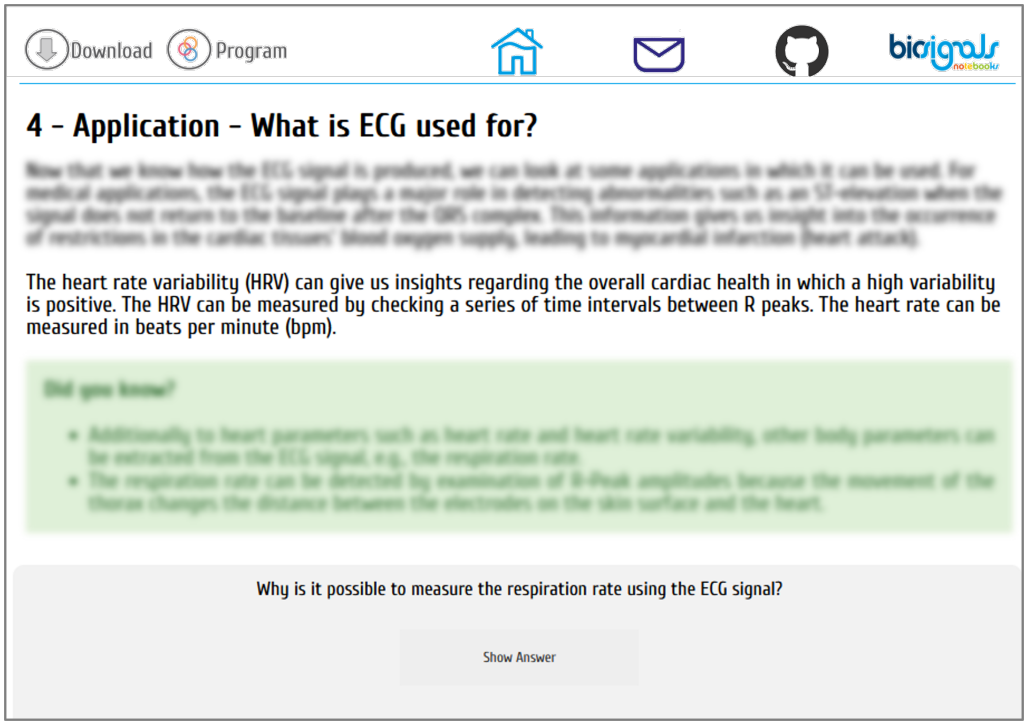

Figure 4.12: Example of a portion of the Introduction to the ECG lesson.

At the end of the lesson, a series of 40 questions was presented, each of which relative to one of the theoretical paragraphs. These questions and the respective answers were generated with the help of GPT-2 [206] and later reviewed and any errors were corrected. These questions were meant to serve as auto-evaluation for the participants, since that, instead of an answer to them, the answer was shown and they were asked if they found it

"Easy", "Hard", "Very Hard", or "Repeat"(this was the case when the participants did not remember to read anything related to the question or respective answer). An example of a question and answer is shown in Figure 4.13.

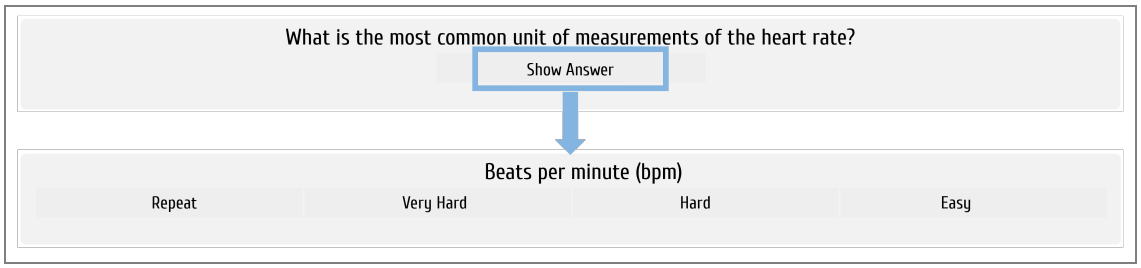

Figure 4.13: Example of a question and respective answer that is shown after clicking "Show Answer".

The answers to the questions would allow us to get a label to each of the read paragraphs. Namely, if the participants found the questions related to a paragraph as "Easy", then it was considered that they were able to learn when reading it and vice-versa otherwise. However, given that most of the recruited participant were biomedical engineering students, they already knew the answers to the questions and, thus, their answers did not reflect their cognitive states while reading and interacting with the lesson and, thus, this task will not be further explored.

4.3.3 Data Acquisition Procedure

Ten volunteers, consisting of 4 females, participated in this study, with ages ranging from 22 to 48 years old (M = 28.2, SD = 7.6). They were recruited from the NOVA School of Science and Technology. All participants were fluent in English, right-handed, reported no psychological disorders, and did not take regular medication. Prior to participation, written informed consent was obtained, and the study strictly adhered to the Ethical Procedures approved by the Ethics Committee of NOVA University of Lisbon.

The biosignals were acquired with a sampling rate of 100 Hz and a resolution of 16 bits, except for the fNIRS that was acquired with a resolution of 24 bits. The change of the sampling rate relative to the study presented in Section 4.2 was due to the fact that a higher sampling rate lead to high computing demands that resulted in loss of data during data acquisition.

4.3.4 Data Pre-Processing, Segmentation, and Labelling

In this study, the focus was on the processing of the fNIRS signals for the detection of cognitive fatigue, as investigated in other studies, e.g., [80–82, 110, 111]. Thus, only the fNIRS signals acquired during the two 2-minute baseline periods were considered.

The first step of signal processing is then to apply the modified Beer-Lambert law as presented in Equations 2.4 and 2.3, to convert the detected values to variation of

concentration of each chromophore of haemoglobin. Since the differential path length is the same in both cases, it was removed and, thus, the variation of the concentration was calculated in order to the factor (unit of mM/DPF). Given that the wavelengths of the emitters used in this work were 660 nm (λ_1) and 860 nm (λ_2), the extinction coefficients were used according to the work of Matcher et al. [207]:

$$\alpha_{Hb}^{\lambda_1} = 3.4408 \quad \text{mM}^{-1}\text{cm}^{-1}$$

$$\alpha_{O_2Hb}^{\lambda_1} = 0.3346 \quad \text{mM}^{-1}\text{cm}^{-1}$$

$$\alpha_{Hb}^{\lambda_2} = 0.7977 \quad \text{mM}^{-1}\text{cm}^{-1}$$

$$\alpha_{O_2Hb}^{\lambda_2} = 1.2071 \quad \text{mM}^{-1}\text{cm}^{-1}$$

Additionally to each chromophore, the variation of total haemoglobin (THb) was calculated, which is simply the sum of the previously mentioned variations of the concentrations of O_2Hb and Hb. After this conversion, the signals were filtered using a second-order band-pass Butterworth finite impulse filter with cut-off frequencies of 0.01–1 Hz, to remove electrical and physiological noise, keeping the most informative frequency band of the fNIRS signals, according to the recommendations in [208].

The baseline signals were then segmented into 10-second windows with no overlap according to the two considered periods—before starting the tasks and after finishing them. A label was attributed to each window, considering if it belonged to the period before the tasks, in which case it was labelled as 'resting state', or if it belonged to the period after all tasks, in which case it was labelled as 'cognitive fatigue'. This processing procedure resulted in 22 time windows for each participant (11 windows per baseline period), as the last 10-second window of each baseline period was discarded to assure that no task was present.

4.4 Multimodal Data Sources Classification for Learning State Monitoring

In the third and last study performed in this work, the approach was broadened and instead of focusing on specific cognitive processes, the main objective was to classify learning states that include various processes. This approach was first proposed in other works, namely, D'Mello and Graesser (2012), to effectively monitor complex learning states involving multiple cognitive processes, as explored in Section 2.1 [209].

The main improvement to the previous study, was the introduction of the self declared ground-truth for classification, namely, considering the strong assumption about how the answers to the questions of the Introduction to the ECG lesson related to cognitive states during learning, which was not accurate. Thus, in this case, the whole procedure was simplified, as illustrated in Figure 4.14.

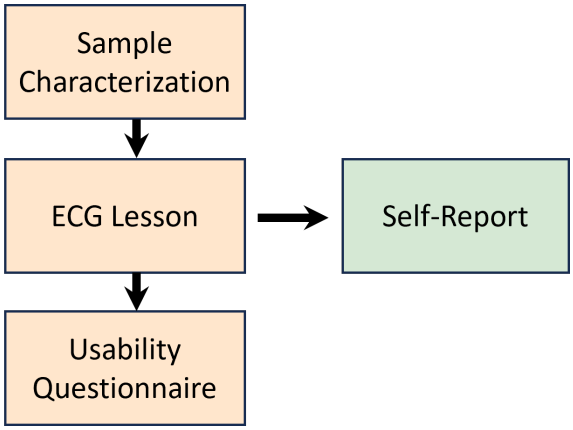

Figure 4.14: Schematic of the procedure followed during the Multimodal Data Sources Classification for Learning State Monitoring. The procedure consists of the modified lesson about the ECG, followed by a self-report stage.

The data acquisition procedure focused solely on a learning task, that consisted of a modified version of the Introduction to the ECG lesson, disposing the cognitive tasks of the previous studies. However, a self-report stage was added to generate a clear ground-truth about the participants cognitive states.

4.4.1 Introduction to the ECG Lesson - Modified Version

In this case, the Introduction to the ECG lesson presented in Section 4.3.2 was modified in a crucial aspect that could help in the self-reporting stage. Although the theoretical contents were the same as before, instead of having a set of questions at the end of the lesson, open-ended questions were added at the end of each of the four sections. These questions were related to the sections the subject was reading and allowed the participants to reflect about their learning process since they would need to recall what they had read to properly answer the questions. Given the need to properly explain and justify their answers, during the self-reporting stage, it was assumed the participants would have a better understanding about their own learning process. An example of such a question is shown in Figure 4.15.

The paragraphs of the lesson and related figures remained blurred unless the mouse cursor hovered over them. Only mouse-tracking metrics, keypresses, and screenshots were saved, related with the variables detailed in Section 4.1.2. Video and audio were captured via a computer-to-computer video call, synchronized with screen recording for subsequent use in the self-reporting phase.

4.4.2 Self-Reporting

During the execution of the modified Introduction to the ECG lesson, a video-call between the laptop that the participant was using and a second computer was made. During this call, the subject was recorded along with the screen the participants were watching. Thus,

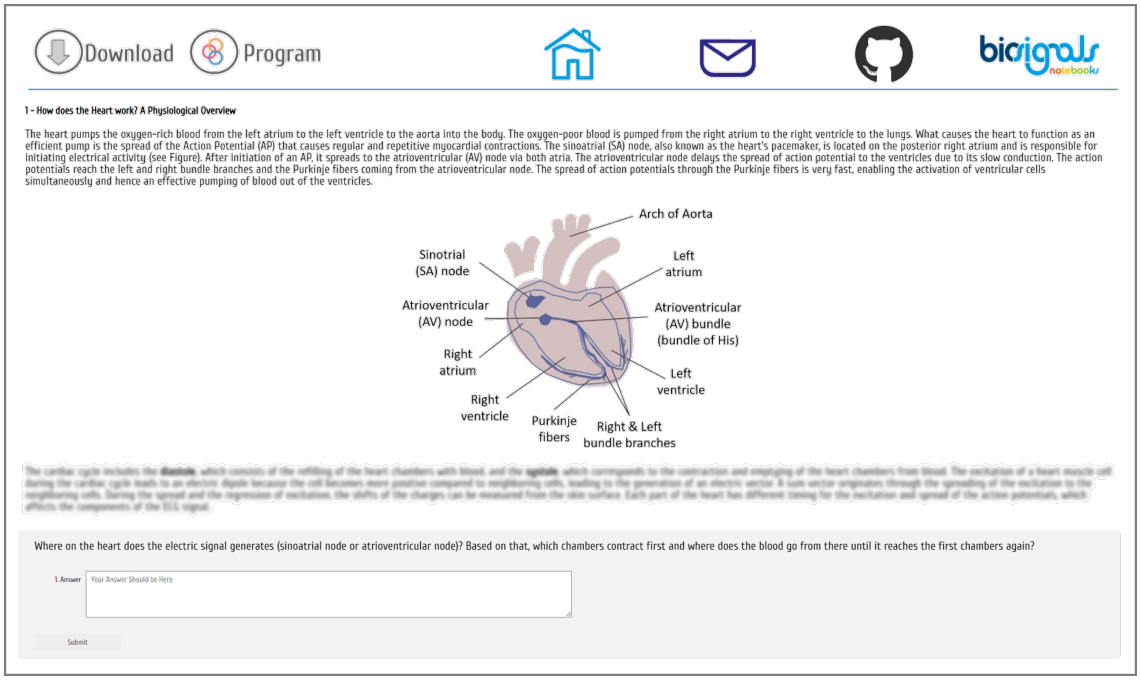

Figure 4.15: First section of the modified version of the Introduction to the ECG lesson.

at the end of the lesson the participants were able to have access to the video as shown in Figure 4.16.

The participants watching the video would be asked to select one of the following complex learning states with the corresponding definition:

Neutral - without discernible emotion or state.

Interest/Flow - a state of interest provoked by involvement in a task.

Surprise - discovering something new.

Boredom - lack of interest in the task.

Distraction - loss of attention due to internal or external factors.

Confusion - provoked by not understanding what is being read/asked.

Eureka - discovering an answer after being in difficulty.

Frustration - a feeling of dissatisfaction or disappointment, usually caused by not achieving an expectation or goal.

For this, the participants were asked to either label changes in the learning process or if 2-minutes passed since the previous label. Since all participants had access to the same list of learning states and the corresponding definition, it was assumed that the reporting of different participants were comparable, even though these are usually viewed as subjective measures.

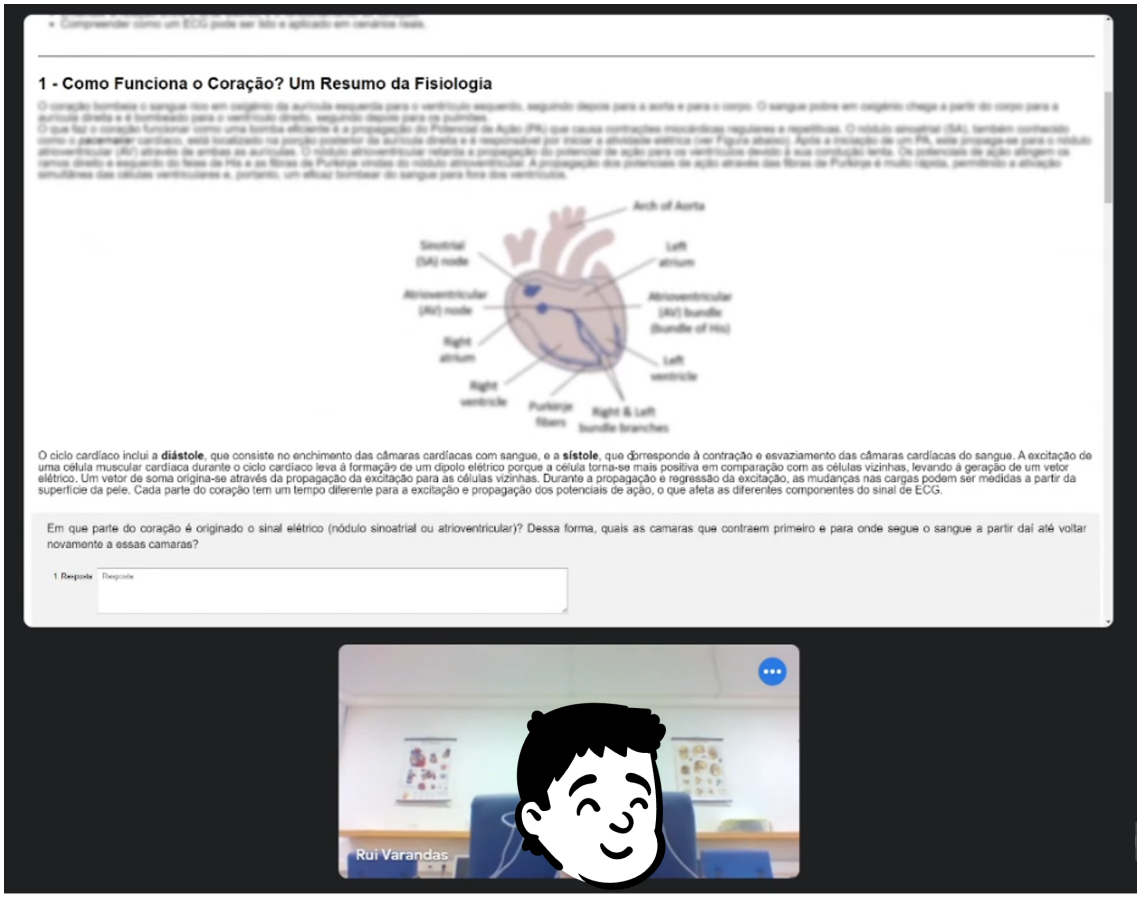

Figure 4.16: Interface of the self-reporting stage.

4.4.3 Data Acquisition Procedure

In this study, 18 volunteers (13 females) participated with ages ranging from 18 to 28 (M = 21.7, SD = 3.2). The participants were recruited from the NOVA School of Science and Technology, they were all biomedical engineering students, they were right-handed, reported no psychological disorders, and did not take regular medication. Prior to participation, written informed consent was obtained, and the study strictly adhered to the Ethical Procedures approved by the Ethics Committee of NOVA University of Lisbon.

The biosignals were acquired with a sampling rate of 100 Hz and a resolution of 16 bits, except for the fNIRS signal that were acquired with a resolution of 24 bits, similarly to the study presented in Section 4.3.

During these acquisitions, one of the EEG channels saturated with electromagnetic noise in one of the participants and two participants had their EDA signals saturated due to excessive sweat and these signals were not considered for further analysis. Moreover, HCI variables were not tracked for one of the participants and, thus, only 17 participants were considered when classifying them.

4.4.4 Data Pre-Processing, Segmentation, and Labelling

This final study aimed to distinguish between complex learning states that involve different cognitive processes. With this intent, the different data sources, namely, all biosignals described in Section 4.1.1 and the mouse-tracking measures were explored as described in Section 4.1.2.

The mouse-tracking signals were pre-processed to match the sampling rate of the biosignals and then synchronized using the timestamps of both data sources. The data was then processed differently for the different biosignals and mouse-tracking, which will be described in detail in Section 4.5.

Then, the signals were segmented according to the timestamps where there were changes in the learning states as described in Section 4.4.2. After this first segmentation step, all signals were then segmented into constant time widows of different periods: 10-second, 12-second, 20-seconds, 30-second, 40-second, 50-second, 60-second, and 90-second. These different time-windows will allow us to study the best segmentation strategy for this application, answering **RQ6**. In every case, if a given state lasts less than the considered time window, it is discarded.

Then, given the different data sources, different features were extracted from each of them, as described in Section 4.5.

On the other hand, four different approaches were followed for classification regarding the considered classes:

Original - All the learning states are considered as mentioned in the self-report stage, thus having eight classes for the classification;

Cascade - Taking into account that one of the classes is more prevalent than all the others, the classification is first performed to detect *Interest/Flow vs. Others* and then the others into the different learning states;

Binary - Consider only the classification between *Interest/Flow vs. Others*, since *Interest/Flow* is the main state where learning is optimal.

Gathered - Given the works of other authors, the learning states were gathered considering the *negative activating* states - Confusion and Frustration -, *negative deactivating* states - Boredom, and Distraction -, *Neutral*, and *positive activating* states - Interest/Flow, Surprise and Eureka.

Finally, the application of different classifiers will be studied: Decision Trees (DT) [210], Random Forest (RF) [211], Multi-Layer Perceptron (MLP) [212], Adaboost (AB) [213], Gaussian Naive Bayes (GNB) [214], Linear Discriminant Analysis (LDA) classifier [215], Gaussian Processes (GP) classifier [216], and a MC using a voting system of the prediction of a RF, an AB classifier, and a MLP. These will be tested to study the impact that different classifiers may have in this work and to help answering **RQ5**.

The full set of studied variables is displayed in Table 4.2.

Table 4.2: List of variables tested regarding the used classifiers, normalisation techniques, time windows for segmentation, and class types for classification.

Classifier	Normalisation	Window	Class Types
- Random Forest	- Z-Score	- 10s	- Original
- Multi-Layer Perceptron	- Min-Max	- 12s	- Gathered
- AdaBoost	- Raw Features	- 20s	- Binary
- Gaussian Naive Bayes		- 30s	-Cascade
- Decision Trees		- 40s	
- Linear Discriminant Analysis		- 50s	
- Gaussian Processes		- 60s	
- Multi-Classifier		- 90s	

4.5 Data Processing

This chapter delves into the core data processing stages. After acquiring and pre-processing data from three studies, the focus now shifts to features extraction. Specific features will be extracted from each dataset and also deriving agnostic features spanning all sources. These techniques find unique attributes within the data, increasing our understanding of underlying patterns and relationships across the datasets, and aid classifiers to distinguish different cognitive and learning states.

4.5.1 Electroencephalography Data Processing

EEG signals correspond to the electrical signals generated by the brain captured on the surface of the head, as explored in Section 2.4.1.1.

First, the EEG signals were filtered using a low-pass Butterworth second-order filter with a cut-off frequency of 49.9 Hz to keep the most relevant band frequencies and removing the 50 Hz frequency that included electromagnetic environmental noise.

Given the frequency bands referred earlier, the first extracted features were their power. Thus, the Welch method was computed, as shown in Figure 4.17 [217] and then calculate the area under each of the frequency bands using the trapezoidal rule. After computing the power of the frequency bands, the ratios between each pair were computed. Still using the Welch method, the dominant frequency was computed, which refers to the frequency corresponding to the maximum power, to be used as another feature.

Finally, the Hu moments of the plots corresponding to the signals and their delayed phase of 40 ms using OpenCV were computed. The transformation from time series into figure is illustrated in Figure 4.18. Then, the Hu moments, which are designed to characterize the shape or contour of an object within an image were computed. These moments describe the shape of an object in a figure, such as its area, centroid, orientation, and higher-order shape features. These moments were also used as features for classification.

These features will then be used as input for classification to potentially separate the classes that indicate different cognitive states during the learning process.

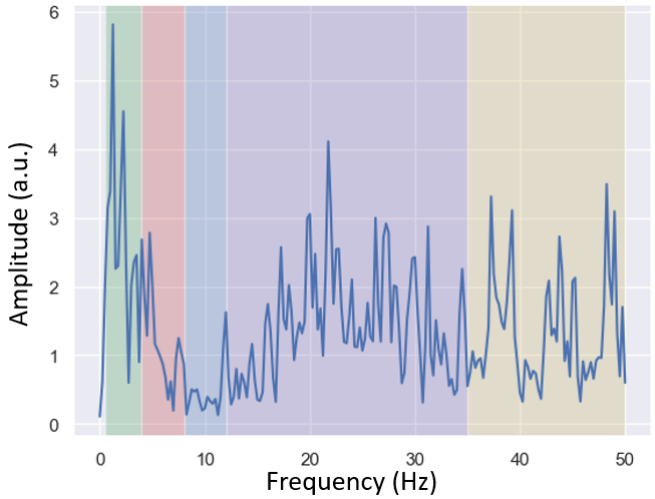

Figure 4.17: Welch plot of an EEG signal. The green shadow represent the Delta band, the red represents the Theta band, the blue represents the Alpha band, the purple represents the Beta band, and the yellow band represents the Gamma band.

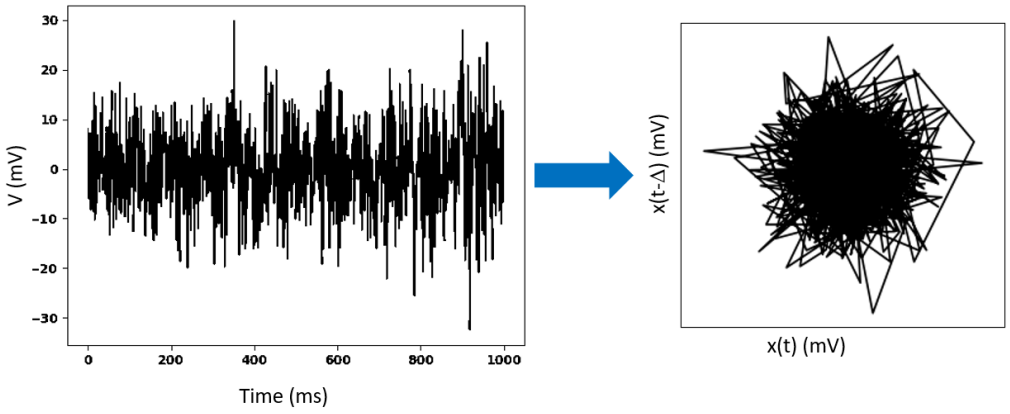

Figure 4.18: Transformation of the EEG signal into a figure from where the Hu moments are calculated. Δ is 40 ms.

4.5.2 Functional Near Infrared Spectroscopy Data Processing

The fNIRS signals correspond to the variation of the light intensity between a pair of emitting LEDs and a detector placed on the surface of the head. This variation is caused by the absorption and scattering of the photons inside the head caused by the interaction with Hb and O_2Hb. In this case, the first processing step is the application of the modified Beer-Lambert law, in Equation 2.1, to calculate the variations of the concentration of both types of haemoglobin, followed by the application of Butterworth second-order filter, as mentioned in Chapter 4. A third component was calculated that corresponded to the THb by the addition of both variations of concentrations. The specific features of the fNIRS signals are agnostic of the chromophore, i.e., all the features are calculated and extracted from the three different variables.

The first extracted feature is the Root Mean Square (RMS), that corresponds to the

squared root of the mean of the squared signal (s):

$$RMS = \sqrt{\overline{s^2}} \tag{4.1}$$

Then, given the neurovascular coupling, it is expected that the changes of the O_2Hb and Hb to be inverse, i.e., an activation leads to an initial decrease in the concentration of the O_2Hb and an increase of the concentration of Hb, followed by an inversion with a high increase of the concentration of O_2Hb. Thus, the slope of a linear regression of the chromophores and also its naive version were computed:

$$NS = s_N - s_0 \tag{4.2}$$

where, s corresponds to the signal, N is the number of data points, and s_0 corresponds to the first data point of the considered segment.

In the same logic, the integral of the signals were calculated which can indicate about their overall power. For this, the composites Simpson rule present in Scipy was used, with a spacing of $1/f_s$, being f_s the sampling frequency. Also, the maximum variation of the signals were calculated by computing the consecutive difference of the smoothed signal, computing the difference between the consecutive data points, and then the maximum value of the result.

Again, these features calculated from the three components of the fNIRS signal will serve as inputs for classification to try and differentiate between cognitive or learning states.

4.5.3 Electrocardiography Data Processing

The ECG signal corresponds to the electrical signals coming from the heart. These signals control the heart beat, that reflects the state of the ANS and, thus, by monitoring it, the ANS function can be monitored indirectly.

The first processing step is the application of a second order Butterworth filter according to Table 4.1. After this, the Pan-Tompkins algorithm is applied to detect the R peaks and then compute the heart rate [218]. An example is shown in Figure 4.19.

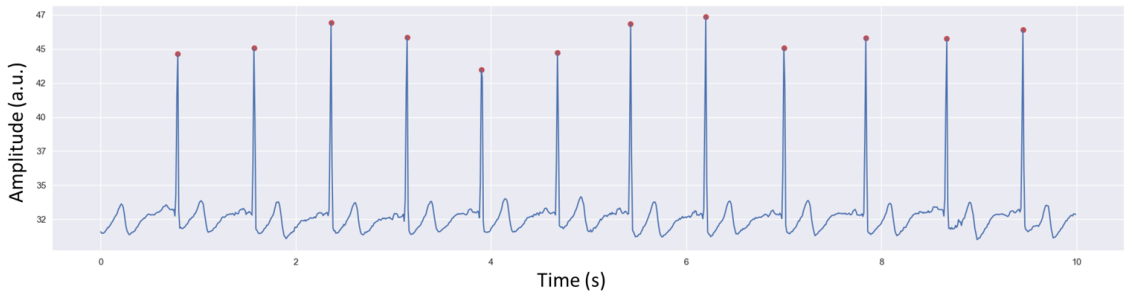

Figure 4.19: Example of an ECG signal with the R peaks marked with red dots, detected by the Pan-Tompkins algorithm.

From this, a new virtual signal is generated corresponding to the heart rate through time, called a tachogram, by calculating the differences in time of the consecutive peaks. Then, temporal, spectral, and non-linear features were computed from the Poincaré plot.

The temporal features correspond to the heart rate variability features in the time domain. Namely, the standard deviation and the RMS of the standard deviation of the tachogram with no ectopic beats, that in this case correspond to the beats with a variation higher than 20% relative to the prior heart beat, were computed. Besides these, the number and percentage of RR intervals longer than 50 ms, and the number and percentage of RR intervals longer than 20 ms as features were also used.

On the other hand, features from the spectral domain that can be used to assess about the sympathetic and parasympathetic nervous system were computed [63]. In this case, the power of the very low frequencies — 0.0033-0.04 Hz — related to physical activity, low frequencies — 0.04-0.15 Hz — related and mediated by the vagal and sympathetic systems, high frequencies — 0.15-0.4 Hz — mediated by the parasympathetic system, and the total from 0.0033-0.4 Hz were extracted [63]. Besides, the activation of each frequency band relative to the overall power was also computed. To highlight the balance between the sympathetic and parasympathetic systems, low frequency and high frequency are assessed in normalized units (nu), that represent the relative value of each power component in relation to the total power minus the very low frequency component [63]. Although some of the considered time windows do not allow these to carry physiologically-relevant information, they were kept as they might still help to distinguish different learning states.

Finally, non-linear features related to the Poincaré plot were computed. This plot corresponds to the plotting of each point of the tachogram against its previous point in time. An example of such a plot is shown in Figure 4.20.

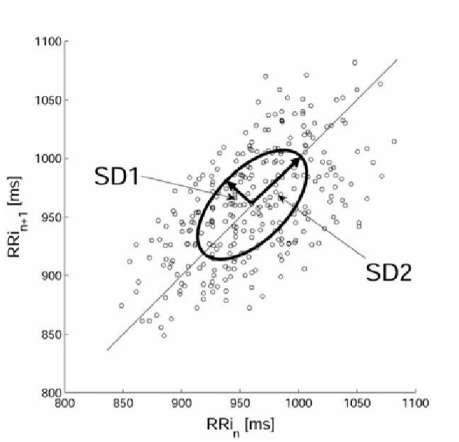

Figure 4.20: Example Poincaré plot with SD1 and SD2 identified. From [219]

From the plot, three features were extracted: SD1, which is the length of the transverse line in Poincaré plot, SD2, the length of the longitudinal line in Poincaré plot, and SD Ratio, ratio that represents SD2/SD1.

Other features could have been extracted, for example from the Detrended Fluctuation Analysis, however, we focused on the most studied and established features from previous works.

4.5.4 Electrodermal Activity Data Processing

The EDA signal corresponds to the measurement of the variations of the conductivity on the surface of the skin, namely, in this work, on the palm of the non-dominant hand. This conductivity varies due to the activity of the sweat glands.

Given this signal is composed by a tonic and a phasic component, as described in Section 2.4.4, features from the raw signal and from those components were extracted. To do so, the Neurokit Python package that includes processing functions for multiple biosignals was used [220].

First, given the interest on the morphology of the signals, standardization was applied using Equation 2.8. Then, the mean, standard deviation, peak-to-peak amplitude, minimum value, and maximum value of the filtered signal, corresponding to the application of a low-pass second order Butterworth filter of 1 Hz, and the two components the EDA signal, decomposed by applying a low-pass/high-pass second order Butterworth filter with a cut-off frequency of 0.05 Hz were extracted.

From the phasic component, it is possible to detect peaks and study each of them individually. In this sense, after detecting every peak of the phasic component, their onsets were detected, the amplitude with and without the tonic component was calculated, the rise time from onset to peak was also calculated, and, finally, the recovery of those peaks was also detected. Given there might be a varying amount of peaks in a time segment, the mean, standard deviation, peak-to-peak amplitude, minimum value, and maximum value of these features were also calculate .

In conclusion, the analysis involved comprehensive feature extraction from both tonic and phasic signal components. Using the Neurokit Python package, the focus was on the separation of the tonic and the phasic component, and the individual peak characteristics were thoroughly explored. This approach might help the classifiers to effectively distinguish between different cognitive or learning states.

4.5.5 Respiratory Inductive Plethysmography Data Processing

The RIP signal corresponds to the variation of the displacement of the rib cage, enabling to monitor respiration cycles and other related features.

First the signals were standardized, applying Equation 2.8 and then detected the peaks of the signal, if any, again using the Neurokit Python package. From this, the maximum amplitude of the detected peaks, the mean amplitude of the peaks, the standard deviation of their amplitude, their RMS, the mean time interval between peaks, and its inverse to get the respiration frequency were calculated. Finally, if a segment had more than 10 peaks, the mean, standard deviation, minimum, maximum, and RMS of the interval between the

last 10 peaks were also calculated, that might indicate a variation considering the overall statistics.

Besides these peaks-related features, features from the raw signal were also extracted. Here, the mean, minimum, maximum, standard deviation, and RMS values were calculated. Furthermore, the number of times the signal passed through zero (zero-crossing rate) was computed, as well as the energy of the signal, and, finally, the integral of the signal computed using the trapezoidal rule.

Given that most of these features depend on the peak detection and on the existence of multiple peaks, these features were only considered in those circumstances. This is important, because for short segments (e.g., segments of 10 seconds), the signal might not include any peak, since the respiration frequency is typically around 0.1 Hz. In such cases, the features considered for the RIP signal were the agnostic ones, presented and described in Section 4.5.8.

4.5.6 Accelerometer Data Processing

The ACC signal is composed of three different axis. The sensor was attached to the side of the head to measure its movements that can possibly indicate or correlate to different cognitive or learning states. First of all, a fourth signal corresponding to the magnitude was calculated:

$$s_{magn} = \sqrt{s_x^2 + s_y^2 + s_z^2} \tag{4.3}$$

and a visual representation is shown in Figure 4.21.

In this case, the sensor measures the signals in g units and the first processing step was then to convert them to m/s^2. Then the signals were standardized using Equation 2.8. From this, the acceleration signal was integrated to calculate the velocity and integrated again to get the displacement of the sensors and, thus, of the head. The mean, minimum, maximum, and standard deviation of the acceleration, velocity, and displacement of each of the axis and the magnitude were then computed.

While this sensor has not been extensively employed in cognitive monitoring studies, we anticipate its utility in distinguishing various cognitive and learning states, as posture could potentially serve as an indicative factor for these states.

4.5.7 Mouse-Tracking Data Processing

Besides the biosignals, the behaviour while participants interacted with the computer during the tasks mentioned in Chapter 4 in the form of mouse-tracking was also monitored.

After processing the data as described in Section 4.4.4, a set of features was extracted from each of the time segments, described in more detail in [63]. The extracted features were: mean of the linear spatial interpolation of the xx axis, mean of the linear spatial interpolation of the yy axis, cumulative distance travelled, straightness of the distance

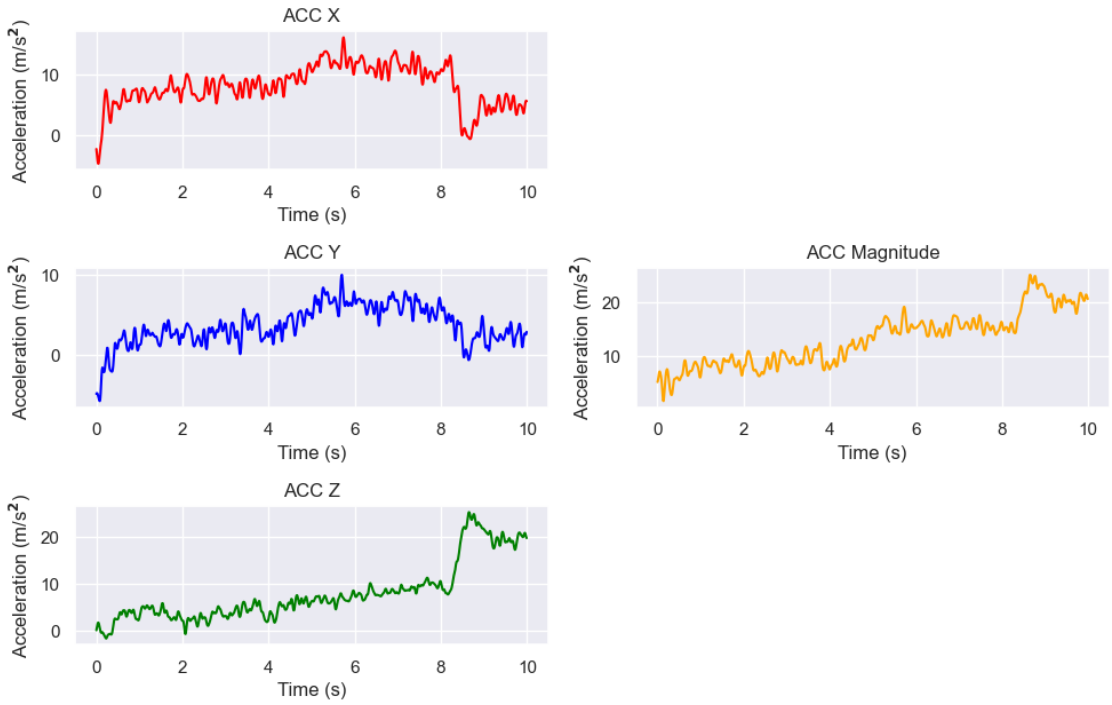

Figure 4.21: Example of the calculus of the magnitude signal based on the three axis of the accelerometer. On the left side, the xx axis, the yy axis, and the zz axis are shown, while on the right side the corresponding magnitude is presented.

travelled, jitter nervous pattern, mean distance between the linear temporal interpolation of the axis, mean distance between the linear spatial interpolation of the axis, angle of the interpolated xx and yy axis, mean angular velocity, mean curvature between the actual path travelled and the direct path between the first and last points, rate of change of curvature, total distance travelled considering the spatial interpolation of both axis, mean of the linear temporal interpolation of the xx axis, mean of the linear temporal interpolation of the yy axis, mean total velocity, mean velocity of the xx axis, mean velocity of the yy axis, mean total acceleration, jerk - a quick, sharp, or sudden movement-, total time, number of pauses longer than 1 second, mean time of pauses longer than 1 second, time of interaction, and number of turns.

With this extensive list of features, we aim to capture the behaviours that identify different states, such as, increased nervousness or calmness, that might relate to cognitive or learning states.

4.5.8 Agnostic Features Extraction

Agnostic features are features that can be extracted and calculated from any kind of signal, such as, ECG, mouse-tracking, or even the magnitude of the ACC sensor. These can be statistical, temporal, spectral, or others. In this step, the TSFEL Python package that implements an extensive list of functions was used, which allows to easily extract

features from time-series data [221]. The features were selected taking into account the computational complexity of the extraction process, since a system like this would benefit from a real-time or close to real-time application.

Starting by the statistical features, the mean, median, minimum, maximum, RMS, standard deviation, and variance were extracted. Besides these standard metrics, the polarity of the signal was also extracted, as shown in Equation 4.4.

$$\text{Polarity} = \frac{s_{max}}{s_{min}} \tag{4.4}$$

Moreover, the skewness and kurtosis of the distribution of the signals, that represent how skewed the data is and also how dispersed it it, respectively, was also calculated. The equations that represent them are:

$$\text{Skewness} = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{s_i - \overline{s}}{\sigma_s} \right)^3 \tag{4.5}$$

$$\text{Kurtosis} = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \left(\frac{s_i - \overline{s}}{\sigma_s} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)} \tag{4.6}$$

where s_i represents the i^{th} point of signal s, $\overline{s}$ is the mean of the signal, σ_s is the standard deviation of signal s, and n is the number of data points.

Yet in the statistical domain, the interquartile range of the distribution of the signals, the mean absolute deviation, the median absolute deviation, the Shannon Entropy, the area under the curve of the signal computed with trapezoid rule. Then, the absolute energy and the total energy of the signals were extracted. Finally, the mean and median of the difference of the signal were calculated.

In the time domain, the peak to peak distance, the slope of a linear regression of the data, and the zero-crossing rate of the signal were computed

Then, a list of features on the spectral domain were also extracted. Namely, the fundamental frequency, that in this case represents the first peak higher than 30% of the maximum amplitude of the frequency spectrum, ignoring 0 frequency was calculated. Furthermore, the maximum frequency - the frequency corresponding to 95% of the cumulative spectrum -, the maximum of the power spectrum, the median frequency - the frequency corresponding to 50% of the cumulative spectrum -, and spectral Kurtosis, Skewness and entropy were extracted. Additionally, the spectral roll-off of the signal which is the frequency where 95% of the signal magnitude is contained below this value, the power spectrum density bandwidth of the signal, corresponding to the width of the frequency band in which 95% of its power is located were computed. Finally, the spectral distance, that corresponds to the distance of the signal's cumulative sum of the FFT elements to the respective linear regression was computed.

In this case, if any of the features failed to compute, then it was filled by a NaN (not a number) value and further processed during classification.

In this study, all these features were extracted from all the signals: ECG, EEG, EDA, RIP, ACC (all three axis and magnitude), Hb, O_2Hb, THb, and mouse-tracking. Although some of these features might overlap with specific features from the specific biosignals or mouse-tracking data, they were still considered and handled during classification.

4.6 Summary

In this chapter, the materials - hardware and software - used during this work to support the data acquisition procedures were described. These procedures allowed us to study different cognitive processes, namely, attention, cognitive fatigue, and others involved in complex learning states. The incremental nature of these procedures proved crucial in refining subsequent studies, aligning them with this work's objectives for measurement and classification.

The biosignalsplux and all the used sensors, OpenSignals, and Latent that helped us acquire and synchronize the biosignals and mouse-tracking data were presented.

Then, the different studies performed during this work, as well as, all the cognitive tasks and e-learning lessons that support them were described. Namely, in the first study, the N-Back task with increasing difficulty levels, a mental subtraction task, and a Programming Language lesson to detect attention were applied.

In the second study, the tasks were changed. A digital version of the Corsi block task, a concentration test, and an Introduction to the ECG lesson, that allowed us to acquire more mouse-tracking data than in the first study were employed. However, an assumption was made that did not hold true for most participants of the study, leading us to focus on the cognitive fatigue detection.

In the third study the procedure was simplified to focus on the detection of complex learning states during a modified version of the Introduction to the ECG lesson. In this case, a self-report stage was added that allowed us to get a common ground-truth from all the participants. With this, the impact that different classification pipelines might have on the classification results will be tested, answering Questions 4 and 5 presented in Section 1.3.

Finally, in this chapter, the features extracted from the multiple data-sources used during this work were also explored. It was evident that different signals required different processing aspects and that features extraction needs to be thought of in order to represent the signals concisely and effectively.

5

Results

This chapter will present the results achieved in the performed studies described in Chapter 4. Namely, it will describe how the classification was performed, which classifiers were used and what types of variables were studied.

The results of the Attention Classification Based on Biosignals and Cognitive Fatigue Detection Based on Biosignals studies were published in [201] and [205], respectively.

5.1 Attention Classification Based on Biosignals

In this study, the main objective was to classify different attention states based on the execution of cognitive tasks. In this case, instead of using the comprehensive set of features described in Section 4.5, the TSFEL [221] was used to extract the following features: maximum, minimum, polarity, mean, variance, standard deviation, kurtosis, skewness, mean of differences, total energy, area under the curve, absolute energy, peak to peak distance, entropy, slope of linear regression, fundamental frequency, maximum frequency, power bandwidth, spectral distance, median frequency, and spectral entropy.

Then, a RF algorithm was used to perform the model training and classify the attention events [222]. In this case, the train and test sets were selected by stratified k-folds cross-validator. This cross-validator separates the training and test sets maintaining the proportions of classes of the original data in both sets. The models were trained with individual biosignals or their combinations. The different combinations of biosignals depend on the general results of each sensor. Mouse-tracking data was not considered here due to the lack of computer interaction during the execution of the cognitive tasks.

Moreover, when the datasets are imbalanced, i.e., the predictive classes do not exist in the same quantity, the models could produce biased results. In this scenario, the two classes were never exactly in the same quantity, and therefore, to handle this issue, a random undersampling was applied, that consists of sampling from the majority class to keep only a part of it. This way, the majority and the minority classes have the same dataset size [223].

In this case, the features were first standardized, according to Equation 2.8, and, thus,

the final set of values had a mean value of zero and a standard deviation of one. To keep the test set unseen during the training phase, the normalization is first performed to the training set of features, recording each mean and standard deviation. The normalization of the testing set is based on the recorded mean and standard deviation of the training set.

The Random Forest classifier is a supervised learning algorithm consisting of an ensemble of DTs, that produces a more accurate and stable prediction merging multiple DTs. To overcome the sensitivity of the DTs, in a RF each tree is trained on different sets of data through bagging, a method that randomly samples a data set with replacement. The features considered in each node are from a random subset of features to create an uncorrelated forest of trees whose prediction is more accurate than that of any individual tree. The subset of random features should be smaller than the set of features to avoid this problem. Every decision tree consists of decision nodes, leaf nodes, and a root node. The leaf node of each tree is the final output produced by that specific DT. For the final decision, the RF classifier aggregates the results of individual trees. The selection of the outcome follows the majority-voting system leading to the RF classifier exhibiting good generalization [224]. In this case, the subset of features used at each node was limited to the logarithm of the number of features. The number of trees was limited to 200 due to computational time cost.

The model training procedure generated 13 models for each subject, considering each biosignal separately, all biosignals combined, and the following combinations:

- fNIRS-EEG
- EEG-ECG
- EEG-ACC
- EEG-ECG-ACC
- fNIRS-EEG-ECG
- fNIRS-EEG-ECG-EDA

Regarding the 8 participants and the above combination of biosignals, in the end, this task generated 104 different models. After the biosignals processing part, the application of the filter for uneven sampling to the segments of 10 seconds removed one single segment from a total of 810.

The segmentation procedure, that differentiates periods of *task* and *baseline*, was performed for each participant and the total number of samples belonging to each class is presented in Table 5.1. As already mentioned, to use these classes to perform model training, it might be important to have a balanced dataset. After randomly downsampling the classes to keep only the original data, the total number of tasks will have exactly the same number as the *baseline*, given that *task* is the majority class for all the participants and assuming that the number of samples is sufficient to train the classifiers.

Table 5.1: Initial number of samples for the considered classes for each participant.

Participant	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
# task	54	60	37	56	74	56	52	66
# baseline	48	48	19	47	45	50	50	47

Ten folds were selected to apply the stratified k-folds cross-validator given the previously presented size of the final subsets. This way, the minimum test set by fold contains three samples (participant A), and the maximum test set by fold contains ten samples (participant F or G). Therefore, the evaluation of each model is performed in 10 different test folds so, ten distinct accuracy scores are calculated for each model. The final accuracy is calculated based on the mean of the 10-folds accuracy scores.

Considering all the individual models built from the different combinations of sensors, Table 5.2 presents the accuracy scores of the models based on each sensor separately to evaluate the performance of each biosignal to predict attention.

Table 5.2: Individual accuracy to predict attention based on each sensor. Results marked with * can be inverted and the scores would be $\text{score}_{inv} = 1 - \text{score}$ because this is a binary classification case.

Participant	**fNIRS**	**EEG**	**ECG**	**EDA**	**RIP**	**ACC**
A	0.57	0.60	0.73	0.52	0.72	0.53
B	0.62	0.72	0.59	0.53	0.65	0.66
C	0.76	0.72	0.57	0.50	0.64	0.69
D	0.48*	0.58	0.62	0.44*	0.48*	0.54
E	0.51	0.51	0.79	0.44*	0.58	0.63
F	0.52	0.64	0.57	0.61	0.51	0.68
G	0.68	0.73	0.56	0.66	0.52	0.60
H	0.51	0.64	0.78	0.63	0.63	0.55

Results lower than 0.5, indicated with * in Table 5.2, can be inverted and the scores would be $\text{score}_{inv} = 1 - \text{score}$. This is possible because the datasets are balances and it is a binary classification problem. Thus, if a score lower than 0.5 would be detected, i.e., worse than the random chance, then the predicted labels should be inverted by classifiers to get the best classification possible.

Except for one participant (A), all the others achieve better results with training based on more than one biosignal. Table 5.3 presents the best accuracy scores reached for each participant. To reduce the required equipment, the accuracy of the models trained with a maximum of two sensors is also presented in Table 5.3.

The number of considered sensors is important because it can have an impact on the setup of a possible solution to monitor students outside the laboratory, thus, to keep ecological validity. Namely, the higher the number of sensors, the more complex the setup, which can have an impact on the utilization of such a system. Moreover, an increased

Table 5.3: Individual accuracy to predict attention based on combined biosignals. The accuracy scores achieved using a maximum of up to two sensors are marked with *.

Participant	Sensors	Accuracy
A	ECG	0.73
B	fNIRS-EEG	0.73
C	fNIRS-EEG-ECG	0.81
C*	EEG-ACC	0.78
D	fNIRS-EEG-ECG-ACC-RIP-EDA	0.68
D*	ECG	0.62
E	EEG-ECG	0.80
F	fNIRS-EEG-ECG-ACC-RIP-EDA	0.76
F*	EEG-ACC	0.71
G	fNIRS-EEG	0.77
H	EEG-ECG	0.78

number of sensors can also have an impact on the comfort of the utilization of such a system and be detrimental for learning.

5.2 Cognitive Fatigue Detection Based on Biosignals

In this study, the objective was to distinguish the cognitive states of participants prior and after highly demanding cognitive tasks, assuming that the state after them reflected cognitive fatigue. In this case, the focus was on the use of the fNIRS sensor since it is one of the most used sensors in this area of study.

The duration of each task, the total duration of the agglomerate of the tasks, and the average time to complete each task is presented in Table 5.4. Given that the tasks did not have time limits, their duration varied among individuals.

Table 5.4: Time each participant spent on each individual task.

Participant	ECG Lesson	Corsi-Block	Concentration	Total
A	16m 05s	07m 12s	26m 32s	49m 50s
B	23m 42s	05m 02s	19m 51s	48m 36s
C	10m 12s	05m 49s	17m 25s	33m 26s
D	11m 55s	06m 18s	20m 24s	38m 38s
E	19m 56s	06m 12s	19m 27s	45m 35s
F	18m 48s	04m 38s	30m 24s	53m 50s
G	17m 02s	04m 47s	20m 00s	41m 49s
H	18m 34s	04m 41s	23m 18s	46m 33s
I	20m 05s	05m 58s	22m 16s	48m 19s
J	20m 40s	04m 47s	28m 54s	54m 21s
Average	17m 42s	05m 32s	22m 51s	46m 05s

After converting the values to the respective variations of concentration of O_2Hb and

Hb, the signals were filtered using a second-order band-pass Butterworth finite impulse filter with cut-off frequencies of 0.01-1 Hz, to remove electrical and physiological noise, keeping the most informative frequency band of the fNIRS signals, according to the recommendations of [208].

The baseline signals were then segmented into 10-second windows with no overlap [225] according to the two considered periods - before starting the tasks and after finishing them. A label was attributed to each window, considering if it belonged to the period before the tasks, in which case it was labelled as *absence of cognitive fatigue*, or if it belonged to the period after all tasks, in which case it was labelled as *cognitive fatigue*. This processing procedure resulted in 22 time windows for each participant, with 11 windows per baseline period. The last 10-second window of each baseline period was excluded to ensure that it did not overlap with any task period, thus preserving the integrity of the baseline data.

After segmentation and labelling, a set of features were extracted from each of the time windows. These features were a mix of agnostic features extracted using TSFEL [221] and features specifically tailored to represent fNIRS signals. The list of used features in this study is: maximum, minimum, polarity, mean, variance, standard deviation, kurtosis, skewness, mean of differences, total energy, area under the curve, absolute energy, peak to peak distance, entropy, slope of linear regression, zero crossing count, fundamental frequency, maximum frequency, power bandwidth, spectral distance, median frequency, spectral entropy, root mean square, slope of the naive linear regression, maximum variation, and minimum variation.

Thus, given that the fNIRS had two channels and that each channel had three metrics, i.e., Hb, O_2Hb and THb, the final number of features was 26×6 features or 156 features for each data sample. Given the relation between that number and the number of samples for each participant, feature selection is important to reduce the number of parameter training in the classifiers.

As feature selection method, the Recursive Feature Elimination with Cross-Validation (RFECV) algorithm was used, where in each classification fold, a cross-validation was performed using only the training set to obtain the features that optimise the models' performance [222]. The algorithm is able to select the best features, thus, the number of selected features varies from fold to fold. The model used in this procedure was similar to the employed classifier.

Again, a RF was used as a classifier model to distinguish between the two states of cognitive fatigue. In this case, the number of DTs was limited to 10, the number of features at the nodes of the trees were limited to the square root of the number of features, the split criterion was Gini, and the minimum number of samples used in node splitting was 2.

Building on the previous study, a user-tuned classification procedure was designed, i.e., different models were built for different participants [201]. For each participant, the models were validated using a stratified 10-fold procedure, in which the whole dataset was divided in ten, keeping the original distribution of classes in both sets. Thus, the results

will be reported for each individual model relative to each participant. Additionally, since the classification periods were of approximately 2 minutes, or around 120 seconds, which were segmented into 10-second windows, each participant ended up with a total of 22 data samples, each containing 156 features. However, after feature selection, each data sample was composed of less features (varied among each cross-validation fold).

Stratified 10-fold divides each dataset into 10 train and test sets, thus, each test set was composed of 1 or 2 data samples. Models' evaluation was performed in 10 different folds, thus, the presented metrics represent the final metrics, i.e., the gathered results of all folds to calculate the metrics at the end of the whole procedure.

The classification results for each individual are shown in Table 5.5 and the respective confusion matrices are illustrated in Figure 5.1 where the values represent the aggregated number of samples in each of the cases.

Table 5.5: Classification results for each individual for the task of detecting cognitive fatigue.

Participant	Accuracy	Precision	Recall	F1-Score	AUC-ROC
A	0.55	0.55	0.55	0.55	0.64
B	0.91	1.00	0.82	0.90	0.96
C	0.86	0.90	0.82	0.86	0.87
D	0.73	0.73	0.73	0.73	0.78
E	0.91	0.91	0.91	0.91	0.94
F	0.59	0.67	0.36	0.47	0.45
G	0.73	0.73	0.73	0.73	0.85
H	0.55	0.54	0.64	0.58	0.49
I	0.59	0.57	0.73	0.64	0.60
J	0.68	0.70	0.64	0.67	0.70
Average	0.71 ± 0.14	0.73 ± 0.15	0.69 ± 0.15	0.70 ± 0.14	0.72 ± 0.17

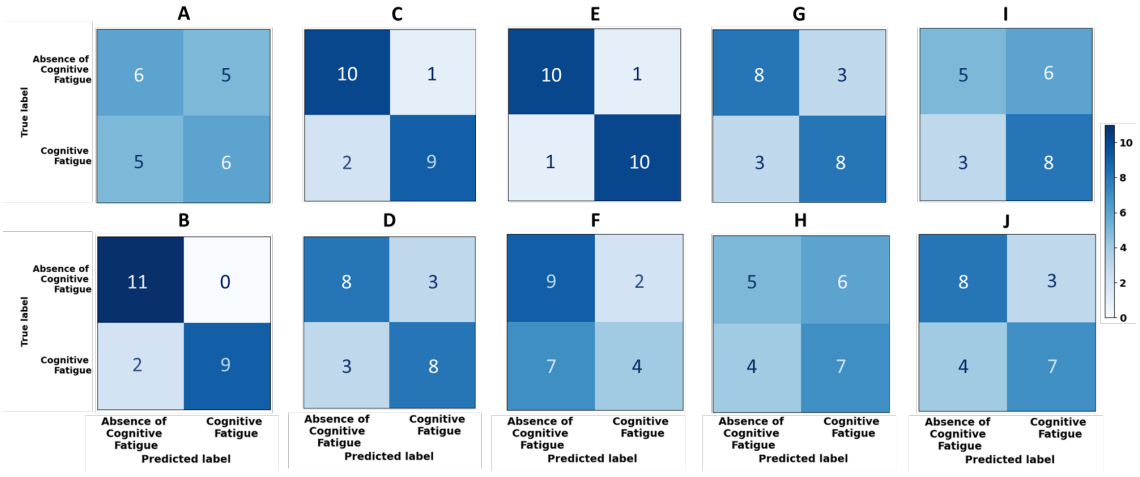

Figure 5.1: Confusion matrices of each participant.

The results represent the performance of using the fNIRS sensor measuring both sides of the dorsolateral PFC, showing that it might be feasible to automatically detect cognitive fatigue using a simple setup with high ecological validity. Our analysis will focus on the results that achieved accuracy scores higher than 0.6, corresponding to the results higher than the random chance that would be 0.5 given the binary classes and balanced datasets.

5.3 Multimodal Data Sources Classification for Learning State Monitoring

Prior studies aimed to reduce participant intrusion by assuming their cognitive states based on tasks. However, it was evident that these assumptions might be inaccurate. This study solely focused on the learning task and added a self-reporting stage for a more accurate understanding of learning states over time.

To analyse the learning states evolution over time, they were aggregated and the transition matrix was plotted in Figure 5.2.

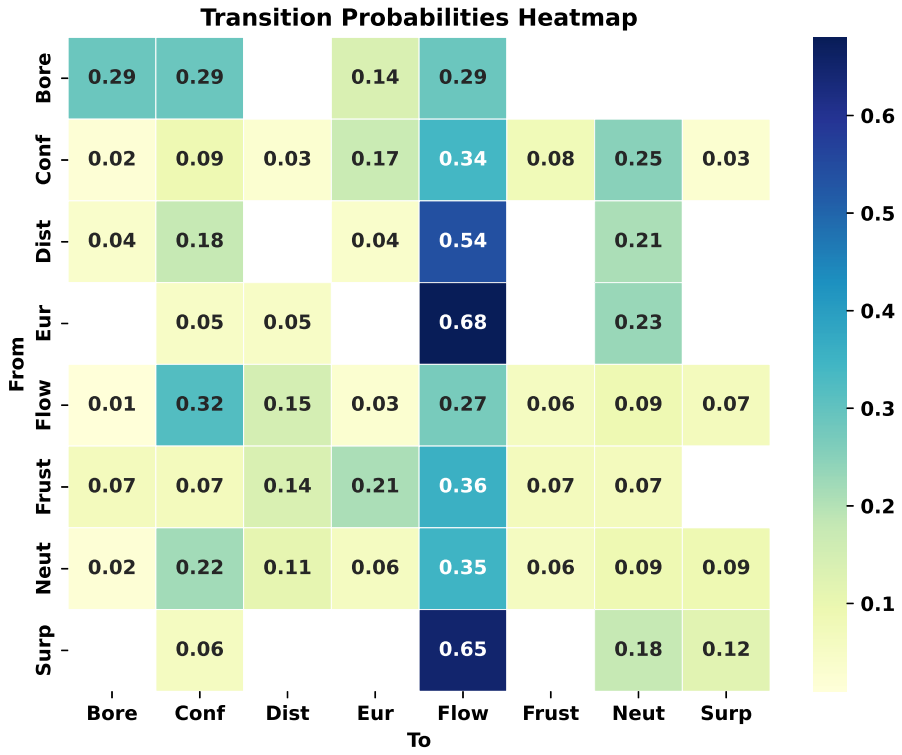

Figure 5.2: Transition matrix of the learning states obtained by the self-reports. The values of each cell correspond to the probability of transitioning from one state to the other. The abbreviations correspond to each of the classes: Bore - *Boredom*; Conf - *Confusion*; Dist - *Distraction*; Eur - *Eureka*; Flow - *Interest/Flow*; Frust - *Frustration*; Neut - *Neutral*; Surp - *Surprise*.

A simplified view is presented in a graph showing the transitions between states and the corresponding transistion probabilities higher than 0.15 in Figure 5.3 [226].

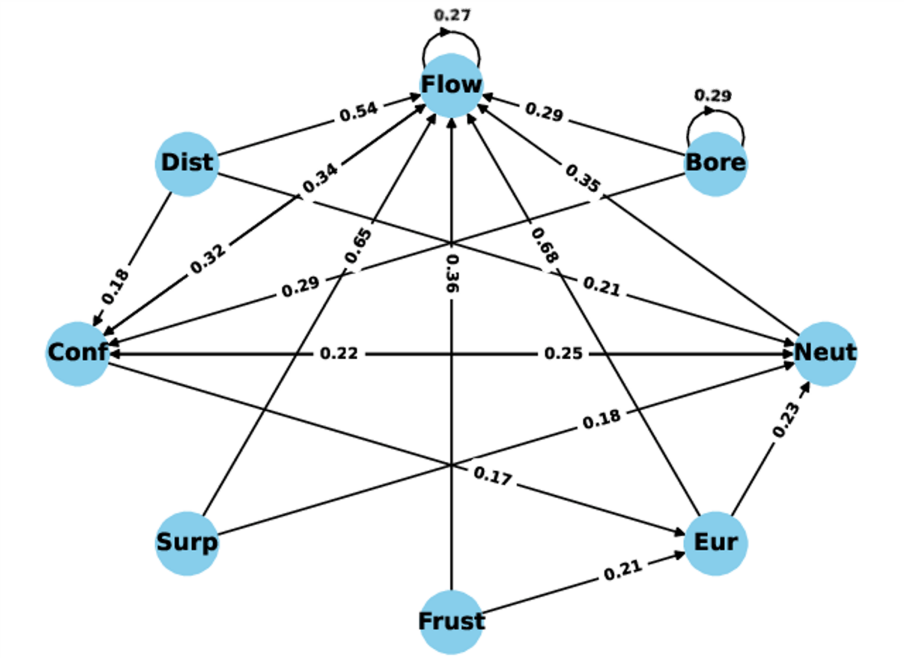

Figure 5.3: Transition graph of the learning states obtained by the self-reports. The probability values correspond to the closest arrow in the bidirectional edges. The abbreviations correspond to each of the classes: Bore - *Boredom*; Conf - *Confusion*; Dist - *Distraction*; Eur - *Eureka*; Flow - *Interest/Flow*; Frust - *Frustration*; Neut - *Neutral*; Surp - *Surprise*.

In the graph, the arrows represent the transition from the first state to the next and the numbers close to the arrow head represent the probability of that transition. Both the matrix and the graph were constructed with the labels acquired during the self-report stage, not considering the segmentation in time windows.

There, the probability to transition from one state to the other is shown, where blank cells represent transitions that never occur. Interestingly, most of the transitions occur between any state and "Interest/Flow", the most prevalent state throughout the acquisitions. Then, the different data sources were processed as previously described in Section 4.4. The number of samples produced by the different time windows segmentation varied given that the lower the time window, the more samples it is possible to produce. The number of samples is shown in Table 5.6.

Table 5.6: Number of samples for each segmentation procedure.

Time Window	**10s**	**12s**	**20s**	**30s**	**40s**	**50s**	**60s**	**90s**
# samples	3107	2572	1487	929	671	499	394	226

The number of samples is significantly higher than that of the previous studies because it was possible to use the whole acquired data for further analysis.

After segmentation, all features described in Section 4.5 were extracted to represent the data samples. This procedure was followed for all the referred time-windows.

In an initial stage, a statistical test was performed to test if the mean values of the features corresponding to each of the learning states are different. For that, a permutation test was computed using the Hotelling's T-squared statistic for multivariate data using 1000 or the number of samples if it was lower (which occurs when the time windows are larger than 20-seconds) as the number of permutations. The statistic and corresponding p-value were computed considering the raw features for each of the data sources, i.e., biosignals and mouse-tracking data, and also their combination, considering the different time windows enumerated before. Besides, this was computed for each of the label types considered (Binary, Cascade, Original, and Gathered), given the data of all participants. The results are shown in Figures 5.4, 5.5, 5.6, and 5.7, where the values represent the test statistic value and the cells in blue correspond to the cases where p-value < 0.05.

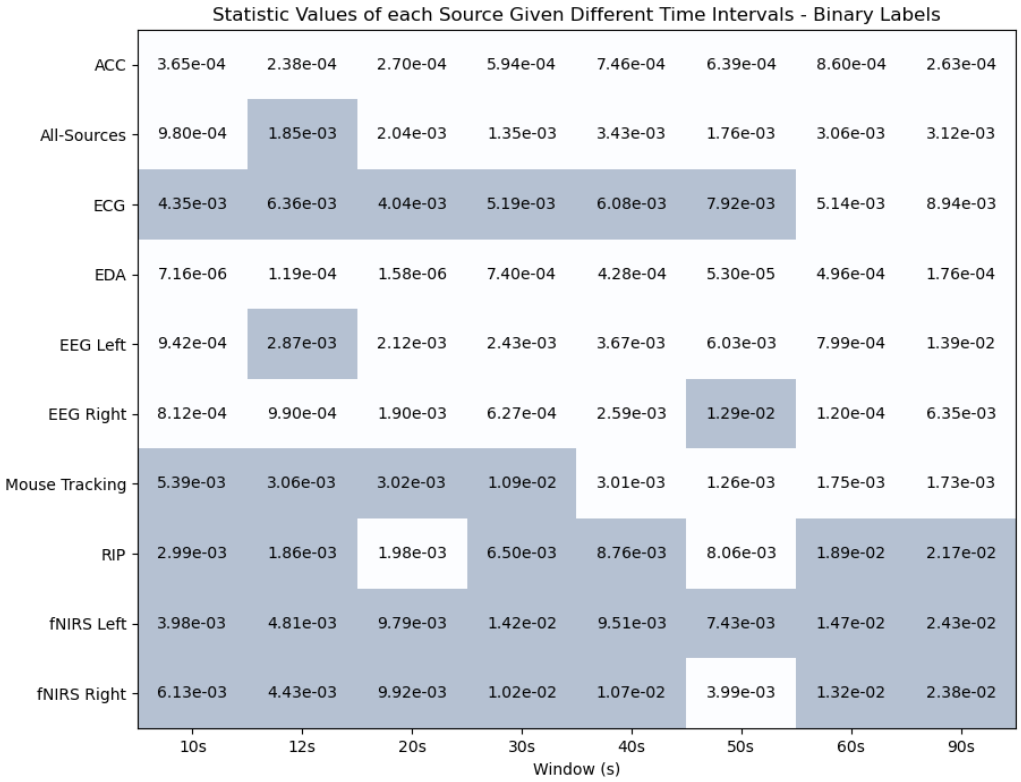

Figure 5.4: Results of the statistical analysis considering the Binary classes. The rows correspond to the results of each data-source for each time-window segmentation. The blue cells correspond to the cases where p-value < 0.05.

From these, it is possible to conclude that the fNIRS and ECG sensors are the most informative as they have the capacity to distinguish the different classes in all types and in most time windows segmentations. On another level, mouse-tracking data can be used in all classes division using specific time windows, while EDA is only appropriate in the Cascade and Original cases. Interestingly, EEG revealed to be the less appropriate sensors

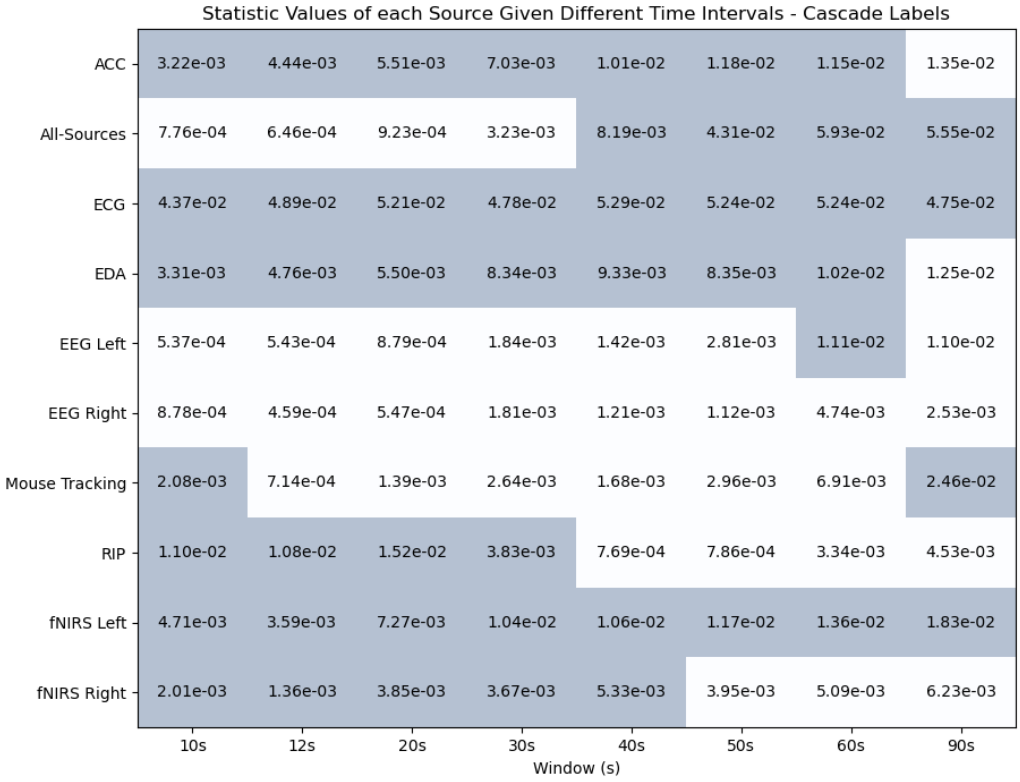

Figure 5.5: Results of the statistical analysis considering the Cascade classes. The rows correspond to the results of each data-source for each time-window segmentation. The blue cells correspond to the cases where p-value < 0.05.

in all cases, which contrasts with the studies found in literature. Lastly, the combination of all data sources revealed poor discriminatory power in most cases, showing how combining improper features can lead to worst results.

Given the possibility to distinguish between the different classes for all the considered types, it should also be possible to use ML to automatically classify the different learning states. Given *RQ5* and *RQ6* of this work, a myriad of classifiers were tested and used to classify the samples of different time windows, both in raw and using the Min-Max Scaling and Z-Score Standardization. In a first approach, even though the poor distinguishing power, the use of the combination of all data-sources for classification was tested. In this case, given that F1-Score can aggregate the performance metrics of precision and recall into a single value, the F1-Score was used to evaluate the classifiers' overall effectiveness across various data processing methods. The results are shown in Figure 5.8.

In the figure, it is shown that the time windows segmentation procedure has an impact on the achieved results. In this case, aside from the LOUO case that correspond to the bottom line of the figure, the others represent the value of the per-subject classification using a repeated stratified k-Fold strategy with 3 splits and 4 repeats that helps to get a more realistic idea of the results. The LOUO results correspond to a cross-validation mechanism in which, in each fold, one of the groups, i.e., participants, is left for testing

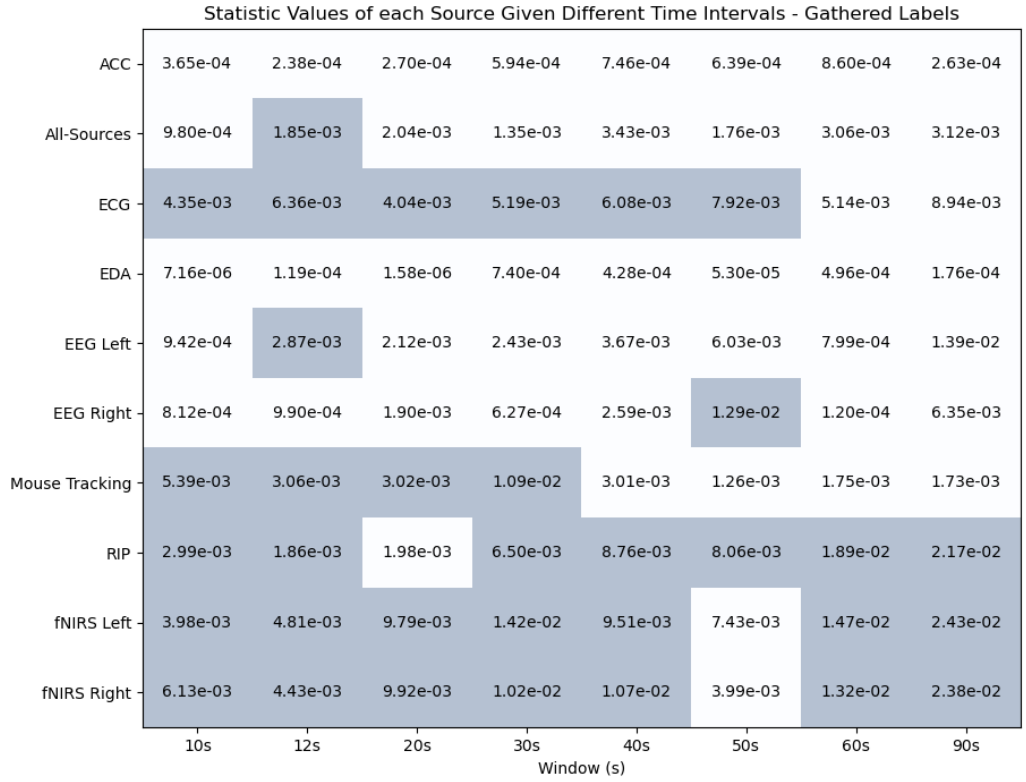

Figure 5.6: Results of the statistical analysis considering the Gathered classes. The rows correspond to the results of each data-source for each time-window segmentation. The blue cells correspond to the cases where p-value < 0.05.

while the others are used for training. The procedure is followed until all participants are used in the testing phase. To summarize the results, Table 5.7 presents the 10 best results based on the F1-Score.

Table 5.7: Best 10 F1-Score results achieved using the features from all data-sources. Only the Binary case has the ROC-AUC score defined.

F1-Score	Accuracy	Precision	Recall	ROC-AUC	Classifier	Norm	Window	Label Type
0.87	0.87	0.87	0.87	0.94	MC	Z-Score	90	Binary
0.80	0.81	0.80	0.81	-	AB	Raw	10	Gathered
0.80	0.81	0.80	0.81	-	AB	Raw	12	Gathered
0.79	0.79	0.79	0.79	-	MLP	Z-Score	90	Gathered
0.79	0.84	0.79	0.84	-	MC	Z-Score	50	Gathered
0.79	0.79	0.78	0.79	-	MLP	Z-Score	50	Gathered
0.78	0.78	0.79	0.78	0.66	MLP	Raw	40	Binary
0.78	0.82	0.79	0.82	-	GNB	Z-Score	60	Gathered
0.78	0.78	0.78	0.78	-	DT	Raw	12	Gathered
0.78	0.79	0.78	0.79	-	AB	Raw	10	Cascade

The table shows that Z-Score standardization and the raw features are the most suitable when dealing with the combination of all data sources. Moreover, most of the best performing classifiers correspond to the MC or to one of the classifiers that compose it,

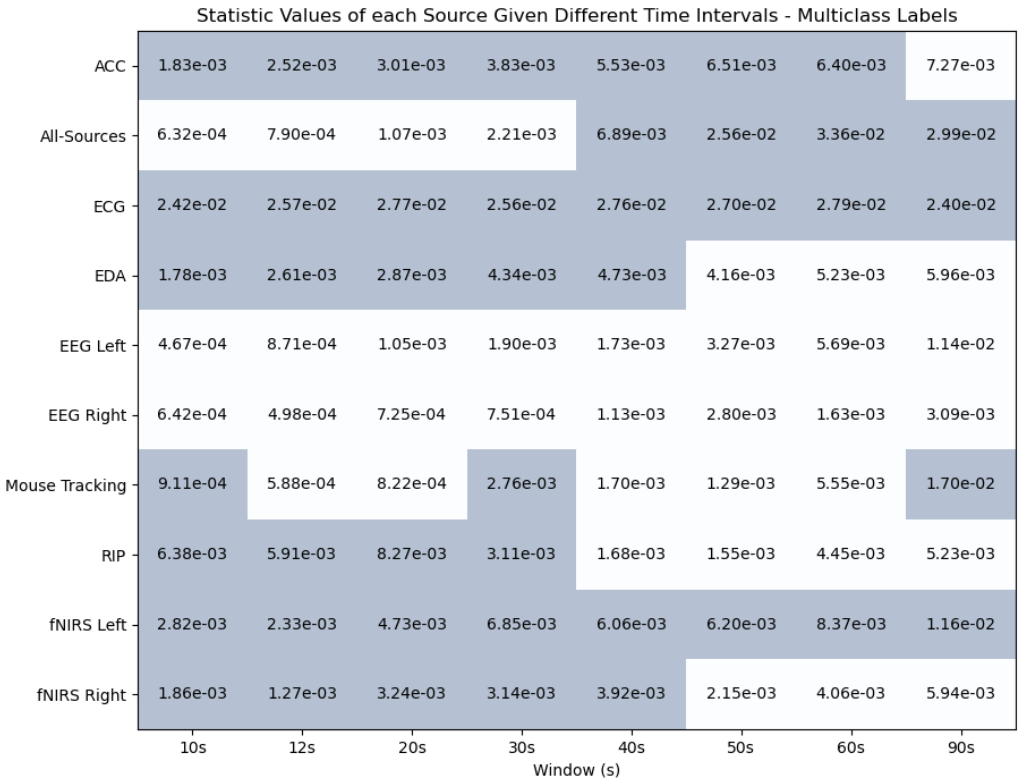

Figure 5.7: Results of the statistical analysis considering the Original classes. The rows correspond to the results of each data-source for each time-window segmentation. The blue cells correspond to the cases where p-value < 0.05.

namely, AB and MLP. Concerning the labels types, it is clear that the Gathered is the most prevalent. Regarding the time windows for segmentation, it is not possible to conclude that there is one best than the others.

Thus, the next step is to analyse the individual sensors and combinations of different data-sources given these conditions: Gathered classes, i.e., the classes that considered the combinations of the original states into similar valence and degree of activation (e.g., *Frustration* with *Confusion*, and *Distraction* with *Boredom*), MC classifier, and raw features. The number of combinations was limited to three data-sources with the objective of keeping the ecological validity. Figures 5.9 and 5.10 show the F1-Scores of the combination of data sources, where the diagonals show the results using the features from each individual data-source.

The presented figures correspond to the matrices of the different time window segmentation. Since that the order of the combination of features is not important, only half of them are filled, and the other half would be symmetrical. Moreover, the dark blue tones correspond to higher F1-Score values while the lighter yellow tones correspond to lower values. However, these can only show the results using individual data-sources or combinations of two. Again, to summarize the results achieved with all the studied combinations, including those of three different data-sources, Table 5.8 show the best

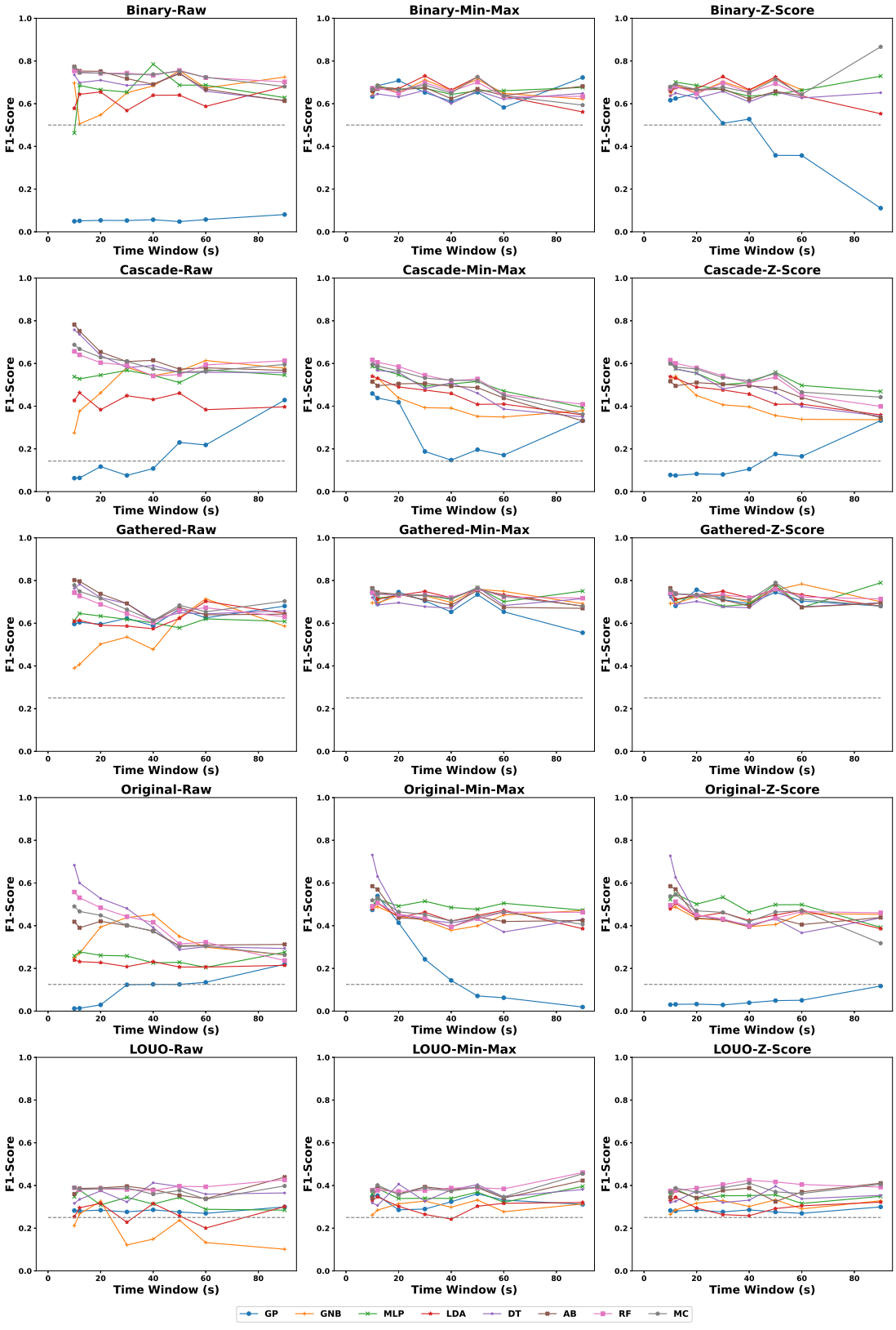

Figure 5.8: F1-Score results for all data sources combined relative to the time-window, type of normalization, and label type. The dashed grey lines represent the expected performance of a random classification.

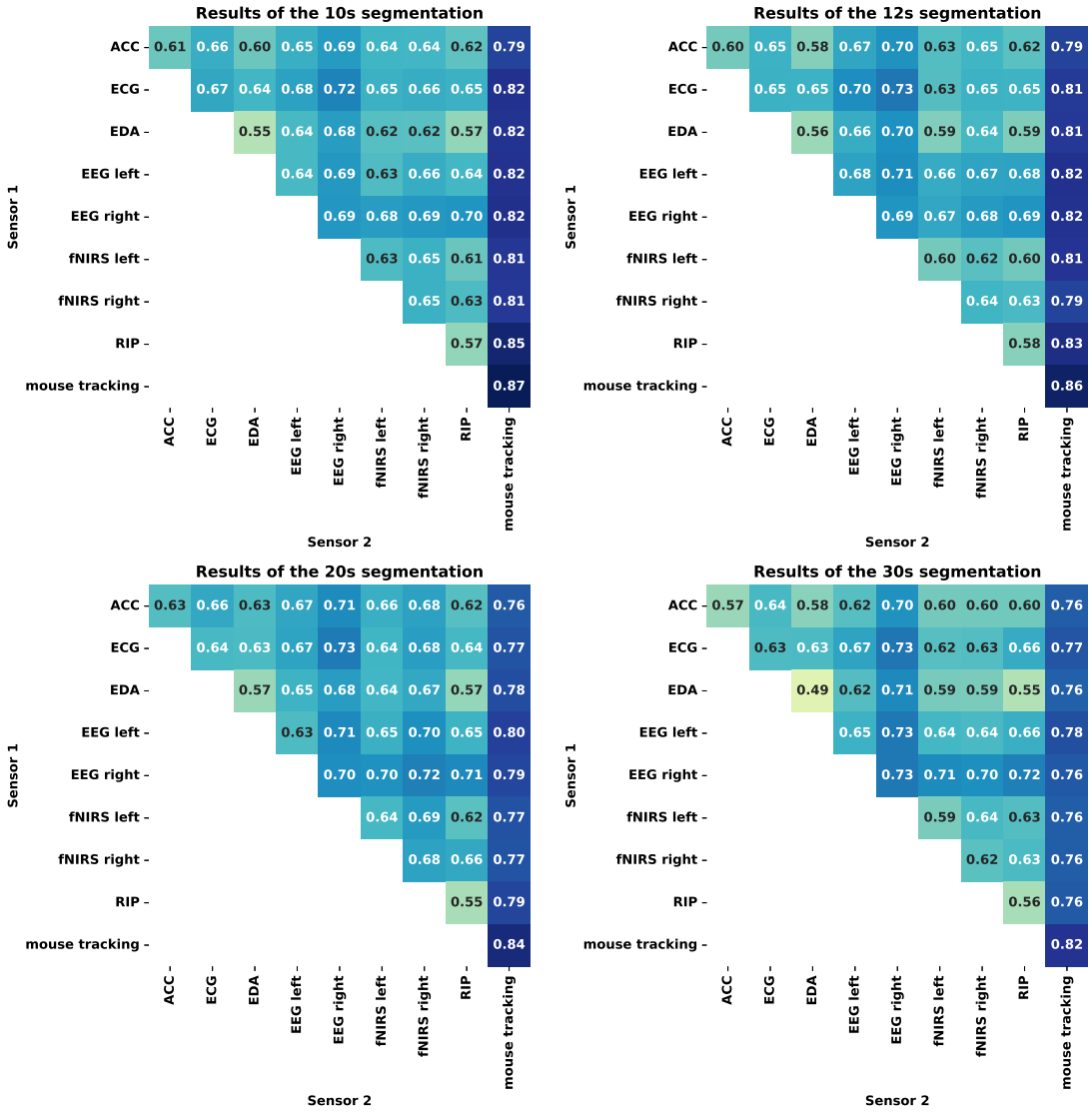

Figure 5.9: F1-Score results for all data sources relative to the time-window, given the raw features, and Gathered classes. Only the time-windows from 10 to 30-seconds are shown.

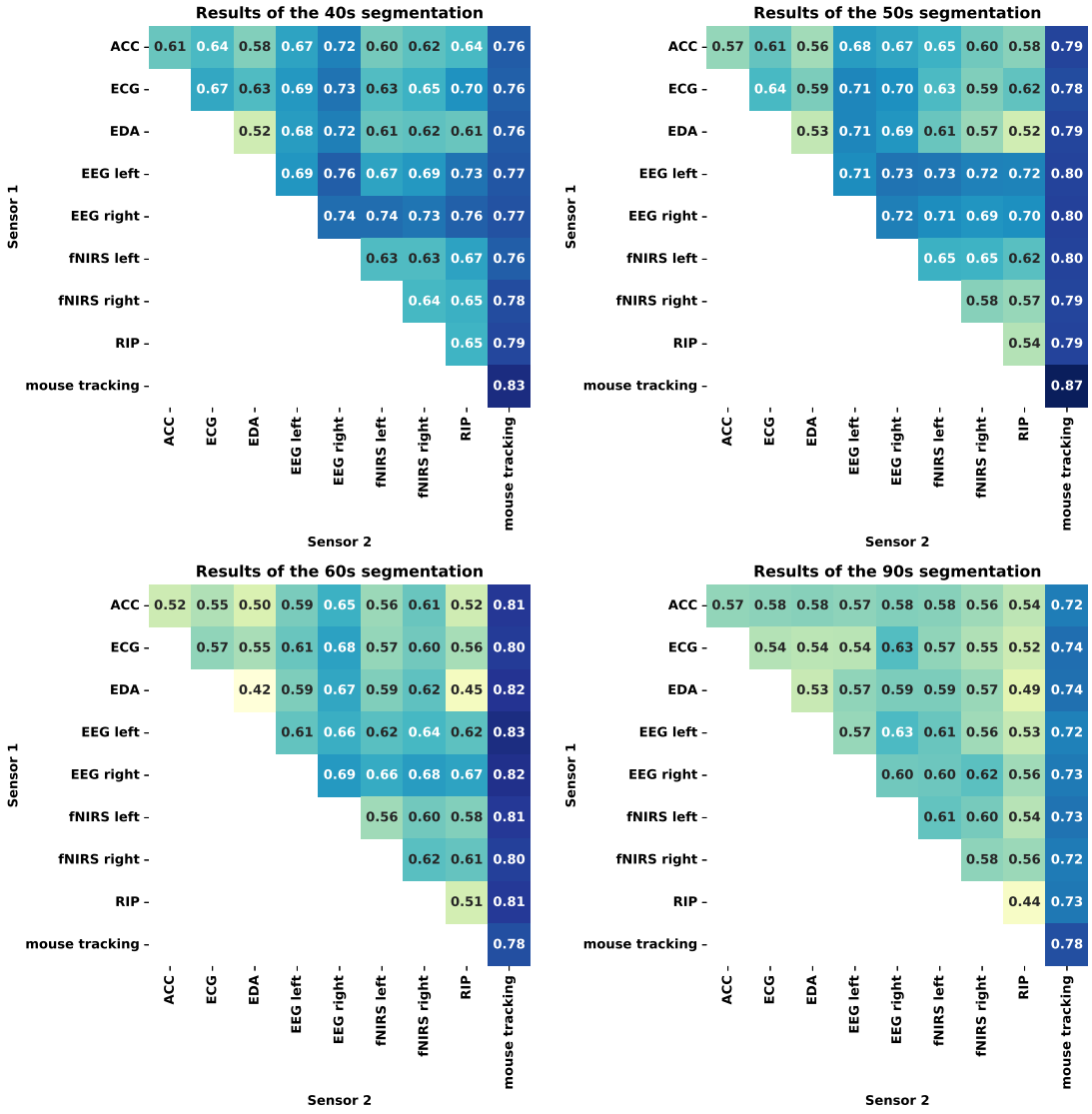

Figure 5.10: F1-Score results for all data sources relative to the time-window, given the raw features, and Gathered classes. Only the time-windows from 40 to 90-seconds are shown.

achieved results, considering F1-Score values.

Table 5.8: Best 10 F1-Score results achieved using the features from all the studied combinations of data-sources considering Gathered labels, MC classifier, and raw features.

Sensors	F1-Score	Accuracy	Precision	Recall	Window
Mouse Tracking	0.87	0.87	0.87	0.87	10
Mouse Tracking	0.87	0.87	0.87	0.87	50
Mouse Tracking	0.86	0.86	0.86	0.86	12
RIP-Mouse Tracking	0.85	0.87	0.86	0.87	10
EEG left-fNIRS left-Mouse Tracking	0.84	0.88	0.81	0.88	60
ACC-EEG left-Mouse Tracking	0.84	0.88	0.81	0.88	60
Mouse Tracking	0.84	0.84	0.84	0.84	20
EEG left-RIP-Mouse Tracking	0.84	0.86	0.84	0.86	10
EEG rigt-RIP-Mouse Tracking	0.84	0.86	0.84	0.86	10
EEG left-fNIRS right-Mouse Tracking	0.84	0.87	0.80	0.87	60

The results show that mouse tracking is the most useful data source to infer about the learning state of participants, namely considering the Gathered labels and raw features, since it is present in every combination of data-sources in the best 10 F1-score values, even when applied alone.

5.4 Summary

This chapter described the final data processing steps followed in each of the studies performed in this work. Then, it also presented the classification methods and the respective results.

The first study demonstrated that it is feasible to infer the cognitive states of participants when engaging with cognitive tasks. It was assumed that engaging with the tasks would require external attention, while baseline periods did not involve attention of any kind.

In the second study, despite the procedural changes to incorporate increased computer interaction and more extensive mouse tracking for study purposes, the protocol design did not take into account all aspects involved in the learning process and, thus, the classification focused on the distinction between the periods before and after demanding cognitive tasks, assuming that these induced cognitive fatigue. It was demonstrated that fNIRS could be used individually to infer the fatigue state of some individuals.

Finally, the third study used the labels provided by the self-reporting stage to classify and detect different learning states while interacting with an e-learning lesson. In this case, all acquired biosignals and the mouse-tracking data were used, when available. The results showed again that the individually tailored classifiers are more effective than general-purpose classifiers in distinguishing different learning states. Those results allowed us to focus on the Gathered classes division, the MC classifier, and on the raw features for

further analysis. Given these, mouse tracking revealed to be the most effective data source considering the different time windows studied in this work.

6

Discussion

Previous chapters outlined the methodology followed in this work and reported results for the detection of attention, cognitive fatigue, and complex learning states. This chapter examines these findings in the context of the research questions posed in Chapter 1.

6.1 Attention Classification Based on Biosignals

The first study aimed to assess the feasibility of detecting attention states using a multitude of sensors during well-known cognitive tasks, answering **RQ1**. With that intent, n-back and mental subtraction were selected as candidate tasks to induce attention on individuals. Given the availability and non-intrusiveness of the mentioned sensors, biosignals data was acquired during those tasks. The biosignals were processed, and features with a low computational complexity were extracted and given as an input to classifiers that automatically distinguished between *task* and *baseline*.

The segmentation and processing steps led to the construction of a dataset containing the number of samples presented in Table 5.1, where it was shown that most participants had similar numbers of samples, except for subject C. That was caused by an unexpected disconnection of the acquiring device during the acquisition procedure, leading to a loss of data relative to the mental subtraction task. Save the combination of sensors with the least number of sources, Table 5.3 reveals that the classification of participant C achieves the best performance comparing with the remaining participants. This happens because the dataset of participant C presents less variation concerning the represented cognitive tasks. Thus, the datasets containing both tasks are more generic, and the cognitive response to the two cognitive tasks may be distinct. Although the performance may be lower with the presence of both tasks, the classifiers would be able to generalize better considering unseen data, as they captured more variability in the data.

On the ecological validity aspect, it would be relevant to identify the most suitable data-sources to detect attention. This should help to simplify the acquisition procedure and reduce the preparation time in future works, besides reducing the interference with individuals that could use such a system. With this intent, classifiers were built for each

participant and for each of the considered biosignals, as shown in Table 5.2. For example, comparing different sensors for subject E reveals that the range of accuracy values is 0.44–0.79, the EDA sensor presenting the lower accuracy and the ECG sensor the highest. On the other hand, it was also illustrated how individual differences can have an impact on the achieved results. For example, the accuracy values achieved using solely fNIRS range between 0.48 - 0.75, suggesting the importance of adopting an individualized approach to achieve optimal results tailored to each individual. However, this approach would always require the execution of controlled tasks to reveal the best sensors for each individual.

The combination of multiple biosignals was studied to optimise the achieved results. Again, the classification of different subjects revealed different accuracy levels and a different optimal combination of sensors, pointing toward the direction of an individual-directed tool. Although the addition of more participants would help to support this thesis, since the followed approach is based on the individual level, the results presented and discussed would not change.

Regarding the most informative biosignals, EEG and ECG have the best mean accuracy scores, i.e., 0.64 and 0.65, respectively, whereas EDA reaches the worst average performance of 0.54. Thus, to apply a single sensor or a combination of few sensors regardless of individual differences, the combination of a low-channel EEG and ECG would be advised. In fact, Table 5.3 shows that the best combinations for each subject always present either EEG or ECG signals. Contrary to our expectations, fNIRS revealed poor performance for most subjects. This might be related to the chosen setup: a wrong choice for the positioning of the sensor, a too simplistic processing of the signals, or an inadequate emitter-detector distance of the used sensor. Considering the results of EEG being higher than those of fNIRS, it might be related to the poor spatial resolution of the EEG sensor, which can be beneficial in this case, since it may be able to detect activations from greater distances to the sensors than the fNIRS sensors. EDA and RIP sensors did not contribute to good results when considering a low number of sensors and, thus, could be removed in future studies. An interesting finding is that the EEG sensor with a low number of channels was able to achieve accurate results, especially in combination with other sensors, which proves that ecological validity and the ease of use can be allied with good results.

To the best of our knowledge, previous studies did not consider the individual differences and developed general models. However, the achieved results demonstrate the importance of user-tuned models. Comparing the methodology of this study with previous studies, some authors solely use cameras [103, 178], or cameras associated with motion sensors [102], and achieve results similar to ours. However, image acquisition might be viewed as being more invasive than sensors data acquisition. To assess attention, the authors that employed biosignals used only EEG, which was enough to find some relationships with attention [99, 101] and achieved good results in predicting attention [100]. However, most of them used 32-channel EEG, which has less ecological validity due to the complex setup than this study's two-channel sensor. No studies were found that combined multiple biosignals in this domain as in this work. Our outcomes demonstrate that

the combination of sensors is better in modulating attention than the individual sensors. An interesting finding is the good performance of ECG, which was never explored in previous works in the context of attention classification and revealed to be meaningful in this context.

6.2 Cognitive Fatigue Detection Based on Biosignals

The main objective of this study was to be able to use HCI data along with the acquired biosignals to detect different cognitive states unobtrusively to help answering **RQ2**. However, the base assumption revealed to be inaccurate and, thus the focus was placed on the detection of cognitive fatigue using only the fNIRS sensor based on the demanding tasks the participants executed, which can be detrimental to performance [227]. fNIRS data was processed and features were extracted and selected to properly represent the two considered classes. Finally, user-tuned classifiers were trained and validated to automatically detect the state of cognitive fatigue.

According to previous work, cognitive processes are subjective to each individual and, thus, classification models should be developed on the individual level, rather than be developed for a general population [114–117, 201]. Therefore, user-tuned models were trained for each participant. The results of classification corroborate the hypothesis that different individuals respond differently to the same demands, since the accuracy results range from 0.55 to 0.91 (mean value 0.71 ± 0.14). The confusion matrices of Figure 5.1 show that when the results for the accuracy are higher than 0.70, the predicted labels are positively correlated with the true labels and, thus the diagonal of the matrices is more populated. Considering only such cases, the mean accuracy is 0.83 ± 0.08, which is in line with other works (e.g., [112]). In cases where the classification is lower than 0.70, the expected labels and the predicted classes do not seem to be related.

Moreover, the models for cognitive fatigue classification of participants B, C, D, E, G, and J (ages from 23 to 34 (M = 27, SD = 4); 2 females) were successful in distinguishing between *cognitive fatigue* vs. *absence of cognitive fatigue*, while the remaining were unable to effectively distinguish the two cognitive states (see Table 5.5). In the cases where the accuracy score is close to 0.50, namely, participants A, F, H, and I (ages ranging from 22 to 48 (M = 29, SD = 11); 2 females), it is also possible to note that neither class is preferable because the values of precision, recall, and AUC-ROC are also closer to 0.50 than the others. On the other hand, in the cases where the models are applicable, precision and recall are balanced, meaning that the models are not biased towards either class.

Time is usually regarded as a crucial aspect for the appearance of cognitive fatigue. However, the achieved results demonstrate that it is not the main factor. For example, while participant C took less time to complete the tasks and participant J took the longest time, the classification results did not reflect that difference (see Table 5.4). Moreover, contrary to the findings of [114], age did not influence the results, since that the groups of participants with accuracy scores higher than 0.70 and close to 0.50 did not reveal relevant

age differences. Moreover, sex was also not a factor that interfered with the prediction of cognitive fatigue.

Given the disparity of results, it seems clear that, although the employed methodology was very effective for inducing and detecting cognitive fatigue in some cases (e.g., participants B and E), automatic cognitive fatigue assessment using the proposed methodology is not applicable in all cases. It is hypothesised that this can be the case due to some aspects: the experimental procedure may not induce sufficient cognitive fatigue in some cases; the extracted and selected features may not be applicable to the cognitive responses for every individual; there may be other factors contributing to cognitive fatigue, e.g., time of day and the regular schedule of work of each individual; how individuals view the tasks (some individuals may try hard to get everything right while others may be less motivated to perform the tasks).

In comparison to other works, the proposed methodology employs fNIRS wearable sensors, easily applicable outside the laboratory, which does not happen with the sensor used, for example, by Trakoolwilaiwan et al. (2017) that uses an fNIRS sensors with 34 channels [113]. Moreover, given that physiological signals were measured over time, the mentioned disadvantages about the subjective and behavioural methods do not apply. However, there is a disadvantage, which is the assumption of the state of the participants, i.e., it is assumed that they present cognitive fatigue at the end of the procedure, which may not be the case, as previously discussed.

6.3 Multimodal Data Sources Classification for Cognitive State Monitoring

The final study took from the previous two and expanded on them by adding the self-report stage and discarding the standard and demanding cognitive tasks. This allowed to focus on complex learning states instead of specific cognitive processes. Moreover, in this study it was possible to acquire and use mouse-tracking data and the mentioned biosignals to analyse the possibility of using them to automatically classify between different learning states allowing to answer **RQ2**. Specific and agnostic features were extracted from the segments of the different data-sources, and a statistical analysis was performed, before applying ML algorithms.

First of all, it was investigated how learning occurs through time, as illustrated in Figure 5.2 and 5.3, contributing to answer **RQ3**. This evolution revealed interesting findings:

Boredom (Bore) - presents an equal probability to transition to itself, to *Confusion*, and to *Interest/Flow*.

Confusion (Conf) - transitions more frequently to *Interest/Flow*, *Neutral*, and *Eureka*, which often reveals the solution of the doubts.

Distraction (Dist) - also leads to *Interest/Flow* and *Neutral* frequently, however, it also leads to *Confusion* because participants lose focus when studying the ECG lesson.

Eureka (Eur) - rarely leads to *Confusion* and *Distraction*, that might reveal that the *Eureka* moment solves a given doubt but might lead to another. Nevertheless, it is more common to lead to *Interest/Flow* or the *Neutral* state.

Frustration (Frust) - a state that is related with *Confusion* in its emotional dimension [228], often leads to *Eureka*, which supports the hypothesis that it marks the solution of doubts. Moreover, it also leads to *Distraction* showing how *Frustration* needs to be handled carefully.

Interest/Flow (Flow) - transitions to *Interest/Flow* very frequently, and also to *Confusion* or *Distraction*. In this case, while *Confusion* indicates something that might not be well explained or the participant might not understand it at first, *Distraction* can simply occur with external or internal stimuli.

Neutral (Neut) - similar to the *Interest/Flow* state, also transitions very often to *Interest/Flow*, and to *Confusion* or *Distraction*.

Surprise (Surp) - only leads to *Interest/Flow*, *Neutral*, *Surprise* again, and *Confusion*.

In summary, while some states exhibit closely correlated transition patterns, such as *Neutral* and *Interest/Flow*, and *Frustration* and *Confusion*, others, notably *Boredom*, *Eureka*, and *Surprise*, present distinct pathways. Notably, the absence of a transition from *Boredom* to *Frustration*, contrary to certain established learning models in the literature [209], highlights potential deviations from expected behavioural sequences. These nuanced state transitions shed light on the intricate dynamics of cognitive and emotional states during learning experiences.

After that analysis, statistical tests were performed considering the multivariate nature of the data and categorical classes for every type of considered label, helping to answer **RQ4**. A permutation test using Hotelling's T-squared statistic was performed and the corresponding results are shown in Figures 5.4, 5.5, 5.6, and 5.7. From those, it is shown that fNIRS and ECG are the most informative sensors since the associated p-values are lower than 0.05 for most time windows in all label types, thus revealing that the means of the different classes in all cases are statistically different. While the fact that the ECG sensor reveals good performance is not new, considering that attention states could also be distinguished using it in the first study (Section 6.1), fNIRS sensors revealed good distinguishable power. This might be related with the fact that a larger amount of features and also some specific features that were not considered in the first study were extracted. On another level, RIP sensor also revealed good discriminative power considering segment windows lower than 30-seconds, the windows where the specific features could not be calculated and only the agnostic ones are present. This contrasts

with the prior results that revealed low classification performance using the RIP sensor, showing that, although it might not be applicable for specific cognitive states detection, namely, attention detection, it might be valid for complex learning states classification. Again, contrary to that study, EEG sensors revealed poor performance in the statistical analysis. It is hypothesised that this happened because, in this statistical analysis, the set of all participants were considered instead of considering a user-tuned analysis. Regarding mouse-tracking, it only revealed some predictive power in the Binary and Gathered labels cases and in some of the studied time windows. Finally, the combination of all data-sources revealed poor discriminative power, indicating that the correct choosing of data-sources is critical to build a system to predict complex learning states.

After attesting the feasibility of using the acquired data to predict the learning states, either using the original classes or one of the combinations of labels, the aim was to assess the possibility of using ML to automatically classify those states using the combination of all data-sources. Furthermore, it also aimed to assess the best time window for segmentation, classifier and type of learning states grouping that should be used, answering **RQ5** and **RQ6**. In this sense, the mentioned set of classification algorithms, labels grouping, and time windows for segmentation were tested using the combination of all data-sources, illustrated in Figure 5.8. To facilitate the analysis, it is also shown the best 10 results in Table 5.7, where the highest F1-score values considering the mentioned combinations are presented. From this, it is possible to conclude that most of the classifiers in this list are either MC or one of the classifiers that compose it. Furthermore, there was no normalization procedure better than the others, given half of the list is composed of raw features, while the other half is composed of the Z-Score standardization. However, given that a real-time system would benefit from lower computational complexity, it makes sense to choose to proceed the analysis with only the raw features. Finally, the label type that is more prevalent in this list, is the Gathered learning states. It is hypothesized this is related to the grouping of states with similar emotional valence and cognitive processes, which can help to better detect the different classes. For example, the Binary type grouped *Confusion* and *Eureka*, which are not comparable. Moreover, the Original labels and the Cascade labels considered classes with few samples compared with others, which might help to explain the lower results. The LOUO grouping represents the case where the classifiers are generic and not user-tuned, leading to the lower results, in line with the findings of the previous studies.

With this, the analysis proceeded by testing the combination of different data-sources to answer **RQ4** and **RQ6**. Figures 5.9 and 5.10 show the F1-score values of the combinations of two-by-two or individual sources considering the aforementioned conditions: MC classifier, Gathered labels, and no normalization. Interestingly, mouse-tracking revealed to be the most appropriate data-source for learning states classification in all the considered time windows. This can be explained by the fact that participants used the video of their interaction with the computer and the video of their faces for self-reporting and, thus, further analysis should be conducted where the ground-truth should not be based on

the HCI variables used for classification. Based on those figures, it is also possible to conclude that the EEG right, the channel in AF8, was the best biosignal considering the F1-score. This happened considering the individual channel and also the combination with the other channels. Regarding the best combinations considering until three different data-sources, Table 5.8 shows the results selected based on the best 10 F1-score values. There, the mouse-tracking source is present in every entry, demonstrating its importance in the learning states classification and detection.

Finally, regarding the time windows for segmentation, increased time windows leads to worse results. This is because, although larger time windows might capture relevant physiological information, this type of segmentation also leads to fewer numbers of samples, which typically has a negative impact on the achieved results, because the classifiers have access to fewer samples during training. These results indicate that it might be possible to build a system working in real-time or close to real-time, once that the time for classification can be as low as 10-seconds according to the performed tests.

6.4 Summary

three different studies were performed in this work to answer the proposed questions in Chapter 1. The results revealed that those studies allowed us to answer them and allowed to explore how biosignals and HCI data can be used to properly classify specific cognitive states, such as, attention and cognitive fatigue, but also complex learning states.

Individually-tailored models showed they can reach high accuracy and F1-score values, while general models might be detrimental for this application. While EEG and ECG revealed to be the best sensors for attention state detection in a binary case, the fNIRS revealed to be appropriate for cognitive fatigue detection, also in a binary case, although only for some participants.

Moreover, algorithms selection was demonstrated to have an impact on the achieved results, as well as, the choosing of the grouping of certain learning states. Mouse-tracking revealed to be the best data-source in all cases of learning states classification considering the Gathered labels, the MC classifier, and the raw features. As for the best biosignal, the EEG sensor single-channel located in the AF8 site of the 10-20 system revealed to be the most appropriate. The time windows should be lower than 60-seconds in order to capture more samples and to allow the application of a learning state classification system in real-time.

7

Conclusions

This closing Chapter will present a short summary of the developed work and achieved results and align them with the research questions and objectives presented in Chapter 1, as well as present the outcomes and publications. It will also present possible applications that this thesis might help to support. Future work suggestions will also be given to drive research in the area of application of biosignals and HCI data in education. The implications and novelty of this work will also be presented.

7.1 Main Results

The main objective of this thesis was to study the possibility of employing biosignals and HCI data to assess the cognitive and learning states of users during e-learning lessons. To do this, statistical tests and ML were employed for the classification of different states using an extensive set of features. This process was performed for the classification of states during standard cognitive tasks, and also during simulated e-learning lessons.

Considering how learning occurs via mental processes and translates into different behaviours, both brain-related biosignals (EEG and fNIRS), and biosignals that help to infer about the cognitive states (EDA, ECG, RIP, and RIP) were monitored. Behaviour monitoring was performed using mouse-tracking data, allowing us to analyse how it may reflect learning states during e-learning lessons. User-tuned models demonstrated to be the best solution to monitor cognitive and learning states, although different sensors lead to different results. Namely, EEG and ECG revealed to be the best suited to detect attention during task performance, while the fNIRS sensor revealed good performance to detect cognitive fatigue, although the performance was solely acceptable for some of the considered users.

The third and final study introduced a self-reporting stage, inexistent during the previous two, to assist in obtaining proper ground-truth for the classification. In this case, instead of focusing on specific cognitive states, the approach widened to classify complex learning states. With this, it was demonstrated that there are different probabilities of transitioning between the considered learning states, being the most prevalent *Interest/Flow*.

It was also demonstrated that it was possible to aggregate different states given their degree of activation (activating vs deactivating states) and its valence relative to learning performance (positive vs negative states), indicated in the literature as being relevant [88, 120]. Statistical tests demonstrated the feasibility of using certain biosignals and mouse-tracking data to predict learning states. Furthermore, the classification of the combination of all the considered data-sources revealed the best conditions to validate the followed approach using the individual or combinations of low numbers of channels. Then, testing user-tuned models to classify the Gathered classes, i.e., the classes that considered the combinations of the original states into similar valence and degree of activation (e.g., *Frustration* with *Confusion*, and *Distraction* with *Boredom*), using the MC classifier and no normalization of the considered features, it was clear that mouse-tracking was the best achieving data-source, given the achieved F1-score results. Finally, low-time windows of less than 30-seconds allowed to achieve good performance results, showing that it should be possible to develop a system working close-to-real-time for the classification of learning states using minimal setups.

Given the proposed research questions of Chapter 1, we will explore how this thesis answered them:

RQ1 - Is it possible to use biosignals to detect cognitive states?

The answer to this question is observed throughout the three performed studies: it is possible to use biosignals to detect different cognitive states. For example, the first study showed how individual biosignals as well as combinations of biosignals allow us to automatically detect attention states related with standard cognitive tasks. In the second study, the fNIRS revealed good performance to detect cognitive fatigue comparing the signal prior and after the execution of demanding cognitive tasks. Finally, the third study also showed that biosignals, individually or in combination, were able to distinguish between different learning states. However, the detection or distinction between different states was only possible using user-tuned ML models, given the high inter-subject variability there is when analysing this topic.

Relative to the state-of-art, different sensors were introduced to measure human activity related to the cognitive states or learning states, namely, the RIP and ACC sensors, and also analysed the combinations of all the considered sensors and biosignals in this context. Some studies employed single sensors, either with multiple channel or single channels [47, 88, 89, 91, 92], while others employed specific combinations, such as, EEG and fNIRS [112] or ECG and EDA [119].

RQ2 - Can HCI data be used to infer about cognitive states related to learning?

In this case, HCI data, namely, mouse-tracking data, was only demonstrated as useful to infer about learning states in the third study. However, the results indicate that, considering the data-sources here represented, it is the most informative one with results of F1-score and accuracy score of 0.87 in distinguishing between Gathered learning states.

Although some studies found in the literature achieved better results, they were focused on specific cognitive states (e.g., attention and cognitive load) and required

external devices [8, 174].

RQ3 - How does learning occur over time?

The final study added a self-report stage to build a ground-truth for the classification procedure, which allowed us to study how learning occurs over time. Figure 5.3 shows the transitions between learning states that occur with a probability higher than 15%.

The proposed methodology allowed to better monitor the learning state paths of participants. Considering that most participants start in a neutral state, the participants initiate the task without discernible emotion or state. Then, they get interested or in flow, where they concentrate on reading the contents of the lesson. Here, they might continue concentrated, go back to the neutral state or they may get confused, especially considering the questions presented throughout the lesson.

When confused, the learner tries to find the lacking information or to re-interpret the contents, reaching an Eureka state and moving onto the interest or neutral states, or directly re-concentrating to reach them. Analogously, if they get frustrated, which rarely happens, they may go back into a state of flow or they get into an Eureka state. If at any time they get distracted, which is not highly probable, then they may get confused, mainly because they lose their train of thought, or directly get back to the interest or neutral states. On the other hand, if they get bored throughout the lesson, then they may get confused, again indicating they lost their train of thought, or they go directly back to a flow state. Finally, surprise, which happens if a new and unexpected information is presented to the participants, only transitions into interest or neutral states.

RQ4 - Which is the best data-source to automatically assess learning?

The results of study 3, presented in Section 5.3, indicated that the most appropriate data-source to automatically assess about learning states is the mouse-tracking data considering the features extracted in this work. However, it might be because the video used for self-reporting included the interaction between the user and the computer, and, thus, the mouse-tracking through time. Other data-source, namely the EEG channel positioned in the AF8 position of the 10-20 system also revealed good results and most of the considered biosignals were able to perform significantly better than the random chance.

RQ5 - Does the ML classification algorithm influence the classification performance?

The results of Figure 5.8 show that the classification algorithm can greatly impact the achieved results, although it is not the only factor. For example, while AB reaches F1-score values of around 0.80 with the raw features and in the Cascade scenario with a time window of 10-seconds, the GP classifier reaches values below the random chance in the same context. Thus, the selection of a classification algorithm should be made carefully to optimise the achieved results.

RQ6 - Does the segmentation of the time series influence the classification performance?

Again referring to the third study of this thesis, the classification performance demonstrated to be influenced by the selection of the time window for the segmentation of the time-series. For example, this is clear in Figure 5.8, in the Binary-Z-Score, Cascade-Raw, Cascade-Z-Score, and Original-Min-Max cases, where the F1-score values present high

variations considering different time windows for classification. While this might be caused by the fluctuation of the number of samples in each case, it might also be related by the variation captured in each time window that increases with the increase of its size.

7.2 Contributions

This thesis allowed to delve into different areas of knowledge, such as, cognitive psychology and engineering. Different data acquisition protocols were developed involving different tasks and methodologies, adapting the signal processing accordingly. Different data-sources were then combined and Machine Learning (ML) algorithms were applied to assess how they allowed to monitor cognitive or learning states automatically for possible applications in the educational domain. The outcomes of this thesis are:

- Development of a Jupyter Notebook extension to support lessons development and implementation in the format of Jupyter Notebooks, allowing to deploy practical classes in engineering courses.
- Development of three databases containing biosignals and HCI data in the context of seven different tasks: N-Back task, Mental Subtraction, programming language tutorial, Corsi-block task, concentration task, ECG lesson, and revised ECG lesson.
- Participation in a research stay at ETH Zurich to support the implementation and development of a study that used mouse-tracking data to study the possibility of predicting different variables related to decision-making.
- Development of a learning model that can be used in future studies to help guide learners in e-learning contexts.
- The developed set of specific features aggregated and thoroughly described, is also considered an outcome to guide future research.
- The methodology to achieve the best performance is an outcome that can help, in future work, to assess the best combinations considering user-tuned applications.

7.3 List of Publications

Journal Publications

- R. Varandas, B. Gonçalves, H. Gamboa, and P. Vieira, 'Quantified Explainability: Convolutional Neural Network Focus Assessment in Arrhythmia Detection', BioMedInformatics, vol. 2, no. 1, pp. 124–138, 2022.
- R. Varandas, R. Lima, S. Bermúdez i Badia, H. Silva, and H. Gamboa, 'Automatic cognitive fatigue detection using wearable fNIRS and machine learning', Sensors, vol. 22, no. 11, p. 4010, 2022.

- P. Gamboa, R. Varandas, J. Rodrigues, C. Cepeda, C. Quaresma, and H. Gamboa, 'Attention Classification Based on Biosignals during Standard Cognitive Tasks for Occupational Domains', Computers, vol. 11, no. 4, p. 49, 2022.

Conference Papers

- R. Varandas, H. Gamboa, I. Silveira, P. Gamboa, and C. Quaresma, 'Automatic Cognitive Workload Classification Using Biosignals for Distance Learning Applications', in Doctoral Conference on Computing, Electrical and Industrial Systems, 2021, pp. 254–261.

- P. Gamboa, C. Quaresma, R. Varandas, and H. Gamboa, 'Benefits, Implications and Ethical Concerns of Machine Learning Tools Serving Mental Health Purposes', in Doctoral Conference on Computing, Electrical and Industrial Systems, 2021, pp. 285–294.

- P. Gamboa et al., 'Design of an Attention Tool Using HCI and Work-Related Variables', in Technological Innovation for Applied AI Systems: 12th IFIP WG 5.5/SOCOLNET Advanced Doctoral Conference on Computing, Electrical and Industrial Systems, DoCEIS 2021, Costa de Caparica, Portugal, July 7–9, 2021, Proceedings 12, 2021, pp. 262–269.

- R. Varandas, G. Ramos, K. Mrotzeck, P. Gamboa, and H. Gamboa, 'BIOSIGNALSNOTEBOOKS: AN INNOVATIVE APPROACH FOR BIOSIGNALS EXPERIENCE SHARING', in INTED2020 Proceedings, 2020, pp. 8194–8200.

- M. Alfaras, R. Varandas, and H. Gamboa, 'Ring-topology echo state networks for ICU sepsis classification', in 2019 Computing in Cardiology (CinC), 2019, p. Page-1.

Presentations and Invited Lectures

- Workshop, "Biosignals Processing using Deep Learning," International Preconference Workshop on "Artificial Intelligence in Healthcare", 2021.

- R. Varandas and H. Gamboa, "System Architecture Proposal for Distance Learning Applications,"in Conference Proceedings Glow, Lisbon, Portugal, Oct. 22, 2021. Accessed: Jan. 9, 2024. [Online]. Available: https://glow.ulusofona.pt/glow21/glow-21-posters-session/

Books

- H. Plácido da Silva, H. Silveira Gamboa, R. Varandas, and G. Ramos, Biosignal Acquisition and Processing: A Project-Based Learning Approach. Learning Materials in Biosciences, 2024.

7.4 Applications

The main focus of this thesis was to study and assess the possibility of integrating biosignals and behavioural metrics in the educational domain. Thus, related applications based on it would be:

Education: the detection of learning states in the educational domain could help assisting in self-learning and development models of learning, as well as support distance or blended learning courses, where tutors face difficulties in assessing the cognitive states of learners.

Driving: multiple works have researched the implications of drivers drowsiness and cognitive fatigue, employing fNIRS as a means to detect them. However, few have employed sensors with as low channels as the sensors employed in this work, which would facilitate the setup of a new system to be applied on-the-fly.

Psychology or Psychiatry research: the work developed in this thesis can be directly transposed into psychology research. Namely, the methodology can be applied to assess other cognitive or emotional states in different contexts. In medical research, it would be interesting to study how this work would translate to detect conditions such as Attention Deficit and Hyperactivity Disorder (ADHD) or others related to learning deficits.

Work related disorders: previous works have investigated how inattention and distraction can have an impact in the life of workers leading to work related disorders [229].

Usability tests: usability is an important aspect in numerous applications, namely, web-design. Its assessment is usually performed with focus groups and questionnaires, however, the work performed during this thesis could help to improve research using biosignals and mouse-tracking to assess users' states during interaction.

7.5 Future Work

Given the work performed during this thesis, future work should help to further improve research and overcome gaps that were left. Furthermore, new research questions arose during the execution of this work:

Increase sample size: although three different studies were conducted in this thesis, the sample sizes were limited, being the larger composed of 18 volunteers. Moreover, the volunteers were mostly biomedical engineering students at the University level and, thus, a more heterogeneous sample might reveal different results.

Improved explainability: explainability consists of the ability of explaining the results in a ML problem. However, this thesis did not delve into the explanation of why

or how classification worked. Future research should investigate what behaviours and features lead to better classification performance while still achieving proper results.

Development of a learning tool: this thesis demonstrated the feasibility of classifying different cognitive and learning states during e-learning lessons. It also showed that it is possible to do so with minimal setups and short time windows that would allow to function in close to real-time. Thus, the development of a learning tool that integrated this classification would benefit learners with real-time feedback and/or adaptation to optimise their learning performance.

Integration of the learning model: algorithms such as neural networks or hidden Markov models have the ability to integrate the probability of transitions between different states and, with that, improve classification performance. This integration might even reveal to be the key to the development of inter-user classifiers.

Addition of mobile devices and virtual reality: this work focused on learning occurring via computer, however, mobile devices, such as smartphones and tablets, are also widely available and can be used to access learning contents. The interaction with those devices involves touches instead of mouse-movements and keyboard presses and, thus, these should be included in future studies. Virtual reality is also a technology that can be used in education in which the interaction is different than the interaction with the computer and, thus, it would be interesting to study how it might help support learning.

7.6 Future Implications

Education today is mainly teacher/tutor-centred, i.e., the knowledge is typically held by them and explained to a class of learners that learns from them in structured manners. However, learners interests are personal and the way by which they learn might be different from one another. This means that the same teaching style might help some learners to achieve their best performance, but might be detrimental to other learners and, thus, a personalized teaching approach should be followed to improve learning. Since the implications of this thesis demonstrated the possibility to assess learning states in real-time simply by using mouse-tracking or in combination with a low number of biosignals, or to assess cognitive states with biosignals, it should be possible to build a learning tool that integrated that data to adapt the learning contents and style in real-time, even in distance learning applications, allowing for a personalized learning experience that potentiated learning for each individual user.

Although mouse usage is ubiquitous, being it one of the main ways to interact with computers, biosignals data acquisition could reveal a greater challenge in building the mentioned tool. However, these are already present in most wearable devices, with

photoplethysmography, EDA, and ACC sensors widely available that should facilitate it. Moreover, since the feasibility of using low numbers of channels of EEG and fNIRS sensors was demonstrated, new smart-helmets could be developed and used in this context, a solution that is already in use in China [230].

However the possibilities, ethical concerns should be taken into account. Namely, the mouse-tracking and biosignals data can be considered personal data and, thus, careful handling is advised. In this sense, the European Union implement the General Data Protection Regulation 2016/679 (GDPR) to protect users' data and privacy by giving them the control over their own data [231]. Moreover, additional ethical considerations involve the ongoing measurement of individuals who may not consent to being monitored. Therefore, there should always be an alternative to opt out of monitoring without facing any repercussions. The European AI Act, aimed at safeguarding European citizens, seeks to regulate the use of AI across various applications. Notably, in the educational domain, which is proposed to be considered as high risk, AI systems require registration before deployment [1]. Emotional recognition, a topic relevant to the presented investigation, is being discussed to be explicitly prohibited in educational institutions according to the same document. As shown in the present thesis, this type of research should still be permitted, advocating for an approach that weighs the potential benefits and risks of each specific application or system, emphasizing the importance of considering the ethical implications associated with each individually. Specifically, ethically well-designed new research should not be banned, enabling the advance of science in these relevant topics.

In conclusion, the incorporation of ML algorithms in this study, characterized by a lack of transparency and explainability, introduces the risk of monitoring and guiding individuals through learning courses based on misleading assumptions. For instance, as highlighted in Section 5.2, the algorithms exhibited ineffectiveness for certain individuals, leading to potential non-optimal recommendations without clear explanations, thereby detrimentally impacting the learning process. It is essential that users retain the ability to question and comprehend the classification decisions made by the employed algorithms. This ensures that the influence of AI on their learning journey remains positive.

[1] EU AI Act: first regulation on artificial intelligence [Accessed in 25/01/2024] https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

Bibliography

[1] J. M. Lourenço. *The NOVAthesis LaTeX Template User's Manual*. NOVA University Lisbon. 2021. URL: https://github.com/joaomlourenco/novathesis/raw/main/template.pdf (cit. on p. iii).

[2] K. A. Renninger and S. Hidi. "Revisiting the Conceptualization, Measurement, and gGneration of Interest". In: *Educational Psychologist* 46.3 (2011), pp. 168–184. ISSN: 00461520. DOI: 10.1080/00461520.2011.587723 (cit. on p. 1).

[3] D. J. Bos et al. "Neural Correlates of Preferred Activities: Development of an Interest-Specific Go/NoGo Task". In: *Social Cognitive and Affective Neuroscience* 12.12 (2017), pp. 1890–1901. ISSN: 17495024. DOI: 10.1093/scan/nsx127 (cit. on p. 1).

[4] M. W. Eysenck and M. T. Keane. *Cognitive Psychology: A Student's Handbook*. New York, NY, US: Psychology Press, 2000. ISBN: 0-86377-550-0 (Hardcover); 0-86377-551-9 (Paperback) (cit. on pp. 1, 8, 9, 11, 12, 16).

[5] R. E. Mayer. "Searching for the Role of Emotions in E-Learning". In: *Learning and Instruction* May (2019), p. 101213. ISSN: 09594752. DOI: 10.1016/j.learninstruc.2019.05.010. URL: https://doi.org/10.1016/j.learninstruc.2019.05.010 (cit. on p. 1).

[6] S.-F. Wu, Y.-L. Lu, and C.-J. Lien. "Detecting Students' Flow States and Their Construct Through Electroencephalogram: Reflective Flow Experiences, Balance of Challenge and Skill, and Sense of Control". In: *Journal of Educational Computing Research* 58.8 (2021), pp. 1515–1540. DOI: 10.1177/0735633120944084. eprint: https://doi.org/10.1177/0735633120944084. URL: https://doi.org/10.1177/0735633120944084 (cit. on p. 1).

[7] B. Kiselev and V. Yakutenko. "An Overview of Massive Open Online Course Platforms: Personalization and Semantic Web Technologies and Standards". In: *Procedia Computer Science* 169.2019 (2020), pp. 373–379. ISSN: 18770509. DOI: 10.1016/j.procs.2020.02.232. URL: https://doi.org/10.1016/j.procs.2020.02.232 (cit. on p. 1).

[8] C.-M. Chen, J.-Y. Wang, and C.-M. Yu. "Assessing the Attention Levels of Students by Using a Novel Attention Aware System Based on Brainwave Signals". In: *British Journal of Educational Technology* 48.2 (2017), pp. 348–369. DOI: https://doi.org/10.1111/bjet.12359. eprint: https://bera-journals.onlinelibrary.wiley.com/doi/pdf/10.1111/bjet.12359. URL: https://bera-journals.onlinelibrary.wiley.com/doi/abs/10.1111/bjet.12359 (cit. on pp. 1, 2, 41, 42, 115).

[9] N.-H. Liu, C.-Y. Chiang, and H.-C. Chu. "Recognizing the Degree of Human Attention Using EEG Signals from Mobile Sensors". In: *Sensors* 13.8 (2013), pp. 10273–10286. ISSN: 1424-8220. DOI: 10.3390/s130810273. URL: https://www.mdpi.com/1424-8220/13/8/10273 (cit. on pp. 2, 41).

[10] R. E. Mayer. "Searching for the role of emotions in e-learning". In: *Learning and Instruction* 70 (2020). Understanding and Measuring Emotions in Technology-Rich Learning Environments, p. 101213. ISSN: 0959-4752. DOI: https://doi.org/10.1016/j.learninstruc.2019.05.010. URL: https://www.sciencedirect.com/science/article/pii/S095947521930324X (cit. on pp. 2, 44).

[11] F. D'Errico, M. Paciello, and L. Cerniglia. "When Emotions Enhance Students' Engagement in E-Learning Processes". In: *Journal of e-Learning and Knowledge Society* 12.4 (2016), pp. 9–23. DOI: 10.20368/1971-8829/1144 (cit. on pp. 2, 44).

[12] M. Ciolacu et al. "Education 4.0 — Fostering Student's Performance with Machine Learning Methods". In: *2017 IEEE 23rd International Symposium for Design and Technology in Electronic Packaging (SIITME)*. 2017, pp. 438–443. DOI: 10.1109/SIITME.2017.8259941 (cit. on p. 2).

[13] J.-I. Castillo-Velazquez and R.-B. Silva-Lopez. "Evolution and Trends in E-Learning Approaches to STEM, Engineering Education and Corporate Learning: A Bibliometric Analysis to 2021". In: *2022 IEEE ANDESCON*. 2022, pp. 1–6. DOI: 10.1109/ANDESCON56260.2022.9989971 (cit. on pp. 2, 14).

[14] A. ZareRavasan and A. Jeyaraj. "Evolution of Information Systems Business Value Research: Topic Modeling Analysis". In: *Journal of Computer Information Systems* 63.3 (2022), pp. 555–573. ISSN: 23802057. DOI: 10.1080/08874417.2022.2085212. URL: https://doi.org/10.1080/08874417.2022.2085212 (cit. on p. 2).

[15] S. Choudhury and S. Pattnaik. "Emerging Themes in E-Learning: A Review from the Stakeholders' Perspective". In: *Computers and Education* 144.September 2018 (2020), p. 103657. ISSN: 03601315. DOI: 10.1016/j.compedu.2019.103657. URL: https://doi.org/10.1016/j.compedu.2019.103657 (cit. on p. 2).

[16] M. Liu and D. Yu. "Towards Intelligent E-learning Systems". In: *Education and Information Technologies* 28 (7 2023), pp. 7845–7876. ISSN: 1573-7608. DOI: 10.1007/s10639-022-11479-6. URL: https://doi.org/10.1007/s10639-022-11479-6 (cit. on pp. 2, 13).

[17] D. S. Evale. "Learning Management System with Prediction Model and Course-content Recommendation Module". In: *Journal of Information Technology Education: Reseacrh* 16 (2017), pp. 437–457. DOI: 10.28945/3883 (cit. on pp. 3, 15).

[18] R. Katarya and O. P. Verma. "Recent Developments in Affective Recommender Systems". In: *Physica A: Statistical Mechanics and its Applications* 461 (2016), pp. 182–190. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2016.05.046. URL: https://www.sciencedirect.com/science/article/pii/S037843711630231X (cit. on p. 3).

[19] S. Gupta, P. Kumar, and R. Tekchandani. "A Machine Learning-Based Decision Support System for Temporal Human Cognitive State Estimation During Online Education Using Wearable Physiological Monitoring Devices". In: *Decision Analytics Journal* 8 (2023), p. 100280. ISSN: 2772-6622. DOI: https://doi.org/10.1016/j.dajour.2023.100280. URL: https://www.sciencedirect.com/science/article/pii/S2772662223001200 (cit. on pp. 3, 43).

[20] J. Díaz-García et al. "Mental Load and Fatigue Assessment Instruments: A Systematic Review". In: *International Journal of Environmental Research and Public Health* 19.1 (2022). ISSN: 1660-4601. DOI: 10.3390/ijerph19010419. URL: https://www.mdpi.com/1660-4601/19/1/419 (cit. on pp. 3, 39).

[21] M. Lansdown. *In The Zone : Helping Children Rise to the Challenge of Learning*. Accession Number: 2194964; OCLC: 1109822688; Language: English. University of Buckingham Press, 2019. ISBN: 9781789551280. URL: https://search.ebscohost.com/login.aspx?direct=true\\&AuthType=ip,shib\\&db=e000bww\\&AN=2194964\\&lang=pt-pt\\&site=eds-live\\&scope=site (cit. on p. 7).

[22] J. R. Eyler. *Introduction*. English. Vol. First edition. Teaching and Learning in Higher Education. Morgantown: West Virginia University Press, 2018. ISBN: 9781946684646. URL: https://search.ebscohost.com/login.aspx?direct=true\\&AuthType=ip,shib\\&db=e000bww\\&AN=1936506\\&lang=pt-pt\\&site=eds-live\\&scope=site (cit. on p. 7).

[23] D. Dias. *Psicologia da Aprendizagem - Paradigmas, Motivação e Dificuldades*. 1st ed. Edições Sílabo, 2018. ISBN: 9789726189794 (cit. on pp. 7–9, 12, 13).

[24] J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson. "Probing the Mechanisms of Attention". In: *Handbook of Psychophysiology*. Ed. by J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson. third. Cambridge University Press, 2007. Chap. 18, pp. 410–432. ISBN: 978-0-511-27907-2 (cit. on p. 8).

[25] J. L. Tracy and D. Randles. "Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt". In: *Emotion Review* 3.4 (2011), pp. 397–405. ISSN: 17540739. DOI: 10.1177/1754073911410747 (cit. on p. 12).

[26] S. Buechel and U. Hahn. "Emotion analysis as a regression problem-dimensional models and their implications on Emotion representation and metrical evaluation". In: *Frontiers in Artificial Intelligence and Applications* 285 (2016), pp. 1114–1122. ISSN: 18798314. DOI: 10.3233/978-1-61499-672-9-1114 (cit. on p. 12).

[27] N. Shin. "Online learner's 'flow' experience an empirical study". In: *British Journal of Educational Technology* 37.5 (2006), pp. 705–720. DOI: 10.1111/j.1467-8535.2006.00641.xVol (cit. on p. 12).

[28] S. Mystakidis, A. Christopoulos, and N. Pellas. "A Systematic Mapping Review of Augmented Reality Applications to Support STEM Learning in Higher Education". In: 27.2 (2022), pp. 1883–1927. ISSN: 15737608. DOI: 10.1007/s10639-021-10682-1. URL: https://doi.org/10.1007/s10639-021-10682-1 (cit. on p. 14).

[29] P. Phanphech et al. "An Analysis of Student Anxiety Affecting on Online Learning on Conceptual Applications in Physics: Synchronous vs. Asynchronous Learning". In: *Education Sciences* 12.4 (2022). ISSN: 2227-7102. DOI: 10.3390/educsci12040278. URL: https://www.mdpi.com/2227-7102/12/4/278 (cit. on p. 14).

[30] J. Fernandes, R. Costa, and P. Peres. "Putting Order into our Universe: The Concept of Blended Learning—A Methodology Within the Concept-Based Terminology Framework". In: *Education Sciences* 6.2 (2016). ISSN: 22277102. DOI: 10.3390/educsci6020015 (cit. on p. 14).

[31] R. Katarya and O. P. Verma. "Recent Developments in Affective Recommender Systems". In: *Physica A: Statistical Mechanics and its Applications* 461 (2016), pp. 182–190. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2016.05.046. URL: https://www.sciencedirect.com/science/article/pii/S037843711630231X (cit. on p. 15).

[32] M. Maravanyika and N. Dlodlo. "An Adaptive Framework for Recommender-Based Learning Management Systems". In: *2018 Open Innovations Conference, OI 2018* (2018), pp. 203–212. DOI: 10.1109/OI.2018.8535816 (cit. on p. 15).

[33] U. C. Apoki et al. "The Role of Pedagogical Agents in Personalised Adaptive Learning: A Review". In: *Sustainability (Switzerland)* 14.11 (2022), pp. 1–17. ISSN: 20711050. DOI: 10.3390/su14116442 (cit. on p. 15).

[34] G. H. Glover. "Overview of Functional Magnetic Resonance Imaging". In: *Neurosurgery Clinics of North America* 22.2 (2011). Functional Imaging, pp. 133–139. ISSN: 1042-3680. DOI: https://doi.org/10.1016/j.nec.2010.11.001. URL: https://www.sciencedirect.com/science/article/pii/S1042368010001129 (cit. on pp. 15, 16).

[35] A. Kawala-Sterniuk et al. "Summary of Over Fifty Years with Brain-Computer Interfaces—A Review". In: *Brain Sciences* 11.1 (2021), pp. 1–41. ISSN: 20763425. DOI: 10.3390/brainsci11010043 (cit. on p. 16).

[36] M. Doulgerakis, A. T. Eggebrecht, and H. Dehghani. "High-Density Functional Diffuse Optical Tomography Based on Frequency-Domain Measurements Improves Image Quality and Spatial Resolution". In: *Neurophotonics* 6.3 (2019), p. 035007. DOI: 10.1117/1.NPh.6.3.035007. URL: https://doi.org/10.1117/1.NPh.6.3.035007 (cit. on p. 16).

[37] F. Biessmann et al. "Analysis of Multimodal Neuroimaging Data". In: *IEEE Reviews in Biomedical Engineering* 4 (2011), pp. 26–58. DOI: 10.1109/RBME.2011.2170675 (cit. on p. 16).

[38] A. Villringer and B. Chance. "Non-Invasive Optical Spectroscopy and Imaging of Human Brain Function". In: *Trends in Neurosciences* 20.10 (1997), pp. 435–442. ISSN: 0166-2236. DOI: https://doi.org/10.1016/S0166-2236(97)01132-6. URL: https://www.sciencedirect.com/science/article/pii/S0166223697011326 (cit. on pp. 16, 20).

[39] J. E. Hall and A. C. Guyton. *Textbook of Medical Physiology*. 12th. Vol. XXXIII. 2. Saunders Elsevier, 2011. Chap. 59. ISBN: 9780874216561. DOI: 10.1007/s13398-014-0173-7.2. arXiv: arXiv:1011.1669v3. URL: http://www.americanbanker.com/issues/179_124/which-city-is-the-next-big-fintech-hub-new-york-stakes-its-claim-1068345-1.html\%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/15003161\%5Cnhttp://cid.oxfordjournals.org/lookup/doi/10.1093/cid/cir991\%5Cnhttp://www.scielo (cit. on pp. 16–19, 25, 27).

[40] P. Adjamian. "The Application of Electro- and Magneto-Encephalography in Tinnitus Research – Methods and Interpretations". In: *Frontiers in Neurology* 5 (2014). ISSN: 1664-2295. DOI: 10.3389/fneur.2014.00228. URL: https://www.frontiersin.org/journals/neurology/articles/10.3389/fneur.2014.00228 (cit. on pp. 17, 20).

[41] C. Q. Lai et al. "Literature Survey: Recording Set Up for Electroencephalography (EEG) Acquisition". In: *2018 IEEE Symposium on Computer Applications Industrial Electronics (ISCAIE)*. 2018, pp. 333–338. DOI: 10.1109/ISCAIE.2018.8405494 (cit. on pp. 19, 20).

[42] A. A. Gaho et al. "EEG Signals Based Brain Source Localization Approaches". In: *International Journal of Advanced Computer Science and Applications* 9.9 (2018). DOI: 10.14569/IJACSA.2018.090934. URL: http://dx.doi.org/10.14569/IJACSA.2018.090934 (cit. on p. 20).

[43] D. P. Subha et al. "EEG Signal Analysis: A Survey". In: *Journal of Medical Systems* 34 (2 2010), pp. 195–212. ISSN: 1573-689X. DOI: 10.1007/s10916-008-9231-z. URL: https://doi.org/10.1007/s10916-008-9231-z (cit. on p. 20).

[44] F. Al-Shargie, T. B. Tang, and M. Kiguchi. "Assessment of Mental Stress Effects on Prefrontal Cortical Activities Using Canonical Correlation Analysis: an fNIRS-EEG Study". In: *Biomed. Opt. Express* 8.5 (2017-05), pp. 2583–2598. DOI: 10.1364/BOE.8.002583. URL: https://opg.optica.org/boe/abstract.cfm?URI=boe-8-5-2583 (cit. on p. 20).

[45] E. S. Louis et al. *Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants*. Ed. by E. K. S. Louis and L. C. Frey. 2016. ISBN: 9780997975604. DOI: 10.5698/978-0-9979756-0-4 (cit. on p. 20).

[46] V. Scarapicchia et al. "Functional Magnetic Resonance Imaging and Functional Near-Infrared Spectroscopy: Insights from Combined Recording Studies". In: *Frontiers in Human Neuroscience* 11 (2017). ISSN: 1662-5161. DOI: 10.3389/fnhum.2017.00419. URL: https://www.frontiersin.org/articles/10.3389/fnhum.2017.00419 (cit. on pp. 21, 23).

[47] C. Herff et al. "Classification of Mental Tasks in the Prefrontal Cortex using fNIRS". In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2013, pp. 2160–2163. DOI: 10.1109/EMBC.2013.6609962 (cit. on pp. 21, 41, 114).

[48] L. Kocsis, P. Herman, and A. Eke. "The Modified Beer–Lambert Law Revisited". In: *Physics in Medicine Biology* 51.5 (2006-02), N91. DOI: 10.1088/0031-9155/51/5/N02. URL: https://dx.doi.org/10.1088/0031-9155/51/5/N02 (cit. on pp. 21, 22).

[49] V. Quaresima and M. Ferrari. "Functional Near-Infrared Spectroscopy (fNIRS) for Assessing Cerebral Cortex Function During Human Behavior in Natural/Social Situations: A Concise Review". In: *Organizational Research Methods* 22.1 (2019), pp. 46–68. DOI: 10.1177/1094428116658959. eprint: https://doi.org/10.1177/1094428116658959. URL: https://doi.org/10.1177/1094428116658959 (cit. on pp. 21, 23, 24).

[50] F. Lange and I. Tachtsidis. "Clinical Brain Monitoring with Time Domain NIRS: A Review and Future Perspectives". In: *Applied Sciences* 9.8 (2019). ISSN: 2076-3417. DOI: 10.3390/app9081612. URL: https://www.mdpi.com/2076-3417/9/8/1612 (cit. on pp. 21–23).

[51] S. C. Harrison and D. E. Hartley. "Shedding Light On The Human Auditory Cortex: A Review Of The Advances In Near Infrared Spectroscopy (NIRS)". In: *Reports in Medical Imaging* 12 (2019), pp. 31–42. DOI: 10.2147/RMI.S174633. eprint: https://www.tandfonline.com/doi/pdf/10.2147/RMI.S174633. URL: https://www.tandfonline.com/doi/abs/10.2147/RMI.S174633 (cit. on p. 21).

[52] W. B. Baker et al. "Modified Beer-Lambert law for blood flow". In: *Biomed. Opt. Express* 5.11 (2014-11), pp. 4053–4075. DOI: 10.1364/BOE.5.004053. URL: https://opg.optica.org/boe/abstract.cfm?URI=boe-5-11-4053 (cit. on p. 22).

[53] D. J. Davies et al. "Frequency-Domain vs Continuous-Wave Near-Infrared Spectroscopy Devices: A Comparison of Clinically Viable Monitors in Controlled Hypoxia". In: *Journal of Clinical Monitoring and Computing* 31 (5 2017), pp. 967–974. ISSN: 1573-2614. DOI: 10.1007/s10877-016-9942-5. URL: https://doi.org/10.1007/s10877-016-9942-5 (cit. on p. 23).

[54] A. Torricelli et al. "Time Domain Functional NIRS Imaging for Human Brain Mapping". In: *NeuroImage* 85 (2014). Celebrating 20 Years of Functional Near Infrared Spectroscopy (fNIRS), pp. 28–50. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2013.05.106. URL: https://www.sciencedirect.com/science/article/pii/S1053811913006095 (cit. on p. 23).

[55] R. K. Almajidy et al. "A Newcomer's Guide to Functional Near Infrared Spectroscopy Experiments". In: *IEEE Reviews in Biomedical Engineering* 13 (2020), pp. 292–308. DOI: 10.1109/RBME.2019.2944351 (cit. on p. 23).

[56] Y. H. M.D. and Y. Yamada. "Overview of Diffuse Optical Tomography and its Clinical Applications". In: *Journal of Biomedical Optics* 21.9 (2016), p. 091312. DOI: 10.1117/1.JBO.21.9.091312. URL: https://doi.org/10.1117/1.JBO.21.9.091312 (cit. on p. 24).

[57] B. K. Bracken et al. "Validation of the fNIRS Pioneer™ , a Portable, Durable, Rugged functional Near-Infrared Spectroscopy (fNIRS) Device". In: 2019, pp. 521–531. ISBN: 9789897583537. DOI: 10.5220/0007471405210531 (cit. on pp. 25, 43).

[58] L. S. Costanzo. *Physiology*. 6th. Elsevier, 2018. ISBN: 9780323478816. URL: http://library1.nida.ac.th/termpaper6/sd/2554/19755.pdf (cit. on pp. 25, 26, 28).

[59] J. Tindle and P. Tadi. "Neuroanatomy, Parasympathetic Nervous System". In: *StatPearls* (2022-10). URL: https://www.ncbi.nlm.nih.gov/books/NBK553141/ (cit. on p. 26).

[60] M. Proença and K. Mrotzeck. *BITalino (r)evolution Lab Guide*. 2020. URL: https://support.pluxbiosignals.com/wp-content/uploads/2022/04/HomeGuide2_ECG.pdf (cit. on p. 27).

[61] W. Lopez-Ojeda et al. *Anatomy, Skin (Integument)*. StatPearls Publishing, 2022-10. URL: http://www.ncbi.nlm.nih.gov/pubmed/28723009 (cit. on p. 28).

[62] B. D. Hodge, T. Sanvictores, and R. T. Brodell. *Anatomy, Skin Sweat Glands*. StatPearls Publishing, Treasure Island (FL), 2022. URL: http://europepmc.org/books/NBK482278 (cit. on p. 28).

[63] C. Cepeda. "Personality Assessment Using Biosignals and Human Computer Interaction applied to Medical Decision Making". PhD. Universidade Nova de Lisboa, 2019. URL: http://hdl.handle.net/10362/93822 (cit. on pp. 28, 29, 32, 46, 80, 82).

[64] H. F. Posada-Quintero and K. H. Chon. "Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic Review". In: *Sensors* 20.2 (2020). ISSN: 1424-8220. DOI: 10.3390/s20020479. URL: https://www.mdpi.com/1424-8220/20/2/479 (cit. on p. 29).

[65] *Respiration (RIP) Sensor Science Hour*. URL: http://notebooks.pluxbiosignals.com/notebooks/Categories/Other/rip_science_hour_rev.html (visited on 2023-08-30) (cit. on p. 30).

[66] R. Varandas. "Evaluation of Spatial-Temporal Anomalies in the Analysis of Human Movement". Master's. NOVA School of Sciences and Technology, 2018. URL: https://run.unl.pt/handle/10362/58215 (cit. on pp. 31, 33).

[67] S. Ansari et al. "Driver Mental Fatigue Detection Based on Head Posture Using New Modified reLU-BiLSTM Deep Neural Network". In: *IEEE Transactions on Intelligent Transportation Systems* 23.8 (2022), pp. 10957–10969. DOI: 10.1109/TITS.2021.3098309 (cit. on p. 31).

[68] F. Ren and Y. Bao. "A Review on Human-Computer Interaction and Intelligent Robots". In: *International Journal of Information Technology and Decision Making* 19.01 (2020), pp. 5–47. DOI: 10.1142/S0219622019300052. eprint: https://doi.org/10.1142/S0219622019300052. URL: https://doi.org/10.1142/S0219622019300052 (cit. on pp. 31, 32).

[69] M. Schoemann et al. "Using Mouse Cursor Tracking to Investigate Online Cognition: Preserving Methodological Ingenuity While Moving Toward Reproducible Science". In: *Psychonomic Bulletin*
Review 28 (3 2021), pp. 766–787. ISSN: 1531-5320. DOI: 10.3758/s13423-020-01851-3. URL: https://doi.org/10.3758/s13423-020-01851-3 (cit. on p. 32).

[70] P. J. Kieslich et al. "Design Factors in Mouse-Tracking: What Makes a Difference?" In: *Behavior Research Methods* 52.1 (2020), pp. 317–341. ISSN: 15543528. DOI: 10.3758/s13428-019-01228-y (cit. on p. 32).

[71] A. K. Jaiswal, P. Tiwari, and M. S. Hossain. "Predicting users' behavior using mouse movement information: an information foraging theory perspective". In: *Neural Computing and Applications* 35 (33 2023), pp. 23767–23780. ISSN: 1433-3058. DOI: 10.1007/s00521-020-05306-7. URL: https://doi.org/10.1007/s00521-020-05306-7 (cit. on p. 33).

[72] J. Shelton et al. "Mouse Tracking, Behavioral Biometrics, and GEFE". In: *2013 Proceedings of IEEE Southeastcon*. 2013, pp. 1–6. DOI: 10.1109/SECON.2013.6567457 (cit. on p. 33).

[73] T. Yamauchi, A. Leontyev, and M. Razavi. "Assessing Emotion by Mouse-cursor Tracking: Theoretical and Empirical Rationales". In: *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. 2019, pp. 89–95. DOI: 10.1109/ACII.2019.8925537 (cit. on p. 33).

[74] I. Arapakis and L. A. Leiva. "Learning Efficient Representations of Mouse Movements to Predict User Attention". In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '20. New York, NY, USA: Association for Computing Machinery, 2020, 1309–1318. ISBN: 9781450380164. DOI: 10.1145/3397271.3401031. URL: https://doi.org/10.1145/3397271.3401031 (cit. on p. 33).

[75] A. Géron. *Hands-on Machine Learning whith Scikit-Learing, Keras and Tensorfow*. 2019, p. 510. ISBN: 978-1-492-03264-9 (cit. on pp. 33, 34).

[76] C.-M. Yang and C.-H. Wu. "The Situational Fatigue Scale: A different approach to measuring fatigue". In: *Quality of Life Research* 14 (5 2005), pp. 1357–1362. ISSN: 1573-2649. DOI: 10.1007/s11136-004-5680-0. URL: https://doi.org/10.1007/s11136-004-5680-0 (cit. on p. 39).

[77] M. Shuman-Paretsky et al. "Development and Validation of the State-Trait Inventory of Cognitive Fatigue in Community-Dwelling Older Adults". eng. In: *Archives of physical medicine and rehabilitation* 98.4 (2017-04), pp. 766–773. ISSN: 1532-821X. DOI: 10.1016/j.apmr.2016.07.024. URL: https://pubmed.ncbi.nlm.nih.gov/27576190https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5651515/ (cit. on p. 39).

[78] E. MeAuley and K. S. Courneya. "The Subjective Exercise Experiences Scale (SEES): Development and Preliminary Validation". In: *Journal of Sport and Exercise Psychology* 16.2 (1994), pp. 163 –177. DOI: 10.1123/jsep.16.2.163. URL: https://journals.humankinetics.com/view/journals/jsep/16/2/article-p163.xml (cit. on p. 40).

[79] L. L. Di Stasi et al. "Towards a Driver Fatigue Test Based on the Saccadic Main Sequence: A Partial Validation by Subjective Report Data". In: *Transportation Research Part C: Emerging Technologies* 21.1 (2012), pp. 122–133. ISSN: 0968-090X. DOI: https://doi.org/10.1016/j.trc.2011.07.002. URL: https://www.sciencedirect.com/science/article/pii/S0968090X1100101X (cit. on p. 40).

[80] V. Kanal et al. "Towards a Serious Game Based Human-Robot Framework for Fatigue Assessment". In: 2020-06, pp. 527–532. ISBN: 9781450377737. DOI: 10.1145/3389189.3398744 (cit. on pp. 40, 42, 71).

[81] S. Ahn et al. "Exploring Neuro-Physiological Correlates of Drivers' Mental Fatigue Caused by Sleep Deprivation Using Simultaneous EEG, ECG, and fNIRS Data". In: *Frontiers in Human Neuroscience* 10 (2016). ISSN: 1662-5161. DOI: 10.3389/fnhum.2

016.00219. URL: https://www.frontiersin.org/articles/10.3389/fnhum.2016.00219 (cit. on pp. 40, 42, 71).

[82] T. Nihashi et al. "Monitoring of Fatigue in Radiologists During Prolonged Image Interpretation Using fNIRS". In: *Japanese Journal of Radiology* 37.6 (2019), pp. 437–448. ISSN: 1867108X. DOI: 10.1007/s11604-019-00826-2. URL: https://doi.org/10.1007/s11604-019-00826-2 (cit. on pp. 40, 42, 71).

[83] B. F. Yuksel et al. "Detecting and Adapting to Users' Cognitive and Affective State to Develop Intelligent Musical Interfaces". In: *New Directions in Music and Human-Computer Interaction*. Ed. by S. Holland et al. Cham: Springer International Publishing, 2019, pp. 163–177. ISBN: 978-3-319-92069-6. DOI: 10.1007/978-3-319-92069-6_11. URL: https://doi.org/10.1007/978-3-319-92069-6_11 (cit. on p. 41).

[84] C. Ferguson, E. L. van den Broek, and H. van Oostendorp. "AI-Induced Guidance: Preserving the Optimal Zone of Proximal Development". In: *Computers and Education: Artificial Intelligence* 3.July 2021 (2022), p. 100089. ISSN: 2666920X. DOI: 10.1016/j.caeai.2022.100089. URL: https://doi.org/10.1016/j.caeai.2022.100089 (cit. on p. 41).

[85] Y. Moriguchi and K. Hiraki. "Prefrontal Cortex and Executive Function in Young Children: a Review of NIRS Studies". In: *Frontiers in Human Neuroscience* 7 (2013). ISSN: 1662-5161. DOI: 10.3389/fnhum.2013.00867. URL: https://www.frontiersin.org/articles/10.3389/fnhum.2013.00867 (cit. on p. 41).

[86] P. Pinti et al. "The Present and Future Use of functional Near-Infrared Spectroscopy (fNIRS) for Cognitive Neuroscience". In: *Annals of the New York Academy of Sciences* 1464.1 (2020), pp. 5–29. DOI: https://doi.org/10.1111/nyas.13948. eprint: https://nyaspubs.onlinelibrary.wiley.com/doi/pdf/10.1111/nyas.13948. URL: https://nyaspubs.onlinelibrary.wiley.com/doi/abs/10.1111/nyas.13948 (cit. on p. 41).

[87] C. Herff et al. "Mental Workload During N-Back Task—Quantified in the Prefrontal Cortex Using fNIRS". In: *Frontiers in Human Neuroscience* 7 (2014). ISSN: 1662-5161. DOI: 10.3389/fnhum.2013.00935. URL: https://www.frontiersin.org/articles/10.3389/fnhum.2013.00935 (cit. on pp. 41, 63).

[88] J. M. Harley et al. "Fostering Positive Emotions and History Knowledge with Location-Based Augmented Reality and Tour-Guide Prompts". In: *Learning and Instruction* 70 (2020). Understanding and measuring emotions in technology-rich learning environments, p. 101163. ISSN: 0959-4752. DOI: https://doi.org/10.1016/j.learninstruc.2018.09.001. URL: https://www.sciencedirect.com/science/article/pii/S095947521730405X (cit. on pp. 41, 44, 114).

[89] M. Mazher et al. "An EEG-Based Cognitive Load Assessment in Multimedia Learning Using Feature Extraction and Partial Directed Coherence". In: *IEEE Access* 5 (2017), pp. 14819–14829. DOI: 10.1109/ACCESS.2017.2731784 (cit. on pp. 41, 114).

[90] F. Benedetti et al. "Attention Training with an Easy–to–Use Brain Computer Interface". In: *Virtual, Augmented and Mixed Reality. Applications of Virtual and Augmented Reality*. Ed. by R. Shumaker and S. Lackey. Cham: Springer International Publishing, 2014, pp. 236–247. ISBN: 978-3-319-07464-1 (cit. on p. 41).

[91] J. Kwon, J. Shin, and C.-H. Im. "Toward a Compact Hybrid Brain-Computer Interface (BCI): Performance Evaluation of Multi-Class Hybrid EEG-fNIRS BCIs with Limited Number of Channels". In: *PLOS ONE* 15.3 (2020-03), pp. 1–14. DOI: 10.1371/journal.pone.0230491. URL: https://doi.org/10.1371/journal.pone.0230491 (cit. on pp. 41, 114).

[92] Z. Liu et al. "A Systematic Review on Hybrid EEG/fNIRS in Brain-Computer Interface". In: *Biomedical Signal Processing and Control* 68 (2021), p. 102595. ISSN: 1746-8094. DOI: https://doi.org/10.1016/j.bspc.2021.102595. URL: https://www.sciencedirect.com/science/article/pii/S1746809421001920 (cit. on pp. 41, 114).

[93] F. Irani et al. "Functional Near Infrared Spectroscopy (fNIRS): An Emerging Neuroimaging Technology with Important Applications for the Study of Brain Disorders". In: *The Clinical Neuropsychologist* 21.1 (2007). PMID: 17366276, pp. 9–37. DOI: 10.1080/13854040600910018. eprint: https://doi.org/10.1080/13854040600910018. URL: https://doi.org/10.1080/13854040600910018 (cit. on p. 41).

[94] L. Lange and R. Osinsky. "Aiming at Ecological Validity—Midfrontal Theta Oscillations in a Toy Gun Shooting Task". In: *European Journal of Neuroscience* 54.12 (2021), pp. 8214–8224. DOI: https://doi.org/10.1111/ejn.14977. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejn.14977. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/ejn.14977 (cit. on p. 41).

[95] P. Aricò et al. "Passive BCI in Operational Environments: Insights, Recent Advances, and Future Trends". In: *IEEE Transactions on Biomedical Engineering* 64.7 (2017), pp. 1431–1436. DOI: 10.1109/TBME.2017.2694856 (cit. on p. 41).

[96] H. A. Maior et al. "Continuous Detection of Workload Overload: An fNIRS Approach". In: *Contemporary Ergonomics and Human Factors 2014* 12.4 (2014), pp. 450–457 (cit. on p. 41).

[97] F. Herold et al. "Applications of Functional Near-Infrared Spectroscopy (fNIRS) Neuroimaging in Exercise–Cognition Science: A Systematic, Methodology-Focused Review". In: *Journal of Clinical Medicine* 7.12 (2018). ISSN: 2077-0383. DOI: 10.3390/jcm7120466. URL: https://www.mdpi.com/2077-0383/7/12/466 (cit. on pp. 41, 59).

[98] B. K. Bracken. et al. "Validation of the fNIRS Pioneer™, a Portable, Durable, Rugged functional Near-Infrared Spectroscopy (fNIRS) Device". In: *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies - RAIDERS*, INSTICC. SciTePress, 2019, pp. 521–531. ISBN: 978-989-758-353-7. DOI: 10.5220/0007471405210531 (cit. on p. 41).

[99] L.-W. Ko et al. "Sustained Attention in Real Classroom Settings: An EEG Study". In: *Frontiers in Human Neuroscience* 11 (2017). ISSN: 1662-5161. DOI: 10.3389/fnhum.2017.00388. URL: https://www.frontiersin.org/articles/10.3389/fnhum.2017.00388 (cit. on pp. 42, 106).

[100] B. Hu et al. "Attention Recognition in EEG-Based Affective Learning Research Using CFS+KNN Algorithm". In: *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 15.1 (2018), pp. 38–45. DOI: 10.1109/TCBB.2016.2616395 (cit. on pp. 42, 106).

[101] F. Fahimi et al. "EEG Predicts the Attention Level of Elderly Measured by RBANS". In: *International Journal of Crowd Science* 2.3 (2018), pp. 272–282. DOI: 10.1108/IJCS-09-2018-0022 (cit. on pp. 42, 106).

[102] X. Zhang et al. "Analyzing Students' Attention in Class Using Wearable Devices". In: *2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. 2017, pp. 1–9. DOI: 10.1109/WoWMoM.2017.7974306 (cit. on pp. 42, 106).

[103] "Predicting Students' Attention in the Classroom from Kinect Facial and Body Features". In: *EURASIP Journal on Image and Video Processing* 2017 (1 2017), p. 80. ISSN: 1687-5281. DOI: 10.1186/s13640-017-0228-8. URL: https://doi.org/10.1186/s13640-017-0228-8 (cit. on pp. 42, 106).

[104] A. F. Abate et al. "Attention Monitoring for Synchronous Distance Learning". In: *Future Generation Computer Systems* 125 (2021), pp. 774–784. ISSN: 0167-739X. DOI: https://doi.org/10.1016/j.future.2021.07.026. URL: https://www.sciencedirect.com/science/article/pii/S0167739X21002880 (cit. on p. 42).

[105] D. Yang et al. "All In One Network for Driver Attention Monitoring". In: *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2020, pp. 2258–2262. DOI: 10.1109/ICASSP40776.2020.9053659 (cit. on p. 42).

[106] J. Batista. "A Drowsiness and Point of Attention Monitoring System for Driver Vigilance". In: *2007 IEEE Intelligent Transportation Systems Conference*. 2007, pp. 702–708. DOI: 10.1109/ITSC.2007.4357702 (cit. on p. 42).

[107] M.-H. Sigari, M. Fathy, and M. Soryani. "A Driver Face Monitoring System for Fatigue and Distraction Detection". In: *International Journal of Vehicular Technology* 2013 (2013). Ed. by C.-R. Dow, p. 263983. ISSN: 1687-5702. DOI: 10.1155/2013/263983. URL: https://doi.org/10.1155/2013/263983 (cit. on p. 42).

[108] D. Wang et al. "Monitoring Workers' Attention and Vigilance in Construction Activities Through a Wireless and Wearable Electroencephalography System". In: *Automation in Construction* 82 (2017), pp. 122–137. ISSN: 0926-5805. DOI: https://doi.org/10.1016/j.autcon.2017.02.001. URL: https://www.sciencedirect.com/science/article/pii/S0926580517301723 (cit. on p. 42).

[109] J. H. Kim et al. "Biosignal-Based Attention Monitoring to Support Nuclear Operator Safety-Relevant Tasks". In: *Frontiers in Computational Neuroscience* 14 (2020). ISSN: 1662-5188. DOI: 10.3389/fncom.2020.596531. URL: https://www.frontiersin.org/articles/10.3389/fncom.2020.596531 (cit. on p. 42).

[110] T. McMorris. "Cognitive Fatigue Effects on Physical Performance: The Role of Interoception". In: *Sports Medicine* 50.10 (2020), pp. 1703–1708. ISSN: 1179-2035. DOI: 10.1007/s40279-020-01320-w. URL: https://doi.org/10.1007/s40279-020-01320-w (cit. on pp. 42, 71).

[111] E. S. Jahanpour et al. "Cognitive Fatigue Assessment in Operational Settings: a Review and UAS Implications". In: *IFAC-PapersOnLine* 53.5 (2020). 3rd IFAC Workshop on Cyber-Physical
Human Systems CPHS 2020, pp. 330–337. ISSN: 2405-8963. DOI: https://doi.org/10.1016/j.ifacol.2021.04.188. URL: https://www.sciencedirect.com/science/article/pii/S2405896321003426 (cit. on pp. 42, 71).

[112] F. Dehais et al. "Monitoring Pilot's Cognitive Fatigue with Engagement Features in Simulated and Actual Flight Conditions Using an Hybrid fNIRS-EEG Passive BCI". In: *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2018, pp. 544–549. DOI: 10.1109/SMC.2018.00102 (cit. on pp. 42, 107, 114).

[113] T. Trakoolwilaiwan et al. "Convolutional Neural Network for High-Accuracy functional Near-Infrared Spectroscopy in a Brain–Computer Interface: Three-Class Classification of Rest, Right-, and Left-Hand Motor Execution". In: *Neurophotonics* 5.1 (2017), p. 011008. DOI: 10.1117/1.NPh.5.1.011008. URL: https://doi.org/10.1117/1.NPh.5.1.011008 (cit. on pp. 43, 108).

[114] K. Sakatani, M. Tanida, and M. Katsuyama. "Effects of Aging on Activity of the Prefrontal Cortex and Autonomic Nervous System During Mental Stress Task". In: *Advances in Experimental Medicine and Biology* 662 (2010), pp. 473–478. ISSN: 00652598. DOI: 10.1007/978-1-4419-1241-1_68 (cit. on pp. 43, 107).

[115] V. Quaresima et al. "Bilateral Prefrontal Cortex Oxygenation Responses to a Verbal Fluency Task: a Multichannel Time-Resolved Near-Infrared Topography Study". In: *Journal of Biomedical Optics* 10.1 (2005), p. 011012. ISSN: 10833668. DOI: 10.1117/1.1851512 (cit. on pp. 43, 107).

[116] H. Zohdi, F. Scholkmann, and U. Wolf. "Individual Differences in Hemodynamic Responses Measured on the Head due to a Long-Term Stimulation Involving Colored Light Exposure and a Cognitive Task: A Spa-fnirs Study". In: *Brain Sciences* 11.1 (2021), pp. 1–16. ISSN: 20763425. DOI: 10.3390/brainsci11010054 (cit. on pp. 43, 107).

[117] L. Holper et al. "Understanding Inverse Oxygenation Responses During Motor Imagery: A functional Near-Infrared Spectroscopy Study". In: *European Journal of Neuroscience* 33.12 (2011), pp. 2318–2328. ISSN: 0953816X. DOI: 10.1111/j.1460-9568.2011.07720.x (cit. on pp. 43, 107).

[118] P. W. Kim. "Real-time Bio-Signal-Processing of Students Based on an Intelligent Algorithm for Internet of Things to Assess Engagement Levels in a Classroom". In: *Future Generation Computer Systems* 86 (2018), pp. 716–722. ISSN: 0167-739X. DOI: https://doi.org/10.1016/j.future.2018.04.093. URL: https://www.sciencedirect.com/science/article/pii/S0167739X17325037 (cit. on p. 43).

[119] M. N. Giannakos et al. "Fitbit for Learning: Towards Capturing the Learning Experience Using Wearable Sensing". In: *International Journal of Human-Computer Studies* 136 (2020), p. 102384. ISSN: 1071-5819. DOI: https://doi.org/10.1016/j.ijhcs.2019.102384. URL: https://www.sciencedirect.com/science/article/pii/S1071581919301508 (cit. on pp. 44, 114).

[120] M. C. Duffy et al. "Emotions in Medical Education: Examining the Validity of the Medical Emotion Scale (MES) Across Authentic Medical Learning Environments". In: *Learning and Instruction* 70 (2020). Understanding and measuring emotions in technology-rich learning environments, p. 101150. ISSN: 0959-4752. DOI: https://doi.org/10.1016/j.learninstruc.2018.07.001. URL: https://www.sciencedirect.com/science/article/pii/S0959475217304255 (cit. on pp. 44, 114).

[121] ISO 9241-210. "Ergonomics of Human-System Interaction — Part 210: Human-Centred Design for Interactive Systems". In: *International Standard* 2 (2019), pp. 1–33. URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-2:v1:enhttps://www.iso.org/standard/77520.html (cit. on p. 44).

[122] A. Hussain et al. "Mobile Experience Evaluation of an E-Reader App". In: *Journal of Telecommunication, Electronic and Computer Engineering* 10.1-10 (2018), pp. 11–15. ISSN: 22898131 (cit. on p. 44).

[123] A. P. O. S. Vermeeren et al. "User Experience Evaluation Methods: Current State and Development Needs". In: NordiCHI '10 (2010), 521–530. DOI: 10.1145/1868914.1868973. URL: https://doi.org/10.1145/1868914.1868973 (cit. on p. 44).

[124] A. C. Ten and F. Paz. "A Systematic Review of User Experience Evaluation Methods in Information Driven Websites". In: (2017). Ed. by A. Marcus and W. Wang, pp. 492–506 (cit. on p. 44).

[125] I. Díaz-Oreiro et al. "Standardized Questionnaires for User Experience Evaluation: A Systematic Literature Review". In: *Proceedings* 31.1 (2019). ISSN: 2504-3900. DOI: 10.3390/proceedings2019031014. URL: https://www.mdpi.com/2504-3900/31/1/14 (cit. on p. 44).

[126] A. Inan Nur, H. B. Santoso, and P. O. Hadi Putra. "The Method and Metric of User Experience Evaluation: A Systematic Literature Review". In: *Proceedings of the 2021 10th International Conference on Software and Computer Applications*. ICSCA '21. New York, NY, USA: Association for Computing Machinery, 2021, 307–317. ISBN: 9781450388825. DOI: 10.1145/3457784.3457832. URL: https://doi.org/10.1145/3457784.3457832 (cit. on pp. 44, 45).

[127] C. Rico-Olarte, D. M. López, and S. Kepplinger. "Towards a Conceptual Framework for the Objective Evaluation of User Experience". In: *Design, User Experience, and Usability: Theory and Practice*. Ed. by A. Marcus and W. Wang. Cham: Springer International Publishing, 2018, pp. 546–559. ISBN: 978-3-319-91797-9 (cit. on p. 45).

[128] D. Xiang and Y. Wu. "Analysis and Research of Internet User Behaviors under the Context of Big Data". In: (2022), pp. 243–247. DOI: 10.1109/BDICN55575.2022.00054 (cit. on p. 45).

[129] R Deepalakshmi and J Amudha. "A Reinforcement Learning based Eye-Gaze Behavior Tracking". In: *2021 2nd Global Conference for Advancement in Technology (GCAT)*. 2021, pp. 1–7. DOI: 10.1109/GCAT52182.2021.9587480 (cit. on p. 45).

[130] Y. Li et al. "Artificial Intelligence-Based Human–Computer Interaction Technology Applied in Consumer Behavior Analysis and Experiential Education". In: *Frontiers in Psychology* 13 (2022). ISSN: 1664-1078. DOI: 10.3389/fpsyg.2022.784311. URL: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.784311 (cit. on p. 45).

[131] Z. Tupikovskaja-Omovie and D. J. Tyler. "Experienced Versus Inexperienced Mobile Users: Eye Tracking Fashion Consumers' Shopping Behaviour on Smartphones". In: *International Journal of Fashion Design, Technology and Education* 15.2 (2022), pp. 178–186. ISSN: 17543274. DOI: 10.1080/17543266.2021.1980614. URL: https://doi.org/10.1080/17543266.2021.1980614 (cit. on p. 45).

[132] F. Zhang et al. "Research of Webpage Complexity Influence on Search Behavior Based on Eye Tracking Experiment". In: *2021 IEEE 16th Conference on Industrial Electronics and Applications (ICIEA)*. 2021, pp. 527–532. DOI: 10.1109/ICIEA51954.2021.9516347 (cit. on p. 45).

[133] P. Sulikowski and T. Zdziebko. "Horizontal vs. Vertical Recommendation Zones Evaluation Using Behavior Tracking". In: *Applied Sciences* 11.1 (2021). ISSN: 2076-3417. DOI: 10.3390/app11010056. URL: https://www.mdpi.com/2076-3417/11/1/56 (cit. on p. 45).

[134] C. C. Armstrong et al. "Youth and Provider Perspectives on Behavior-Tracking Mobile Apps: Qualitative Analysis". In: *JMIR Ment Health* 8.4 (2021-04), e24482. ISSN: 2368-7959. DOI: 10.2196/24482. URL: https://doi.org/10.2196/24482 (cit. on p. 45).

[135] T. Aledavood et al. "Smartphone-Based Tracking of Sleep in Depression, Anxiety, and Psychotic Disorders". In: *Current Psychiatry Reports* 21 (7 2019), p. 49. ISSN: 1535-1645. DOI: 10.1007/s11920-019-1043-y. URL: https://doi.org/10.1007/s11920-019-1043-y (cit. on p. 45).

[136] Z. Jiang et al. "Utilizing Computer Vision for Facial Behavior Analysis in Schizophrenia Studies: A Systematic Review". In: *PLOS ONE* 17.4 (2022-04), pp. 1–22. DOI: 10.1371/journal.pone.0266828. URL: https://doi.org/10.1371/journal.pone.0266828 (cit. on p. 45).

[137] C. Cepeda et al. "Mouse Tracking Measures and Movement Patterns with Application for Online Surveys". In: *Machine Learning and Knowledge Extraction*. Ed. by A. Holzinger et al. Cham: Springer International Publishing, 2018, pp. 28–42. ISBN: 978-3-319-99740-7 (cit. on p. 46).

[138] M. C. Dias. et al. "Predicting Response Uncertainty in Online Surveys: A Proof of Concept". In: *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2019) - BIOSIGNALS*. INSTICC. SciTePress, 2019, pp. 155–162. ISBN: 978-989-758-353-7. DOI: 10.5220/0007381801550162 (cit. on p. 46).

[139] C. Cepeda et al. "Latent: A Flexible Data Collection Tool to Research Human Behavior in the Context of Web Navigation". In: *IEEE Access* 7 (2019), pp. 77659–77673. DOI: 10.1109/ACCESS.2019.2916996 (cit. on pp. 46, 61, 62).

[140] A. ur Rahman et al. "A Neuro-Fuzzy Approach for User Behaviour Classification and Prediction". In: *Journal of Cloud Computing* 8 (1 2019), p. 17. ISSN: 2192-113X. DOI: 10.1186/s13677-019-0144-9. URL: https://doi.org/10.1186/s13677-019-0144-9 (cit. on p. 46).

[141] M. Callara and P. Wira. "User Behavior Analysis with Machine Learning Techniques in Cloud Computing Architectures". In: *2018 International Conference on Applied Smart Systems (ICASS)*. 2018, pp. 1–6. DOI: 10.1109/ICASS.2018.8651961 (cit. on p. 46).

[142] M. Schmidt and R. Huang. "Defining Learning Experience Design: Voices from the Field of Learning Design
Technology". In: *TechTrends* 66 (2 2022), pp. 141–158. ISSN: 1559-7075. DOI: 10.1007/s11528-021-00656-y. URL: https://doi.org/10.1007/s11528-021-00656-y (cit. on p. 46).

[143] A. K. Arora et al. *Evaluating Usability in Blended Learning Programs Within Health Professions Education: a Scoping Review*. Vol. 31. 3. Springer US, 2021, pp. 1213–1246. ISBN: 0123456789. DOI: 10.1007/s40670-021-01295-x. URL: https://doi.org/10.1007/s40670-021-01295-x (cit. on p. 46).

[144] C. M. Baptista, R. M. Silveira, and W. V. Ruggiero. "MSys: An activities Tracking Tool for E-Learning Systems". In: *2008 38th Annual Frontiers in Education Conference*. 2008, S3A–17–S3A–22. DOI: 10.1109/FIE.2008.4720688 (cit. on p. 46).

[145] J. Hardy et al. "Tracking and Visualisation of Student Use of Online Learning Materials in a Large Undergraduate Course". In: *Advances in Web Based Learning – ICWL 2007*. Ed. by H. Leung et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 464–474. ISBN: 978-3-540-78139-4 (cit. on p. 46).

[146] M. May, S. George, and P. Prévôt. "Students' Tracking Data: An Approach for Efficiently Tracking Computer Mediated Communications in Distance Learning". In: *2008 Eighth IEEE International Conference on Advanced Learning Technologies*. 2008, pp. 783–787. DOI: 10.1109/ICALT.2008.53 (cit. on p. 47).

[147] A. L. Webb and S. Choi. "Interactive Radiological Anatomy eLearning Solution for First Year Medical Students: Development, Integration, and Impact on Learning". In: 7.5 (2014), pp. 350–360. ISSN: 19359780. DOI: 10.1002/ase.1428 (cit. on p. 47).

[148] A. F. Baharudin, N. A. Sahabudin, and A. Kamaludin. "Behavioral Tracking in E-Learning by Using Learning Styles Approach". In: *Indonesian Journal of Electrical Engineering and Computer Science* 8.1 (2017), pp. 17–26. ISSN: 25024760. DOI: 10.11591/ijeecs.v8.i1.pp17-26 (cit. on p. 47).

[149] M. Tellakat, R. L. Boyd, and J. W. Pennebaker. "How do Online Learners Study? The Psychometrics of Students' Clicking Patterns in Online Courses". In: *PLOS ONE* 14.3 (2019-03), pp. 1–17. DOI: 10.1371/journal.pone.0213863. URL: https://doi.org/10.1371/journal.pone.0213863 (cit. on p. 47).

[150] K. S. Lukics and Ágnes Lukács. "Tracking Statistical Learning Online: Word Segmentation in a Target Detection Task". In: *Acta Psychologica* 215 (2021), p. 103271. ISSN: 0001-6918. DOI: https://doi.org/10.1016/j.actpsy.2021.103271. URL: https://www.sciencedirect.com/science/article/pii/S0001691821000214 (cit. on p. 47).

[151] C. Lloyd. "Blogging as a Tool for Real-Time Learning in Medical Microbiology". In: *Frontiers in Microbiology* 12.March (2021), pp. 1–5. ISSN: 1664302X. DOI: 10.3389/fmicb.2021.576145. URL: https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2021.576145 (cit. on p. 47).

[152] F. Al-Hawari et al. "Methods to Achieve Effective Web-Based Learning Management Modules: MyGJU versus Moodle". In: *PeerJ Computer Science* 7 (2021-04), e498. ISSN: 2376-5992. DOI: 10.7717/peerj-cs.498. URL: https://doi.org/10.7717/peerj-cs.498 (cit. on p. 47).

[153] M. Shukla et al. "SMART-T: A System for Novel Fully Automated Anticipatory Eye-Tracking Paradigms". In: *Behavior Research Methods* 43 (2 2011), pp. 384–398. ISSN: 1554-3528. DOI: 10.3758/s13428-010-0056-6. URL: https://doi.org/10.3758/s13428-010-0056-6 (cit. on p. 48).

[154] S. M. de Mooij et al. "Should Online Math Learning Environments Be Tailored to Individuals' Cognitive Profiles?" In: *Journal of Experimental Child Psychology* 191 (2020), p. 104730. ISSN: 0022-0965. DOI: https://doi.org/10.1016/j.jecp.2019.104730. URL: https://www.sciencedirect.com/science/article/pii/S002209651930150X (cit. on p. 48).

[155] X. Li et al. "Predicting Spatial Visualization Problems' Difficulty Level from Eye-Tracking Data". In: *Sensors* 20.7 (2020). ISSN: 1424-8220. DOI: 10.3390/s20071949. URL: https://www.mdpi.com/1424-8220/20/7/1949 (cit. on p. 48).

[156] A. A. Jónsdóttir et al. "The Effects of Language Barriers and Time Constraints on Online Learning Performance: An Eye-Tracking Study". In: *Human Factors* 65.5 (2023). PMID: 33945351, pp. 779–791. DOI: 10.1177/00187208211010949. eprint: https://doi.org/10.1177/00187208211010949. URL: https://doi.org/10.1177/00187208211010949 (cit. on p. 48).

[157] C. Calvi, M. Porta, and D. Sacchi. "e5Learning, an E-Learning Environment Based on Eye Tracking". In: *2008 Eighth IEEE International Conference on Advanced Learning Technologies*. 2008, pp. 376–380. DOI: 10.1109/ICALT.2008.35 (cit. on p. 49).

[158] N. Tsianos et al. "Eye-Tracking Users' Behavior in Relation to Cognitive Style within an E-learning Environment". In: (2009), pp. 329–333. DOI: 10.1109/ICALT.2009.110 (cit. on p. 49).

[159] J. Qin et al. "Interaction Design Method Research in E-Learning Based on Eye Tracking". In: *2009 International Conference on Computational Intelligence and Software Engineering*. 2009, pp. 1–4. DOI: 10.1109/CISE.2009.5367054 (cit. on p. 49).

[160] K. Li and Y. Iribe. "Development and Evaluation of Gaze Tracking Integrated E-Learning Contents". In: *International Journal of Emerging Technologies in Learning (iJET)* 6.2 (2011-06), pp. 42–45. ISSN: 1863-0383. URL: https://www.learntechlib.org/p/45220 (cit. on pp. 49, 54).

[161] M. Porta, S. Ricotti, and C. J. Perez. "Emotional E-Learning through Eye Tracking". In: *Proceedings of the 2012 IEEE Global Engineering Education Conference (EDUCON)*. 2012, pp. 1–6. DOI: 10.1109/EDUCON.2012.6201145 (cit. on p. 49).

[162] V. Cantoni et al. "Exploiting Eye Tracking in Advanced E-Learning Systems". In: *Proceedings of the 13th International Conference on Computer Systems and Technologies*. CompSysTech '12. New York, NY, USA: Association for Computing Machinery, 2012, 376–383. ISBN: 9781450311939. DOI: 10.1145/2383276.2383331. URL: https://doi.org/10.1145/2383276.2383331 (cit. on p. 49).

[163] M. Ivanović et al. "Integration of Eye Tracking Technologies and Methods in an E-learning System". In: *Proceedings of the 8th Balkan Conference in Informatics*. BCI '17. New York, NY, USA: Association for Computing Machinery, 2017. ISBN: 9781450352857. DOI: 10.1145/3136273.3136278. URL: https://doi.org/10.1145/3136273.3136278 (cit. on p. 50).

[164] N. Nugrahaningsih, M. Porta, and A. Klašnja-Milićević. "Assessing Learning Styles through Eye Tracking for E-Learning Applications". In: *Computer Science and Information Systems* 18.4 (2021), pp. 1287–1309. ISSN: 24061018. DOI: 10.2298/CSIS201201035N (cit. on p. 50).

[165] C.-W. Kao et al. "The Integrated Gaze, Web and Object Tracking Techniques for the Web-Based E-Learning Platform". In: *Proceedings of 2013 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*. 2013, pp. 720–724. DOI: 10.1109/TALE.2013.6654531 (cit. on p. 50).

[166] S. Charoenpit and M. Ohkura. "Exploring Emotion in an E-Learning System Using Eye Tracking". In: *2014 IEEE Symposium on Computational Intelligence in Healthcare and e-health (CICARE)*. 2014, pp. 141–147. DOI: 10.1109/CICARE.2014.7007846 (cit. on p. 50).

[167] S. Liu, X. Tao, and Q. Gui. "Research on Emotional State in Online Learning by Eye Tracking Technology". In: *Proceedings of the 4th International Conference on Intelligent Information Processing*. ICIIP '19. New York, NY, USA: Association for Computing Machinery, 2020, 471–477. ISBN: 9781450361910. DOI: 10.1145/3378065.3378154. URL: https://doi.org/10.1145/3378065.3378154 (cit. on p. 50).

[168] I. Joe Louis Paul et al. *Eye Gaze Tracking-Based Adaptive E-learning for Enhancing Teaching and Learning in Virtual Classrooms*. Ed. by S. Fong, S. Akashe, and P. N. Mahalle. Singapore: Springer Singapore, 2019, pp. 165–176. ISBN: 978-981-13-0586-3 (cit. on p. 51).

[169] F. van der Sluis and E. L. van den Broek. *Feedback Beyond Accuracy: Using Eye-Tracking to Detect Comprehensibility and Interest During Reading*. 2023. DOI: https://doi.org/10.1002/asi.24657. eprint: https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.24657. URL: https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.24657 (cit. on p. 51).

[170] K. Nosu and T. Kurokawa. "Facial Tracking for an Emotion-Diagnosis Robot to Support E-Learning". In: *2006 International Conference on Machine Learning and Cybernetics*. 2006, pp. 3811–3816. DOI: 10.1109/ICMLC.2006.258689 (cit. on p. 51).

[171] T. Robal et al. "Webcam-based Attention Tracking in Online Learning: A Feasibility Study". In: IUI '18 (2018), 189–197. DOI: 10.1145/3172944.3172987. URL: https://doi.org/10.1145/3172944.3172987 (cit. on p. 51).

[172] A. Khurshid and J. Scharcanski. "An Adaptive Face Tracker with Application in Yawning Detection". In: *Sensors* 20.5 (2020). ISSN: 1424-8220. DOI: 10.3390/s20051494. URL: https://www.mdpi.com/1424-8220/20/5/1494 (cit. on p. 52).

[173] N. T. Ramaha et al. "Towards Webcam-Based Face Direction Tracking to Detect Learners' Attention within Asynchronous E-Learning Environment". In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 46.4/W5-2021 (2021), pp. 445–449. ISSN: 16821750. DOI: 10.5194/isprs-Archives-XLVI-4-W5-2021-445-2021 (cit. on p. 52).

[174] A. R. Khan et al. "EXECUTE: Exploring Eye Tracking to Support E-learning". In: *2022 IEEE Global Engineering Education Conference (EDUCON)*. 2022, pp. 670–676. DOI: 10.1109/EDUCON52537.2022.9766506 (cit. on pp. 52, 115).

[175] F. Pernkopf. "Tracking of Multiple Targets Using Online Learning for Reference Model Adaptation". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 38.6 (2008), pp. 1465–1475. DOI: 10.1109/TSMCB.2008.927281 (cit. on p. 52).

[176] Y. Bandung, K. Mutijarsa, and L. B. Subekti. "Design and Implementation of Video Conference System with Object Tracking for Distance Learning". In: *2017 International Symposium on Electronics and Smart Devices (ISESD)*. 2017, pp. 6–11. DOI: 10.1109/ISESD.2017.8253296 (cit. on p. 52).

[177] L. Zhao et al. "A Visual Tracker Offering More Solutions". In: *Sensors* 20.18 (2020). ISSN: 1424-8220. DOI: 10.3390/s20185374. URL: https://www.mdpi.com/1424-8220/20/18/5374 (cit. on p. 52).

[178] A. F. Abate et al. "Attention Monitoring for Synchronous Distance Learning". In: *Future Generation Computer Systems* 125 (2021), pp. 774–784. ISSN: 0167-739X. DOI: https://doi.org/10.1016/j.future.2021.07.026. URL: https://www.sciencedirect.com/science/article/pii/S0167739X21002880 (cit. on pp. 52, 106).

[179] D. I. Sutanto et al. "Auto-Tracking Camera System for Remote Learning Using Face Detection and Hand Gesture Recognition Based on Convolutional Neural Network". In: 1 (2021), pp. 451–457. DOI: 10.1109/ICCSAI53272.2021.9609744 (cit. on p. 52).

[180] A. van Drunen et al. "Exploring workload and attention measurements with uLog mouse data". In: *Behavior Research Methods* 41.3 (2009), pp. 868–875. ISSN: 1554-3528. DOI: 10.3758/BRM.41.3.868. URL: https://doi.org/10.3758/BRM.41.3.868 (cit. on p. 53).

[181] T. Mailhot et al. "Using a Wireless Electroencephalography Device to Evaluate E-Health and E-Learning Interventions". In: *Nursing Research* 67 (1 2018). ISSN: 0029-6562. URL: https://journals.lww.com/nursingresearchonline/fulltext/2018/01000/using_a_wireless_electroencephalography_device_to.7.aspx (cit. on p. 53).

[182] M. Taub et al. "Tracking Changes in Students' Online Self-Regulated Learning Behaviors and Achievement Goals Using Trace Clustering and Process Mining". In: *Frontiers in Psychology* 13 (2022). ISSN: 1664-1078. DOI: 10.3389/fpsyg.2022.813514. URL: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.813514 (cit. on p. 53).

[183] K. L. Wan et al. "Visual-Spatial Dimension Integration in Digital Pathology Education Enhances Anatomical Pathology Learning". In: *BMC Medical Education* 22 (1 2022), p. 587. ISSN: 1472-6920. DOI: 10.1186/s12909-022-03545-x. URL: https://doi.org/10.1186/s12909-022-03545-x (cit. on p. 53).

[184] I. Porter-Szucs and B. DeCicco. "TriHy: Teaching an MA TESOL Class Face-to-Face, Synchronously Online, and Asynchronously Online". In: *SN Social Sciences* 2.8 (2022). ISSN: 2662-9283. DOI: 10.1007/s43545-022-00434-4. URL: https://doi.org/10.1007/s43545-022-00434-4 (cit. on pp. 53, 55).

[185] G. M. Notaro and S. G. Diamond. "Development and Demonstration of an Integrated EEG, Eye-Tracking, and Behavioral Data Acquisition System to Assess Online Learning". In: *Proceedings of the 10th International Conference on Education Technology and Computers*. ICETC '18. New York, NY, USA: Association for Computing Machinery, 2018, 105–111. ISBN: 9781450365178. DOI: 10.1145/3290511.3290526. URL: https://doi.org/10.1145/3290511.3290526 (cit. on p. 54).

[186] B. A. Zardari et al. "QUEST E-Learning Portal: Applying Heuristic Evaluation, Usability Testing and Eye Tracking". In: *Universal Access in the Information Society* 20 (3 2021), pp. 531–543. ISSN: 1615-5297. DOI: 10.1007/s10209-020-00774-z. URL: https://doi.org/10.1007/s10209-020-00774-z (cit. on p. 54).

[187] E. A. Nelson. "E-Learning: A Practical Solution for Training and Tracking in Patient-Care Settings". In: *Nursing Administration Quarterly* 27.1 (2003), pp. 29–32. ISSN: 15505103. DOI: 10.1097/00006216-200301000-00007 (cit. on p. 54).

[188] A. Booth et al. "Applying Findings from a Systematic Review of Workplace-Based E-Learning: Implications for Health Information Professionals". In: *Health Information & Libraries Journal* 26.1 (2009), pp. 4–21. DOI: https://doi.org/10.1111/j.1471-1842.2008.00834.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1471-1842.2008.00834.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1471-1842.2008.00834.x (cit. on p. 54).

[189] J. Koch et al. "Nursing Students' Perception of a Web-Based Intervention to Support Learning". In: *Nurse Education Today* 30.6 (2010), pp. 584–590. ISSN: 0260-6917. DOI: https://doi.org/10.1016/j.nedt.2009.12.005. URL: https://www.sciencedirect.com/science/article/pii/S0260691709002354 (cit. on p. 55).

[190] F. Oeffner et al. "Interactive E-Learning Courses in Human Genetics: Usage and Evaluation by Science and Medical Students at the Faculty of Medicin". en. In: *GMS Z Med Ausbild* 28.3 (2011-08), Doc38 (cit. on p. 55).

[191] D. A. Cook et al. "Educational Technologies for Physician Continuous Professional Development: A National Survey". In: *Academic Medicine* 93.1 (2018), pp. 104–112. ISSN: 1938808X. DOI: 10.1097/ACM.0000000000001817 (cit. on p. 55).

[192] P. Sinnayah, A. Salcedo, and S. Rekhari. "Reimagining Physiology Education with Interactive Content Developed in H5P". In: *Advances in Physiology Education* 45.1 (2021), pp. 71–76. ISSN: 15221229. DOI: 10.1152/ADVAN.00021.2020 (cit. on p. 55).

[193] M. Co et al. "Distance Education for Anatomy and Surgical Training – A Systematic Review". In: *The Surgeon* 20.5 (2022), e195–e205. ISSN: 1479-666X. DOI: https://doi.org/10.1016/j.surge.2021.08.001. URL: https://www.sciencedirect.com/science/article/pii/S1479666X21001335 (cit. on p. 55).

[194] A. R. Shah et al. "Remote versus In-Person Health Education: Feasibility, Satisfaction, and Health Literacy for Diverse Older Adults". In: *Health Education & Behavior* 50.3 (2023). PMID: 36124443, pp. 369–381. DOI: 10.1177/10901981221121258. eprint: https://doi.org/10.1177/10901981221121258. URL: https://doi.org/10.1177/10901981221121258 (cit. on p. 55).

[195] L. Z. J. W. L. C. P. B. L. A. K. T. Danielle DAmico Iris Yusupov and S. Vandermorris. "Feasibility, Acceptability, and Impact of a Self-Guided E-Learning Memory and Brain Health Promotion Program for Healthy Older Adults". In: *Clinical Gerontologist* 47.1 (2024). PMID: 35713408, pp. 4–16. DOI: 10.1080/07317115.2022.2088325. eprint: https://doi.org/10.1080/07317115.2022.2088325. URL: https://doi.org/10.1080/07317115.2022.2088325 (cit. on p. 56).

[196] "Systematic Review of Research on Artificial Intelligence Applications in Higher Education – Where are the Educators?" In: *International Journal of Educational Technology in Higher Education* 16.1 (2019). ISSN: 23659440. DOI: 10.1186/s41239-019-0171-0 (cit. on p. 56).

[197] H. Wang et al. "Examining the Applications of Intelligent Tutoring Systems in Real Educational Contexts: A Systematic Literature Review from the Social Experiment Perspective". In: *Education and Information Technologies* 28 (7 2023), pp. 9113–9148. ISSN: 1573-7608. DOI: 10.1007/s10639-022-11555-x. URL: https://doi.org/10.1007/s10639-022-11555-x (cit. on pp. 56, 57).

[198] N. S. Raj and V. G. Renumol. "A Systematic Literature Review on Adaptive Content Recommenders in Personalized Learning Environments from 2015 to 2020". In: *Journal of Computers in Education* 9 (1 2022), pp. 113–148. ISSN: 2197-9995. DOI: 10.1007/s40692-021-00199-4. URL: https://doi.org/10.1007/s40692-021-00199-4 (cit. on p. 57).

[199] H. F. Silveira Gamboa. "Multi-Modal Behavioral Biometrics Based on HCI and Electrophysiology". PhD thesis. Universidade Técnica de Lisboa, 2008 (cit. on pp. 59, 68, 69).

[200] J. N. Acharya et al. "American Clinical Neurophysiology Society Guideline 2: Guidelines for Standard Electrode Position Nomenclature". In: *Journal of Clinical Neurophysiology* 33.4 (2016). ISSN: 0736-0258. DOI: 10.1097/WNP.0000000000000316. URL: https://journals.lww.com/clinicalneurophys/Fulltext/2016/08000/American_Clinical_Neurophysiology_Society.4.aspx (cit. on p. 60).

[201] P. Gamboa et al. "Attention Classification Based on Biosignals during Standard Cognitive Tasks for Occupational Domains". In: *Computers* 11.4 (2022). ISSN: 2073-431X. DOI: 10.3390/computers11040049. URL: https://www.mdpi.com/2073-431X/11/4/49 (cit. on pp. 61, 87, 91, 107).

[202] J. Peirce et al. "PsychoPy2: Experiments in Behavior Made Easy". In: *Behavior Research Methods* 51.1 (2019), pp. 195–203. ISSN: 1554-3528. DOI: 10.3758/s13428-018-01193-y. URL: https://doi.org/10.3758/s13428-018-01193-y (cit. on p. 63).

[203] I. Zyma et al. "Electroencephalograms During Mental Arithmetic Task Performance". In: *Data* 4.1 (2019). ISSN: 2306-5729. DOI: 10.3390/data4010014. URL: https://www.mdpi.com/2306-5729/4/1/14 (cit. on p. 63).

[204] A. P. L. J. K. Roy P. C. Kessels Martine J. E. van Zandvoort and E. H. F. de Haan. "The Corsi Block-Tapping Task: Standardization and Normative Data". In: *Applied Neuropsychology* 7.4 (2000). PMID: 11296689, pp. 252–258. DOI: 10.1207/S15324826AN0704_8. eprint: https://doi.org/10.1207/S15324826AN0704_8. URL: https://doi.org/10.1207/S15324826AN0704_8 (cit. on p. 68).

[205] R. Varandas et al. "Automatic Cognitive Fatigue Detection Using Wearable fNIRS and Machine Learning". In: *Sensors* 22.11 (2022). ISSN: 1424-8220. DOI: 10.3390/s22114010. URL: https://www.mdpi.com/1424-8220/22/11/4010 (cit. on pp. 68, 69, 87).

[206] A. Radford et al. "Language Models are Unsupervised Multitask Learners". In: (2019). URL: https://api.semanticscholar.org/CorpusID:160025533 (cit. on p. 70).

[207] S. Matcher et al. "Performance Comparison of Several Published Tissue Near-Infrared Spectroscopy Algorithms". In: *Analytical Biochemistry* 227.1 (1995), pp. 54–68. ISSN: 0003-2697. DOI: https://doi.org/10.1006/abio.1995.1252. URL: https://www.sciencedirect.com/science/article/pii/S0003269785712523 (cit. on p. 72).

[208] M. A. Yücel et al. "Best Practices for fNIRS Publications". In: *Neurophotonics* 8.1 (2021), p. 012101. DOI: 10.1117/1.NPh.8.1.012101. URL: https://doi.org/10.1117/1.NPh.8.1.012101 (cit. on pp. 72, 91).

[209] S. D'Mello and A. Graesser. "Dynamics of Affective States During Complex Learning". In: *Learning and Instruction* 22.2 (2012), pp. 145–157. ISSN: 0959-4752. DOI: https://doi.org/10.1016/j.learninstruc.2011.10.001. URL: https://www.sciencedirect.com/science/article/pii/S0959475211000806 (cit. on pp. 72, 109).

[210] J. R. Quinlan. "Induction of decision trees". In: *Machine Learning* 1.1 (1986), pp. 81–106. ISSN: 1573-0565. DOI: 10.1007/BF00116251. URL: https://doi.org/10.1007/BF00116251 (cit. on p. 76).

[211] L. Breiman. "Random Forests". In: *Machine Learning* 45.1 (2001), pp. 5–32. ISSN: 1573-0565. DOI: 10.1023/A:1010933404324. URL: https://doi.org/10.1023/A:1010933404324 (cit. on p. 76).

[212] P. Marius-Constantin et al. "Multilayer perceptron and neural networks". In: *WSEAS Transactions on Circuits and Systems* 8.7 (2009), pp. 579–588. ISSN: 11092734 (cit. on p. 76).

[213] Y. Freund and R. E. Schapire. "A Short Introduction to Boosting". In: *Journal of Japanese Society for Artificial Intelligence* 14.5 (1999), pp. 771–780 (cit. on p. 76).

[214] Rish I. "An empirical study of the naive bayes classifier". In: *IJCAI 2001 workshop on empirical methods in artificial intelligence* January 2001 (2001), pp. 41–46. URL: http://www.cc.gatech.edu/home/isbell/classes/reading/papers/Rish.pdf (cit. on p. 76).

[215] P. Xanthopoulos, P. M. Pardalos, and T. B. Trafalis. "Linear Discriminant Analysis BT - Robust Data Mining". In: ed. by P. Xanthopoulos, P. M. Pardalos, and T. B. Trafalis. New York, NY: Springer New York, 2013, pp. 27–33. ISBN: 978-1-4419-9878-1. DOI: 10.1007/978-1-4419-9878-1_4. URL: https://doi.org/10.1007/978-1-4419-9878-1_4 (cit. on p. 76).

[216] C. E. Rasmussen. "Gaussian Processes in Machine Learning". In: ed. by O. Bousquet, U. von Luxburg, and G. Rätsch. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 63–71. ISBN: 978-3-540-28650-9. DOI: 10.1007/978-3-540-28650-9_4. URL: https://doi.org/10.1007/978-3-540-28650-9_4 (cit. on p. 76).

[217] P. Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17 (3 2020), pp. 261–272. ISSN: 1548-7105. DOI: 10.1038/s41592-019-0686-2. URL: https://doi.org/10.1038/s41592-019-0686-2 (cit. on p. 77).

[218] J. Pan and W. J. Tompkins. "A Real-Time QRS Detection Algorithm". In: *IEEE Transactions on Biomedical Engineering* BME-32.3 (1985), pp. 230–236. DOI: 10.1109/TBME.1985.325532 (cit. on p. 79).

[219] J. Piskorski and P. Guzik. "Filtering Poincaré Plots". In: *Computational Methods in Science and Technology* 11.1 (2005), pp. 39–48. ISSN: 15050602. DOI: 10.12921/cmst.2005.11.01.39-48 (cit. on p. 80).

[220] D. Makowski et al. "NeuroKit2: A Python Toolbox for Neurophysiological Signal Processing". In: *Behavior Research Methods* 53.4 (2021-02), pp. 1689–1696. DOI: 10.3758/s13428-020-01516-y. URL: https://doi.org/10.3758%2Fs13428-020-01516-y (cit. on p. 81).

[221] M. Barandas et al. "TSFEL: Time Series Feature Extraction Library". In: *SoftwareX* 11 (2020), p. 100456. ISSN: 2352-7110. DOI: https://doi.org/10.1016/j.softx.2020.100456. URL: https://www.sciencedirect.com/science/article/pii/S2352711020300017 (cit. on pp. 84, 87, 91).

[222] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cit. on pp. 87, 91).

[223] M. Galar et al. "EUSBoost: Enhancing Ensembles for Highly Imbalanced Data-Sets by Evolutionary Undersampling". In: *Pattern Recognition* 46.12 (2013), pp. 3460–3471. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2013.05.006. URL: https://www.sciencedirect.com/science/article/pii/S0031320313002100 (cit. on p. 87).

[224] L. Breiman. "Random Forests". In: *Machine Learning* 45 (1 2001), pp. 5–32. ISSN: 1573-0565. DOI: 10.1023/A:1010933404324. URL: https://doi.org/10.1023/A:1010933404324 (cit. on p. 88).

[225] X. Hu et al. "fNIRS Evidence for Recognizably Different Positive Emotions". In: *Frontiers in Human Neuroscience* 13 (2019). ISSN: 1662-5161. DOI: 10.3389/fnhum.2019.00120. URL: https://www.frontiersin.org/articles/10.3389/fnhum.2019.00120 (cit. on p. 91).

[226] A. A. Hagberg, D. A. Schult, and P. J. Swart. "Exploring Network Structure, Dynamics, and Function using NetworkX". In: *Proceedings of the 7th Python in Science Conference*. Ed. by G. Varoquaux, T. Vaught, and J. Millman. Pasadena, CA USA, 2008, pp. 11 –15 (cit. on p. 94).

[227] M. Yung, R. Manji, and R. P. Wells. "Exploring the Relationship of Task Performance and Physical and Cognitive Fatigue During a Daylong Light Precision Task". In: *Human Factors* 59.7 (2017). PMID: 28658591, pp. 1029–1047. DOI: 10.1177/0018720817717026. eprint: https://doi.org/10.1177/0018720817717026. URL: https://doi.org/10.1177/0018720817717026 (cit. on p. 107).

[228] J. A. DeFalco et al. "Detecting and Addressing Frustration in a Serious Game for Military Training". In: *International Journal of Artificial Intelligence in Education* 28 (2 2018), pp. 152–193. ISSN: 1560-4306. DOI: 10.1007/s40593-017-0152-1. URL: https://doi.org/10.1007/s40593-017-0152-1 (cit. on p. 109).

[229] P. Gamboa et al. "Design of an Attention Tool Using HCI and Work-Related Variables". In: *Technological Innovation for Applied AI Systems*. Ed. by L. M. Camarinha-Matos, P. Ferreira, and G. Brito. Cham: Springer International Publishing, 2021, pp. 262–269. ISBN: 978-3-030-78288-7 (cit. on p. 118).

[230] C. Tai. *Under AI's Watchful Eye, China Wants to Raise Smarter Students*. 2019. URL: https://www.wsj.com/video/series/in-depth-features/under-ais-watchful-eye-china-wants-to-raise-smarter-students/C4294BAB-A76B-4569-8D09-32E9F2B62D19?mod=Searchresults_pos3\\&page=1 (cit. on p. 120).

[231] European Parliament and Council of the European Union. *Regulation (EU) 2016/679 of the European Parliament and of the Council*. of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). 2016-05-04. URL: https://data.europa.eu/eli/reg/2016/679/oj (visited on 2023-04-13) (cit. on p. 120).

I

Annex 1 - Questionnaires

The initial questionnaire performed in the beginning in the three studies is shown in Figure I.1.

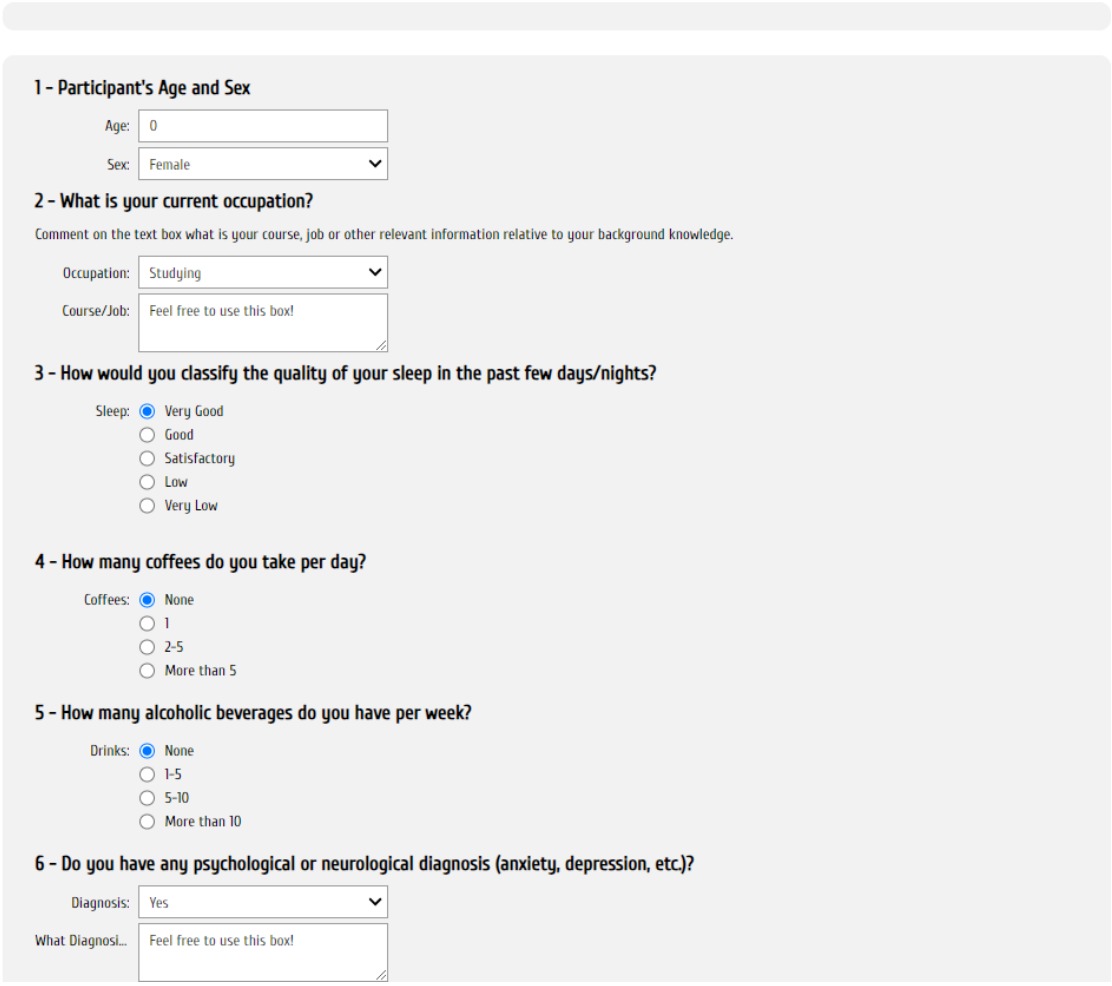

Figure I.1: Initial Questionnaire.

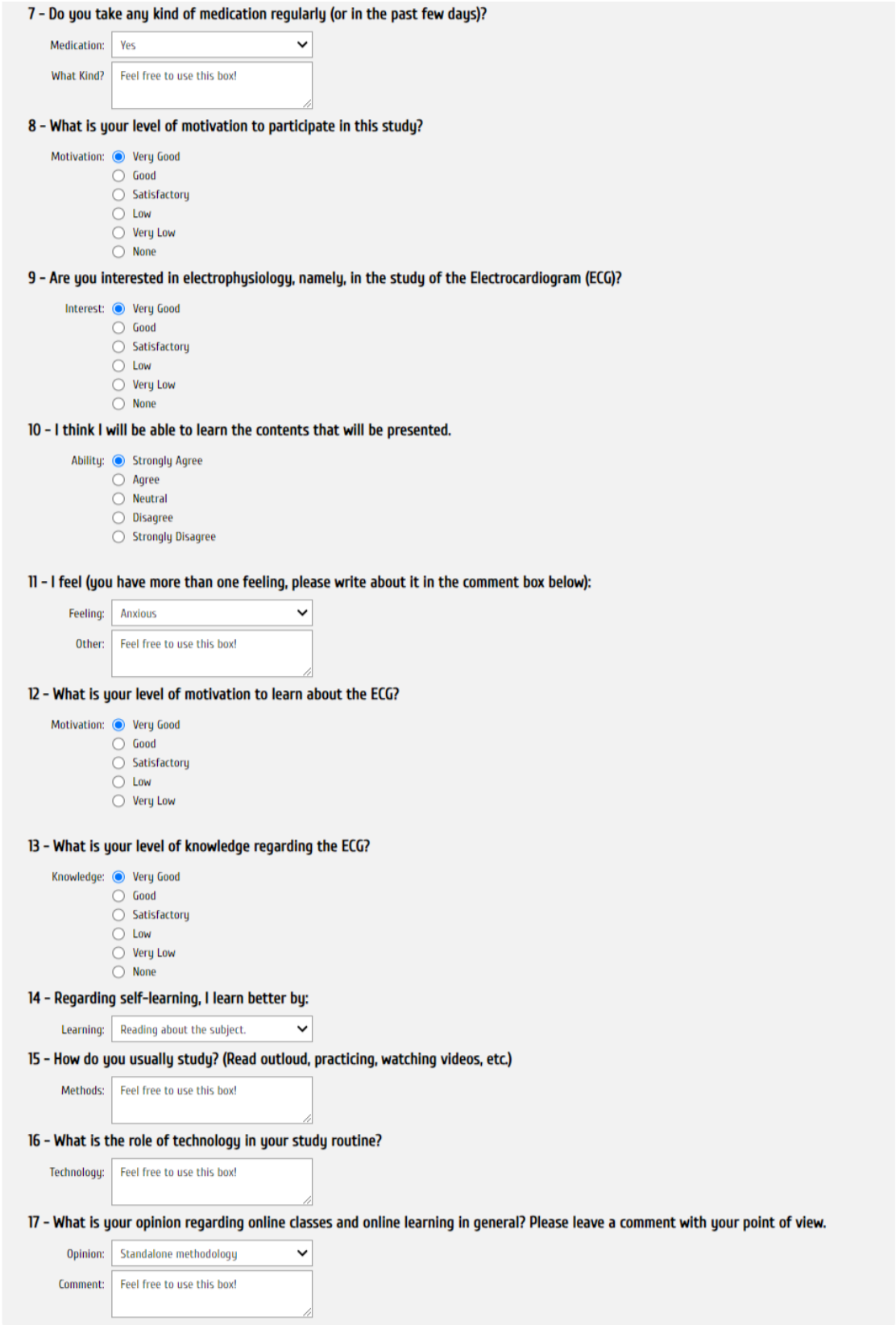

Figure I.1: Initial Questionnaire (continuation).

The final questionnaire performed in the end of the three studies is shown in Figure I.2.

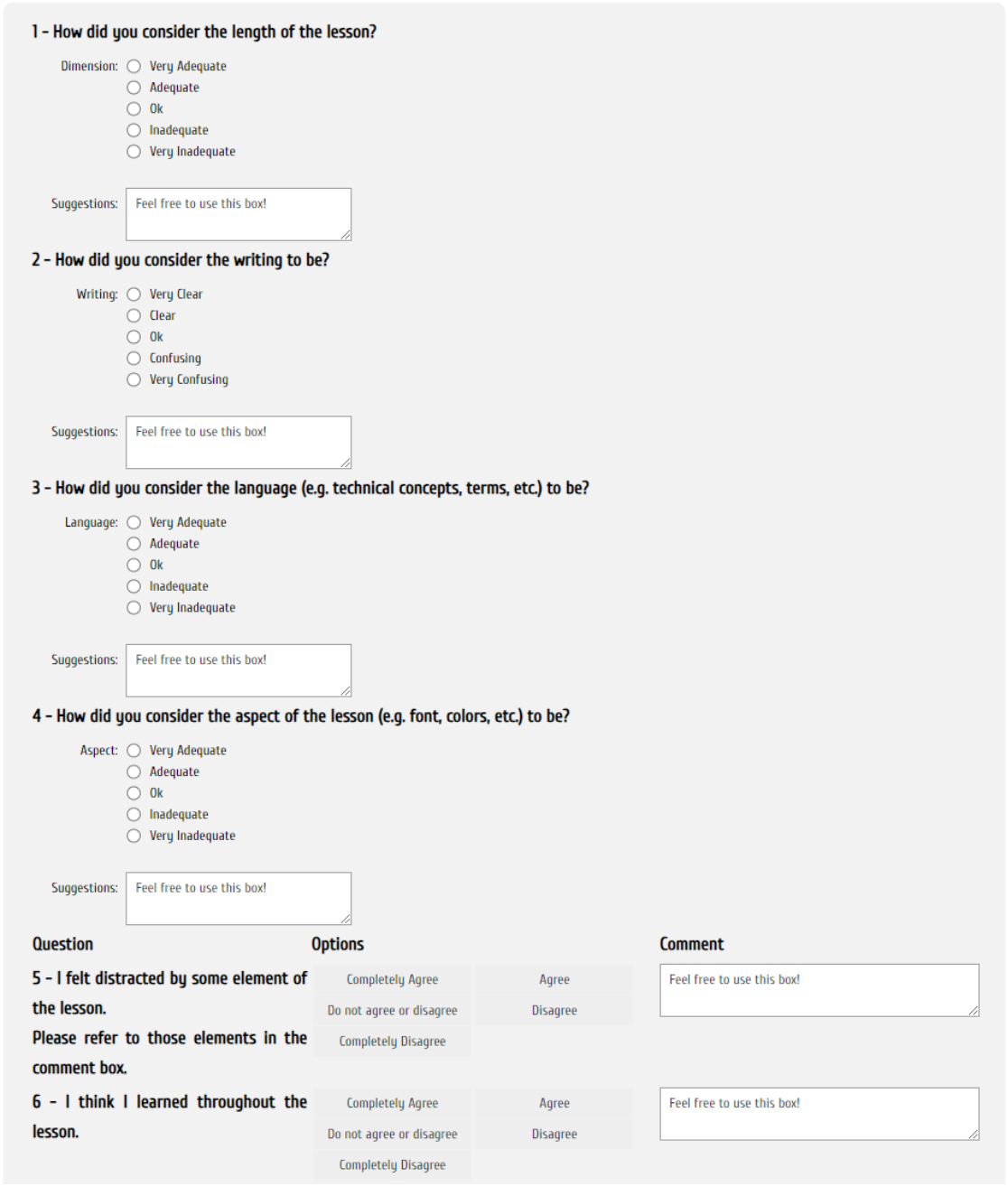

Figure I.2: Final Questionnaire.

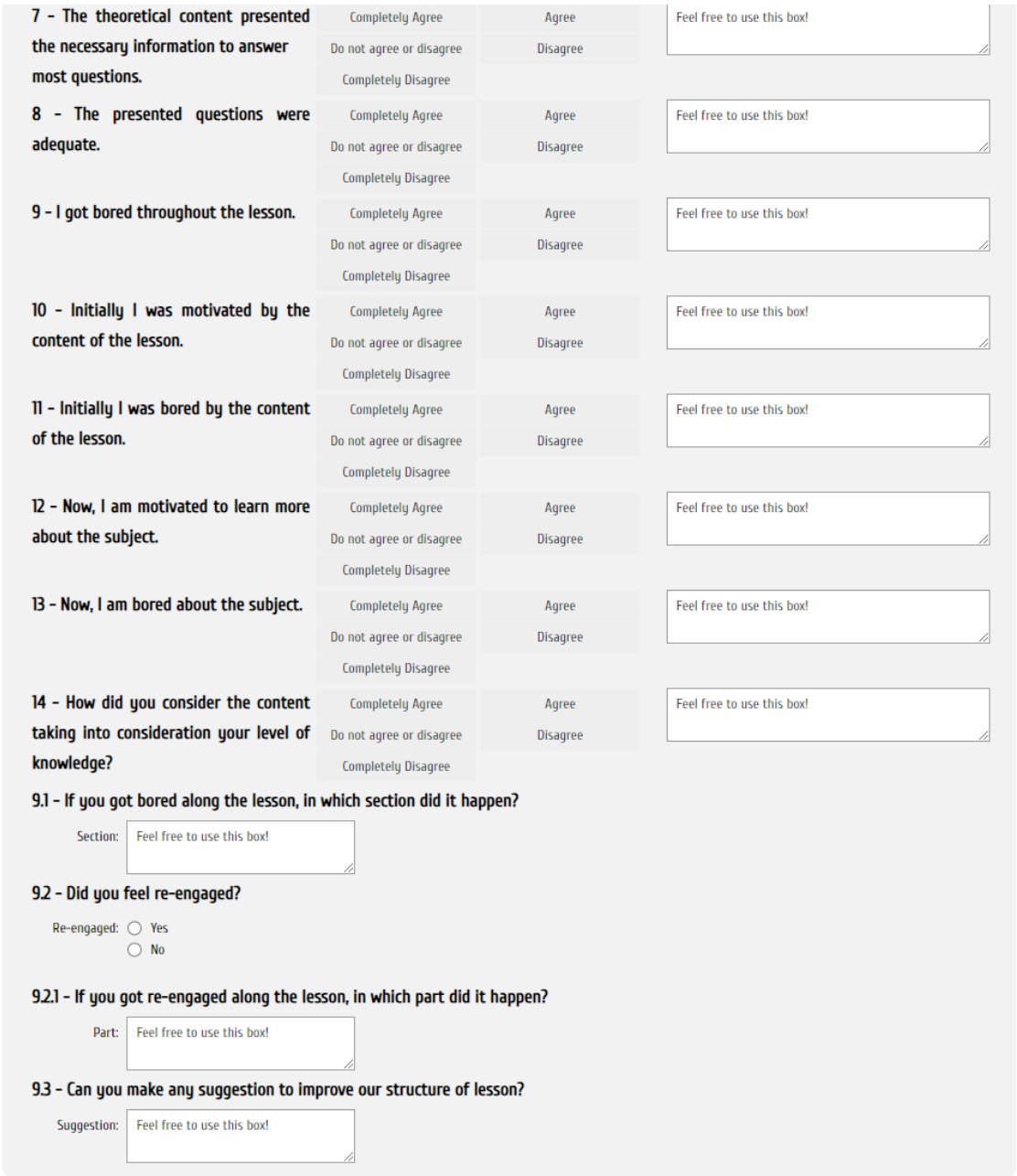

Figure I.2: Final Questionnaire (continuation).

II

Results All Sources

Table II.1: Results of classification considering the combinations of all data sources taking into account the Binary paradigm.

F1-Score	Accuracy	Precision	Recall	ROC-AUC	Classifier	Norm	Window (s)
0.73	0.73	0.74	0.73	0.54	DT	Raw	10
0.75	0.83	0.73	0.83	0.56	RF	Raw	10
0.46	0.41	0.74	0.41	0.52	MLP	Raw	10
0.77	0.80	0.76	0.80	0.62	AB	Raw	10
0.70	0.68	0.71	0.68	0.46	GNB	Raw	10
0.58	0.52	0.72	0.52	0.50	LDA	Raw	10
0.77	0.80	0.76	0.80	0.59	MC	Raw	10
0.05	0.17	0.03	0.17	0.50	GP	Raw	10
0.63	0.63	0.64	0.63	0.51	DT	Min-Max	10
0.67	0.75	0.68	0.75	0.54	RF	Min-Max	10
0.67	0.69	0.65	0.69	0.58	MLP	Min-Max	10
0.66	0.68	0.65	0.68	0.56	AB	Min-Max	10
0.64	0.62	0.67	0.62	0.57	GNB	Min-Max	10
0.66	0.67	0.64	0.67	0.54	LDA	Min-Max	10
0.67	0.72	0.65	0.72	0.58	MC	Min-Max	10
0.63	0.63	0.64	0.63	0.49	GP	Min-Max	10
0.64	0.63	0.64	0.63	0.51	DT	Z-Score	10
0.67	0.75	0.68	0.75	0.55	RF	Z-Score	10
0.68	0.70	0.66	0.70	0.60	MLP	Z-Score	10
0.67	0.69	0.65	0.69	0.56	AB	Z-Score	10
0.64	0.62	0.67	0.62	0.57	GNB	Z-Score	10
0.66	0.67	0.65	0.67	0.54	LDA	Z-Score	10
0.68	0.74	0.67	0.74	0.59	MC	Z-Score	10
0.62	0.60	0.64	0.60	0.50	GP	Z-Score	10

Table II.1 continued from previous page

F1-Score	Accuracy	Precision	Recall	ROC-AUC	Classifier	Norm	Window (s)
0.70	0.69	0.71	0.69	0.49	DT	Raw	12
0.75	0.82	0.71	0.82	0.47	RF	Raw	12
0.69	0.68	0.69	0.68	0.47	MLP	Raw	12
0.75	0.78	0.74	0.78	0.60	AB	Raw	12
0.51	0.45	0.72	0.45	0.52	GNB	Raw	12
0.64	0.60	0.71	0.60	0.50	LDA	Raw	12
0.75	0.81	0.71	0.81	0.52	MC	Raw	12
0.05	0.17	0.03	0.17	0.50	GP	Raw	12
0.65	0.64	0.65	0.64	0.52	DT	Min-Max	12
0.67	0.74	0.66	0.74	0.62	RF	Min-Max	12
0.68	0.70	0.67	0.70	0.62	MLP	Min-Max	12
0.67	0.69	0.65	0.69	0.56	AB	Min-Max	12
0.69	0.68	0.69	0.68	0.60	GNB	Min-Max	12
0.68	0.69	0.66	0.69	0.58	LDA	Min-Max	12
0.68	0.73	0.66	0.73	0.62	MC	Min-Max	12
0.68	0.69	0.68	0.69	0.55	GP	Min-Max	12
0.65	0.65	0.65	0.65	0.53	DT	Z-Score	12
0.68	0.75	0.68	0.75	0.62	RF	Z-Score	12
0.70	0.71	0.69	0.71	0.67	MLP	Z-Score	12
0.68	0.70	0.66	0.70	0.57	AB	Z-Score	12
0.69	0.68	0.70	0.68	0.60	GNB	Z-Score	12
0.68	0.69	0.66	0.69	0.58	LDA	Z-Score	12
0.69	0.74	0.68	0.74	0.65	MC	Z-Score	12
0.62	0.60	0.66	0.60	0.50	GP	Z-Score	12
0.71	0.71	0.71	0.71	0.51	DT	Raw	20
0.74	0.82	0.70	0.82	0.53	RF	Raw	20
0.67	0.65	0.68	0.65	0.44	MLP	Raw	20
0.75	0.79	0.73	0.79	0.50	AB	Raw	20
0.55	0.49	0.73	0.49	0.54	GNB	Raw	20
0.66	0.62	0.70	0.62	0.46	LDA	Raw	20
0.74	0.81	0.72	0.81	0.47	MC	Raw	20
0.05	0.18	0.03	0.18	0.50	GP	Raw	20
0.63	0.63	0.64	0.63	0.51	DT	Min-Max	20
0.65	0.74	0.61	0.74	0.53	RF	Min-Max	20
0.67	0.69	0.65	0.69	0.60	MLP	Min-Max	20
0.66	0.70	0.65	0.70	0.49	AB	Min-Max	20
0.65	0.66	0.64	0.66	0.52	GNB	Min-Max	20
0.67	0.71	0.65	0.71	0.58	LDA	Min-Max	20

Table II.1 continued from previous page

F1-Score	Accuracy	Precision	Recall	ROC-AUC	Classifier	Norm	Window (s)
0.66	0.73	0.65	0.73	0.55	MC	Min-Max	20
0.71	0.70	0.71	0.70	0.63	GP	Min-Max	20
0.63	0.62	0.63	0.62	0.50	DT	Z-Score	20
0.65	0.74	0.60	0.74	0.52	RF	Z-Score	20
0.69	0.70	0.68	0.70	0.63	MLP	Z-Score	20
0.66	0.70	0.64	0.70	0.50	AB	Z-Score	20
0.65	0.66	0.64	0.66	0.52	GNB	Z-Score	20
0.67	0.71	0.65	0.71	0.57	LDA	Z-Score	20
0.67	0.73	0.65	0.73	0.58	MC	Z-Score	20
0.65	0.62	0.70	0.62	0.50	GP	Z-Score	20
0.69	0.67	0.70	0.67	0.48	DT	Raw	30
0.74	0.82	0.68	0.82	0.45	RF	Raw	30
0.65	0.64	0.68	0.64	0.44	MLP	Raw	30
0.72	0.74	0.70	0.74	0.49	AB	Raw	30
0.65	0.62	0.68	0.62	0.48	GNB	Raw	30
0.57	0.51	0.70	0.51	0.49	LDA	Raw	30
0.74	0.80	0.69	0.80	0.41	MC	Raw	30
0.05	0.18	0.03	0.18	0.50	GP	Raw	30
0.66	0.66	0.67	0.66	0.52	DT	Min-Max	30
0.70	0.77	0.70	0.77	0.52	RF	Min-Max	30
0.67	0.70	0.65	0.70	0.50	MLP	Min-Max	30
0.68	0.70	0.66	0.70	0.47	AB	Min-Max	30
0.71	0.75	0.70	0.75	0.55	GNB	Min-Max	30
0.73	0.77	0.73	0.77	0.56	LDA	Min-Max	30
0.69	0.75	0.67	0.75	0.51	MC	Min-Max	30
0.65	0.65	0.66	0.65	0.47	GP	Min-Max	30
0.66	0.66	0.66	0.66	0.51	DT	Z-Score	30
0.70	0.77	0.70	0.77	0.51	RF	Z-Score	30
0.66	0.67	0.66	0.67	0.52	MLP	Z-Score	30
0.67	0.69	0.65	0.69	0.46	AB	Z-Score	30
0.70	0.74	0.69	0.74	0.54	GNB	Z-Score	30
0.73	0.77	0.72	0.77	0.56	LDA	Z-Score	30
0.68	0.73	0.65	0.73	0.52	MC	Z-Score	30
0.51	0.46	0.62	0.46	0.50	GP	Z-Score	30
0.69	0.70	0.69	0.70	0.48	DT	Raw	40
0.73	0.81	0.67	0.81	0.53	RF	Raw	40
0.78	0.78	0.79	0.78	0.66	MLP	Raw	40
0.69	0.70	0.68	0.70	0.39	AB	Raw	40

Table II.1 continued from previous page

F1-Score	Accuracy	Precision	Recall	ROC-AUC	Classifier	Norm	Window (s)
0.68	0.66	0.71	0.66	0.52	GNB	Raw	40
0.64	0.61	0.69	0.61	0.48	LDA	Raw	40
0.74	0.80	0.71	0.80	0.63	MC	Raw	40
0.06	0.18	0.03	0.18	0.50	GP	Raw	40
0.60	0.59	0.61	0.59	0.46	DT	Min-Max	40
0.66	0.74	0.62	0.74	0.41	RF	Min-Max	40
0.64	0.70	0.61	0.70	0.49	MLP	Min-Max	40
0.62	0.64	0.61	0.64	0.40	AB	Min-Max	40
0.66	0.70	0.64	0.70	0.51	GNB	Min-Max	40
0.66	0.73	0.64	0.73	0.49	LDA	Min-Max	40
0.65	0.72	0.61	0.72	0.42	MC	Min-Max	40
0.61	0.64	0.59	0.64	0.54	GP	Min-Max	40
0.61	0.61	0.61	0.61	0.47	DT	Z-Score	40
0.65	0.74	0.61	0.74	0.42	RF	Z-Score	40
0.63	0.66	0.62	0.66	0.48	MLP	Z-Score	40
0.62	0.64	0.61	0.64	0.40	AB	Z-Score	40
0.66	0.70	0.64	0.70	0.51	GNB	Z-Score	40
0.66	0.73	0.64	0.73	0.49	LDA	Z-Score	40
0.65	0.72	0.62	0.72	0.43	MC	Z-Score	40
0.53	0.49	0.62	0.49	0.50	GP	Z-Score	40
0.74	0.74	0.75	0.74	0.56	DT	Raw	50
0.76	0.83	0.69	0.83	0.56	RF	Raw	50
0.69	0.67	0.71	0.67	0.47	MLP	Raw	50
0.74	0.74	0.75	0.74	0.53	AB	Raw	50
0.75	0.75	0.76	0.75	0.56	GNB	Raw	50
0.64	0.59	0.72	0.59	0.50	LDA	Raw	50
0.75	0.81	0.72	0.81	0.54	MC	Raw	50
0.05	0.17	0.03	0.17	0.50	GP	Raw	50
0.66	0.66	0.66	0.66	0.50	DT	Min-Max	50
0.70	0.77	0.68	0.77	0.50	RF	Min-Max	50
0.66	0.70	0.63	0.70	0.41	MLP	Min-Max	50
0.67	0.68	0.66	0.68	0.41	AB	Min-Max	50
0.71	0.77	0.70	0.77	0.41	GNB	Min-Max	50
0.73	0.78	0.68	0.78	0.28	LDA	Min-Max	50
0.73	0.78	0.68	0.78	0.18	MC	Min-Max	50
0.65	0.64	0.67	0.64	0.52	GP	Min-Max	50
0.66	0.65	0.66	0.65	0.54	DT	Z-Score	50
0.69	0.76	0.66	0.76	0.54	RF	Z-Score	50

Table II.1 continued from previous page

F1-Score	Accuracy	Precision	Recall	ROC-AUC	Classifier	Norm	Window (s)
0.64	0.66	0.63	0.66	0.46	MLP	Z-Score	50
0.66	0.67	0.65	0.67	0.46	AB	Z-Score	50
0.72	0.77	0.71	0.77	0.46	GNB	Z-Score	50
0.73	0.78	0.68	0.78	0.28	LDA	Z-Score	50
0.71	0.76	0.68	0.76	0.17	MC	Z-Score	50
0.36	0.30	0.59	0.30	0.50	GP	Z-Score	50
0.66	0.65	0.66	0.65	0.33	DT	Raw	60
0.72	0.80	0.66	0.80	0.33	RF	Raw	60
0.69	0.68	0.69	0.68	0.50	MLP	Raw	60
0.67	0.67	0.67	0.67	0.50	AB	Raw	60
0.67	0.71	0.64	0.71	0.50	GNB	Raw	60
0.59	0.54	0.71	0.54	0.54	LDA	Raw	60
0.72	0.78	0.69	0.78	0.43	MC	Raw	60
0.06	0.18	0.03	0.18	0.50	GP	Raw	60
0.62	0.62	0.62	0.62	0.36	DT	Min-Max	60
0.63	0.70	0.60	0.70	0.36	RF	Min-Max	60
0.66	0.70	0.64	0.70	0.55	MLP	Min-Max	60
0.64	0.65	0.63	0.65	0.55	AB	Min-Max	60
0.65	0.71	0.63	0.71	0.55	GNB	Min-Max	60
0.63	0.68	0.62	0.68	0.56	LDA	Min-Max	60
0.64	0.68	0.62	0.68	0.52	MC	Min-Max	60
0.58	0.59	0.57	0.59	0.48	GP	Min-Max	60
0.63	0.62	0.63	0.62	0.38	DT	Z-Score	60
0.64	0.70	0.61	0.70	0.38	RF	Z-Score	60
0.66	0.69	0.65	0.69	0.58	MLP	Z-Score	60
0.64	0.65	0.63	0.65	0.58	AB	Z-Score	60
0.66	0.72	0.65	0.72	0.58	GNB	Z-Score	60
0.64	0.68	0.62	0.68	0.56	LDA	Z-Score	60
0.64	0.69	0.63	0.69	0.54	MC	Z-Score	60
0.36	0.34	0.51	0.34	0.50	GP	Z-Score	60
0.62	0.60	0.63	0.60	0.50	DT	Raw	90
0.70	0.79	0.63	0.79	0.50	RF	Raw	90
0.63	0.61	0.65	0.61	0.54	MLP	Raw	90
0.61	0.59	0.63	0.59	0.54	AB	Raw	90
0.72	0.78	0.71	0.78	0.54	GNB	Raw	90
0.68	0.78	0.60	0.78	0.55	LDA	Raw	90
0.68	0.78	0.60	0.78	0.47	MC	Raw	90
0.08	0.22	0.05	0.22	0.5	GP	Raw	90

Table II.1 continued from previous page

F1-Score	Accuracy	Precision	Recall	ROC-AUC	Classifier	Norm	Window (s)
0.65	0.64	0.66	0.64	NaN	DT	Min-Max	90
0.63	0.72	0.57	0.72	NaN	RF	Min-Max	90
0.68	0.71	0.66	0.71	0.57	MLP	Min-Max	90
0.68	0.67	0.69	0.67	NaN	AB	Min-Max	90
0.62	0.69	0.56	0.69	NaN	GNB	Min-Max	90
0.56	0.62	0.51	0.62	0.41	LDA	Min-Max	90
0.59	0.68	0.53	0.68	0.92	MC	Min-Max	90
0.72	0.71	0.84	0.71	0.5	GP	Min-Max	90
0.65	0.64	0.67	0.64	NaN	DT	Z-Score	90
0.73	0.72	0.74	0.72	0.67	MLP	Z-Score	90
0.55	0.60	0.51	0.60	0.38	LDA	Z-Score	90
0.87	0.87	0.87	0.87	0.94	MC	Z-Score	90
0.11	0.26	0.07	0.26	0.5	GP	Z-Score	90

Table II.2: Results of classification considering the combinations of all data sources taking into account the Cascade paradigm.

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.76	0.76	0.76	0.76	DT	Raw	10
0.66	0.72	0.69	0.72	RF	Raw	10
0.54	0.55	0.53	0.55	MLP	Raw	10
0.78	0.79	0.78	0.79	AB	Raw	10
0.27	0.29	0.53	0.29	GNB	Raw	10
0.43	0.39	0.54	0.39	LDA	Raw	10
0.69	0.74	0.72	0.74	MC	Raw	10
0.06	0.17	0.47	0.17	GP	Raw	10
0.62	0.62	0.62	0.62	DT	Min-Max	10
0.62	0.65	0.66	0.65	RF	Min-Max	10
0.59	0.60	0.59	0.60	MLP	Min-Max	10
0.52	0.55	0.52	0.55	AB	Min-Max	10
0.51	0.54	0.51	0.54	GNB	Min-Max	10
0.54	0.56	0.54	0.56	LDA	Min-Max	10
0.60	0.63	0.63	0.63	MC	Min-Max	10
0.46	0.48	0.51	0.48	GP	Min-Max	10
0.62	0.62	0.62	0.62	DT	Z-Score	10
0.62	0.65	0.67	0.65	RF	Z-Score	10

Table II.2 continued from previous page

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.60	0.61	0.60	0.61	MLP	Z-Score	10
0.52	0.55	0.52	0.55	AB	Z-Score	10
0.52	0.54	0.52	0.54	GNB	Z-Score	10
0.54	0.55	0.53	0.55	LDA	Z-Score	10
0.60	0.64	0.63	0.64	MC	Z-Score	10
0.08	0.22	0.05	0.22	GP	Z-Score	10
0.74	0.74	0.74	0.74	DT	Raw	12
0.64	0.71	0.68	0.71	RF	Raw	12
0.53	0.53	0.53	0.53	MLP	Raw	12
0.75	0.77	0.74	0.77	AB	Raw	12
0.38	0.37	0.58	0.37	GNB	Raw	12
0.46	0.43	0.54	0.43	LDA	Raw	12
0.67	0.72	0.70	0.72	MC	Raw	12
0.06	0.17	0.47	0.17	GP	Raw	12
0.57	0.57	0.57	0.57	DT	Min-Max	12
0.61	0.64	0.68	0.64	RF	Min-Max	12
0.58	0.60	0.58	0.60	MLP	Min-Max	12
0.50	0.52	0.50	0.52	AB	Min-Max	12
0.53	0.57	0.53	0.57	GNB	Min-Max	12
0.53	0.55	0.53	0.55	LDA	Min-Max	12
0.59	0.63	0.64	0.63	MC	Min-Max	12
0.44	0.45	0.46	0.45	GP	Min-Max	12
0.57	0.57	0.57	0.57	DT	Z-Score	12
0.60	0.64	0.67	0.64	RF	Z-Score	12
0.57	0.59	0.58	0.59	MLP	Z-Score	12
0.50	0.52	0.50	0.52	AB	Z-Score	12
0.54	0.57	0.53	0.57	GNB	Z-Score	12
0.53	0.55	0.53	0.55	LDA	Z-Score	12
0.58	0.62	0.63	0.62	MC	Z-Score	12
0.08	0.21	0.05	0.21	GP	Z-Score	12
0.64	0.64	0.64	0.64	DT	Raw	20
0.60	0.69	0.64	0.69	RF	Raw	20
0.55	0.54	0.55	0.54	MLP	Raw	20
0.65	0.68	0.64	0.68	AB	Raw	20
0.46	0.44	0.62	0.44	GNB	Raw	20
0.38	0.35	0.52	0.35	LDA	Raw	20
0.63	0.70	0.66	0.70	MC	Raw	20
0.12	0.20	0.52	0.20	GP	Raw	20

Table II.2 continued from previous page

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.56	0.56	0.56	0.56	DT	Min-Max	20
0.59	0.63	0.60	0.63	RF	Min-Max	20
0.55	0.58	0.56	0.58	MLP	Min-Max	20
0.50	0.53	0.49	0.53	AB	Min-Max	20
0.44	0.50	0.42	0.50	GNB	Min-Max	20
0.49	0.54	0.49	0.54	LDA	Min-Max	20
0.57	0.61	0.59	0.61	MC	Min-Max	20
0.42	0.42	0.53	0.42	GP	Min-Max	20
0.55	0.56	0.55	0.56	DT	Z-Score	20
0.58	0.63	0.60	0.63	RF	Z-Score	20
0.55	0.58	0.56	0.58	MLP	Z-Score	20
0.51	0.53	0.50	0.53	AB	Z-Score	20
0.45	0.51	0.45	0.51	GNB	Z-Score	20
0.49	0.54	0.49	0.54	LDA	Z-Score	20
0.57	0.62	0.60	0.62	MC	Z-Score	20
0.08	0.22	0.26	0.22	GP	Z-Score	20
0.58	0.58	0.58	0.58	DT	Raw	30
0.59	0.69	0.57	0.69	RF	Raw	30
0.57	0.56	0.57	0.56	MLP	Raw	30
0.61	0.62	0.60	0.62	AB	Raw	30
0.58	0.55	0.65	0.55	GNB	Raw	30
0.45	0.41	0.52	0.41	LDA	Raw	30
0.61	0.68	0.61	0.68	MC	Raw	30
0.08	0.17	0.63	0.17	GP	Raw	30
0.48	0.48	0.49	0.48	DT	Min-Max	30
0.54	0.61	0.56	0.61	RF	Min-Max	30
0.49	0.56	0.50	0.56	MLP	Min-Max	30
0.51	0.54	0.49	0.54	AB	Min-Max	30
0.39	0.46	0.36	0.46	GNB	Min-Max	30
0.48	0.55	0.48	0.55	LDA	Min-Max	30
0.53	0.60	0.55	0.60	MC	Min-Max	30
0.19	0.26	0.45	0.26	GP	Min-Max	30
0.48	0.48	0.49	0.48	DT	Z-Score	30
0.54	0.60	0.55	0.60	RF	Z-Score	30
0.50	0.54	0.51	0.54	MLP	Z-Score	30
0.50	0.53	0.48	0.53	AB	Z-Score	30
0.41	0.47	0.38	0.47	GNB	Z-Score	30
0.48	0.55	0.48	0.55	LDA	Z-Score	30

Table II.2 continued from previous page

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.53	0.59	0.55	0.59	MC	Z-Score	30
0.08	0.21	0.30	0.21	GP	Z-Score	30
0.59	0.59	0.59	0.59	DT	Raw	40
0.54	0.65	0.50	0.65	RF	Raw	40
0.54	0.55	0.54	0.55	MLP	Raw	40
0.61	0.63	0.60	0.63	AB	Raw	40
0.54	0.53	0.56	0.53	GNB	Raw	40
0.43	0.40	0.49	0.40	LDA	Raw	40
0.57	0.66	0.58	0.66	MC	Raw	40
0.11	0.19	0.46	0.19	GP	Raw	40
0.51	0.51	0.51	0.51	DT	Min-Max	40
0.52	0.58	0.52	0.58	RF	Min-Max	40
0.50	0.56	0.49	0.56	MLP	Min-Max	40
0.49	0.53	0.48	0.53	AB	Min-Max	40
0.39	0.47	0.37	0.47	GNB	Min-Max	40
0.46	0.53	0.45	0.53	LDA	Min-Max	40
0.52	0.58	0.53	0.58	MC	Min-Max	40
0.15	0.25	0.31	0.25	GP	Min-Max	40
0.50	0.50	0.50	0.50	DT	Z-Score	40
0.51	0.57	0.51	0.57	RF	Z-Score	40
0.51	0.54	0.50	0.54	MLP	Z-Score	40
0.50	0.53	0.48	0.53	AB	Z-Score	40
0.40	0.47	0.37	0.47	GNB	Z-Score	40
0.46	0.53	0.45	0.53	LDA	Z-Score	40
0.52	0.58	0.52	0.58	MC	Z-Score	40
0.11	0.23	0.30	0.23	GP	Z-Score	40
0.56	0.56	0.56	0.56	DT	Raw	50
0.55	0.67	0.49	0.67	RF	Raw	50
0.51	0.51	0.51	0.51	MLP	Raw	50
0.57	0.60	0.55	0.60	AB	Raw	50
0.56	0.59	0.54	0.59	GNB	Raw	50
0.46	0.45	0.48	0.45	LDA	Raw	50
0.56	0.64	0.55	0.64	MC	Raw	50
0.23	0.26	0.46	0.26	GP	Raw	50
0.46	0.45	0.47	0.45	DT	Min-Max	50
0.53	0.58	0.53	0.58	RF	Min-Max	50
0.52	0.56	0.50	0.56	MLP	Min-Max	50
0.49	0.50	0.47	0.50	AB	Min-Max	50

Table II.2 continued from previous page

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.35	0.42	0.31	0.42	GNB	Min-Max	50
0.41	0.47	0.38	0.47	LDA	Min-Max	50
0.52	0.57	0.56	0.57	MC	Min-Max	50
0.20	0.28	0.33	0.28	GP	Min-Max	50
0.46	0.46	0.46	0.46	DT	Z-Score	50
0.54	0.59	0.54	0.59	RF	Z-Score	50
0.56	0.59	0.55	0.59	MLP	Z-Score	50
0.48	0.51	0.47	0.51	AB	Z-Score	50
0.36	0.43	0.32	0.43	GNB	Z-Score	50
0.41	0.47	0.38	0.47	LDA	Z-Score	50
0.55	0.60	0.58	0.60	MC	Z-Score	50
0.18	0.27	0.30	0.27	GP	Z-Score	50
0.56	0.55	0.57	0.55	DT	Raw	60
0.59	0.71	0.51	0.71	RF	Raw	60
0.57	0.57	0.57	0.57	MLP	Raw	60
0.58	0.60	0.56	0.60	AB	Raw	60
0.61	0.65	0.59	0.65	GNB	Raw	60
0.38	0.34	0.48	0.34	LDA	Raw	60
0.56	0.67	0.50	0.67	MC	Raw	60
0.22	0.25	0.41	0.25	GP	Raw	60
0.39	0.38	0.39	0.38	DT	Min-Max	60
0.45	0.54	0.44	0.54	RF	Min-Max	60
0.47	0.54	0.44	0.54	MLP	Min-Max	60
0.44	0.45	0.43	0.45	AB	Min-Max	60
0.35	0.42	0.31	0.42	GNB	Min-Max	60
0.41	0.49	0.36	0.49	LDA	Min-Max	60
0.45	0.53	0.44	0.53	MC	Min-Max	60
0.17	0.25	0.31	0.25	GP	Min-Max	60
0.40	0.40	0.40	0.40	DT	Z-Score	60
0.45	0.54	0.43	0.54	RF	Z-Score	60
0.50	0.53	0.48	0.53	MLP	Z-Score	60
0.44	0.46	0.42	0.46	AB	Z-Score	60
0.34	0.40	0.31	0.40	GNB	Z-Score	60
0.41	0.49	0.36	0.49	LDA	Z-Score	60
0.47	0.53	0.46	0.53	MC	Z-Score	60
0.17	0.24	0.29	0.24	GP	Z-Score	60
0.56	0.55	0.56	0.55	DT	Raw	90
0.61	0.73	0.53	0.73	RF	Raw	90

Table II.2 continued from previous page

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.55	0.59	0.51	0.59	MLP	Raw	90
0.57	0.60	0.54	0.60	AB	Raw	90
0.58	0.66	0.51	0.66	GNB	Raw	90
0.40	0.36	0.45	0.36	LDA	Raw	90
0.60	0.71	0.51	0.71	MC	Raw	90
0.43	0.43	0.43	0.43	GP	Raw	90
0.35	0.35	0.35	0.35	DT	Min-Max	90
0.41	0.50	0.36	0.50	RF	Min-Max	90
0.39	0.43	0.37	0.43	MLP	Min-Max	90
0.33	0.37	0.30	0.37	AB	Min-Max	90
0.38	0.49	0.32	0.49	GNB	Min-Max	90
0.36	0.39	0.33	0.39	LDA	Min-Max	90
0.36	0.46	0.30	0.46	MC	Min-Max	90
0.33	0.36	0.33	0.36	GP	Min-Max	90
0.35	0.36	0.35	0.36	DT	Z-Score	90
0.40	0.50	0.34	0.50	RF	Z-Score	90
0.47	0.50	0.44	0.50	MLP	Z-Score	90
0.35	0.38	0.32	0.38	AB	Z-Score	90
0.34	0.42	0.28	0.42	GNB	Z-Score	90
0.36	0.39	0.33	0.39	LDA	Z-Score	90
0.44	0.54	0.42	0.54	MC	Z-Score	90
0.33	0.36	0.33	0.36	GP	Z-Score	90

Table II.3: Results of classification considering the combinations of all data sources taking into account the Gathered paradigm.

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.76	0.76	0.76	0.76	DT	Raw	10
0.74	0.78	0.79	0.78	RF	Raw	10
0.61	0.60	0.62	0.60	MLP	Raw	10
0.80	0.81	0.80	0.81	AB	Raw	10
0.39	0.42	0.71	0.42	GNB	Raw	10
0.61	0.61	0.61	0.61	LDA	Raw	10
0.78	0.80	0.80	0.80	MC	Raw	10
0.60	0.72	0.51	0.72	GP	Raw	10
0.72	0.72	0.72	0.72	DT	Min-Max	10

Table II.3 continued from previous page

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.74	0.82	0.73	0.82	RF	Min-Max	10
0.76	0.79	0.75	0.79	MLP	Min-Max	10
0.76	0.79	0.75	0.79	AB	Min-Max	10
0.70	0.68	0.72	0.68	GNB	Min-Max	10
0.75	0.77	0.74	0.77	LDA	Min-Max	10
0.76	0.82	0.77	0.82	MC	Min-Max	10
0.74	0.77	0.72	0.77	GP	Min-Max	10
0.72	0.72	0.72	0.72	DT	Z-Score	10
0.74	0.82	0.72	0.82	RF	Z-Score	10
0.75	0.77	0.73	0.77	MLP	Z-Score	10
0.76	0.79	0.75	0.79	AB	Z-Score	10
0.69	0.67	0.71	0.67	GNB	Z-Score	10
0.75	0.77	0.73	0.77	LDA	Z-Score	10
0.75	0.82	0.75	0.82	MC	Z-Score	10
0.74	0.76	0.72	0.76	GP	Z-Score	10
0.78	0.78	0.78	0.78	DT	Raw	12
0.73	0.77	0.78	0.77	RF	Raw	12
0.65	0.66	0.64	0.66	MLP	Raw	12
0.80	0.81	0.80	0.81	AB	Raw	12
0.41	0.43	0.71	0.43	GNB	Raw	12
0.61	0.61	0.62	0.61	LDA	Raw	12
0.75	0.79	0.79	0.79	MC	Raw	12
0.60	0.72	0.52	0.72	GP	Raw	12
0.69	0.68	0.69	0.68	DT	Min-Max	12
0.74	0.82	0.69	0.82	RF	Min-Max	12
0.72	0.75	0.69	0.75	MLP	Min-Max	12
0.74	0.77	0.73	0.77	AB	Min-Max	12
0.69	0.69	0.70	0.69	GNB	Min-Max	12
0.71	0.74	0.69	0.74	LDA	Min-Max	12
0.74	0.80	0.71	0.80	MC	Min-Max	12
0.69	0.69	0.69	0.69	GP	Min-Max	12
0.69	0.68	0.69	0.68	DT	Z-Score	12
0.74	0.82	0.70	0.82	RF	Z-Score	12
0.71	0.73	0.69	0.73	MLP	Z-Score	12
0.74	0.76	0.72	0.76	AB	Z-Score	12
0.69	0.69	0.70	0.69	GNB	Z-Score	12
0.71	0.74	0.69	0.74	LDA	Z-Score	12
0.74	0.80	0.71	0.80	MC	Z-Score	12

Table II.3 continued from previous page

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.68	0.68	0.68	0.68	GP	Z-Score	12
0.72	0.72	0.72	0.72	DT	Raw	20
0.69	0.74	0.73	0.74	RF	Raw	20
0.63	0.64	0.63	0.64	MLP	Raw	20
0.74	0.75	0.73	0.75	AB	Raw	20
0.50	0.49	0.66	0.49	GNB	Raw	20
0.59	0.58	0.60	0.58	LDA	Raw	20
0.72	0.76	0.74	0.76	MC	Raw	20
0.60	0.71	0.51	0.71	GP	Raw	20
0.70	0.69	0.70	0.69	DT	Min-Max	20
0.73	0.81	0.69	0.81	RF	Min-Max	20
0.73	0.77	0.70	0.77	MLP	Min-Max	20
0.74	0.78	0.71	0.78	AB	Min-Max	20
0.73	0.75	0.72	0.75	GNB	Min-Max	20
0.73	0.79	0.70	0.79	LDA	Min-Max	20
0.73	0.80	0.71	0.80	MC	Min-Max	20
0.75	0.78	0.73	0.78	GP	Min-Max	20
0.70	0.70	0.70	0.70	DT	Z-Score	20
0.73	0.81	0.71	0.81	RF	Z-Score	20
0.73	0.75	0.71	0.75	MLP	Z-Score	20
0.73	0.77	0.70	0.77	AB	Z-Score	20
0.72	0.74	0.71	0.74	GNB	Z-Score	20
0.73	0.79	0.70	0.79	LDA	Z-Score	20
0.73	0.80	0.71	0.80	MC	Z-Score	20
0.76	0.80	0.75	0.80	GP	Z-Score	20
0.69	0.69	0.69	0.69	DT	Raw	30
0.64	0.72	0.65	0.72	RF	Raw	30
0.62	0.62	0.62	0.62	MLP	Raw	30
0.69	0.71	0.68	0.71	AB	Raw	30
0.54	0.51	0.67	0.51	GNB	Raw	30
0.59	0.57	0.60	0.57	LDA	Raw	30
0.66	0.72	0.66	0.72	MC	Raw	30
0.62	0.74	0.54	0.74	GP	Raw	30
0.68	0.67	0.68	0.67	DT	Min-Max	30
0.73	0.82	0.67	0.82	RF	Min-Max	30
0.73	0.79	0.70	0.79	MLP	Min-Max	30
0.71	0.74	0.69	0.74	AB	Min-Max	30
0.73	0.77	0.70	0.77	GNB	Min-Max	30

Table II.3 continued from previous page

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.75	0.82	0.76	0.82	LDA	Min-Max	30
0.73	0.81	0.68	0.81	MC	Min-Max	30
0.70	0.75	0.67	0.75	GP	Min-Max	30
0.68	0.67	0.68	0.67	DT	Z-Score	30
0.73	0.81	0.67	0.81	RF	Z-Score	30
0.68	0.69	0.67	0.69	MLP	Z-Score	30
0.71	0.74	0.69	0.74	AB	Z-Score	30
0.73	0.77	0.71	0.77	GNB	Z-Score	30
0.75	0.82	0.76	0.82	LDA	Z-Score	30
0.72	0.79	0.67	0.79	MC	Z-Score	30
0.71	0.77	0.66	0.77	GP	Z-Score	30
0.62	0.62	0.62	0.62	DT	Raw	40
0.60	0.67	0.58	0.67	RF	Raw	40
0.60	0.62	0.59	0.62	MLP	Raw	40
0.61	0.62	0.60	0.62	AB	Raw	40
0.48	0.46	0.57	0.46	GNB	Raw	40
0.57	0.57	0.58	0.57	LDA	Raw	40
0.61	0.67	0.60	0.67	MC	Raw	40
0.59	0.71	0.50	0.71	GP	Raw	40
0.67	0.67	0.67	0.67	DT	Min-Max	40
0.72	0.80	0.69	0.80	RF	Min-Max	40
0.71	0.78	0.67	0.78	MLP	Min-Max	40
0.69	0.70	0.67	0.70	AB	Min-Max	40
0.70	0.75	0.65	0.75	GNB	Min-Max	40
0.72	0.79	0.68	0.79	LDA	Min-Max	40
0.72	0.80	0.66	0.80	MC	Min-Max	40
0.65	0.66	0.64	0.66	GP	Min-Max	40
0.67	0.67	0.67	0.67	DT	Z-Score	40
0.72	0.80	0.68	0.80	RF	Z-Score	40
0.69	0.72	0.67	0.72	MLP	Z-Score	40
0.68	0.70	0.67	0.70	AB	Z-Score	40
0.70	0.75	0.66	0.75	GNB	Z-Score	40
0.72	0.79	0.68	0.79	LDA	Z-Score	40
0.71	0.78	0.66	0.78	MC	Z-Score	40
0.69	0.74	0.65	0.74	GP	Z-Score	40
0.65	0.65	0.65	0.65	DT	Raw	50
0.66	0.73	0.65	0.73	RF	Raw	50
0.58	0.57	0.59	0.57	MLP	Raw	50

Table II.3 continued from previous page

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.67	0.68	0.67	0.68	AB	Raw	50
0.63	0.61	0.67	0.61	GNB	Raw	50
0.62	0.62	0.63	0.62	LDA	Raw	50
0.68	0.73	0.66	0.73	MC	Raw	50
0.67	0.77	0.59	0.77	GP	Raw	50
0.77	0.77	0.77	0.77	DT	Min-Max	50
0.76	0.83	0.74	0.83	RF	Min-Max	50
0.77	0.83	0.75	0.83	MLP	Min-Max	50
0.75	0.75	0.76	0.75	AB	Min-Max	50
0.76	0.83	0.72	0.83	GNB	Min-Max	50
0.76	0.83	0.70	0.83	LDA	Min-Max	50
0.76	0.83	0.75	0.83	MC	Min-Max	50
0.73	0.73	0.73	0.73	GP	Min-Max	50
0.76	0.76	0.76	0.76	DT	Z-Score	50
0.76	0.83	0.70	0.83	RF	Z-Score	50
0.79	0.79	0.78	0.79	MLP	Z-Score	50
0.77	0.77	0.77	0.77	AB	Z-Score	50
0.76	0.82	0.71	0.82	GNB	Z-Score	50
0.76	0.83	0.70	0.83	LDA	Z-Score	50
0.79	0.84	0.79	0.84	MC	Z-Score	50
0.75	0.80	0.70	0.80	GP	Z-Score	50
0.64	0.64	0.65	0.64	DT	Raw	60
0.67	0.74	0.67	0.74	RF	Raw	60
0.62	0.62	0.62	0.62	MLP	Raw	60
0.64	0.64	0.64	0.64	AB	Raw	60
0.71	0.71	0.72	0.71	GNB	Raw	60
0.70	0.72	0.69	0.72	LDA	Raw	60
0.65	0.70	0.64	0.70	MC	Raw	60
0.63	0.74	0.54	0.74	GP	Raw	60
0.68	0.67	0.69	0.67	DT	Min-Max	60
0.72	0.81	0.66	0.81	RF	Min-Max	60
0.70	0.76	0.66	0.76	MLP	Min-Max	60
0.67	0.67	0.68	0.67	AB	Min-Max	60
0.75	0.79	0.73	0.79	GNB	Min-Max	60
0.73	0.81	0.70	0.81	LDA	Min-Max	60
0.73	0.80	0.66	0.80	MC	Min-Max	60
0.65	0.67	0.64	0.67	GP	Min-Max	60
0.68	0.67	0.68	0.67	DT	Z-Score	60

Table II.3 continued from previous page

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.72	0.80	0.66	0.80	RF	Z-Score	60
0.67	0.70	0.65	0.70	MLP	Z-Score	60
0.67	0.68	0.67	0.68	AB	Z-Score	60
0.78	0.82	0.79	0.82	GNB	Z-Score	60
0.73	0.81	0.70	0.81	LDA	Z-Score	60
0.71	0.78	0.66	0.78	MC	Z-Score	60
0.70	0.76	0.66	0.76	GP	Z-Score	60
0.66	0.65	0.66	0.65	DT	Raw	90
0.63	0.68	0.60	0.68	RF	Raw	90
0.61	0.60	0.61	0.60	MLP	Raw	90
0.64	0.63	0.66	0.63	AB	Raw	90
0.59	0.61	0.56	0.61	GNB	Raw	90
0.65	0.64	0.65	0.64	LDA	Raw	90
0.70	0.75	0.69	0.75	MC	Raw	90
0.68	0.78	0.60	0.78	GP	Raw	90
0.71	0.72	0.71	0.72	DT	Min-Max	90
0.72	0.80	0.74	0.80	RF	Min-Max	90
0.75	0.80	0.76	0.80	MLP	Min-Max	90
0.67	0.67	0.67	0.67	AB	Min-Max	90
0.69	0.76	0.64	0.76	GNB	Min-Max	90
0.68	0.78	0.60	0.78	LDA	Min-Max	90
0.68	0.78	0.60	0.78	MC	Min-Max	90
0.56	0.56	0.56	0.56	GP	Min-Max	90
0.70	0.70	0.70	0.70	DT	Z-Score	90
0.71	0.80	0.64	0.80	RF	Z-Score	90
0.79	0.79	0.79	0.79	MLP	Z-Score	90
0.69	0.69	0.69	0.69	AB	Z-Score	90
0.70	0.77	0.64	0.77	GNB	Z-Score	90
0.68	0.78	0.60	0.78	LDA	Z-Score	90
0.68	0.78	0.60	0.78	MC	Z-Score	90
0.68	0.78	0.60	0.78	GP	Z-Score	90

Table II.4: Results of classification considering the combinations of all data sources taking into account the Multiclass paradigm.

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.68	0.68	0.68	0.68	DT	Raw	10
0.56	0.60	0.58	0.60	RF	Raw	10
0.26	0.25	0.27	0.25	MLP	Raw	10
0.42	0.47	0.47	0.47	AB	Raw	10
0.25	0.26	0.35	0.26	GNB	Raw	10
0.24	0.22	0.31	0.22	LDA	Raw	10
0.49	0.53	0.51	0.53	MC	Raw	10
0.01	0.08	0.07	0.08	GP	Raw	10
0.73	0.73	0.73	0.73	DT	Min-Max	10
0.49	0.58	0.54	0.58	RF	Min-Max	10
0.49	0.52	0.48	0.52	MLP	Min-Max	10
0.58	0.59	0.59	0.59	AB	Min-Max	10
0.48	0.49	0.47	0.49	GNB	Min-Max	10
0.48	0.50	0.47	0.50	LDA	Min-Max	10
0.52	0.59	0.55	0.59	MC	Min-Max	10
0.47	0.48	0.48	0.48	GP	Min-Max	10
0.73	0.73	0.73	0.73	DT	Z-Score	10
0.50	0.59	0.56	0.59	RF	Z-Score	10
0.52	0.54	0.51	0.54	MLP	Z-Score	10
0.58	0.59	0.59	0.59	AB	Z-Score	10
0.48	0.49	0.47	0.49	GNB	Z-Score	10
0.48	0.50	0.47	0.50	LDA	Z-Score	10
0.54	0.61	0.58	0.61	MC	Z-Score	10
0.03	0.13	0.02	0.13	GP	Z-Score	10
0.60	0.60	0.60	0.60	DT	Raw	12
0.53	0.57	0.55	0.57	RF	Raw	12
0.28	0.28	0.28	0.28	MLP	Raw	12
0.39	0.46	0.45	0.46	AB	Raw	12
0.27	0.29	0.35	0.29	GNB	Raw	12
0.23	0.22	0.29	0.22	LDA	Raw	12
0.47	0.51	0.49	0.51	MC	Raw	12
0.01	0.08	0.08	0.08	GP	Raw	12
0.63	0.63	0.63	0.63	DT	Min-Max	12
0.51	0.59	0.57	0.59	RF	Min-Max	12
0.53	0.55	0.52	0.55	MLP	Min-Max	12
0.57	0.58	0.56	0.58	AB	Min-Max	12

Table II.4 continued from previous page

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.49	0.50	0.48	0.50	GNB	Min-Max	12
0.51	0.53	0.50	0.53	LDA	Min-Max	12
0.53	0.60	0.59	0.60	MC	Min-Max	12
0.54	0.54	0.55	0.54	GP	Min-Max	12
0.63	0.63	0.63	0.63	DT	Z-Score	12
0.51	0.60	0.59	0.60	RF	Z-Score	12
0.55	0.56	0.54	0.56	MLP	Z-Score	12
0.57	0.58	0.57	0.58	AB	Z-Score	12
0.49	0.50	0.48	0.50	GNB	Z-Score	12
0.50	0.53	0.50	0.53	LDA	Z-Score	12
0.55	0.61	0.59	0.61	MC	Z-Score	12
0.03	0.13	0.02	0.13	GP	Z-Score	12
0.53	0.53	0.53	0.53	DT	Raw	20
0.48	0.52	0.49	0.52	RF	Raw	20
0.26	0.26	0.26	0.26	MLP	Raw	20
0.42	0.47	0.46	0.47	AB	Raw	20
0.39	0.40	0.40	0.40	GNB	Raw	20
0.23	0.21	0.25	0.21	LDA	Raw	20
0.45	0.48	0.47	0.48	MC	Raw	20
0.03	0.08	0.11	0.08	GP	Raw	20
0.45	0.45	0.45	0.45	DT	Min-Max	20
0.45	0.56	0.50	0.56	RF	Min-Max	20
0.49	0.53	0.48	0.53	MLP	Min-Max	20
0.44	0.47	0.42	0.47	AB	Min-Max	20
0.45	0.49	0.44	0.49	GNB	Min-Max	20
0.44	0.49	0.42	0.49	LDA	Min-Max	20
0.47	0.56	0.50	0.56	MC	Min-Max	20
0.41	0.39	0.54	0.39	GP	Min-Max	20
0.45	0.45	0.45	0.45	DT	Z-Score	20
0.45	0.56	0.49	0.56	RF	Z-Score	20
0.50	0.52	0.49	0.52	MLP	Z-Score	20
0.44	0.46	0.42	0.46	AB	Z-Score	20
0.43	0.47	0.42	0.47	GNB	Z-Score	20
0.44	0.49	0.42	0.49	LDA	Z-Score	20
0.47	0.56	0.49	0.56	MC	Z-Score	20
0.03	0.14	0.02	0.14	GP	Z-Score	20
0.48	0.48	0.48	0.48	DT	Raw	30
0.44	0.47	0.46	0.47	RF	Raw	30

Table II.4 continued from previous page

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.26	0.26	0.26	0.26	MLP	Raw	30
0.40	0.43	0.43	0.43	AB	Raw	30
0.44	0.45	0.45	0.45	GNB	Raw	30
0.21	0.19	0.24	0.19	LDA	Raw	30
0.40	0.43	0.42	0.43	MC	Raw	30
0.12	0.16	0.12	0.16	GP	Raw	30
0.43	0.43	0.43	0.43	DT	Min-Max	30
0.44	0.55	0.44	0.55	RF	Min-Max	30
0.51	0.57	0.50	0.57	MLP	Min-Max	30
0.43	0.46	0.41	0.46	AB	Min-Max	30
0.43	0.51	0.41	0.51	GNB	Min-Max	30
0.46	0.55	0.47	0.55	LDA	Min-Max	30
0.45	0.56	0.44	0.56	MC	Min-Max	30
0.24	0.24	0.53	0.24	GP	Min-Max	30
0.43	0.42	0.43	0.42	DT	Z-Score	30
0.43	0.55	0.44	0.55	RF	Z-Score	30
0.53	0.55	0.52	0.55	MLP	Z-Score	30
0.43	0.45	0.41	0.45	AB	Z-Score	30
0.43	0.51	0.40	0.51	GNB	Z-Score	30
0.46	0.55	0.46	0.55	LDA	Z-Score	30
0.46	0.55	0.45	0.55	MC	Z-Score	30
0.03	0.13	0.02	0.13	GP	Z-Score	30
0.39	0.39	0.39	0.39	DT	Raw	40
0.42	0.46	0.46	0.46	RF	Raw	40
0.23	0.23	0.22	0.23	MLP	Raw	40
0.38	0.41	0.37	0.41	AB	Raw	40
0.45	0.47	0.44	0.47	GNB	Raw	40
0.23	0.22	0.25	0.22	LDA	Raw	40
0.37	0.41	0.37	0.41	MC	Raw	40
0.13	0.16	0.13	0.16	GP	Raw	40
0.41	0.41	0.41	0.41	DT	Min-Max	40
0.39	0.52	0.36	0.52	RF	Min-Max	40
0.49	0.54	0.47	0.54	MLP	Min-Max	40
0.39	0.41	0.38	0.41	AB	Min-Max	40
0.38	0.47	0.36	0.47	GNB	Min-Max	40
0.42	0.52	0.41	0.52	LDA	Min-Max	40
0.42	0.53	0.41	0.53	MC	Min-Max	40
0.14	0.20	0.61	0.20	GP	Min-Max	40

Table II.4 continued from previous page

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.40	0.41	0.40	0.41	DT	Z-Score	40
0.40	0.52	0.35	0.52	RF	Z-Score	40
0.46	0.48	0.45	0.48	MLP	Z-Score	40
0.40	0.41	0.38	0.41	AB	Z-Score	40
0.40	0.47	0.38	0.47	GNB	Z-Score	40
0.42	0.52	0.42	0.52	LDA	Z-Score	40
0.42	0.51	0.40	0.51	MC	Z-Score	40
0.04	0.14	0.05	0.14	GP	Z-Score	40
0.29	0.29	0.29	0.29	DT	Raw	50
0.31	0.36	0.32	0.36	RF	Raw	50
0.23	0.23	0.23	0.23	MLP	Raw	50
0.31	0.33	0.30	0.33	AB	Raw	50
0.35	0.38	0.33	0.38	GNB	Raw	50
0.21	0.19	0.23	0.19	LDA	Raw	50
0.30	0.33	0.31	0.33	MC	Raw	50
0.13	0.17	0.11	0.17	GP	Raw	50
0.43	0.43	0.43	0.43	DT	Min-Max	50
0.43	0.56	0.38	0.56	RF	Min-Max	50
0.48	0.54	0.45	0.54	MLP	Min-Max	50
0.44	0.46	0.42	0.46	AB	Min-Max	50
0.40	0.51	0.35	0.51	GNB	Min-Max	50
0.45	0.56	0.43	0.56	LDA	Min-Max	50
0.44	0.55	0.40	0.55	MC	Min-Max	50
0.07	0.16	0.61	0.16	GP	Min-Max	50
0.43	0.43	0.44	0.43	DT	Z-Score	50
0.44	0.57	0.39	0.57	RF	Z-Score	50
0.50	0.50	0.49	0.50	MLP	Z-Score	50
0.44	0.45	0.42	0.45	AB	Z-Score	50
0.41	0.51	0.39	0.51	GNB	Z-Score	50
0.45	0.56	0.43	0.56	LDA	Z-Score	50
0.46	0.56	0.43	0.56	MC	Z-Score	50
0.05	0.15	0.04	0.15	GP	Z-Score	50
0.30	0.30	0.30	0.30	DT	Raw	60
0.32	0.35	0.33	0.35	RF	Raw	60
0.20	0.21	0.20	0.21	MLP	Raw	60
0.31	0.35	0.31	0.35	AB	Raw	60
0.30	0.34	0.29	0.34	GNB	Raw	60
0.21	0.19	0.24	0.19	LDA	Raw	60

Table II.4 continued from previous page

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.31	0.34	0.31	0.34	MC	Raw	60
0.14	0.17	0.15	0.17	GP	Raw	60
0.37	0.36	0.39	0.36	DT	Min-Max	60
0.46	0.60	0.41	0.60	RF	Min-Max	60
0.51	0.57	0.48	0.57	MLP	Min-Max	60
0.42	0.44	0.40	0.44	AB	Min-Max	60
0.45	0.57	0.38	0.57	GNB	Min-Max	60
0.47	0.59	0.44	0.59	LDA	Min-Max	60
0.46	0.57	0.41	0.57	MC	Min-Max	60
0.06	0.15	0.35	0.15	GP	Min-Max	60
0.37	0.35	0.38	0.35	DT	Z-Score	60
0.47	0.60	0.42	0.60	RF	Z-Score	60
0.50	0.50	0.50	0.50	MLP	Z-Score	60
0.41	0.42	0.39	0.42	AB	Z-Score	60
0.46	0.56	0.39	0.56	GNB	Z-Score	60
0.47	0.60	0.44	0.60	LDA	Z-Score	60
0.47	0.56	0.42	0.56	MC	Z-Score	60
0.05	0.14	0.33	0.14	GP	Z-Score	60
0.29	0.30	0.29	0.30	DT	Raw	90
0.24	0.30	0.21	0.30	RF	Raw	90
0.28	0.27	0.28	0.27	MLP	Raw	90
0.31	0.32	0.34	0.32	AB	Raw	90
0.26	0.36	0.34	0.36	GNB	Raw	90
0.21	0.21	0.24	0.21	LDA	Raw	90
0.26	0.30	0.26	0.30	MC	Raw	90
0.22	0.23	0.24	0.23	GP	Raw	90
0.43	0.42	0.44	0.42	DT	Min-Max	90
0.46	0.58	0.39	0.58	RF	Min-Max	90
0.47	0.54	0.42	0.54	MLP	Min-Max	90
0.42	0.46	0.40	0.46	AB	Min-Max	90
0.47	0.61	0.38	0.61	GNB	Min-Max	90
0.39	0.50	0.31	0.50	LDA	Min-Max	90
0.41	0.53	0.33	0.53	MC	Min-Max	90
0.02	0.03	0.02	0.03	GP	Min-Max	90
0.44	0.44	0.43	0.44	DT	Z-Score	90
0.46	0.58	0.38	0.58	RF	Z-Score	90
0.39	0.38	0.41	0.38	MLP	Z-Score	90
0.44	0.48	0.41	0.48	AB	Z-Score	90

Table II.4 continued from previous page

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.45	0.57	0.38	0.57	GNB	Z-Score	90
0.39	0.50	0.31	0.50	LDA	Z-Score	90
0.32	0.37	0.28	0.37	MC	Z-Score	90
0.12	0.08	0.60	0.08	GP	Z-Score	90

Table II.5: Results of classification considering the combinations of all data sources taking into account the LOUO paradigm.

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.32	0.31	0.33	0.31	DT	Raw	10
0.39	0.45	0.37	0.45	RF	Raw	10
0.35	0.33	0.37	0.33	MLP	Raw	10
0.36	0.35	0.37	0.35	AB	Raw	10
0.21	0.21	0.36	0.21	GNB	Raw	10
0.25	0.24	0.34	0.24	LDA	Raw	10
0.39	0.40	0.38	0.40	MC	Raw	10
0.28	0.45	0.21	0.45	GP	Raw	10
0.32	0.31	0.33	0.31	DT	Min-Max	10
0.38	0.43	0.34	0.43	RF	Min-Max	10
0.36	0.38	0.35	0.38	MLP	Min-Max	10
0.34	0.33	0.35	0.33	AB	Min-Max	10
0.26	0.24	0.36	0.24	GNB	Min-Max	10
0.33	0.31	0.36	0.31	LDA	Min-Max	10
0.38	0.39	0.37	0.39	MC	Min-Max	10
0.35	0.41	0.36	0.41	GP	Min-Max	10
0.32	0.31	0.33	0.31	DT	Z-Score	10
0.37	0.43	0.37	0.43	RF	Z-Score	10
0.36	0.36	0.35	0.36	MLP	Z-Score	10
0.34	0.34	0.35	0.34	AB	Z-Score	10
0.26	0.24	0.36	0.24	GNB	Z-Score	10
0.33	0.31	0.35	0.31	LDA	Z-Score	10
0.37	0.38	0.36	0.38	MC	Z-Score	10
0.28	0.45	0.21	0.45	GP	Z-Score	10
0.34	0.33	0.34	0.33	DT	Raw	12
0.38	0.44	0.35	0.44	RF	Raw	12
0.38	0.38	0.38	0.38	MLP	Raw	12

Table II.5 continued from previous page

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.39	0.40	0.39	0.40	AB	Raw	12
0.27	0.29	0.31	0.29	GNB	Raw	12
0.30	0.29	0.33	0.29	LDA	Raw	12
0.38	0.40	0.39	0.40	MC	Raw	12
0.28	0.45	0.20	0.45	GP	Raw	12
0.31	0.30	0.32	0.30	DT	Min-Max	12
0.38	0.44	0.36	0.44	RF	Min-Max	12
0.39	0.39	0.38	0.39	MLP	Min-Max	12
0.39	0.40	0.40	0.40	AB	Min-Max	12
0.28	0.25	0.34	0.25	GNB	Min-Max	12
0.35	0.33	0.36	0.33	LDA	Min-Max	12
0.40	0.42	0.39	0.42	MC	Min-Max	12
0.35	0.45	0.40	0.45	GP	Min-Max	12
0.33	0.32	0.34	0.32	DT	Z-Score	12
0.38	0.43	0.34	0.43	RF	Z-Score	12
0.38	0.39	0.38	0.39	MLP	Z-Score	12
0.38	0.39	0.38	0.39	AB	Z-Score	12
0.28	0.25	0.34	0.25	GNB	Z-Score	12
0.34	0.33	0.36	0.33	LDA	Z-Score	12
0.39	0.41	0.38	0.41	MC	Z-Score	12
0.28	0.45	0.20	0.45	GP	Z-Score	12
0.37	0.38	0.37	0.38	DT	Raw	20
0.38	0.44	0.35	0.44	RF	Raw	20
0.31	0.30	0.33	0.30	MLP	Raw	20
0.39	0.39	0.39	0.39	AB	Raw	20
0.33	0.39	0.35	0.39	GNB	Raw	20
0.32	0.29	0.37	0.29	LDA	Raw	20
0.38	0.39	0.39	0.39	MC	Raw	20
0.28	0.45	0.21	0.45	GP	Raw	20
0.41	0.42	0.40	0.42	DT	Min-Max	20
0.37	0.42	0.35	0.42	RF	Min-Max	20
0.34	0.36	0.33	0.36	MLP	Min-Max	20
0.36	0.35	0.38	0.35	AB	Min-Max	20
0.31	0.28	0.39	0.28	GNB	Min-Max	20
0.30	0.28	0.34	0.28	LDA	Min-Max	20
0.36	0.38	0.35	0.38	MC	Min-Max	20
0.29	0.44	0.22	0.44	GP	Min-Max	20
0.37	0.39	0.37	0.39	DT	Z-Score	20

Table II.5 continued from previous page

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.39	0.44	0.43	0.44	RF	Z-Score	20
0.34	0.35	0.33	0.35	MLP	Z-Score	20
0.34	0.33	0.36	0.33	AB	Z-Score	20
0.32	0.29	0.40	0.29	GNB	Z-Score	20
0.29	0.27	0.33	0.27	LDA	Z-Score	20
0.37	0.38	0.36	0.38	MC	Z-Score	20
0.28	0.45	0.21	0.45	GP	Z-Score	20
0.32	0.32	0.33	0.32	DT	Raw	30
0.38	0.43	0.34	0.43	RF	Raw	30
0.34	0.35	0.37	0.35	MLP	Raw	30
0.40	0.42	0.38	0.42	AB	Raw	30
0.12	0.14	0.55	0.14	GNB	Raw	30
0.23	0.21	0.35	0.21	LDA	Raw	30
0.39	0.41	0.38	0.41	MC	Raw	30
0.28	0.45	0.20	0.45	GP	Raw	30
0.33	0.32	0.33	0.32	DT	Min-Max	30
0.38	0.42	0.34	0.42	RF	Min-Max	30
0.34	0.36	0.34	0.36	MLP	Min-Max	30
0.39	0.41	0.38	0.41	AB	Min-Max	30
0.33	0.30	0.39	0.30	GNB	Min-Max	30
0.26	0.24	0.34	0.24	LDA	Min-Max	30
0.39	0.42	0.37	0.42	MC	Min-Max	30
0.29	0.41	0.27	0.41	GP	Min-Max	30
0.32	0.31	0.33	0.31	DT	Z-Score	30
0.40	0.46	0.37	0.46	RF	Z-Score	30
0.35	0.36	0.34	0.36	MLP	Z-Score	30
0.38	0.40	0.36	0.40	AB	Z-Score	30
0.33	0.30	0.39	0.30	GNB	Z-Score	30
0.26	0.24	0.33	0.24	LDA	Z-Score	30
0.39	0.42	0.37	0.42	MC	Z-Score	30
0.28	0.45	0.20	0.45	GP	Z-Score	30
0.41	0.40	0.43	0.40	DT	Raw	40
0.38	0.43	0.34	0.43	RF	Raw	40
0.31	0.32	0.32	0.32	MLP	Raw	40
0.38	0.40	0.37	0.40	AB	Raw	40
0.15	0.15	0.34	0.15	GNB	Raw	40
0.32	0.29	0.48	0.29	LDA	Raw	40
0.36	0.38	0.35	0.38	MC	Raw	40

Table II.5 continued from previous page

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.29	0.46	0.21	0.46	GP	Raw	40
0.38	0.38	0.38	0.38	DT	Min-Max	40
0.39	0.44	0.35	0.44	RF	Min-Max	40
0.34	0.35	0.33	0.35	MLP	Min-Max	40
0.38	0.40	0.37	0.40	AB	Min-Max	40
0.30	0.27	0.35	0.27	GNB	Min-Max	40
0.24	0.23	0.31	0.23	LDA	Min-Max	40
0.37	0.40	0.36	0.40	MC	Min-Max	40
0.32	0.41	0.31	0.41	GP	Min-Max	40
0.33	0.32	0.35	0.32	DT	Z-Score	40
0.42	0.48	0.39	0.48	RF	Z-Score	40
0.35	0.36	0.35	0.36	MLP	Z-Score	40
0.39	0.41	0.37	0.41	AB	Z-Score	40
0.30	0.28	0.35	0.28	GNB	Z-Score	40
0.26	0.24	0.33	0.24	LDA	Z-Score	40
0.41	0.43	0.40	0.43	MC	Z-Score	40
0.29	0.46	0.21	0.46	GP	Z-Score	40
0.39	0.39	0.40	0.39	DT	Raw	50
0.40	0.45	0.36	0.45	RF	Raw	50
0.34	0.35	0.35	0.35	MLP	Raw	50
0.35	0.39	0.34	0.39	AB	Raw	50
0.24	0.23	0.38	0.23	GNB	Raw	50
0.26	0.23	0.34	0.23	LDA	Raw	50
0.38	0.40	0.36	0.40	MC	Raw	50
0.28	0.45	0.20	0.45	GP	Raw	50
0.40	0.40	0.41	0.40	DT	Min-Max	50
0.39	0.44	0.35	0.44	RF	Min-Max	50
0.37	0.38	0.36	0.38	MLP	Min-Max	50
0.39	0.42	0.38	0.42	AB	Min-Max	50
0.33	0.31	0.37	0.31	GNB	Min-Max	50
0.30	0.28	0.36	0.28	LDA	Min-Max	50
0.39	0.42	0.38	0.42	MC	Min-Max	50
0.36	0.41	0.33	0.41	GP	Min-Max	50
0.40	0.39	0.40	0.39	DT	Z-Score	50
0.42	0.47	0.38	0.47	RF	Z-Score	50
0.36	0.36	0.35	0.36	MLP	Z-Score	50
0.32	0.37	0.31	0.37	AB	Z-Score	50
0.33	0.31	0.38	0.31	GNB	Z-Score	50

Table II.5 continued from previous page

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.29	0.27	0.34	0.27	LDA	Z-Score	50
0.37	0.40	0.36	0.40	MC	Z-Score	50
0.28	0.45	0.20	0.45	GP	Z-Score	50
0.36	0.36	0.36	0.36	DT	Raw	60
0.39	0.44	0.36	0.44	RF	Raw	60
0.29	0.29	0.30	0.29	MLP	Raw	60
0.34	0.35	0.33	0.35	AB	Raw	60
0.13	0.13	0.33	0.13	GNB	Raw	60
0.20	0.17	0.27	0.17	LDA	Raw	60
0.34	0.35	0.33	0.35	MC	Raw	60
0.27	0.44	0.19	0.44	GP	Raw	60
0.35	0.35	0.35	0.35	DT	Min-Max	60
0.38	0.43	0.35	0.43	RF	Min-Max	60
0.32	0.33	0.31	0.33	MLP	Min-Max	60
0.34	0.36	0.33	0.36	AB	Min-Max	60
0.28	0.27	0.29	0.27	GNB	Min-Max	60
0.32	0.30	0.34	0.30	LDA	Min-Max	60
0.35	0.37	0.33	0.37	MC	Min-Max	60
0.33	0.39	0.30	0.39	GP	Min-Max	60
0.34	0.34	0.34	0.34	DT	Z-Score	60
0.40	0.45	0.37	0.45	RF	Z-Score	60
0.32	0.33	0.31	0.33	MLP	Z-Score	60
0.37	0.39	0.35	0.39	AB	Z-Score	60
0.29	0.28	0.31	0.28	GNB	Z-Score	60
0.30	0.29	0.33	0.29	LDA	Z-Score	60
0.36	0.38	0.34	0.38	MC	Z-Score	60
0.27	0.44	0.19	0.44	GP	Z-Score	60
0.37	0.36	0.37	0.36	DT	Raw	90
0.43	0.48	0.39	0.48	RF	Raw	90
0.28	0.31	0.27	0.31	MLP	Raw	90
0.44	0.47	0.41	0.47	AB	Raw	90
0.10	0.09	0.29	0.09	GNB	Raw	90
0.30	0.28	0.33	0.28	LDA	Raw	90
0.40	0.44	0.37	0.44	MC	Raw	90
0.30	0.47	0.22	0.47	GP	Raw	90
0.38	0.37	0.40	0.37	DT	Min-Max	90
0.46	0.51	0.42	0.51	RF	Min-Max	90
0.40	0.41	0.38	0.41	MLP	Min-Max	90

Table II.5 continued from previous page

F1-Score	Accuracy	Precision	Recall	Classifier	Norm	Window (s)
0.42	0.45	0.40	0.45	AB	Min-Max	90
0.31	0.30	0.34	0.30	GNB	Min-Max	90
0.32	0.30	0.34	0.30	LDA	Min-Max	90
0.45	0.48	0.43	0.48	MC	Min-Max	90
0.31	0.46	0.28	0.46	GP	Min-Max	90
0.35	0.35	0.36	0.35	DT	Z-Score	90
0.39	0.44	0.36	0.44	RF	Z-Score	90
0.35	0.35	0.35	0.35	MLP	Z-Score	90
0.41	0.44	0.39	0.44	AB	Z-Score	90
0.33	0.31	0.36	0.31	GNB	Z-Score	90
0.32	0.30	0.34	0.30	LDA	Z-Score	90
0.41	0.43	0.39	0.43	MC	Z-Score	90
0.30	0.47	0.22	0.47	GP	Z-Score	90