

UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingenieros de Telecomunicación



**Holistic system for personalized management
and predictive modeling of symptomatic crises
in health**

DOCTORAL THESIS

Submitted for the degree of Doctor by:

Javier Gálvez Goicuría
Telecommunications Engineer

Madrid, 2024



UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingenieros de Telecomunicación

Doctoral Degree in Electronic Systems Engineering

**Holistic system for personalized management
and predictive modeling of symptomatic crises
in health**

DOCTORAL THESIS

Submitted for the degree of Doctor by:

Javier Gálvez Goicuría
Telecommunications Engineer

Under the supervision of:
Dr. Jose Manuel Moya Fernández
Josué Pagán Ortiz

Madrid, 2024

Title: Holistic system for personalized management and predictive modeling of symptomatic crises in health

Author: Javier Gálvez Goicuría

Doctoral Programme: Electronic Systems Engineering

Thesis Supervision:

Dr. José Manuel Moya Fernández, , Profesor Contratado Doctor, Universidad Politécnica de Madrid (Supervisor)

Dr. Josué Pagán Ortiz, Profesor Contratado Doctor, Universidad Politécnica de Madrid

External Reviewers:

Thesis Defense Committee:

Thesis Defense Date:

This thesis has been partially supported by Comunidad de Madrid, with reference number: IND2019/BMD-17276.

Quiero dedicar esta tesis a todas esas personas que me han acompañado en el transcurso de la misma. A Sara EnfermeraMaña por acompañarme, apoyarme y darme una perspectiva de la salud mucho más humana, a mis directores de tesis especialmente a Josué y Jose, que me le han dedicado mucho tiempo a este trabajo y han aguantado todos mis defectos. Al despacho 229, las Lauras, Dani, Kevin, Rodrigo y Paula. La gente del grupo Green-LSI de la UPM, Roman y Patri. A mis acompañantes durante mi estancia en Dinamarca, Jonas, Luciana, Stinna, Ryad, Amir, Edith, Eugenio, Yane, Carol y muchos más. Y a mi familia.

Acknowledgement

Aunque ya los he mencionado anteriormente, creo que este trabajo no habría sido posible sin Josué Pagán Ortiz y José Luis Ayala, he tenido mucha suerte al elegirlos cómo acompañantes en este trabajo y con que me dieran la oportunidad. Por hacer grupo, en un grupo de investigación formado por gente muy capaz. Por todas esas reuniones en las que compartir conocimiento y descubrimientos juntos. Todos estos compañeros, Laura, María, Carlos, Fernando, Paula, Sandra y muchos más...

Gracias por el apoyo clínico a la Doctora Ana Gago y Rybel Wix y por su trabajo. Por las reuniones analizando datos, corrigiendo errores y su apoyo y buen hacer.

Gracias a Adam Hulman por darme la oportunidad de realizar una maravillosa estancia en Dinamarca, disponiendo de una instalaciones increíbles y un equipo incipiente con un potencial enorme de investigación en el Hospital Universitario de Aarhus. Por su hospitalidad y la de todos los de allí. A los compañeros de despacho, Jonas, Sissel, Fernando... Y a los compañeros de edificio, Omar entre otros...

Abstract

The use of Artificial Intelligence (AI) in healthcare has led to an increase in the use of wearable devices in this sector. This work focuses on improving the personalization of digital health models using Machine Learning (ML) and Deep Learning (DL) technologies. The present manuscript describes a methodology to generate knowledge from data gathered from multiple sources applied to symptomatic crises of chronic diseases such as migraine. The objective is to (i) take a small step forward in research related to creating population models that can be transferred to individuals and thus personalize them, and (ii) integrate state-of-the-art unobtrusive and low-cost wearable devices to acquire knowledge to enhance the performance of population prediction models. Migraines affect quality of life, leading to stigmatization of patients, and can increase the risk of suicidal behaviors in sufferers.

This work uses a real-time collection of migraine-related biomedical signals and can help prevent migraine pain with greater precision during analgesic intake. This work focuses on migraine as a specific use case for information acquisition to improve the prediction of symptomatic crises in chronic diseases. The goal is to validate these theories.

Pain is a common complication of migraine. It is characterized by a variety of types of pain, ranging from mild to severe, with a high frequency of episodes of pain. The presented holistic system contains different modules that all have the aim of giving the patient more information about the disease and their status. For the development of these modules, several experiments were conducted: various studies recording hemodynamic variables are made through a wrist-worn device. These studies focus on temperature, accelerometry, Photoplethysmography (PPG), Heart Rate (HR), and Electrodermal Activity (EDA) signals. The research of the classification of migraine pain curves is made based on their morphology, type of pain, and severity, together with an analysis of various variables related to migraine. A recommender is developed, an innovative mobile application for migraine patients, using the same technologies used to improve patient education and support. In addition, a sleep state detection model is implemented due to the close relationship between sleep quality and migraine crises. All of this culminates in the migraine pain prediction model.

The results report a novel classification of pain curves and show that migraines can be clustered into 4 different types based on 8 features of the pain curves. A recommender system for advice and self-learning about the disease to improve patient adherence to treatment. Two experiments aimed at predicting sleep postures and states using wearable devices and advanced ML techniques. And finally, the migraine pain predictor experiment was able to predict pain attacks in 7/8 (86%) of the patients, with a sensitivity of 100% and no false negatives among them using Long Short-Term Memory Network (LSTM) Neural Network (NN)s.

This work presents a comprehensive research of the development and implementation of a holistic system for the management and prediction of symptomatic crises in chronic diseases using advanced AI techniques. The system is based on different modules that are related to migraine. The results pave the way for future innovations and improvements in a-AI-driven healthcare solutions, ultimately aiming to enhance the quality of life of patients suffering from chronic diseases.

Resumen

El uso de la Inteligencia Artificial (IA) en la salud ha aumentado el uso de dispositivos portátiles. Este trabajo se centra en mejorar la personalización de los modelos de predicción de salud digital utilizando tecnologías de ML y DL. Describe una metodología para generar conocimiento a partir de datos recopilados de múltiples fuentes aplicados a enfermedades crónicas con crisis sintomáticas, como la migraña. Los objetivos son (i) avanzar en la investigación sobre la creación de modelos poblacionales que puedan personalizarse para individuos e (ii) integrar dispositivos portátiles avanzados, no intrusivos y de bajo costo para mejorar los modelos de predicción poblacionales. Las migrañas afectan la calidad de vida de los pacientes, llevando a la estigmatización y aumentando el riesgo de suicidio.

Este trabajo utiliza la recolección en tiempo real de señales biomédicas relacionadas con la migraña y puede ayudar a prevenir el dolor de migraña con mayor precisión durante la ingesta de analgésicos. Este trabajo se centra en la migraña como un caso de uso específico para la adquisición de información para mejorar la predicción de crisis sintomáticas en enfermedades crónicas. El objetivo es validar estas teorías.

El dolor es una complicación común de la migraña, variando de leve a severo y con alta frecuencia de episodios. El sistema holístico presentado incluye módulos para brindar al paciente más información sobre la enfermedad y su estado. Para desarrollar estos módulos se realizaron varios experimentos: estudios con variables hemodinámicas usando un dispositivo de muñeca, centrándose en señales de temperatura, acelerometría, PPG, HR y EDA; clasificación de curvas de dolor de migraña según su morfología, tipo y severidad, y análisis de variables relacionadas; desarrollo de un recomendador, una aplicación móvil para mejorar la educación y apoyo al paciente; y un modelo de detección del estado del sueño por su relación con las crisis de migraña. Todo culmina en el modelo de predicción del dolor de migraña.

Los resultados muestran una nueva clasificación de las curvas de dolor y que las migrañas se pueden agrupar en 4 tipos diferentes basados en 8 características de las curvas de dolor. También muestran un sistema de recomendación para consejos y autoaprendizaje sobre la enfermedad para mejorar la adherencia del paciente al tratamiento. Dos experimentos destinados a predecir posturas y estados de sueño utilizando dispositivos portátiles y técnicas avanzadas de ML. Y finalmente, el experimento del modelo de predicción de dolor de migraña que pudo predecir ataques de dolor en 7/8 (86%) de los pacientes, con una sensibilidad del 100% y sin falsos negativos entre ellos, utilizando LSTM NNs.

Este trabajo presenta una investigación integral sobre el desarrollo e implementación de un sistema holístico para la gestión y predicción de crisis sintomáticas en enfermedades crónicas utilizando técnicas avanzadas de IA. El sistema se basa en diferentes módulos relacionados con la migraña. Los resultados de esta investigación abren el camino para futuras innovaciones y mejoras en soluciones de salud impulsadas por la IA, con el objetivo final de mejorar la calidad de vida de los pacientes que sufren de enfermedades crónicas.

Table of Contents

Acknowledgement	v
Abstract	vi
Resumen	vii
List of Figures	xi
List of Tables	xv
Abbreviations and acronyms	xix
1 Introduction	1
1.1 Machine Learning and Deep Learning in healthcare	5
1.2 Wearable devices in healthcare	6
1.3 Motivation	7
1.4 Objectives of this work	9
1.5 List of publications	10
2 State of the art	11
2.1 Wearable devices	11
2.2 Artificial Intelligence in health	14
2.2.1 Machine Learning in health	14
2.2.2 Deep Learning in health	16
2.2.3 Transfer Learning in health	18
2.3 Migraine disease	19
2.3.1 Sleep	22
2.3.1.1 Sleep posture	23
2.3.1.2 Sleep prediction	24
2.3.2 Prediction	25
3 Methodologies and experiments	27
3.1 Pain curve characterization	27
3.1.1 Experimental setup	29
3.1.2 Pain evolution marking	29
3.1.3 Data pre-processing	30
3.1.4 Processing and modeling methodology	31
3.1.5 Generation of features	34
3.1.5.1 Features related to the acme	35
3.1.5.2 Features related to the duration	35

	Features related to the area	36
	Features related to the intensity	37
3.1.6	Removal of outliers	37
3.1.7	Feature selection	38
3.1.8	Data transformation	39
3.1.9	Unsupervised clustering	41
3.1.10	Model selection and model curve	41
	Model curve: schematic representation of types of pain	42
3.1.11	Group clustering by pain type	43
3.2	Recommender	43
3.2.1	System components	43
3.2.2	Input data and advice	44
3.2.3	Advice selection	46
3.3	Sleep model	50
3.3.1	Posture detector model	51
	Data acquisition	52
	Data labeling	53
	Feature generation and feature selection	54
	Modeling	55
3.3.2	Sleep model implementation	55
	Dataset	56
	Signal preprocessing and feature extraction	56
	Different experiments conducted	57
	Transfer Learning	57
3.4	Migraine pain predictor	58
3.4.1	Study population	58
3.4.2	Study protocol	60
3.4.3	Statistical analysis	61
	Data processing, feature generation and selection	62
3.4.4	Predictive system	65
4	Results	69
4.1	Pain curve characterization	69
4.1.1	Results of the unsupervised pain clustering model	69
	Clustering cohesion index	69
	Distribution of pain curves per clustering model	70
	Distribution of features per cluster	71
	Characterization of the pain curves per cluster	73
4.1.2	Supervised feature selection	74
4.1.3	Supervised classification model	74
4.2	Group clustering by pain type	77
4.2.1	Patient cohorts	77
4.2.2	ESPA recommendation system	80
4.2.3	Study of the impact of migraine pain type information	80
4.2.4	Demonstration of a use case of the ESPA recommendation system	82

4.2.5	Clinical analysis for patient cohorts	84
4.3	Feasibility of sleep modeling for migraine patients	85
4.3.1	Posture detector model	86
	Analysis of feature selection	86
	Record-wise methodology on new subjects data	86
	Subject-wise models and detection of torso turns	87
4.3.2	Sleep model implementation	90
	Transfer Learning	92
4.4	Migraine pain alarm	92
5	Discussion	101
5.1	Pain curve characterization	101
5.1.1	Group clustering by pain type	102
5.2	Feasibility of sleep modeling for migraine patients	103
5.2.1	Posture detector model	103
5.2.2	Sleep model implementation	103
5.3	Migraine pain alarm	104
5.3.1	LSTM and alarm detector	105
6	Conclusions	107
6.1	Feasibility and effectiveness of AI models	107
6.2	Analysis and confrontation from different fronts	107
6.3	Improvement in model performance	108
6.4	Practical implications for clinical use	108
6.5	Problems and complications encountered during the course of the work . . .	109
6.6	Future directions and potential improvements	109
6.7	Application of AI methodologies to the field of healthcare	110
	Bibliography	111
	Appendix A: Tables	127
	Appendix B: Ethics committee agreement	129

List of Figures

1.1	Relation between Artificial Intelligence, Machine Learning and Deep Learning.	2
1.2	Diagram of the methodology and objectives of the thesis proposed.	3
1.3	Number active wearable devices by region [192].	8
2.1	Empatica E4 wearable monitoring device.	13
3.1	A schematic representation of all modules revolving around and interrelating with migraines includes sociodemographic information of migraine patients, sleep characteristics, hemodynamic signals extracted using wearable devices, and the implication of prior knowledge of migraine onset through an alarm, along with additional related information. This representation is utilized to extract information aimed at improving the understanding and management of the disease.	28
3.2	Data pre-processing carried out in the project. Finally, 345 migraine episodes from 51 different patients are used.	30
3.3	Evolution of the pain marked in the smartphone app. The ordinate indicates the subjective pain evolution and the abscissa indicates the number of points marked.	31
3.4	Example of pain marks collected and repaired from a migraine, and the pain reconstruction by a two Gaussian approximations.	32
3.5	Methodology flow followed to find the clustering model that describes the types of pain curves.	33
3.6	Morphology of the pain curve and features.	34
3.7	To compute the areas—or <i>amounts of pain</i> —, the curves are decomposed in rectangles and triangles between the pain marks.	36
3.8	Scattering and distributions of $Acme$, $\delta_{acme\uparrow}$ and D , and their transformations. <i>Migraines (id)</i> identify each migraine.	40
3.9	Example of the schematic representation of a curve model based on morphological features.	42
3.10	System Diagram: This schematic delineates the entirety of the system, elucidating the information flow from the databases to the adviser, from the adviser to the patient, and the process of gathering patient information.	44

3.11	Architecture for Episode-Specific Personalized Advice (ESPA) selector. The input layer of the Recurrent Neural Network (RNN) is partitioned into three distinct blocks. The first block e_p representing the combination of symptoms from all the patient's migraine occurrences, subject to a decay function. The second block e_c is the mean of all the cohort migraines, including those of the patient under consideration. In the figure, the red row are from the patient E_p and the orange are the rest of the cohort patients. The third block c_i is defining the advice to be assessed and scored.	49
3.12	Methodology diagram.	51
3.13	Example with the accelerometry values of 4 postures and turns.	54
3.14	Network architecture with the automatic feature extraction model.	56
3.15	Screenshot of the BrainGuard application specifically designed for migraine studies.	60
3.16	Data pre-processing methodology framework: Methodology and data flow from data collection through edwar to feature generation.	65
3.17	Methodology framework: The methodology developed based on the steps. . .	65
3.18	Generic architecture of the structure for LSTM models.	66
4.1	(<i>minus</i>) k -means score for the k -means clustering algorithm with k varying from 2 to 9 clusters.	70
4.2	Distribution of features per class for $k = 4$ clusters.	72
4.3	Schematic curves of each migraine in gray. The average curve of each cluster is drawn in color overlay.	73
4.4	Classification of types of pain using the supervised LMT model with 99.0% accuracy. Plotting three of four features needed.	76
4.5	The graphs depicts the Mean Squared Error (MSE) or loss of the RNN using the best combination of parameters throughout the epochs for both the training (train) and validation (val) datasets. The x-axis represents the epochs, which denote the number of times the training algorithm goes through the entire dataset during training. The y-axis represents the MSE, which is a measure of the difference between the predicted and actual outputs.	82
4.6	Timeline demo. Representation of ESPA advice for an example patient across an 8-week period, encompassing three migraine episodes, during which the patient's cohort assignment dynamically changes.	83
4.7	Confusion matrix of the RF trained on the dataset with turns labeled. . . .	88
4.8	Confusion matrix of the RF trained on the dataset with turns removed. . . .	88
4.9	Manual Feature Extraction (FE) plus Multilayer Perceptron (MLP) 50 neurons with softmax output. Blue color represent the train-accuracy and cyan color represent validation-accuracy. Awake-sleep model only with patients with EDA and PPG signals with the least amount of noise.	91
4.10	Convolutional Neural Network (CNN) 8 neurons plus MLP 50 neurons with softmax output. Intense color represent the train-accuracy and light color represent validation-accuracy. Awake-sleep model only with patients with EDA and PPG signals with the least amount of noise.	92

4.11	CNN 8 neurons plus MLP 50 neurons with softmax output. Intense colors represent the train-accuracy and light colors represent validation-accuracy. Blue - Periodic Limb Movement disorder (PLM) patients, Red - insomnia patients, Green - Sleep Apnea-Hypopnea Syndrome (SAHS) patients and Yellow control patients	93
4.12	Manual FE plus MLP 50 neurons with softmax output. Intense color represent the train-accuracy and light color represent validation-accuracy. Awake-sleep models with the clusters obtained with the k-means 4 cluster classification. .	94
4.13	Manual FE plus MLP 50 neurons with softmax output. Intense color represent the train-accuracy and light color represent validation-accuracy. Awake-sleep models with the clusters obtained with the k-means 7 cluster classification. .	95
4.14	CNN 8 neurons plus MLP 50 neurons with softmax output. Purple color represent the train-accuracy and green color represent validation-accuracy. Sleep state detection models.	96
4.15	Dataset of the patients included in the study. Red: Pain period. Green: no pain period. Dark green: noisy data.	96
4.16	Variables registered with the Empatica E4 the wearable devices. A) Example of a monitoring period lasting 10 hours with a migraine crisis lasting roughly 2.5 hours. B) Example of the monitoring of a no pain period. The examples show the prediction of the LSTM and the output of the alarm filter. HR: Heart rate. Temperature (°C).	98
4.17	Boxplot of the distribution of the Shapley values of the features for all the models trained for all patients.	99

List of Tables

2.1	Comparison of some examples of Wearable devices and the respective signals they measure, based on their body position, in this case, Head vs. Limbs. . .	12
2.2	Best ML algorithm for disease type [55].	15
3.1	Correlation between features where values ≥ 0.8 are highlighted. Highly correlated features are candidates to be removed.	38
3.2	ESPA topics	46
3.3	Migraine dimensions	46
3.4	BPA target basals (1/2)	48
3.5	Target basals of BPAs (2/2)	48
3.6	Relevant works sleep position detection for comparison	53
3.7	Data labeling distribution	54
3.8	Baseline characteristics of the patients used in this study. P:Patent; F: Female; M:Male; HIT-6: Headache Impact Test. Level of education according to the International Standard Classification of Education (ISCED); 4: Post-secondary non-tertiary education; 5: short-tertiary education. - Part 1	59
3.9	Baseline characteristics of the patients used in this study. P:Patent; BMI: Body Mass Index; CDS: Candesartan; Mg:Magnesium; OnabotA: OnabotulinumtoxinA; TPM:Topiramate; AMT:Amitriptyline. - Part 2	59
3.10	The table documents the notifications sent to patients via the BrainGuard app throughout their involvement in the research.	61
3.11	Algorithm 1 and 2 based on the different state-of-the-art methodologies. . . .	62
4.1	Data distribution for each cluster $c_{k,i}$ for each model C_k	71
4.2	Accuracy, precision, recall and F-measure of the supervised models trained for migraine type classification.	75
4.3	Coefficients of the linear equations of the Logistic Model Tree (LMT) model at the end leafs for each type of migraine.	75
4.4	Pain per cohort	77
4.5	Statistical significance of cohorts regarding sociodemographic variables. The numbering in the columns of each group A, B, C, D, and E refers to how they should be regrouped. All those with a value of 0 will be in the same group, those with a value of 1 in another, and so on.	78

4.6	Statistical significance of patient types regarding triggers. The numbering in the columns of each group A, B, C, D, and E refers to how they should be regrouped. All those with a value of 0 will be in the same group, those with a value of 1 in another, and so on.	79
4.7	Batch number comparison	80
4.8	Arch search	81
4.9	Arch search	81
4.10	Migraines of each type per patient cohort.	84
4.11	List of feature selected.	86
4.12	Mean F-value of record-wise method on new subjects data.	86
4.13	Average f-values of subject-wise 5-cv over 10 runs of different test subjects.	87
4.14	Hyperparameters of RF algorithm.	87
4.15	Hyperparameters of SVM algorithm	89
4.16	Hyperparameters of KNN algorithm	89
4.17	Comparison of metrics from different posture detection studies with the study conducted, indicating the improvement index relative to each.	90
4.18	Monitoring sessions with migraine attacks, TP (True Positive) detection per migraine.	93
4.19	Monitoring sessions in basal periods (± 24 h from migraine), FP (False Positives) detection per minute.	97
1	Distribution of Shapley values by type of sensor	127
2	Distribution of Shapley values by past time window	127
3	Distribution of Shapley values by type of statistical feature	127

Abbreviations and acronyms

ACC	Accelerometry
AE	Auto-Encoder
AI	Artificial Intelligence
ANS	Autonomous Nervous System
APH	Absolute Prediction Horizon
AUC	Area Under the Curve
BCE	Binary Cross Entropy
BCI	Brain-Computer Interface
BMI	Body Mass Index
BPA	Baseline Personalized Advice
CC	Cloud Computing
CNN	Convolutional Neural Network
CV	cross-validation
CWT	Continuous Wavelet Transformation
DARPA	Defense Advanced Research Projects Agency
DL	Deep Learning
DT	Decision Tree
DWT	Discrete Wavelet Transform
ECG	Electrocardiogram
EDA	Electrodermal Activity
EEG	Electroencephalogram
EHR	Electronic Health Records
EMG	Electromyogram
EOG	Electrooculography
ESPA	Episode-Specific Personalized Advice
FC	Fog Computing

fCWT	fast Continuous Wavelet Transformation
FE	Feature Extraction
FL	Federated Learning
FN	False Negatives
FNN	Fuzzy Neural Networks
FP	False Positives
FS	Feature Selection
GA	General Advice
HIT-6	Headache Impact Test
HNS	Hybrid Neural System
HR	Heart Rate
HRV	Heart Rate Variability
IBI	Interbeat Interval
IoT	Internet of Things
IQR	Inter-Quartile Ranges
ITU	International Telecommunication Union
KNN	K-Nearest Neighbor
LDA	Linear Discriminant Analysis
LMT	Logistic Model Tree
LOO	Leave-One-Out
LR	Logistic Regression
LSTM	Long Short-Term Memory Network
MAE	Mean Absolute Error
MLP	Multilayer Perceptron
ML	Machine Learning
MP	Propriospinal myoclonus
MSE	Mean Squared Error
NLP	Natural Language Processing

NN	Neural Network
PLM	Periodic Limb Movement disorder
PPG	Photoplethysmography
PSG	Polysomnography
REM	Rapid Eye Movement
RF	Random Forest
RNN	Recurrent Neural Network
RR	Respiratory Rate
ROC	Receiver Operating Characteristic
SAHS	Sleep Apnea-Hypopnea Syndrome
SCL	Skin Conductance Level
SCR	Skin Conductance Response
SD	Standard Deviation
SMNA	Sudomotor Nerve Activity
<i>SpO₂</i>	Oxygen saturation
SVM	Support Vector Machine
TEMP	Temperature
TL	Transfer Learning
TP	True Positive
WCSS	Within-Cluster Sum of Squares

Chapter 1

Introduction

This work embraces the problem of prediction modeling in the context of digital health. More specifically, this work focuses on the improvement of personalization of digital health prediction models using AI.

There is no consensus as to the definition of AI and even it has changed over the decades. But John McCarthy in [139] defined it as “the science and engineering of making intelligent machines, especially intelligent computer programs”. He says that “it is related to the similar task of using computers to understand human intelligence”, however, it is important to note that “AI does not have to confine itself to methods that are biologically observable”. The AI encloses Machine Learning (ML) and DL (Figure 1.1). ML is the study of computer algorithms that improve automatically through experience [142]. The performance of these algorithms improves as they are exposed to more and more data over time. DL, as a subspace of ML, is defined by [17] as “those methods that are formed by the composition of multiple nonlinear transformations with the goal of yielding more abstract representations”. As an example, DL contemplates different types of the already well known algorithms called NN to solve complex problems.

In any case, despite all being encompassed within the field of AI, we have ultimately failed to create a machine or software that truly embodies intelligence. Additionally, defining appropriate metrics to discern the intelligence or consciousness of an entity proves challenging. Hence, the importance of data selection and the established relationship between the input data for our model and its consequent output cannot be overstated. It is commonly known that ML models train better with more data, but this does not imply that any data is valid. Referencing the term "Garbage-in, Garbage-out" coined in the realm of computer systems, the meticulous selection of data and its relationship with the model's output is crucial [101].

It is de facto assumed that when ML and DL models are created, the data used for their generation and the data used in their use will be on the same feature space. However, although this is true in most cases, sometimes, through Transfer Learning (TL) techniques, we can use ML and DL models in different feature spaces [158]. TL is a technique to transfer knowledge from one domain to another defined by Defense Advanced Research Projects Agency (DARPA) as “the ability of a system to recognize and apply knowledge and skills learned in previous

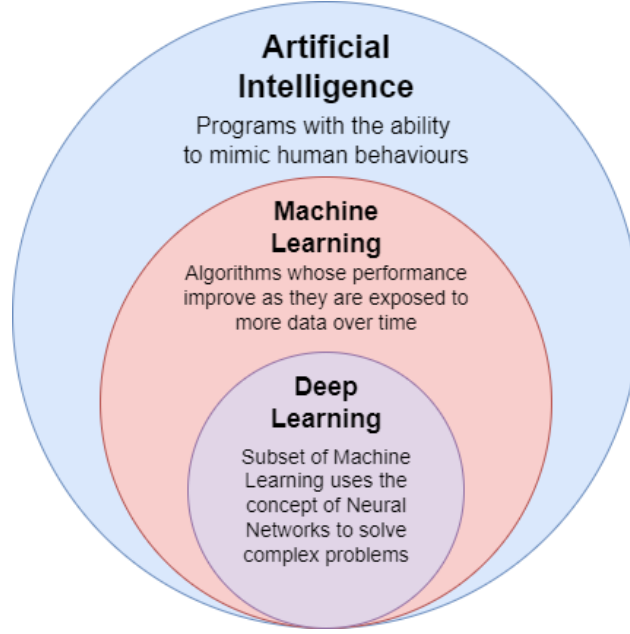


Figure 1.1: Relation between Artificial Intelligence, Machine Learning and Deep Learning.

tasks to novel tasks.” [70].

TL represents ML models as domains and tasks. Where a *domain* consists of a feature space and his marginal probability distribution. And a *task* consists of a label space and a objective predictive function. Given a source domain, a source learning task, a target domain and a target learning task, TL aims to help improve the learning of the target predictive function using the knowledge from the source and domain and task. The source and target domains have different feature spaces and the source and target task have different label spaces [123].

The present research aims to take advantage of the benefits of ML by applying it to the field of AI in health.

The objective of the present manuscript (see Figure 1.2) depicts a methodology to generate knowledge from data gathered from multiple sources applied to a chronic diseases cursing with symptomatic crises such as the migraine disease. Based on this knowledge, predictions can be made and actions taken in order to alleviate or eliminate the symptoms of the crises. This methodology can be divided into parts:

1. The first part consists of the use of data mining techniques for the extraction of biomedical information. This work is going to use: data acquisition from distributed systems for information extraction, FE methodologies for knowledge acquisition, and Feature Selection (FS) algorithms to improve the model performance, the time of execution and thus costs (*i.e.* energy). Data mining is defined by Maimon *et al.* [130] as “the core of the knowledge discovery in databases process, involving the inferring of algorithms that explore the data, develop the model and discover previously unknown patterns”. AI can be employed in data mining processes, utilizing machine learning methodologies to explore data migraine related and discover new patterns. These

patterns, in turn, would enhance the performance of machine learning models used for prediction, whether for symptomatic prevention or diagnosis.

2. The second part involves the training of the model and knowledge acquisition, aiming to find the architecture and hyperparameters that yield the best metrics as outcomes.
3. In the third part is where an actuation is going to be predicted and proposed, and this is closely related to clinical intervention. This is reason why the development and implementation of the methodology of this Thesis is intended to be in collaboration with clinicians.

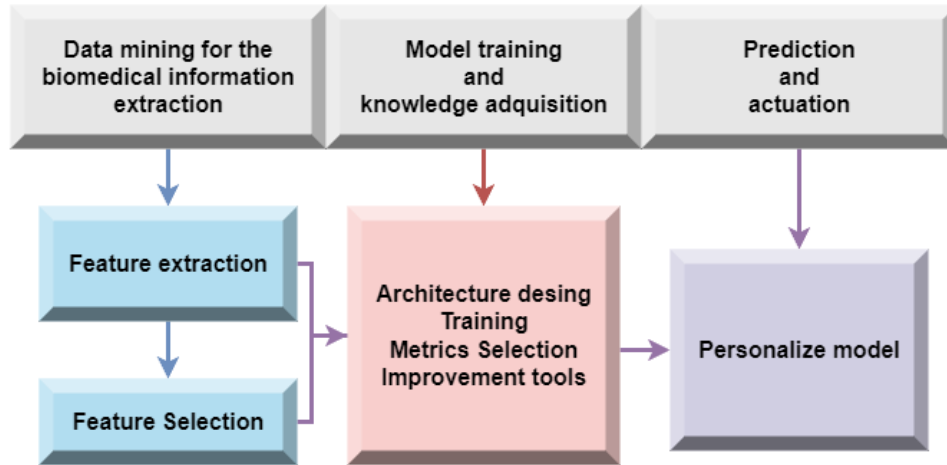


Figure 1.2: Diagram of the methodology and objectives of the thesis proposed.

Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL) algorithms are becoming more and more popular, and these algorithms are continuously appearing as basic tools in different areas of research. These algorithms give the opportunity to bring knowledge closer to each person. Consequently, the number of peer-reviewed AI publications is increasing. The “The AI Index 2023 Annual Report”, of the Stanford University [135] shows that the total number of AI related publications were around 200,000 in 2010 while in 2021 the number increased up to 500,000. So in the past decade, the volume of publications in the field of AI has more than doubled.

The evolution of the Internet of Things (IoT) makes possible the application of AI in different areas. The IoT is defined by the International Telecommunication Union (ITU) [1] as the “global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies”. The IoT has penetrated into all spheres of our society: health, agriculture, industry... Angelopoulos *et al.* in [9] presented a survey discussing about the fourth industrial revolution, which bases on the Industrial IoT and the effects of AI and ML in this industry. L. Chen *et al.* [37] mentioned that the use of platforms and tools with AI have enabled and improved teaching effectiveness and efficiency, which make the instructional quality better and richer. G. Bannerjee *et al.* in [16] show more than 100 important contributions of AI in different fields. What is more, the presence of AI becomes a

topic in the daily life. As an example, R. S. Lee discusses about the AI in the daily life in his book “*Artificial intelligence in daily life*” [112]. This thesis will focus on the use of AI on IoT in healthcare.

The application of AI in healthcare is synchronized with the transformations of the medicine in the last decades. The medicine has evolved rapidly from what has been called P0 medicine—strongly focused on the physician—to P7 medicine. Originally mentioned by Hood L. *et al.* in [86], they authors describe the P4 medicine as: Predictive, Preventive, Personalized and Participatory. This P4 medicine is the most widespread and established modern medicine in developed countries. Currently, we are in a situation to bring P5 medicine to the entire population in a democratic way. P5 medicine adds the Psychological point to the equation. P5 medicine is the bridge to the democratization (Public) of modern precision medicine—what we call 6P medicine—. P6 medicine can be found actually implemented in a number of cases [27], and although it has only reached a small part of the population, there are already new proposals for the definition of modern medicine P7.

Recently, James A. described the P7 medicine in [131]. P7 medicine tries to answer the critics about the *personalized* P in P4-P6 medicine, regarding the misconception of *personalized*. In words of James A.: “the patient *qua* person is not part of the algorithmic framework but rather the patient simply becomes a calculated entity”. This is something that Horwitz *et al.* already stated in [87] and called is the impersonal or (de)personal medicine. James A. reorder the *Ps*, but the most important aspect lies in his definition of the P7 medicine to humanize the medicine. To do so, he proposes to divide the *Personalized P* in (i) *Physico-biomolecular* which emphasizes the genomic and biological dimensions of the patient and contemplates the *personalization*, and (ii) the *Personal* medicine “which involves the notion of personhood in a humanistic rather than simply a technical way”.

As mentioned above, P4 medicine is the description of modern medicine that most closely matches reality, and it is the one this thesis will focus on. The most important aspects of P4 medicine that can lead to synergies with AI are: Predictive, Personalized (or Physico-biomolecular and Personal in P7 medicine), and Participatory.

The information that patients provide to the AI systems enables the personalization of their treatment, prediction of symptomatic crises, or even the recommendation of guidelines and behaviours in order to prevent certain pathology [51], [154], [222]. The wealth of the data that are generated in healthcare poses considerable challenges for their analysis. According to [85], some of these challenges are: (i) integration interactive data by data fusion and pre-selection, (ii) interactive sampling, data cleaning, preprocessing and mapping, (iii) advance in data mining technologies by pattern discovery, and (iv) visualization, decision support and analytic. These are some of the aspects that will be addressed in this thesis.

The importance of applying AI techniques to chronic diseases has already been suggested in this manuscript, and these are the conditions to address. In particular, and mainly this work will focus on the migraine disease which will be described later, in Chapter 2.3. Chronic diseases affect to a large percentage of the population worldwide [219]. This leads to high economic costs for healthcare systems. The migraine is a chronic disease and consists in a recurrent headache disorder. Currently there is no cure for migraine. Migraine sufferers

take preventive treatments to reduce the number of monthly crises, however these does not disappear totally. For the remaining crises diagnosed patients take what are called *rescue treatments* or medication for the acute phase. However, the treatment only is effective if the pills are taken before the pain starts, which is something that patients do not know when it will occur. In addition, the symptomatic crises of migraine might be different from one patient to another. This disparity, common in chronic diseases, leads to complexity on the predictive models.

Therefore this work aims to develop a methodology to generate personalized predictive models in migraine in order to warn the patient before the onset of pain to take the pill and avoid the pain or reduce the intensity considerably. Migraine pain modeling and prediction of crises was already demonstrated [155]. Previous work that objective pain models can be generated and these can be predicted using biomedical variables. On this basis and in short, this Thesis aims to (i) take a small step forward in research related to creating population models that can be transferred to individuals and thus personalize them, and (ii) integrate state-of-the-art unobtrusive and low-cost wearable devices, (iii) acquire knowledge to enhance the performance of population prediction models.

1.1 Machine Learning and Deep Learning in healthcare

As it was mentioned, the use of ML algorithms has gained popularity in healthcare over the last decade; there are thousands of examples. The use of classic ML algorithms can be found in studies such as that the ones of Stokes *et al.* [194] for supporting prognosis of bronchitis and pneumonia; Richard *et al.* [179] to predict risk of cancer; or the one of Leighton *et al.* [114] for the classification and prediction of the one-year outcome after a first episode of psychosis. ML models work well for datasets obtained from surveys, socio-demographic data or qualitative dataset in general. However, if this work aims to use time-series data — such as biosignals (*e.g.* Electroencephalogram (EEG), Electrocardiogram (ECG) or Electromyogram (EMG))—, DL models perform better.

In the use of DL models, many works can be highlighted as well. As an example, Syed *et al.* [8] present the use of CNN—a DL algorithm—for detection and classification of pathology on the monitoring of EEG. DL algorithms can be used for text recognition and drawing movements applied to healthcare. To this end Gil *et al.* [71] proposed a system based on CNN for the detection of Parkinson’s disease from drawing movements. Related to text, but oriented to typewriting, Elena Davcheva in [46] uses LSTM—a DL algorithm—for classifying mental health conditions by identifying symptoms getting data from posts in social media.

More complex methodologies combining ML and DL are also very common. The authors [144] claim the use of ML neuro fuzzy-based algorithms and DL Boltzmann belief networks for the prediction of diseases based on the monitoring of multiple vital signs of patients within the combination of information from electronic health records.

The combination of ML and DL algorithms is also suitable for the problem of feature extraction as shown by Guyon *et al.* [77]. The feature extraction aims to select the variables that better describe a problem before training a model. This process (i) can be guided manually by using

classic ML algorithms [180] such as: Decision Tree (DT), Random Forest (RF), AdaBoost, Linear Discriminant Analysis (LDA), K-Nearest Neighbor (KNN), or (ii) can be automatic by using DL algorithms by including the process of feature selection while creating the model, such as Wan *et al.* do in [208] using CNN.

Dealing with TL problems, the work of Akbari *et al.* [4] poses two goals: (i) to use TL under constrained data, and (ii) boosting accuracy by means of TL. The use of constrained data could be categorized into (i) the instance-based transfer, (ii) feature representation transfer, (iii) classifier-based transfer, or (iv) hybrid knowledge transfer. Classifier-based TL approaches try to learn a new model that minimizes the generalization error in the target domain. This methodology could be used to resolve the the problem of personalization of models. Yiquiang Chen *et al.* in [39] present a framework to apply this methodology. They represent the general model as the source task and the personalized model with the data of the patient as the target task and domain.

A further layer in the field of knowledge transfer is the use of Federated Learning (FL). FL is a decentralized approach compatible with Fog Computing (FC), leaving the data on the origin and learning a shared model by aggregating updates improving the privacy sensitivity and in a first approach for data security. The isolation of the data due to laws in healthcare is a concern. Laws like the General Data Protection Regulation (GDPR) [45] enforced by the European Union on May 25, 2018, do not allow to share all the data recorded. This problem was already thought and a solution was proposed by Qiang Yang *et al.* in [218]. They use a FL framework—firstly proposed by Google in 2016 [141]—, to tackle this problem. This study explains the paradigm of Federated-Transfer Learning to bring a solution for the security and model personalization problems. However, the security concern is not fully resolved with this solution. Chunyi Zhoy *et al.* showed in [226] that FL do not resist collusion attacks launched by multiple malicious entities. But they resolve this problem by using a combination of FL and FC.

1.2 Wearable devices in healthcare

As it was said, wearable devices are key elements for data generation and knowledge-enablers on IoT applied to mobile health. Lukasz Piwek *et al.* in [161] showed how wearable devices have established as utilities in lifestyle technologies or gaming and entertainment industries, as an example. However, the authors stated that in the healthcare environment the usefulness of wearable devices may be questioned. They point out that mass-marketed gadgets that could “drift into obscurity” if both untrained practitioner and patient face with these new technologies where data is prone to error and somehow unreliable. But they also conclude that “if frameworks are in place allowing wearable devices to be integrated into health care systems, this could, in turn, kick-start the development of validation programmes that would sit alongside appropriate training for health care professionals” and “wearable technology can become a valuable asset for health care in the 21st century”.

Indeed, Matthew Smuck *et al.* in [191] indicated in 2021 that the use of wearable devices in healthcare environments is expected to increase. Hence wearable devices are being more and more commonly used in health analysis, assessment, monitoring and schedule. They offer not

only a personalized assistance, upgrading the health lifestyle, but also monitoring metabolic clinic signals, diagnostic and treatment actively in an unobtrusive way.

If there is one thing that characterizes wearable devices, it is that they are non-intrusive (they do not bother or interfere with your daily life while wearing them), although they may or may not be invasive—they involve an implant, such as a pacemaker, or a puncture, such as some diabetes monitoring devices—. Wearable devices can be smart (they process the data they monitor and generate knowledge), or not (they only report the physical magnitudes they measure). There are a lot of kind of wearable devices; some of them are flexible, stretchable or printable on skin [225]. These characteristics allow them to monitor physiological and biochemical users' biosignals in an unobtrusive way.

A wristband is one the most suitable and popular form factors for the acquisition of biomedical signals [42]. Wristband devices are easy to find in the mass market as *smartwatches* or other lifestyle products, but they are also very common in research. The main differences between *commercial* and *research-oriented* wearable devices are: (i) the look-and-feel: while the former are oriented to a mass market with a very polished marketing design, the latest are oriented to be functional keep the esthetic aspects aside; and (ii) the access to raw data: this is a key point for researchers, while accessing to data generated by *commercial* devices may be hard or impossible in most of the cases, gathering the raw data from research-oriented wearable monitoring devices is a must. Examples of companies that have developed research-oriented devices for healthcare are: Shimmer¹, Biovotion (not research oriented anymore)² [221], or Empatica E4³. These devices might be equipped with electrodermal activity (EDA) sensors, photoplethysmography (PPG) sensors, or temperature and kinematic sensors among others. Accessing and processing the data gathered by these sensors to apply ML techniques to extract knowledge, drives the increasingly real implementation of the P4 medicine.

The development of this Thesis will involve the use of wearable devices such as the Empatica E4.

1.3 Motivation

Industry 4.0 is the transformation of the industry of the 20th century to the novel industry of the 21th century. This transformation integrates (i) cyber-physical systems, (ii) IoT services, (iii) cloud-based business processes, and (iv) big data and AI for the extraction of knowledge. The extension of Industry 4.0 to the health sector is what is called the Healthcare 4.0 [116].

The increase of the IoT devices connected around the world opens the opportunity to collect data in real time and process them with novel analytic techniques thanks to the capabilities of Cloud Computing (CC) on the management of these data. Figure 1.3 shows an example of this fast growth for wearable devices only.

¹Shimmer device: <https://shimmersensing.com/>. Accessed: July 15, 2022

²Original Biovotion device: <https://www.wearable-technologies.com/2015/04/sustainability-is-key/>. Current company that bought the technology of Biovotion for a different market strategy: <https://biofourmis.com/>. Accessed: July 15, 2022

³Empatica E4: <https://www.empatica.com/en-eu/research/e4/>. Accessed: July 15, 2022

The growth of the amount of data collected is not stopping and the capabilities of CC are not enough to tackle the problems that show up like [14]: (i) the increase of latency, (ii) low capacity (how much workload the system can hold), (iii) network failures, or (iv) security. The problems of high latency, low capacity and network failures are suitable to be resolved by means of the new paradigm of the FC bringing data processing and storage locally, to the network edge. A priori, the FC environment is inherently less secure than CC due to its decentralized nature, which poses challenges for applying centralized security and privacy mechanisms [14].

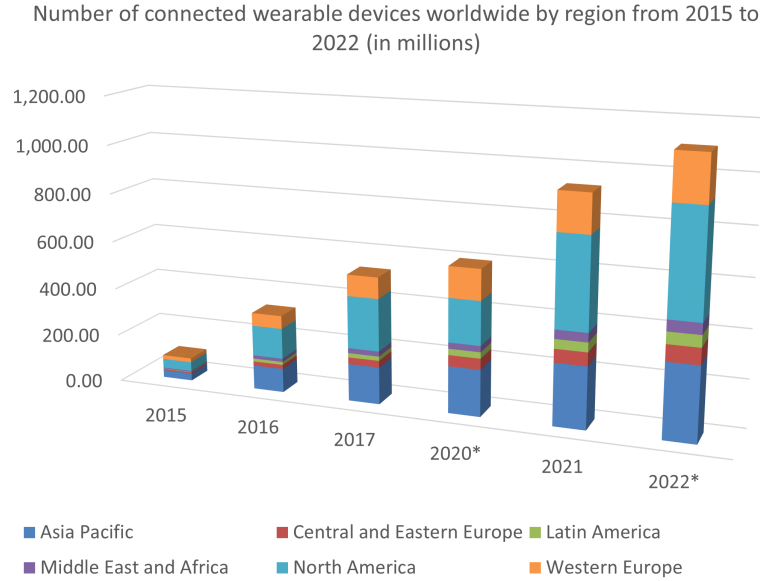


Figure 1.3: Number active wearable devices by region [192].

Just as AI, ML, and DL techniques find application in the healthcare sector, the proliferation of IoT devices has also influenced the healthcare landscape. In this context, one direct application would be the monitoring of chronic diseases.

Chronic conditions are the primary causes of disability in the world. Chronic diseases are, as an example, heart disease, many neurological diseases, or diabetes. Some of these diseases curse with symptomatic crises, and it becomes important to predict the onset of symptomatic crises to improve the quality of live of patients and reduce costs. 70% to 80% of all healthcare costs in the EU—estimated at €700 billion [171]—are currently spent on chronic diseases, while in the United States this cost reaches \$3.8 trillion [132]. The migraine is a neurological disease that affects around 12% of the population worldwide [219] and leads to high economic costs for private and national health systems [23], [72]. Only in the European Community, the migraine costs grow up to more than €27 billion per year [195].

The migraine is a recurrent headache disorder manifesting in attacks typically lasting from 4 to 72 hours of moderate or severe intensity, and aggravated by routine physical activity presenting nausea and/or photophobia and phonophobia [12]. It is long recognized that migraine is more than just the attacks and that migraine should be understood as a cycling brain disorder with at least 4 phases: (i) prodromal, (ii) aura, (iii) headache and (iv) postdromal [20]. Among

chronic daily headach patients, those with migraine (chronic migraine, transformed migraine, and chronic daily headach with migraine) had the lowest quality of life. The findings suggested that chronic daily headach leads to greater productivity loss compared to episodic headache, especially when migraine features or medication overuse are present. These affects the quality of life of migraine sufferers [109] leading to the stigmatization of the patients [134], and can increasing the risk of suicidal behaviours in sufferers [35]. Currently, there is no cure for the migraine, and patients take the pills when it is too late, dismissing the effectiveness of the treatments what can lead to over-medication [19] and thus the chronification of the disease [207]. Therefore, the prediction of pain crisis could advance the intake of the medication and avoid the pain before it happens [155]. Stating that this prediction may be complex is justified, as pain is defined as a subjective magnitude from the patient’s perspective and, consequently, difficult to measure. Meanwhile, the temporal variables to be used for its prediction are entirely objective, making this task a challenge for research.

In view of the introduction that explains the context in which this Thesis is developed, the most relevant motivations of this dissertation are: (i) the use of low-cost **wearable devices** (IoT elements) in the Healthcare 4.0 sector, (ii) to which data can be applied **ML and DL algorithms for predictive modeling** of symptomatic crises in chronic diseases—such as **migraine**—(iii) where to introduce **TL for the transfer of knowledge** from population models to **personalized models** and (iv) employing ML and DL techniques to enhance our understanding and/or information pertaining to the disease, thereby facilitating improved disease management.

1.4 Objectives of this work

Chronic diseases, particularly those involving recurring symptomatic crises, significantly impact patients’ quality of life. Understanding and managing these symptoms, especially pain, is crucial for developing effective treatment strategies. Recent advancements in health technology offer promising avenues for improving patient support and outcomes. This study focuses on leveraging personalized approaches to better understand and predict symptoms, particularly migraines, by combining symptom characterization, sleep analysis, and hemodynamic monitoring. The following objectives guide our exploration:

- To characterize and classify different types of pain experienced during symptomatic crises in chronic diseases.
- To develop a personalized recommendation system that supports patients based on their unique symptoms and needs.
- To create a model that detects sleep states and body posture, enabling analysis of sleep quality and its relationship to migraine occurrences.
- To predict migraine crises through the analysis of hemodynamic signals, improving early intervention and management strategies.

This comprehensive approach aims to enhance patient care by offering tailored insights into pain, sleep, and migraine prediction.

1.5 List of publications

In this section, the publications made during the doctoral thesis are listed. Considering the interdisciplinary nature of the work, technical papers related to AI as well as clinical papers related to the studies conducted could be found.

List of main publications from this work:

- Cluster-then-classify methodology for the identification of pain episodes in chronic diseases [66]. JCR (7.3)
- Reproducible and accurate subject-wise sleep posture detection by detecting and removing turns [65]

List of publications related to this work:

- A study awarded by the Madrid Neurology Association. “¿Cómo es la curva de dolor en una crisis de migraña?” XIX Reunión anual de la Asociación Madrileña de Neurología, (2021): Póster. [63]
- A paper in the annual meeting of Asociación Madrileña de neurología: “¿Es posible predecir una crisis de Migraña? Estudio prospectivo para valorar la capacidad de predicción en tiempo real, a través de variables hemodinámicas, mediante un dispositivo portátil de muñeca.” [62]
- A paper in the annual meeting of Sociedad Española de neurología: “Predicción de una crisis de migraña utilizando variables hemodinámicas registradas en tiempo real mediante dispositivos portátiles de muñeca.” [74]
- Prediction of a migraine crisis in real-time using a wearable device in Journal Of Headache And Pain 23 (SUPPL 1) [75]
- Classification of the pain curve and semiology of a migraine crisis in Journal Of Headache And Pain 23 (SUPPL 1) [73]
- Monitoring differences in the function of the autonomic nervous system in patients with chronic insomnia using a wearable device [212] JCR (3.8)

Chapter 2

State of the art

This chapter is going to make a more detailed review regarding (i) the monitoring devices, (ii) ML and DL algorithms, and (iii) FC architectures to be used in this Thesis and mentioned in Chapter 1.

Wearable devices are an easy, cheap, unobtrusive and ambulatory solution to record biosignals. There are a lot of different kinds of wearable devices. Wearable devices could be classified by their location on the body or by the biosignals they record. In Section 2.1, some wearable devices are compared regarding the categories mentioned. Once the data are collected, they have to be cleaned (filtered) and formatted, and thus the processed data could be used as input to feed the AI models. The format depends on the algorithm and structure of the model. In Section 2.2 there are explained some of the most common algorithms used in healthcare for prediction or diagnosis. In addition, some techniques and methodologies for the improvement of these algorithms are analyzed. Section 2.3 will briefly discuss the chosen chronic disease as the focal point of our thesis, providing justification for its selection and delving into its characteristics, which will be more thoroughly examined in the subsequent Chapter 3.

2.1 Wearable devices

Let's recall that the significance of wearable devices in this study lies in their role in facilitating continuous ambulatory monitoring of patients in the most precise and non-intrusive manner possible. Hence, the potential locations of various devices and the corresponding signals they can measure will be analyzed and compared.

Table 2.1 shows some examples of wearable devices grouped by its location on body. This table contains the different biosignals measured by each device. The location of the wearable devices is classified in two big groups: the head and the limbs; and each group is divided as well. The head group contains: glasses, earplug and hearables, and the limbs group contains: armband and wristband. The signals referred are PPG, EDA, ECG, EEG, EMG, Electrooculography (EOG), HR, Temperature (TEMP) and accelerometry.

This work focuses on the use of the data to predict symptomatic crises in chronic diseases. The use case scenario is the migraine disease. Migraine crises are related with the Autonomous

Table 2.1: Comparison of some examples of Wearable devices and the respective signals they measure, based on their body position, in this case, Head vs. Limbs.

Signals	Wearable				
	Head			Limbs	
	Glasses	Earplug	Hearables	Armband	Wristband
PPG	✓		✓		✓
EDA	✓			✓	✓
ECG	✓				
EEG		✓	✓		
EMG		✓	✓		
EOG		✓	✓		
HR				✓	
TEMP				✓	✓
Motion	✓		✓	✓	✓

Nervous System (ANS) [21], which is related with the body TEMP [82], the HR and the EDA [120]. Previous researchers already used these features to monitor and create the former migraine predictive proto-models [89], [153]. Consequently, devices that integrate these signals will be used. As shown in Table 2.1, the EDA, HR and TEMP can be measured in the armband [119]. However it is also possible to compute the HR from the PPG [7]. In order to use more comfortable wearable devices, wristband-like devices are the ones to be considered. J. Cho *et al.* in [41] listed the preferred locations for wearing a wearable device, being the most preferred location the wrist (65%). One advantage of the wristbands—that also have armbands and earplugs [146]—over glasses and hearables [162], [204], is the comfort and easy wearing during sleep, which is something to be considered in our studies.

After an exhaustive search of the desirable wristband device integrating the preferred sensors, the Empatica E4 has been chosen. Figure 2.1 shows a picture of the Empatica E4 wristband device with the axes of Accelerometry (ACC) on it. Many studies corroborate the feasibility and utility of Empatica E4 wristband device [25], [138], [181]. A. Borrego *et al.* in [25] show the reliability of the Empatica E4 device to measure EDA. The authors pointed out that one of the weak points of the Empatica E4 is that the sampling rate of the sensors is not configurable and for some applications might be low, but not for ours. The Empatica E4 implements the following sensors with these sampling rate: TEMP (4 *Hz*), EDA (4 *Hz*), PPG (64 *Hz*), and ACC (32 *Hz*). Numerous studies corroborate its effectiveness as a migraine prediction device [103], which aligns with one of the primary objectives of this thesis. The Empatica E4 wristband allows for the monitoring of ANS-controlled variables, including distal skin temperature, HR via PPG, EDA, and 3-axis ACC, even in suboptimal locations like the wrist.



Figure 2.1: Empatica E4. The most comfortable and accurate wristband to monitor physiological signals in real time. A great tool for research applications in daily life.

The PPG is a non-invasive technique with light source and a sensor typically in the frequency bands of red, green and infrared (IR). The light measurement indicates the blood flow under the skin. As mentioned above, from the PPG can be computed the HR as well as the Heart Rate Variability (HRV), but also other features like the Respiratory Rate (RR), blood pressure, or Oxygen saturation (SpO_2) (two channels are needed: red and IR). McCarthy *et al.* in [138] validated the PPG signal gathered from the Empatica E4 compared to an ECG signal obtained from the General Electric's SEER Light Extend Recorder Holter, a portable electrocardiograph commonly used in the clinical practice and considered as the ground truth for ambulatory monitoring of ECG. The authors proved that the performance of both devices was consistent on 85% of the analyzed segments. A. Schuurmans *et al.* in [181] compared both HR and HRV signals computed from the Empatica E4 against VU University Ambulatory Monitoring System (VU-AMS). They conclude that the Empatica E4 monitoring device is suitable to measure HRV in an unobtrusive way, and it is a valid tool for research on HR and HRV in ambulatory scenarios—performing better in non-movement conditions—.

The EDA can be decomposed into 2 signals: the tonic (Skin Conductance Level (SCL)) and the phasic components (Skin Conductance Response (SCR)). The phasic component (SCR) represents the slow variations of the EDA, while the tonic component (SCL) represents the rapid variations [98], [201]. The Empatica device measures the EDA, which is the response measured on the skin (sweating) of the activation (peaks) of the Sudomotor Nerve Activity (SMNA). Previous research [76], proposed an optimization method to extract both SCR and SCL, and the SMNA too, from the EDA.

Notice the reader that the skin temperature measured on the wrist (distal temperature) does not represent the body-core temperature and it is affected by external (atmospheric) changes. The relative variations do will reflect changes of interest, however the absolute value would not have to be considered as the real status of temperature of the subjects.

2.2 Artificial Intelligence in health

ML and DL are powerful tools for data analytic, classification, regression and other applications on complex information extraction. The use of these tools ML and DL are already a reality and they are working in many areas such as the Industry 4.0 leading to increase the quality of services, ease the human processing and predicting the results of certain tasks.

ML techniques are helping in interpretation, diagnosis and decision support for healthcare professionals whose most common tools are currently based on simple statistical analyses. “ML-based analyses often seek nonlinear relationships among hundreds or thousands of covariates” Wiens *et al.* [211] say. To find complex and relationships DL-based artificial neural networks are pretty much accurate than classic statistical analyses. These technique allow the development of data-driven solutions as well as the automatic generation of features, reducing the human intervention and avoiding the human error factor [169]. One such technology is digital twins. A digital twin is a virtual representation of a physical object, enriched with real-time data captured through sensors or data analytics technologies. This technology integrates the IoT, AI, ML, and data analysis [5].

2.2.1 Machine Learning in health

Since ML techniques started to be used, they have been applied to healthcare [102]. ML can be applied for data analysis, diagnosis, or prediction [55]. ML algorithms can be classified from different points of view, and one of them is by the the way that the optimization kernels of the algorithms use to tune their hyperparameters, the training process. ML algorithms can be (i) supervised: when you train the model using labeled data, (ii) unsupervised: when you let the model to cluster the data by its own based on data statistics without providing labeling, (iii) semi-supervised: a mix of supervised and unsupervised where only a few samples of the dataset are labeled, and (iv) reinforced: an iterative process that takes feedback from the environment to improve the performance based on the achievement of targets [48]. Unsupervised ML algorithms are normally use for data analysis while supervised ML algorithms are normally use for diagnosis and prediction.

There can be found many examples of knowledge acquisition by means of the analysis of biomedical data using unsupervised ML can be found in the genetics field. However, examples of this field will not be mentioned because it has many particularities. Some examples of unsupervised ML whose datasets can be more similar to some of those to be used in this thesis are: the one presented by Bose *et al.* in [26] where the authors try to find groups (clusters) monitoring health of patients at home using *telehealth*, and their analysis revealed patterns of association between certain features like obesity, mental health and pulmonary disorders. Another example of unsupervised ML is the one shown by Wang *et al.* [209]. The authors implement a modified Latent Dirichlet Allocation algorithm—an unsupervised algorithm commonly used in Natural Language Processing (NLP) problems—to discover in an osteoporosis cohort study that the psychosis leads to substantial difference in the survival curves of the participants. This disclosure implied that the *risedronate* (a commonly used medication for the treatment of osteoporosis) could trigger psychiatric side effects, and it might be reconsidered its use as treatment.

For diagnosis and prediction it is commonly used supervised ML. Krishnan *et al.* [104] presented a KNN-based model—a supervised ML algorithm—to find the relationship between genes and the autism spectrum disorder. The study of Krishnan *et al.* helped in the forecasting of risk in the autism disorder [31]. Ferdous *et al.* in [55] elaborate a survey with the best ML algorithms for different diseases. Table 2.2 shows a summary of different algorithms for each disease thanks to the work of the authors.

Disease	Best algorithm
Liver Disease	Logistic Regression (LR)
Diabetes Mellitus	RF
Breast Cancer	Support Vector Machine (SVM)
Sickle Cell Disease	Sequential Minimal Optimization
Heart Disease	Fuzzy Neural Networks (FNN)
Cardiac Arrest	NN
Thyroid Disease	SVM
Alzheimer’s disease	NN
Chronic Kidney Disease	Naive Bayes
Ear Disorder	SVM
Eczema	AdaBoost
General Disease Prediction	CNN
Chronic Disease	RNN

Table 2.2: Best ML algorithm for disease type [55].

An important part of designing ML models is the data, its quality shape, and nature. There are a lot of different kinds of datasets. In healthcare there can be found specific of data type like clinical, omics, and sensor (time-series) data. Clinical data are typically enclosed as Electronic Health Records (EHR). EHR contain socio-demographic information and the results of the various medical tests, and the treatments of patients. Omics data include genomics, proteomics, transcriptomics, pharmacogenomics and phenomics, proteins RNAs, genome drugs effect and phenotype information. The volume and dimension of omics data are large. Genomics studies the structure and function of genomes, while transcriptomics examines all messenger RNA molecules in an organism. Proteomics investigates proteins, their functions, and changes, and metabolomics explores metabolites, the end products of cellular processes. The goal of omics sciences is to identify, characterize, and quantify all biological molecules involved in the functioning and dynamics of cells, tissues, or organisms. These are advanced theories for personalized healthcare and emphasize the importance of holistic methodologies in the prevention and management of symptoms and chronic diseases [205].

Sensor data comes from different kind of sensor recording signals like ACC, ECG, EDA, EEG, EMG, PPG... Dhillon and Singh in [48] presented a survey analyzing the different types of medical data. They prepared a summary according to the type of data and listed the different metrics that can be used to validate the models usually. Examples of these metrics are: Cohen’s kappa, F1 score, area under the Receiver Operating Characteristic (ROC) curve

or Area Under the Curve (AUC), accuracy, sensitivity, specificity and precision. This research is going to focus on sensor data mainly whereas it might be used data gathered from EHR as well.

Each kind of dataset needs a different pre-process methodology. One step after recording/-gathering the data is cleaning and formatting them. The objective of formatting the data is to avoid false information due to noise, improbable values and missing values. The treatment of missing values and outliers are general techniques applicable to any dataset. Missing value imputation and outlier detection techniques could be classified in two types: (i) statistical and (ii) ML-based. Lin *et al.* [122] mentioned that the most widely used statistical techniques to apply missing value imputation are: expectation-maximization, linear regression, least squares and mean/mode. And the ML based techniques most widely used for missing value imputation are: clustering, decision-trees, KNN and RF. Gaddam *et al.* in [61] summarized the different outlier detection techniques and discussed pros and cons of each one of them.

Signal data processing is another step in the path of data preprocessing. This step is specific for each signal. For example a signal of our concern is the EDA. The EDA is a complex signal which has been studied over the last decades (see 2.1 for further details). Posada-Quintero *et al.* [164] provided a review on the processing of the EDA. Each one of the components of the EDA signal contains different information. For example, in some cases, the amplitude of event-related SCR signal could infer sympathetic arousal, and using a peak detector algorithm over the SCR signal, the bursts of SMNA could be detected.

Finally, after cleaning and filtering the data, it is commonly needed in time-series signals to perform a FE. This due to the fact that most of ML algorithms are timeless. This is why in most cases, wrapping and FE techniques are used. For example, Sundararajan *et al.* [197], and Sano *et al.* [177] performed sleep detection/classification models using ML. In these works the authors manually extracted different features on windows of the monitored signals. The extracted features can belong: to the time domain such as the mean, Standard Deviation (SD), maximum and minimum values *etc.*, to the frequency domain like spectral power or energy, frequency peaks *etc.*, or to any other domain (*e.g.* Wavelet domain).

Other examples of manual FE are shown in the studies conducted by Romine *et al.* [173] and Aqajari *et al.* [10]. In the former, the authors measured students' cognitive load using the mean and SD of the EDA, HR and distal temperature. In the latest, the authors used the EDA signal for detection of pain in uncommunicative postoperative patients. They based their FE process in another study [11] in which study they used the EDA to predict stress. They performed two processes (i) a *manual* FE where they decomposed the EDA signal in its two components, phasic and tonic using the framework *cvxEDA* [76] to extract statistical features: the number of peaks, the mean of the EDA, and the max peak amplitude, and (ii) an automatic FE process in which they used DL algorithms.

2.2.2 Deep Learning in health

DL is a sub-field of ML and “NNs make up the backbone of deep learning algorithms. In fact, it is the number of node layers, or depth, of neural networks that distinguishes a single neural

network from a deep learning algorithm, which must have more than three”¹. DL is used to find complex relationships between input variables. NNs mimic the human neural system. NNs are composed of interconnected cells where each cell is a function in which inputs are the outputs of other cells, and its output is the input for subsequent cells. The architecture of the network is the diagram of all the cells and the mentioned connections. There are different types of NNs classified depending on how the connections are made and the functions of the cells.

As mentioned above, in [11], the NNs were used to perform automatic FE of time-series signals. Långkvist *et al.* in [108] proposed a sleep stage classification using EEG and EOG signals as raw inputs, and they defined a Deep Belief Network as classification model without computing features manually.

An important network architecture is the one of Auto-Encoder (AE)s. The AE network consists of two parts: (i) an encoder, and (ii) a decoder. The first part encodes the networks inputs reducing their dimension, creating a representation of the data in an arbitrary domain that contains the latent information of the original inputs. This encoded data feed the second part of the AE which decodes the encoded output, recovering the original signal. The purpose of this second part is to compare the decoded signal with the original one so that the encoder part of the AE encodes the information in the best possible way without much loss of information. Thus, one of the various functionalities of the AEs is to perform automatic FE—where the features are the intermediate encoded vector of information—. Thirukovalluru *et al.* perform an experiment and analysis on 5 datasets to show how it is possible to obtain good feature representations and good classification performances using a modified AE [203].

As already mentioned, a functionality of DL algorithms is to analyze complex non-linear problems. For example, Faust *et al.* in [53] used LSTM for automated detection of atrial fibrillation. LSTM networks are part of RNN. Generally speaking, simple ML models assume that the inputs and the outputs of the data of time-series signals do not have inter-dependence (the position where it occurs does not matter). RNN began to gain interest when this assumption ceased to be true. Problems like NLP do not fulfill this assumption. RNN are a kind of networks which can retain information about the previous *states* of the signal. Normally, the first RNN retain long-term information. LSTM models appear to retain long and short past events information. In the study of Faust *et al.* a DL algorithm manage time dependency of the data and with that FE is not needed whit this approach.

The CNN are other kind of RNN. While RNNs are useful for time series in predicting the next output, CNNs do not have the ability to find the relationship between elements of a series (as in a sentence) and are more commonly used for classification. CNNs are usually used to analyze 2D data. In [170] and [136] the authors use CNNs to analyze computed tomography and magnetic resonance images to detect and classify head strokes and diagnosis of liver pathology respectively. CNN can be also used with 1D signals.

¹IBM AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What’s the Difference?: <https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks>. Accessed: July 15, 2022.

Studies using CNNs on 1D time series data are not as common, but there are also numerous examples. Much work can be found in applications using EEG brain signals—large amounts of data—such as for the classification of epileptic seizures [176], the classification of motor imaginations tasks in Brain-Computer Interface (BCI) applications [127], or ACC activity signals—complex outputs with multiple possible classifications—such as the authors in [79] did for gait parameter extraction, or the work presented by Sathyanarayana *et al.* in [178] for the classification of sleep quality and sleep efficiency. In other research fields it is also frequent to find some studies as the real-time motor fault detection proposed by Ince *et al.* [92]. More examples can be found from the review of Bolhasani *et al.* in [22].

2.2.3 Transfer Learning in health

TL uses the knowledge obtained in a ML procedure and it applies this knowledge in other but similar procedure. A *procedure* can be understood as an application domain. TL could be used to improve the performance of the procedure as well. For example, in cases where there are not enough data, TL allows to use the knowledge obtained from a similar procedure with more data recorded. In [2] Jaoude *et al.* implemented an automated sleep stage classification model using as input the EEG signal. In order to train a model that would receive as input data the EEG from an ambulatory scalp for which they did not have much data, they trained a system with EEG collected in a clinical environment with a larger dataset—6,341 patients—and then applied TL and fine tuned the parameters using the small dataset—112 patients—from the ambulatory process. Their results showed that the TL methodology improved the results for the small ambulatory dataset compared to applying the trained network directly on the small dataset without applying fine tuning with TL techniques.

Another example of use TL in a FE module is showed Li *et al.* in [117]. They created a first model extracting cardio-respiratory coupling (CRC) spectrogram image from ECG signals, and then they applied this knowledge to another dataset with PPG signals to feed their CNN model. Phan *et al.* in [160] used TL for personalized automatic sleep classifier. They used a pre-trained model which was fine-tuned with a single-night data of each patient. They repeated the personalization 75 times, one for each patient. They applied the Kullback–Leibler (KL) divergence regularized method to avoid overfitting in the personalization process. They improved the accuracy of the model in 4.5 percentage points with this method.

TL is a powerful concept that encloses multiple methodologies which can improve ML models. These methodologies can be applied in the data preprocessing part (domain) or in the model (task). TL can use knowledge from the preprocessing methodology of different model which use similar data in order to improve the starting point of our model to train. TL allows to use the knowledge of feature extraction from an equivalent dataset used for other model to obtain features from our dataset to feed our model. And TL is able to reuse the knowledge acquired from a general model to train a new personalized model for each person. Part of these information were extracted from the review of Ebbehøj *et al.* in [50]. If the production and deployment phase is further explored, it's essential to highlight a technique for training models through Transfer Learning. An effective methodology for this purpose would be the application of FL.

FL was presented by Google less than a decade ago. McMahan *et al.* in [140] presented a new decentralized learning approach that reduces the communication time, but its most important feature is to make a decentralized learning without any exchange of subject data—personal data, data from individuals—. This could be use as a solution for privacy problems. Since then and until now FL has been the basis of multiple methodologies and has lead feasible solutions in the paradigm of distributed AI technology and IoT. Nguyen *et al.* in [147] made a comprehensive survey about FL for IoT with more than 200 references. Their survey includes a section dedicated to healthcare. This section is divided into two parts: (i) one for the EHR, and (ii) another named *Healthcare Cooperation*. In *Healthcare Cooperation* the studies focus on the improvement of models using FL, and it seems that the combination between FL and *Blockchain* technologies could be the solution to reduce the data leakage due to curious third-parties (intruders).

In the review of Nguyen *et al.* there can be found studies that implement the concept of *Federated Transfer Learning* and Edge Computing with FL for healthcare problems [39], [215]. In [39] the authors faced two challenges: (i) first the isolation of the healthcare data makes difficult the aggregation of data without compromising privacy, and (ii) second the problem of personalization while training models in the *Cloud*. They proposed *FedHealth*, a federated transfer learning framework. This framework aggregates the data with FL and personalizes the model with TL. In [215], the authors tackle heterogeneity issues in IoT environments. They study a personalized FL method as a solution. Their *PerFit* framework integrates multiple personalized FL methods and enhanced performance for devices in IoT applications. In both studies mentioned above the authors used activity recognition models to validate their theories. Our goal is to validate these theories with a real application of healthcare using the prediction model of migraine crises as a case study.

This work focuses primarily on time-series forecasting, which involves certain challenges such as temporal dependence, making analysis more complex due to autocorrelation between observations. Many series are non-stationary, requiring adjustments to handle trends and seasonalities that change over time. In addition, noise, outliers, and missing data complicate the identification of significant patterns. The inclusion of external variables and heteroscedasticity, which entails variable variance, also adds complexity to the modeling process. Finally, long-term forecasting increases uncertainty and reduces the accuracy of the model. We have reviewed various techniques to address these challenges, and throughout this work we will also show how these challenges have been addressed.

2.3 Migraine disease

This thesis focuses on migraine as a specific use case for information acquisition to enhance the prediction of symptomatic episodes. Real data will be utilized to feed our models and extract knowledge.

Migraine is distinguished from other types of headaches by its pulsating pain, often localized on one side of the head, moderate to severe intensity, and duration lasting between 4 and 72 hours. It is frequently accompanied by nausea, vomiting, and sensitivity to light (photophobia) and sound (phonophobia). Unlike other headaches, of which there are hundreds recognized,

such as tension-type headaches (characterized by mild to moderate bilateral pressure), cluster headaches (severe, unilateral pain around the eye occurring in patterns or "clusters"), or sinus headaches (associated with sinusitis and pressure in the facial area), migraines can also present with an aura, which involves visual or sensory disturbances [91]. Additionally, medication-overuse headaches arise from excessive use of pain relief drugs, leading to a cycle of recurring headaches. These distinct features contribute to the World Health Organization's classification of migraine as one of the most disabling conditions globally. Migraine stands as one of the most incapacitating neurological conditions impacting approximately 12% of the global population [219]. Migraine exacts substantial economic tolls on both private and national health systems [23], [72]. Marked by severe headaches predominantly localized on one side, the duration of pain can range from a few hours to several days [12]. The disease exhibits a higher incidence in women than in men, with prevalence peaking between the ages of 25 and 55, considered the zenith of human productivity [125]. Overall, women significantly outnumbered men in terms of migraine representation.[91]. Recent studies in Spain and England estimate costs exceeding 5,000 euros per patient per year due to migraine [69], [151]. As a chronic condition, there is currently no cure, prompting medical efforts to focus on pain reduction and the prevention of its occurrence. This preventive measure entails the administration of medication to alleviate pain. The critical aspect lies in the effectiveness of this treatment, which is most pronounced when analgesics are ingested hours before the onset of pain. Consequently, the accurate prediction of pain becomes paramount. The suboptimal administration of medication, or more precisely, when pain has already manifested, can result in over-medication [19], [137], and thus the chronification of the disease [207].

There are different events related to migraine episodes:

- Triggers: Stimuli that interact with the patient and initiate a pain episode. Triggers can be related to diet, environment, or psychological factors, among others.
- Premonitory symptoms: Subjective sensations experienced by the patient preceding the pain. Some symptoms may appear up to 48 hours in advance, although the average is around 10 hours before the onset of pain [64]. Examples include sound sensitivity, fatigue, or dizziness.
- Aura: A transient focal neurological phenomenon occurring usually less than an hour before the onset of pain. Depending on the affected body part, aura can manifest as visual —the most frequency—, sensory, language, or motor disturbances.
- Concurrent symptoms: Subjective sensations experienced by the patient during the pain phase. Many of these symptoms may overlap with premonitory symptoms.
- Postdromal symptoms: Subjective sensations experienced by the patient after the pain phase, often involving fatigue or discomfort.

Migraine can be classified in various ways. One classification distinguishes between “with aura” or “without aura” based on whether the patient has experienced aura in any of their pain episodes. It can also be categorized as episodic (less than 15 headache days per month) or chronic (exceeding this threshold). These being the main classifications of migraine types [91]. Where there is no classification of migraines by type of pain.

Several of these events occur prior to the onset of pain or even before the migraine itself. Therefore, both triggers and premonitory symptoms, including aura, could offer valuable insights and information regarding the imminent onset of pain. It is crucial to understand the characteristics of our sample, considering that if a majority of patients are classified as non-aura cases or if even those with aura do not consistently experience it with every migraine, relying solely on this event for the entire prediction may not be appropriate.

In order to maintain a record of these events, participants are required to mark the events they experience along with any relevant characteristics, making it a challenging and poorly incentivized task. In this context, an information recommendation system could serve as a significant incentive for many volunteers to engage more extensively or for a longer duration in the study. In other words, it could gamify the experience to make it more appealing and fulfill the study requirements [97].

Recommendation algorithms have been present for a long time since the publication of the first research articles in the mid-1990s. While there are ongoing research efforts to refine results and explore new paradigms, the foundations of recommenders are well-established. The rise of digital content platforms and online shopping has further fueled the field of recommendations, as companies strive to capture user interests. External competitions, like the Netflix Prize, have promoted the search for better algorithms. The recommendation process aims to satisfy user needs by providing the most interesting content, with the system taking the initiative.

Recommendation algorithms can be classified into three types: collaborative filtering, content-based, and hybrid approaches. Collaborative filtering algorithms leverage user evaluations to compare recommended items and find similarities between users. Content-based methods compare items based on their descriptions and attributes, overcoming the cold start problem. Hybrid models combine collaborative filtering and content-based techniques to address their individual limitations. These systems are already used in health related recommendation systems [220]. Learning-to-Rank techniques, while not directly focused on recommendations, play a role in ordering elements for optimal presentation and are more commonly used in a professional environment, so this technique will be used in the methodology [36], [167].

In the context of a migraine recommendation system, therapeutic recommendations and informative trivia would be encompassed. Collaborative filtering would be highly effective for disseminating trivia, such as sharing insights regarding International Migraine Day, yet is not suitable as it relies on user opinions rather than migraine characteristics. Content-based approaches are more appropriate, allowing the organic incorporation of migraine features. Hybrid models, with a stronger emphasis on content-based components, can also be considered. Learning to Rank algorithms offer insights into comparing elements directly without intermediate ratings.

Overall, since the aim is to make a personalised system, a suitable approach for the migraine recommendation system would involve comparing the features of the individual migraine episodes with available advice to determine their relevance to each patient.

Furthermore, the development of a robust and intelligent recommendation system that uses patient classification as a foundation is proposed. This system will leverage ML algorithms to analyze vast amounts of patient data, including medical records, self-reported pain scales,

and lifestyle factors, to generate personalized recommendations for symptom management.

Other important factor in migraine is sleep. Sleep plays a crucial role in migraine, as highlighted in the work of [56]. In their study, they emphasize that some headache disorders exhibit a significant correlation with sleep, with certain types exclusively occurring in association with sleep patterns. Specifically for migraine, poor sleep quality or insufficient duration can act as a trigger for migraine attacks [121], and migraine patients with inadequate sleep tend to experience a higher frequency of attacks [227]. Conversely, headache disorders can exert an impact on sleep, while sleep itself may contribute to the partial alleviation of headaches. The comprehensive assessment of sleep in a large population of headache patients, utilizing objective measurements such as Polysomnography (PSG) or actigraphy, holds the potential to elucidate various pathophysiological aspects. As recently demonstrated [128], actigraphy provides a distinct advantage by enabling the examination of patients in their natural environments over extended periods. Or, what is more significant in the studies [126], [157], it has been indicated that improvement in treatments for sleep initiation could have beneficial effects on headaches. Finally in [188], they find a correlation between sleep and the onset of migraine, using the same wearable device as us with seven subjects and a coarse-grain analysis on the signals, extracting features from the entire night, always emphasizing that it is only appreciated by making subject-wise models and due to the small number of patients, unable to find an explication of the significant difference in accuracy between one patient and another.

In recent years, many studies have been conducted on the factors that promote good sleep quality. Some authors have shown that sleep posture is an important factor in sleep quality [44], [143]. Many studies demonstrate the relationship between sleep posture and sleep disorders or other diseases—physical or neurological—such as: obstructive sleep apnea [94] and how sleeping in lateral decubitus can decrease the symptoms of sleep disorders in people with apnea-hypoapnea syndrome (OSAHS) [111], neck muscle activity [113], non-specific spinal cord symptoms [33], renal diseases [187], [228], and neurological diseases such as migraine [15]. Sleep and headaches, particularly migraines, are intricately connected, with migraine attacks influenced by both excessive and insufficient sleep. The body’s circadian rhythm, which regulates sleep-wake cycles through hormonal changes in response to light, plays a crucial role in this relationship. Migraines often occur in the early morning, indicating a potential link to sleep patterns and circadian rhythms. Disruptions, such as shift work or jet lag, can trigger migraines, while sleep may alleviate symptoms. Maintaining a balanced sleep-wake cycle through good sleep hygiene—such as consistent sleep schedules, appropriate sleep duration, exposure to natural light, and creating a restful environment—can help reduce migraine frequency. Additionally, avoiding stimulants like caffeine and nicotine before bed is advisable to improve sleep quality. This work focus on demonstrating whether it is possible to use low-cost commercial wearable devices for high-quality and reproducible sleep posture detection.

2.3.1 Sleep

Sleep hygiene is defined as a set of behavioral and environmental recommendations intended to promote healthy sleep [193]. The importance of not having a correct sleep hygiene is a

public health problem because it leads to sleep disorders, such as insomnia, that affect our daily life and worsen the symptoms of other diseases [93], including migraines.

Two problems that can be mentioned about sleep studies are: (i) on the one hand, patients only spend one night in the hospital, and from that night a diagnosis is made without the possibility of continuous ambulatory and recurrent monitoring for follow-up. And (ii) on the other hand, all the sensors used are uncomfortable and disturb the patient's rest, in a sleep session that is not representative. It would be very interesting for professionals, and would open up a new avenue of study, to be able to use comfortable non-invasive ambulatory monitoring devices that do not disturb patient's sleep. Many studies are being carried out along these lines.

As one of the factors influencing sleep quality is posture, a study on its detection has been conducted [65].

Sleep posture

In recent years, many studies have been conducted on the factors that promote good sleep quality. Some authors have shown that sleep posture is an important factor in sleep quality [44], [143]. Many studies demonstrate the relationship between sleep posture and sleep disorders or other diseases—physical or neurological—such as: obstructive sleep apnea [94] and how sleeping in lateral decubitus can decrease the symptoms of sleep disorders in people with apnea-hypopnea syndrome (OSAHS) [111], neck muscle activity [113], non-specific spinal cord symptoms [33], renal diseases [187], [228], and neurological diseases such as migraine [15]. This section tries to answer the following question: Is it feasible to detect sleep posture using a commercially available low-cost wearable device?

The study of posture during sleep is performed professionally in hospitals, in sleep units. In these spaces, patients who come for suspected illnesses are monitored overnight: restless leg syndrome, insomnia, sleep apnea, etc. The sleep study is called PSG, and multiple biomedical signals are recorded, such as blood pressure, ECG, EEG, respiration, temperature, or sleep position.

Currently, posture during PSG is measured with a sensor placed on the chest. Four postures of the torso are recorded: prone, supine, left lateral and right lateral. Numerous studies are exploring alternative, less obtrusive methods for posture detection with different technologies.

- Works such as the one of Enayati *et al.* [52], or the study of Hsiao *et al.* [88] use pressure sensors on the bed. The solution proposed by Hsiao *et al.* reaches an accuracy of 0.881 for posture detection using resistive force sensors placed on the top of the bed to measure and an array of infrared sensor.
- Other authors have implemented video image analysis. The authors in [150] achieve an overall accuracy of 0.916 using 13 body postures (not only torso postures). In another study [118], the authors achieve an average accuracy of 0.909 distinguishing between 10 body postures. The use of cameras has the drawback of compromising user privacy and consuming more computational resources and energy than other technologies. On the positive side, by looking at the movements of the whole body, they are able to detect

more postures than the aforementioned torso postures used in clinical practice. This study will focus only on the four torso postures.

- Other studies analyze inertial signals such as accelerometry with single or multiple wearable devices [24], [96], [206]. In the work shown by Parastoo Alinia *et al.* [6] four body positions have been studied and nine locations have been evaluated for the placement of the accelerometer. The results have shown F-values between 0.952 to 0.978, with the best positions for accelerometer placement on the thighs and the chest.

Sleep prediction

Consumer sleep-tracking devices have garnered widespread usage and are progressively advancing in technological sophistication, evoking significant interest among researchers and clinicians as potential alternatives to conventional actigraphy. In an effort to evaluate the efficacy of several state-of-the-art consumer sleep-tracking devices, [40]. conducted a comprehensive assessment, juxtaposing their performance with actigraphy and the gold-standard sleep assessment technique, PSG. Their findings affirm that consumer sleep-tracking devices exhibit commendable performance in sleep detection, with a majority demonstrating comparable or superior accuracy to actigraphy in wake detection. However, assessments of sleep stages by these devices exhibited inconsistencies. The results suggest that numerous contemporary sleep-tracking devices hold promise in accurately tracking sleep and wake patterns. To substantiate their broader validity and utility, rigorous testing in diverse populations and settings is warranted. Other studies as [172] corroborate the sleep and wake tracking results.

The question at hand revolves around the feasibility of utilizing non-commercial devices to detect various sleep states while addressing the dual challenges outlined at the outset of this chapter. Several studies [80], [115], [198] substantiate this prospect, leveraging devices capable of measuring the EEG. However, despite the convenience of home-based sleep monitoring, the persisting discomfort for patients arises from the necessity of affixing electrodes to their face during sleep.

In exploring less intrusive alternatives for sleep state detection, exemplified by a wrist-worn wearable in this study, numerous investigations come into play. A private enterprise, employing the Firstbeat Bodyguard 2 device [105], [163], demonstrates potential, albeit with results diverging significantly from those obtained through EEG. Concurrently, the research group led by Pedro Fonseca *et al.* has undertaken extensive studies in this domain [165], [166], [216]. Employing a sophisticated algorithm and harnessing TL, they strive to enhance preceding outcomes, yet these endeavors still fall short of the benchmarks set by studies employing EEG.

While EDA signals are not incorporated in these studies, [83] underscores the signal's relevance to distinct sleep states. Their findings culminate in the creation of a wakefulness/asleep model, comparable in efficacy to models grounded in actigraphy.

2.3.2 Prediction

In any scenario, all these symptoms inherently possess a notable subjective element, rendering a system solely based on this information unreliable for prediction. Given that, in general, the majority of prediction systems hinge on statistical methodologies, inherently imprecise in nature, relying exclusively on a similarly imprecise foundation would inevitably compromise precision in the predictive process. In any case, all these factors and means can always help by providing additional information and enabling improvements. For example, in [84], they employ a linear regression model, a fundamental technology in ML, to predict and enhance migraine attacks in patients using previously identified triggers. However, leveraging technology allows us to explore the hemodynamic signals of the body, granted there is substantive evidence of their correlation with the disease.

In a similar manner, ML models have been utilized for predicting crises in chronic diseases, as demonstrated by [60]. In subsequent studies, more advanced models incorporating DL techniques have been employed to attempt to predict crises [224]. Or even studies to predict the patient's response to certain migraine treatments, which could assist in treatment decision-making [28], [29]. These techniques are not only applied to epilepsy but also to other crises in chronic diseases, such as cancer pain [152].

Additionally, relying on data collected directly by a sensor, which removes the need for patient input, enhances patient adherence to the system. This enables ambulatory monitoring in any healthcare setting. These advancements have only become possible in the past decade due to improvements in wearable devices.

In this instance, as detailed in Chapter 2.1, the Empatica E4 wristband emerges as the most suitable device for monitoring hemodynamic signals associated with migraine. Furthermore, drawing upon findings from other studies within our research group, the potential for migraine pain prediction utilizing this particular device is demonstrated [156].

Based on these results, this work aim to develop an innovative prediction model that, by integrating state-of-the-art DL technologies and real-time collection of migraine related biomedical signals already mention in Chapter 2.1, can assist in preventing migraine pain with enhanced accuracy during analgesic intake.

Chapter 3

Methodologies and experiments

This chapter will elucidate various modules investigated to ultimately enhance the prediction of migraine pain, as outlined in the proposed methodology. Here it is going to proceed to (i) explain different methodologies or materials common for all the modules (ii) characterize and classify migraine pain curves, envisioning subsequent classification based on pain types in patients, coupled with an analysis of diverse migraine-related variables. This aimed at extracting additional personalized information through classification. As a concluding effort, it is proposed the development of a recommender—an innovative mobile application for migraine patients—, leveraging the same technologies employed in this work to enhance patient education and support.

The research trajectory then extended to (iii) a comprehensive sleep study, incorporating a posture detection system. This involved modeling for the identification of distinct sleep phases and subsequent metrics for sleep quality. The final phase undertook (iv) the modeling of migraine pain prediction, consolidating all acquired knowledge from the preceding planning stages into this comprehensive model.

These studies aim to analyze and extract information from all clinical and social aspects of migraine patients to enhance the understanding and management of the disease. This allows the acquired information to be processed and extrapolated for personalized treatment for each patient. A schematic representation is shown in Figure 3.1.

For all the models, the CESVIMA supercomputing center at the Polytechnic University of Madrid was used, which has the following characteristics: Magerit is a cluster composed of 72 Lenovo ThinkSystem SD530 nodes, each equipped with 2 Intel® Xeon® Gold 6230 processors with 20 cores @ 2.10 GHz (1.344 GFLOPS) and 192 GiB of RAM, and 48 Lenovo ThinkSystem SD530 nodes (each equipped with 2 Intel® Xeon® Gold 6240R processors with 24 cores @ 2.40 GHz (1.344 GFLOPS) and 768 GiB of RAM).

3.1 Pain curve characterization

This initial module focuses on the classification of migraine pain. This research yielded the following article [66]. The methodology carried out is going to be explained.

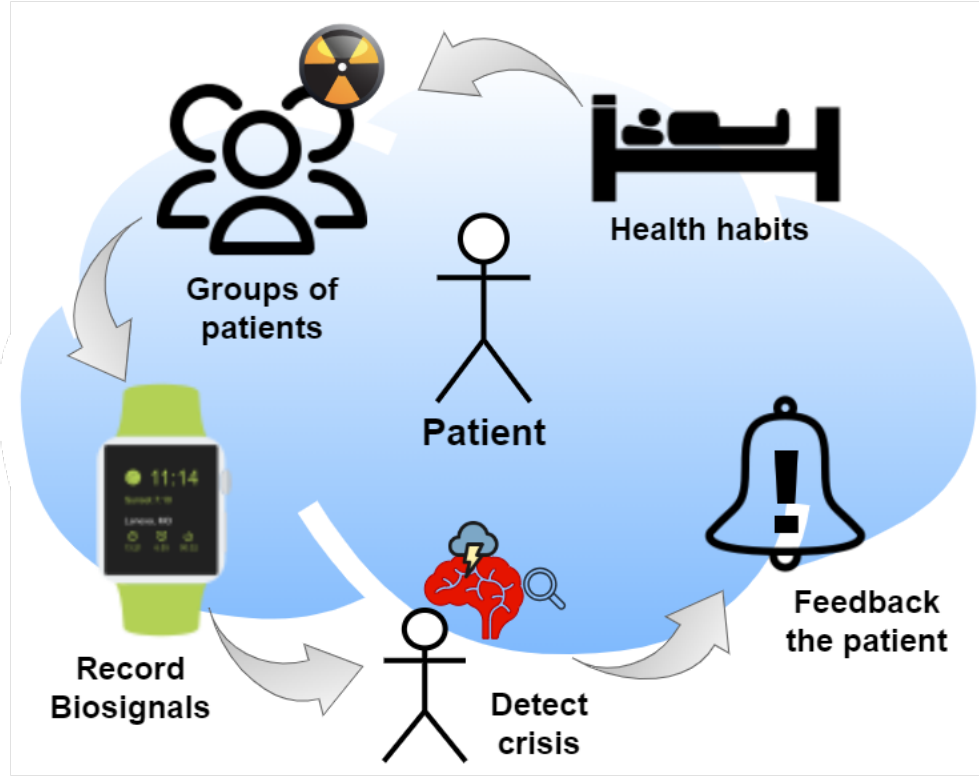


Figure 3.1: A schematic representation of all modules revolving around and interrelating with migraines includes sociodemographic information of migraine patients, sleep characteristics, hemodynamic signals extracted using wearable devices, and the implication of prior knowledge of migraine onset through an alarm, along with additional related information. This representation is utilized to extract information aimed at improving the understanding and management of the disease.

The definition of types of migraines is a novel approach to better understand the predictability of migraine onset of pain, and the effect of medication on the subjective perception of pain. The study of the characteristics of the pain evolution has been already studied in a broader aspect though. In [124] it is compared the shape of migraine pain between the treated and the untreated patient with triptan. There, the known time-intensity relation of migraine pain [183], [213] was confirmed. Because the objective of these works was to study the fluctuations in pain intensity levels and changes in symptoms onset between treated and untreated crises, no studies were found in the research conducted in this work to compare with this method for classifying pain types based on their evolution.

Rappaport *et al.* [168] already used the temporary variables of duration and intensity of the pain for classification of types of headache; among them, the migraine. In order to study the progression from episodic to chronic migraine, Bigal *et al.* [18] used temporal definitions to characterize the migraine curves. In that study the authors showed that 15% of the migraines reached the maximum in less than 5 minutes, the 70% took up to 2 hours, and the rest reached the maximum after 2 hours. They also showed that migraines last from 4 to 72 hours in 90% of the cases.

In this study, these characteristics are considered as a starting point for the search for the different types of migraine pain curves. Once the different types of migraine pains are clustered, a representative curve of each cluster will be defined to visualize their morphology.

3.1.1 Experimental setup

The dataset used in this module comes from data collected by a research group in a clinical study at the Hospital Universitario de la Princesa in Madrid with protocol number PI-879 from 2015 to 2019, the documentation is in Annex 6.7. The clinical study was managed by the Neurology Unit of the Hospital Universitario de la Princesa de Madrid, in Spain. The patients recruited had a diagnosis of episodic migraine (less than 15 days of pain per month). Each patient participated in the study for a period of 4 to 6 weeks, depending on the number of crises reported. Patients abandoned the study if no migraines were reported within three weeks.

The patients were trained in the management of a mobile application developed to collect the real-time evolution of the pain of a migraine attack. The patients were asked to use the app to mark the beginning and end of a pain attack, as well as the subjective evolution of the pain in real time. Thus, for each episode, there is a tuple of time-pain values to work with.

The original dataset consisted of 464 migraines from 56 patients. Each patient reported 8.3 crises on average. After data pre-processing, 51 patients were selected. 90.2% were women ($n=46$) and 9.8% were men ($n=5$), with ages between 39.0 ± 11.1 and 40.4 ± 9.3 respectively.

6.7% women suffered from hypertension, 2.2% had diabetes, 11.1% had dyslipidemia, and 17.8% were smokers. The cognitive reserve was 16.1 ± 3.4 . The 73.33% had a level 6 European Qualifications Framework (EQF) or higher, the 24.4% had a level 5 of EQF and the 2.2% had a level 4 of EQF or fewer. The 53.33% had aura. 6.7% had low frequency episodic migraines, the 80.0% had episodic migraines, 6.7% had high frequency episodic migraines, 4.4% had chronic migraines and the remaining 2.2% is missing value.

None of the men suffered from diabetes nor hypertension, 20.0% had dyslipidemia and none of them are smokers. Their cognitive reserve was 16.8 ± 3.9 . 20.0% of them had a level 5 of EQF, while 80.0% had a level 6 EQF or higher. No men had aura. 80.0% had episodic migraines and 20.0% had high frequency episodic migraines.

The total number of migraines recorded was 345. The average number of migraines per patient was 6.8 ± 5.0 . The data pre-processing diagram is shown in Figure 3.2, and it is further described.

3.1.2 Pain evolution marking

In the clinical practice there are different ways to indicate the subjective feeling of pain. The most widely used is the VAS scale, which represents the pain in a range from 0 to 10 [106]. However, because of its limitations to the marking of pain levels in real time, the unbounded relative pain scale proposed in [153] has been used.

Patients were asked to indicate the onset of pain at the lowest perceived intensity in an app.

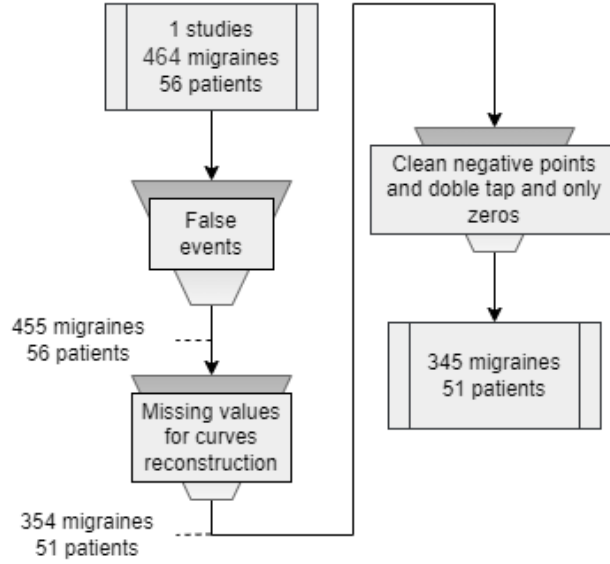


Figure 3.2: Data pre-processing carried out in the project. Finally, 345 migraine episodes from 51 different patients are used.

They were also asked to indicate as many points of pain in real time as their condition allowed them. If the pain was over, they indicated the end of the pain. If the pain did not progress, the patients could indicate it as a false pain event. Figure 3.3 shows a snapshot of the used app, where an example of the evolution of a migraine pain is presented. The reader should notice that the ordinate indicates the subjective pain evolution, and the scale has no limits. The abscissa indicates the number of points marked, and for the sake of simplicity in the representation of the app, they are equally spaced.

Using the (+) and (-) buttons in the app, the patients are able to indicate positive values if the pain goes up, negative values if it goes down, or zero if the pain level remains unchanged. When the patients indicate the end of the pain attack it is assumed that, at that time, the absolute pain level is 0, and thus negative values are not allowed. This unbounded scale was the preferred by patients who also were previously monitored using the VAS scale in real time.

At the end of each pain crisis the patients fulfill a final questionnaire where they classify the overall pain intensity level using the VAS scale from 0 to 10. The VAS scale can be used in this case because it indicates, once the pain has passed, how intense the pain was among all those suffered by the patient.

3.1.3 Data pre-processing

Figure 3.2 depicts the data pre-processing and cleaning pipeline carried out before the actual processing and modeling phases. The original dataset consists of 464 migraines from 56 patients. 9 migraines events tagged as false migraines are removed. A false positive is a pain event reported as no pain by the patient afterwards. After removing the false positive

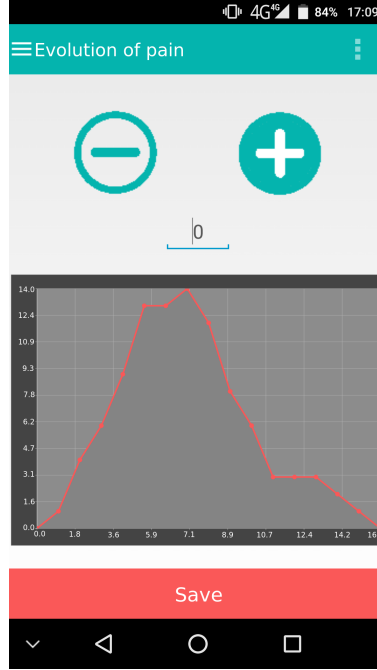


Figure 3.3: Evolution of the pain marked in the smartphone app. The ordinate indicates the subjective pain evolution and the abscissa indicates the number of points marked.

events, the pain curves are normalized to their maximum and scaled according to the overall pain intensity level using the VAS scale of the final questionnaire. The migraines without evolution pain marks are removed as well. If a patient indicates more than one point level within one minute interval, only the latest value is considered and the others are removed. After this, if there are no other pain marks than the start and the end points, the migraine is removed. These cleaning processes removed 110 migraines. Eventually, there have been removed 5 patients. On average it has been removed 1.5 pain events per patient. Finally, the dataset remains with 345 migraines from 51 patients.

Figure 3.4 illustrates the result of the data pre-processing. The figure shows the evolution of a pain crisis lasting 14 hours approximately. The dots represent the original marks indicated by the patient in real time. The values are normalized and scaled by the overall pain intensity level (in the case of the figure is 10/10). The crosses represent the useful data points after removing the consecutive points within a minute interval. The line connecting the points represents the curve on which the data processing and modeling will be applied.

3.1.4 Processing and modeling methodology

This section describes the methodology workflow for processing and modeling of pain types based on characteristics of the pain curve. The workflow is shown in Figure 3.5 and consists of 7 processes, among which there are highlights: feature generation, data processing and analysis, and modeling.

The modeling consists on finding the types of pain through grouping techniques to use the labels of the grouped types to make a supervised feature selection and classification as final

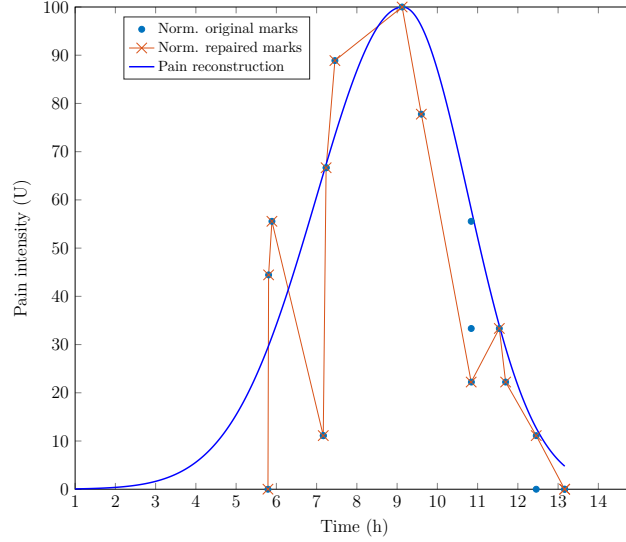


Figure 3.4: Example of pain marks collected and repaired from a migraine, and the pain reconstruction by a two Gaussian approximations.

model, as shown in Figure 3.5. This “cluster-then-classify” approach allows simplifying the final model and improving its performance compared to the one achieved by the unsupervised clustering.

1. Generation of features: there is no clinical description of pain curves for chronic diseases. There are, however, some characteristics that can be observed easily, such as the maximum amplitude or total duration. In addition to these simple clinical characteristics, a set of features based on basic mathematical descriptions of curves has been described. To allow the comparison between crises, a linear interpolation of the pain curves has been performed with an equal pain mark rate for all curves.
2. Removal of outliers: to improve the accuracy of the clustering method, the second proposed process is the detection and elimination of outliers. Based on the Z – score standardization, every pain curve with any feature detected as outlier is removed.
3. Unsupervised feature selection: an unsupervised feature selection has been performed using the Pearson’s correlation coefficient. Highly correlated features have been removed.
4. Data transformation: normalization of variables and logarithmic transformation of features with non-linear behaviour are applied. These transformations lead to improve the clustering results [34]. The transformation to be applied to each feature is chosen in a loop-based clustering algorithm, as shown in Figure 3.5.
5. Unsupervised clustering and model curve: Unsupervised clustering is performed since there are no labels for pain types. GMM and k -means clustering models have been tested, with the latter showing the best results and presented in this manuscript. Finally a schematic definition of the curves of each cluster is constructed. This could lead to clinical discussions based on the general morphology of the pain evolution of each cluster, to find characteristics in order to improve the effectiveness of the treatments, to correlate

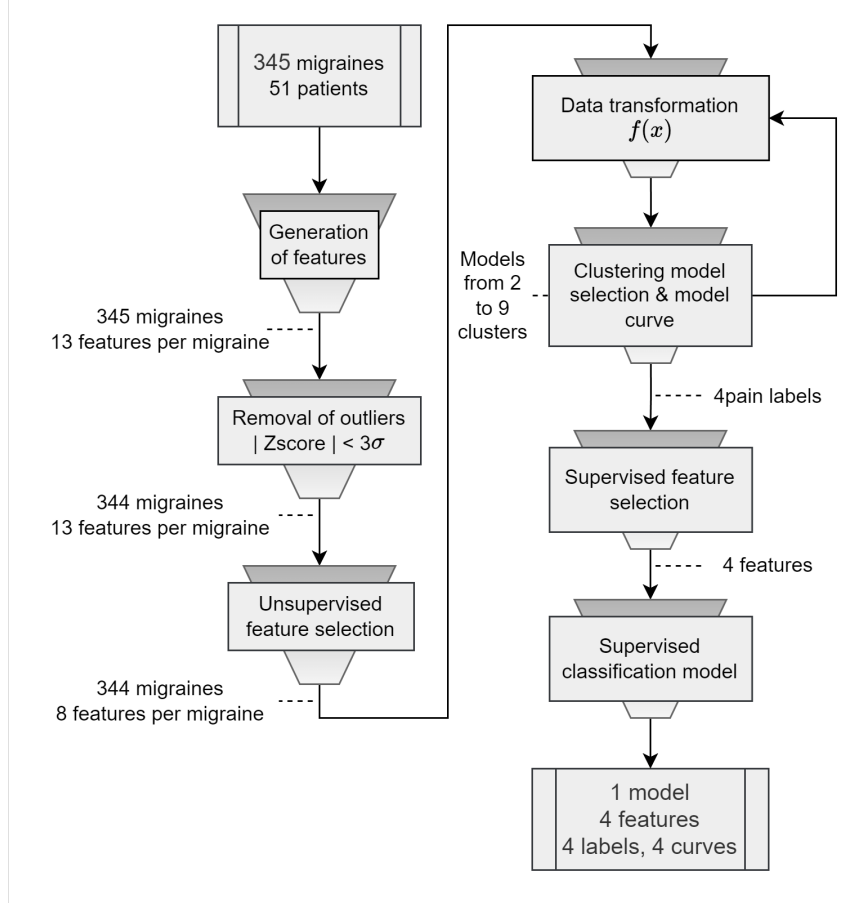


Figure 3.5: Methodology flow followed to find the clustering model that describes the types of pain curves.

with symptoms of the disease, *etc.* This clustering model is the primary goal of this part. However, as the model is intended for use in a real environment, optimization is sought, and, employing the cluster-then-classify strategy, a discussion on the next steps of the methodology, including supervised feature selection and classification modeling, is proposed at the end of this paper.

6. Supervised feature selection: The result of the previous step provides a set of labels that associate each pain with the type that has been assigned by the clustering model. It is more simple to identify the most relevant features of the dataset thanks to these groups. To do that, a correlation-based algorithm has been used.
7. Supervised classification model: In order to find the classification model underlying the unsupervised group assignment, tree-type of classification models have been studied in view of the distribution of samples after the feature selection. Specifically, the algorithms J48 tree, J48-graft tree, RandomForest, and Logistic Model Trees have been studied, showing the latest the best results. This step and the previous one are detailed in Chapter: 4.

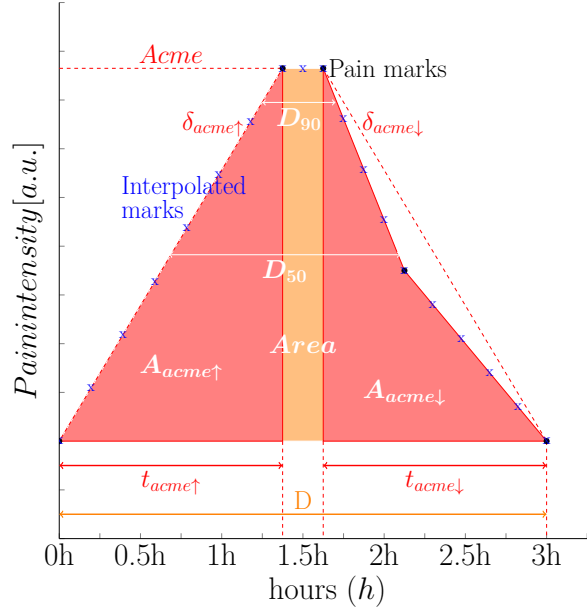


Figure 3.6: Morphology of the pain curve and features.

3.1.5 Generation of features

A set of features has been defined in order to allow the classification of the different types of chronic pain. The features hereafter presented are based on basic mathematical descriptions of curves. In order to help the reader to better understand these definitions, Figure 3.6 has been depicted. All the features are based on the subjective pain marks that the patients indicate in real time. These points are shown as black dots in Figure 3.6. As shown in Figure 3.4, these points have been repaired following the aforementioned methodology. After that, the curves are interpolated linearly with a sampling period $T_s = 10[\frac{\text{points}}{\text{hour}}]$ to standardize the rate of points between curves. The interpolated pain marks are represented with crosses in Figure 3.6.

13 features have been described and grouped according to the following relationships:

- features related to the *Acme*: the maximum level of intensity $Acme$; the time to the highest point $t_{acme\uparrow}$; the time from the highest point to the end $t_{acme\downarrow}$; the slope to reach the highest point $\delta_{acme\uparrow}$; and the release slope of pain from the highest point $\delta_{acme\downarrow}$.
- features related to the duration of the pain: the total duration D ; and duration of the pain with an intensity higher than 50% and 90% (D_{50} and D_{90} respectively).
- features related to the area under the curve, or *amount of pain*: the total amount of pain A ; and the amounts of pain suffered before and after the *Acme* $A_{acme\uparrow}$ and $A_{acme\downarrow}$ respectively.
- features related to the intensity of the curve: the energy E ; and the power P . These features do not appear in Figure 3.6.

The definitions and equations of the features are shown in the following paragraphs.

Features related to the acme

In medicine, the acme refers to the period of highest intensity of a symptom of any disease. Thus, for us, the acme is the maximum level of pain intensity of a migraine. Based on the acme, a group of features related to it has been defined: the acme itself $Acme$ in Equation 3.1; $t_{acme\uparrow}$ and $t_{acme\downarrow}$ in Equations 3.2 and 3.3 which are respectively the time to reach the $Acme$ and the decay time of the pain after the end of the $Acme$; and $\delta_{acme\uparrow}$ and $\delta_{acme\downarrow}$ in Equations 3.4 and 3.5 which are the upward and downward slopes of the pain to and from the $Acme$ as shown in Figure 3.6.

The acme is not an isolated point in time, on the contrary, the acme lasts for minutes or hours. The duration and intensity will depend on the medication that the patient has taken, if any. Unfortunately, in the experiments the patients could not be able to mark the evolution of the pain at a fixed rate, and because of this, sometimes the $Acme$ will be represented with a single point.

$$Acme \equiv \text{highest intensity of the migraine } [pain], \quad (3.1)$$

$t_{acme\uparrow}$ gives us an idea of the time it takes for the pain to reach the $Acme$. On the other hand, $t_{acme\downarrow}$ gives us an idea of how long the patient continues suffering pain once the pain starts to decrease from its peak. In Equations 3.2 and 3.3 t_{acme_0} is the time when the acme starts; $t_{acme_{end}}$ is the time when the acme ends—if the acme is only one point $t_{acme_{end}} = t_{acme_0}$ —; and t_0 and t_N represent the times when the pain starts and ends indicated by the patient.

$$t_{acme\uparrow} = t_{acme_0} - t_0 \text{ [hour]}, \quad (3.2)$$

$$t_{acme\downarrow} = t_N - t_{acme_{end}} \text{ [hour]}, \quad (3.3)$$

$\delta_{acme\uparrow}$ explains the speed of the onset of pain. It represents how fast the pain reaches its maximum level. On the other hand, $\delta_{acme\downarrow}$ shows how fast the pain vanishes. $\delta_{acme\downarrow}$ is defined in absolute value and thus, all features are positive.

$$\delta_{acme\uparrow} = \frac{Acme}{t_{acme\uparrow}} \left[\frac{pain}{hour} \right], \quad (3.4)$$

$$\delta_{acme\downarrow} = \left| \frac{-Acme}{t_{acme\downarrow}} \right| \left[\frac{pain}{hour} \right]. \quad (3.5)$$

In Equations 3.4 and 3.5 $Acme$ is the value of the maximum pain intensity; t_{acme_0} and $t_{acme_{end}}$ are the times as defined above.

Features related to the duration

The duration D represents the time between the beginning and the end of the pain t_0 and t_N respectively:

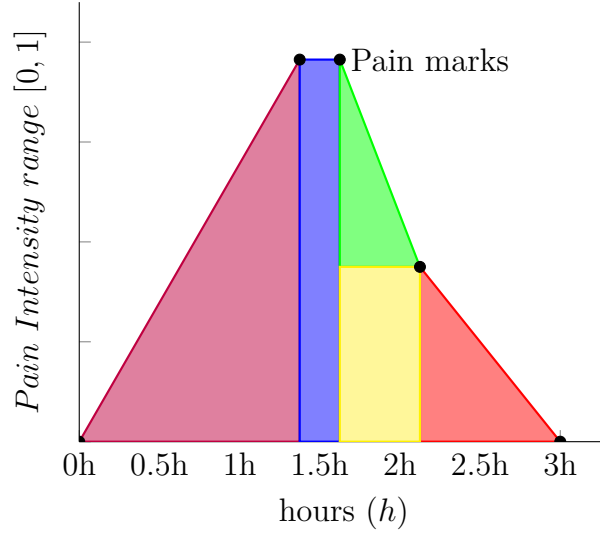


Figure 3.7: To compute the areas—or *amounts of pain*—, the curves are decomposed in rectangles and triangles between the pain marks.

$$D = t_N - t_0 \text{ [hour]}. \quad (3.6)$$

In order to understand the pain better, the width of the pain curve at different pain intensities is going to be described. To this end, D_{50} and D_{90} have been defined, which cover the time while the pain is above 50% and 90% of the *Acme* respectively. These times are computed using the interpolated lines as shown in Figure 3.6.

$$D_{50} = t_{50_{end}} - t_{50_0} \text{ [hour]}, \quad (3.7)$$

$$D_{90} = t_{90_{end}} - t_{90_0} \text{ [hour]}. \quad (3.8)$$

D_{50} is shown in Equation 3.7, where t_{50_0} and $t_{50_{end}}$ are the first and last times where the curve reaches a pain intensity of the 50% of the *Acme*. D_{50} represents the amount of time that the patient has suffered a pain level higher than the half of its *Acme* during the episode. Similarly it is define D_{90} in Equation 3.8. D_{90} represents the amount of time that the patient has suffered an intense pain level during the episode.

Features related to the area

Another feature that defines the morphology of the curves is its area. The linear interpolation between the points of the curve allows us to compute the area as the sum of the areas of its decomposition in rectangles and triangles as shown in Figure 3.7.

The total area of the curve, A , refers to the total amount of pain and it is computed as shown in Equation 3.9:

$$A = \sum_{i=0}^{N-1} \left(\frac{x_{i+1} + x_i}{2} \right) * (t_{i+1} - t_i) \text{ [pain * hour]}, \quad (3.9)$$

where x_i represent the i – th pain value at time t_i ; and N indicates the latest point of the curve marked by the patient.

The areas $A_{acme\uparrow}$ and $A_{acme\downarrow}$ are defined similarly to A in Equations 3.10 and 3.11:

$$A_{acme\uparrow} = \sum_{i=0}^{t_{a0}-1} \left(\frac{x_{i+1} + x_i}{2} \right) * (t_{i+1} - t_i) \text{ [pain * h]}, \quad (3.10)$$

$$A_{acme\downarrow} = \sum_{i=t_{aend}}^{N-1} \left(\frac{x_{i+1} + x_i}{2} \right) * (t_{i+1} - t_i) \text{ [pain * h]}, \quad (3.11)$$

where t_{acme_0} and $t_{acme_{end}}$ are start and end instants of the *Acme*; x_i represents the i – th pain value at time t_i ; and N indicates the latest point of the curve marked by the patient.

Features related to the intensity

The pain crises can reach high and brief acmes, or long lasting and low acmes. In order to distinguish these cases, the following parameters have been defined: energy, E , and power, P , of the curves, based on the re-sampling carried out. These features are defined as shown in Equations 3.12 and 3.13 respectively:

$$E = \sum_{i=0}^N x_i^2 \text{ [pain}^2], \quad (3.12)$$

$$P = \frac{E}{D} \left[\frac{\text{pain}^2}{\text{hour}} \right], \quad (3.13)$$

where N is the number of points of the interpolated curves; and t_0 and t_N are the times of the start and the end of the pain. E and D in Equation 3.13 are the energy and duration of the curve respectively.

As a reminder, the number of points x_i shown in all the equations is not the number of pain marks indicated by the patients but an oversampling at rate $f_s = \frac{1}{T_s}$. With this, the energy and power of the pains are comparable between the pain episodes independently if the patient is proactive in indicating the evolution of the pain or not.

3.1.6 Removal of outliers

In any statistical study there are samples that fall out from the distribution. In this study, patients are being asked to continuously and quantitatively reflect something as subjective as pain during their daily lives. The intrinsic error of the measurement tool—human subjectivity—means that some times the curves drawn contain errors. In order to remove those pain curves

that could have been clearly bad annotated, in this step of the workflow it is performed a detection of outliers and removal of curves that distort the dataset.

The scattering plots and histograms of the features $Acme$, $\delta_{acme\uparrow}$, and D . These plots are representative of the scattering plots of the remaining features—being $A_{acme\uparrow}$ and $t_{acme\uparrow}$ similar to $\delta_{acme\uparrow}$ and the rest similar to D —. Isolated points are features that could be labeled as outliers. These plots could be seen at: [Scattering and distribution figure](#).

A statistical methodology to remove the outlier migraines has been carried out in order to get better clustering results. The technique used to this end is the Z – score standardization shown in Equation 3.14:

$$Z(x) = \frac{x - \mu_x}{\sigma_x}, \quad (3.14)$$

where x represents each one of the features of the pain curve, and μ_x and σ_x are the mean and the standard deviation of each feature.

A value $Z(x_i)$ is considered an outlier if $|Z(x_i)| > 3\sigma_{Z(x)}$. Being $\sigma_{Z(x)}$ the standard deviation of the Z – score. This threshold keeps 99.73% of the elements of the distribution. The criterion of $\pm 3\sigma$ showed the best results for the dataset. But other methodologies such as the inter-quartile interval were also considered. A pain episode is removed if any of the values of its standardized features is detected as an outlier. Only 1 migraine has been removed, and thus, the number of events available is 344 from 51 patients.

3.1.7 Feature selection

The number of defined features is not too high and an unsupervised feature selection based on Pearson’s correlation coefficient has been used. Features with Pearson’s correlation equal or higher than 0.80 have been removed to reduce the complexity of the study.

	$Acme$	$t_{acme\uparrow}$	$t_{acme\downarrow}$	$\delta_{acme\uparrow}$	$\delta_{acme\downarrow}$	D	D_{50}	D_{90}	A	$A_{acme\uparrow}$	$A_{acme\downarrow}$	E	P
$Acme$	1.00	-	-	-	-	-	-	-	-	-	-	-	-
$t_{acme\uparrow}$	0.22	1.00	-	-	-	-	-	-	-	-	-	-	-
$t_{acme\downarrow}$	0.21	0.25	1.00	-	-	-	-	-	-	-	-	-	-
$\delta_{acme\uparrow}$	0.08	-0.21	-0.03	1.00	-	-	-	-	-	-	-	-	-
$\delta_{acme\downarrow}$	0.04	-0.02	-0.12	0.00	1.00	-	-	-	-	-	-	-	-
D	0.26	0.71	0.83	-0.12	-0.08	1.00	-	-	-	-	-	-	-
D_{50}	0.05	0.39	0.05	0.01	0.07	0.43	1.00	-	-	-	-	-	-
D_{90}	0.16	0.61	0.31	-0.07	0.00	0.67	0.85	1.00	-	-	-	-	-
A	0.32	0.76	0.65	-0.11	-0.05	0.92	0.56	0.82	1.00	-	-	-	-
$A_{acme\uparrow}$	0.25	0.96	0.25	-0.19	-0.03	0.71	0.46	0.69	0.81	1.00	-	-	-
$A_{acme\downarrow}$	0.26	0.44	0.76	-0.03	-0.03	0.76	0.11	0.44	0.72	0.46	1.00	-	-
E	0.13	0.63	0.47	-0.07	-0.03	0.74	0.61	0.73	0.79	0.69	0.49	1.00	-
P	0.18	0.68	0.66	-0.10	-0.05	0.89	0.54	0.72	0.88	0.72	0.64	0.96	1.00

Table 3.1: Correlation between features where values ≥ 0.8 are highlighted. Highly correlated features are candidates to be removed.

According to the results in Table 3.1, it can be seen that:

- The power P is the highest correlated feature. P correlates with the duration D of the pain episode, the area A , and the energy E . P has been removed.
- A and $A_{acme\uparrow}$ are also highly correlated with two features each and thus, they have been removed.
- D_{50} correlates with D_{90} and thus D_{50} has been removed as well. It is interesting to note that it has been decided to keep D_{90} and remove A and D_{50} because D_{90} describes the amount of time that the patient is suffering high intensive pain, which is a more specific, restrictive and interesting feature related to the severity of the pain.
- $t_{acme\downarrow}$ is highly correlated with D and it has been removed. In this case D is kept because of the simplicity of the concept of the duration of a pain episode.

The 8 remaining features are: $Acme$, $t_{acme\uparrow}$, $\delta_{acme\uparrow}$, $\delta_{acme\downarrow}$, D , D_{90} , $A_{acme\downarrow}$, and E . Nevertheless, further discussion will show that not all of them are needed for accurate supervised classification of the types of pain.

3.1.8 Data transformation

Some of the features depict distributions similar to the ones in [Scattering and distribution figure](#), which are non-linear distributions with high concentration of occurrences in small ranges of magnitude. A data transformation based on the normalization of variables and the logarithmic transformation of features with non-linear behaviour have been carried out.

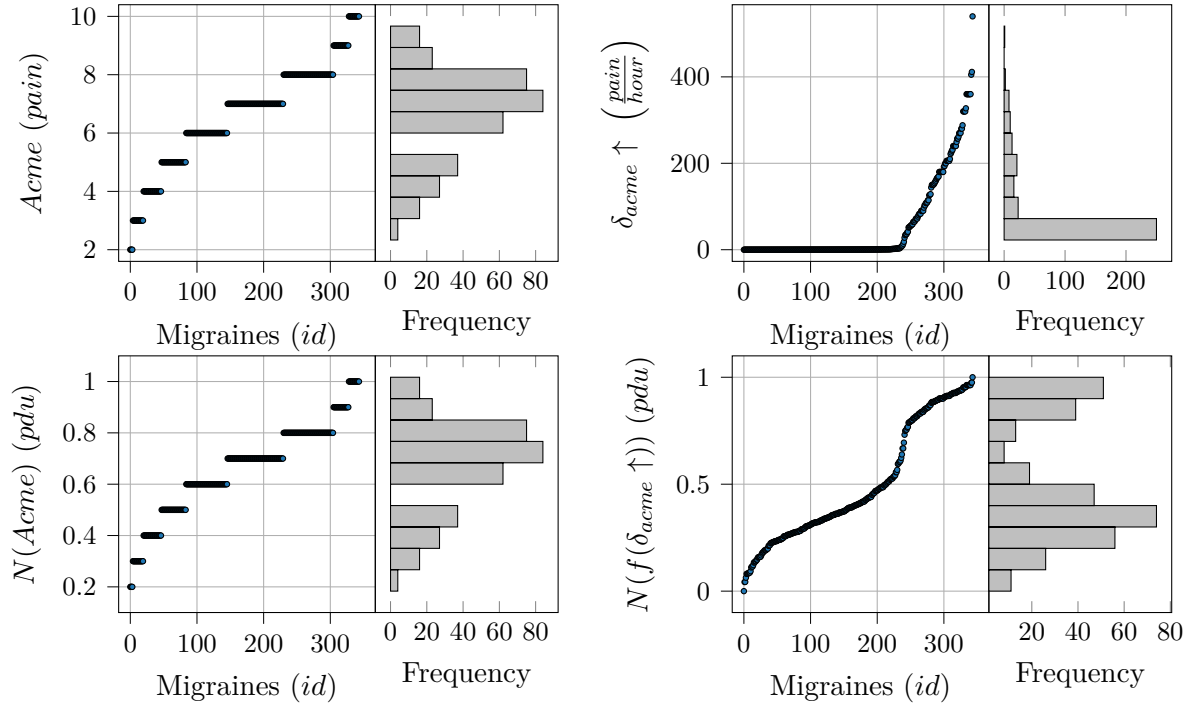
The logarithmic transformation shown in Equation 3.15 helps to reduce data concentration, improve sparsity of the distributions and gain linearity. It has been applied only to some features. The choice of the features to be transformed using the natural logarithm has been made in a closed loop based on the results of the clustering as shown in Figure 3.5. Nevertheless, all features have been normalized by $N(f(x))$ using Equation 3.16.

$$f(x) = \ln(x) - \min\{\ln(x)\}. \quad (3.15)$$

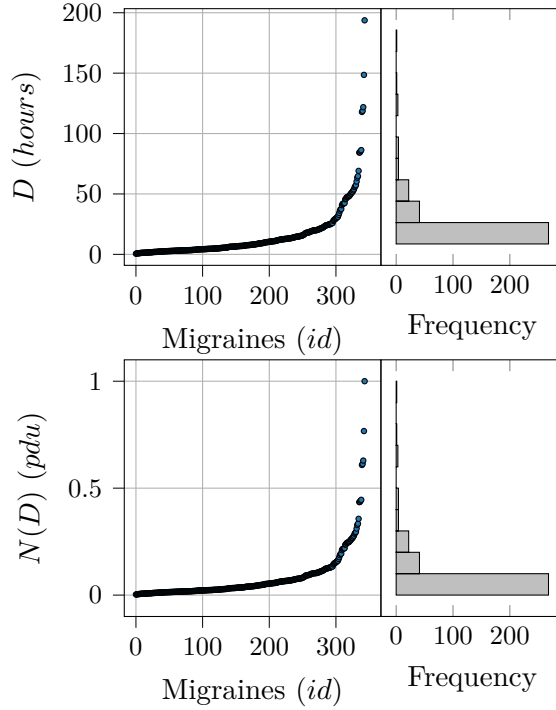
In Equation 3.15, x represents any of the features. The transformation in Equation 3.15 can be divided into two parts: (i) a natural logarithm $\ln$ that flattens high values to gain linearity, and (ii) an offset. All the features are positive by definition. The offset raises or lowers all the values to 0 in order to avoid negative values due to the $\ln$ of values within the range $0 - 1$.

$$N(y) = \frac{y}{\max\{y\}}, \quad (3.16)$$

In Equation 3.16, y represents either a feature or a transformation $f(x)$; and $\max$ and $\min$ refer to the maximum and the minimum values observed. Normalization sets all features to the same range $0 - 1$. This favors the convergence in time and accuracy of the gradient descent optimization engines of the clustering algorithms used further.



(a) Scattering and distribution of $Acme$ and $N(Acme)$ (b) Scatter and frequency of $\delta_{acme} \uparrow$ and $N(f(\delta_{acme} \uparrow))$



(c) Scatter and frequency of D and $N(D)$

Figure 3.8: Scattering and distributions of $Acme$, $\delta_{acme} \uparrow$ and D , and their transformations. *Migraines (id)* identify each migraine.

The Figure 3.8 shows the results of the transformations. These plots show the normalization applied to the features $Acme$ and D respectively. The $Acme$ plot shows the transformation applied to the $\delta_{acme\uparrow}$ that combines Equations 3.15 and 3.16. At a glance, if it is considered the histograms of both row $\delta_{acme\uparrow}$ and D , it seems that both have a similar distributions and ranges. However the curve of D is softer than the abrupt elbow drawn by $\delta_{acme\uparrow}$, where it can be seen that most of the pain episodes seem to have low $\delta_{acme\uparrow}$. This is the reason why the transformations of both features are different. While the logarithm over the $\delta_{acme\uparrow}$ leads to a distribution that favors the clustering, there is no improvement when it is applied over the duration D . Thus, only normalization has been used in the duration.

3.1.9 Unsupervised clustering

The novelty of this research resides in the clustering of the different types of the evolution of pain that a patient suffering from a chronic with symptomatic crises reports in real time. This is a novel proposal and, as aforementioned, the technique in a case of study of chronic disease: migraine will be evaluated.

Because this is a brand new clustering application and there is no reference to compare with nor an expert eye to supervise a machine learning algorithm, it has been implemented an unsupervised learning algorithm to find the different types of migraine pains based on their morphology.

A k -means algorithm which tends to find the centroids of k clusters in a N -dimensional space has been selected. The Euclidean distance has been used to assign pain episodes—the vectors of their morphological features—to each cluster. To find the optimum number of clusters, there have been run 10 iterations of randomized starting centroid locations, for k varying from 2 to 9 clusters.

3.1.10 Model selection and model curve

The last step of the methodology shown in Figure 3.5 consists in modeling the different types of pain of each cluster. To select the right number of clusters k , the *score* of the k -means has been used as cluster-cohesion metric. This metric is useful, because it indicates the difficulty for breaking a cluster into two as it is explain in [49]. The *score* indicates the internal coherence of the clusters, which minimizes the sum of the squares of the distances of the points within a cluster to its centroid. This metric is called the Within-Cluster Sum of Squares (WCSS), and it is computed as shown in Equation 3.17:

$$WCSS(C_k) = \sum_{i=1}^k \sum_{x \in c_i} \|x - \mu_i\|^2, \quad (3.17)$$

where C_k is the k -th clustering composed of k clusters $c_0, c_1, \dots, c_{k-1}$; x is each point (pain event) in the space of the cluster c_i ; and μ_i is the centroid of the cluster c_i .

The higher the number of clusters, the better the k -means score. However, there is a compromise to choose the best solution in order to find the optimal number of clusters

avoiding the overfitting. To achieve this, the elbow-method has been used.

Model curve: schematic representation of types of pain

Our hypothesis establishes that pain curves of symptomatic episodes of chronic diseases can be clustered based on features of their morphology, even though the evolution over time of the pain curves are all different. To get a graphic representation of their morphology, it has been decided to make a schematic representation using some of the described features. It has been assumed that the schematic representation of a pain curve can be drawn using three lines as shown in Figure 3.9. The first line (y_1 in the figure) represents the ascent of the pain. The second one (y_2 in the figure) represents the pain at 90% of the maximum intensity level. The third line (y_3 in the figure) represents the descent of the pain until its end.

It is important explain that this representation has no effect on the classification nor on the model. This simple representation is proposed to facilitate the interpretation of the classification results. This representation shows the assumptions of the characteristics of a pain curve, more specifically the pain suddenness ($\delta_{acme\uparrow}$), the time when the pain starts decreasing, ($\delta_{acme\downarrow}$), the duration (D) and the duration at high pain intensity (D_{90}). If the maximum pain ($Acme$) has not selected, it is because this value is usually represented by a single point.

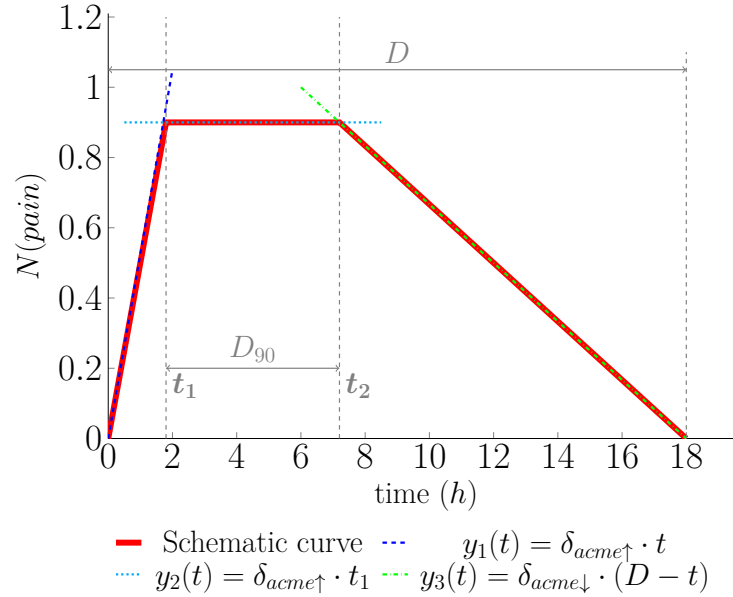


Figure 3.9: Example of the schematic representation of a curve model based on morphological features.

The features needed for this symbolic representation are: $\delta_{acme\uparrow}$, $\delta_{acme\downarrow}$, D , and D_{90} . In addition, two other parameters are needed: t_1 , and t_2 , calculated as shown in Equation 3.18:

$$\left. \begin{array}{l} t_1 = t_2 - D_{90} \\ y_1(t_1) = y_3(t_2) \end{array} \right\}$$

$$t_2 = \frac{\delta_{acme\downarrow} \cdot D + \delta_{acme\uparrow} \cdot D_{90}}{\delta_{acme\uparrow} + \delta_{acme\downarrow}} \quad (3.18)$$

The schematic representation of each crisis will allow to study the shape of the average representation of the pain of each group. Thus, in a simple manner, it will be possible to see the differences between types of pain, which can be interesting from the clinical point of view and proceed with further research.

3.1.11 Group clustering by pain type

Continuing with the previous research, the natural course is to verify if this classification allows patient categorization and if such patient classification could exhibit common characteristics within groups, thus validating the aforementioned classification. To address this, the continuation of the analysis was proposed, and the results and methodology would lead to the following work by Álvaro Martínez Petit [133].

Following the classification of migraines based on the type of pain, the next step involves categorizing patients according to the specific types of pain experienced during their migraines. To achieve this, the same methodology employed in the previous module is applied, utilizing the k-means algorithm as outlined in Chapter 3.1.10.

To implement this approach, each patient is represented by an n-dimensional vector, with each dimension corresponding to a specific type of migraine. The value for each dimension is calculated by summing the number of migraines of the corresponding type experienced by the patient and dividing it by the total number of migraines suffered for normalization. Subsequently, all these vectors are aggregated into a comprehensive matrix, referred to as the patient matrix, which will be used to input data into the k-means algorithm for cohort segmentation.

After identifying the distinct patient groups, a rigorous statistical analysis will be performed to investigate potential relationships with various migraine patient characteristics. This analysis will be based on metrics such as the p-value and alpha to ensure robust statistical significance.

3.2 Recommendator

Drawing from the outcomes of prior groupings, a plan for implementing a clinical advice recommendation system has also been devised. A segment of the findings is detailed in the aforementioned work by Álvaro Martínez Petit [133]. The methodology employed for this system is outlined below.

3.2.1 System components

The recommendation system and its parts will be explained here.

The diagram in Figure 3.10 depicts the components of the recommendation system, including external elements and the data they utilize. The system relies on databases to retrieve

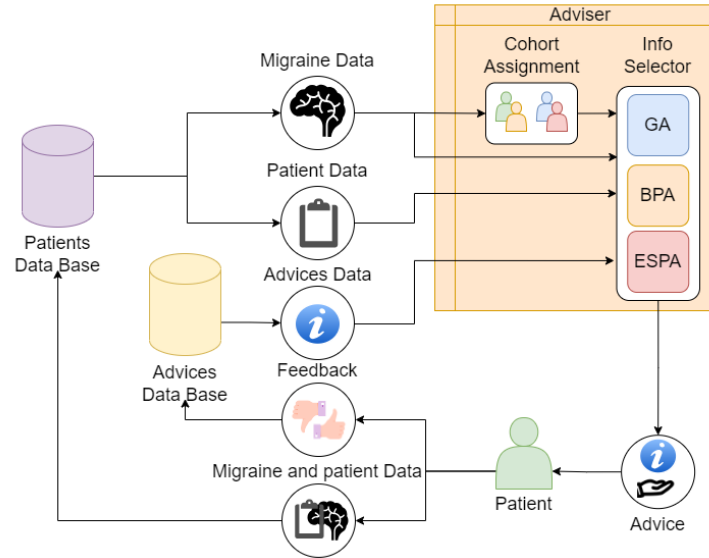


Figure 3.10: System Diagram: This schematic delineates the entirety of the system, elucidating the information flow from the databases to the adviser, from the adviser to the patient, and the process of gathering patient information.

patient-specific data and a collection of recommended advice. The recommendation system consists of two main functions: assigning patients to specific patient cohorts and selecting relevant information based on the patient’s cohort, migraine condition, and sociodemographic data. Patients play an active role by providing essential data for the system’s functioning and can also offer feedback on the received recommendations. The feedback component of the system retrieves data pertaining to previously administered advice to a specific patient, thereby enabling the system to avoid redundant recommendations.

3.2.2 Input data and advice

The system operation heavily relies on two distinct types of patient data:

- **Baseline data:** These are patient-specific characteristics that do not change frequently. They are typically provided once at the beginning of the study and include information such as age, sex, and chronic conditions the patients may have.
- **Pain episode data:** This includes details about each migraine episode experienced by the patients, such as duration, pain intensity, and subjective sensations. A single patient may provide data on multiple pain episodes.

The information intended to be provided to the patients consists of pre-stored advice, developed and/or supervised by clinical personnel or experts in the targeted disease. The advice can be in various formats, including short videos, informative text, or links to external resources such as websites, articles, or other videos.

Before continuing, it is found necessary to clarify certain aspects of the terminology used. Firstly, the term “advice” refers to the different pieces of information that can be shown to a

patient. Secondly, “recommendation” is the act of delivering specific advice to the patient. Thus, two recommendations may consist of the same advice, or multiple pieces of advice may be provided in the same recommendation.

The advice has been classified into three categories based on their content:

- General Advice (GA): This contains interesting information relevant to any user at any time. Although they are not personalized, they are still valuable in patient education. In the case of migraine, general advice may include tips for prevention, technological advancements related to the condition, or interesting facts about the disease.
- Baseline Personalized Advice (BPA): This contains information specifically relevant to individuals based on their baseline characteristics. The relevance of this advice for each patient remains relatively stable over time. Some examples of targeted baseline characteristics are sociodemographic data like sex, age, weight, disease-related factors such as prescribed treatment, or other chronic conditions like high cardiovascular risk.
- ESPA: This category includes information related to the specific characteristics of the disease, such as symptoms or triggers of pain episodes. The relevance of each advice within this category may vary over time, depending on the evolution of the patient’s episodes and the patient cohort they belong to (which will be described later). In the case of migraine, episode-specific advice may address topics like sensitivity to light or lack of sleep.

The recommendation system selects the tips based on their content regardless of their display format, video, text, web link... In order for the system to access this content, each piece of advice is accompanied by a description that specifies its details. The information included in these descriptions varies depending on the type of advice.

GA descriptions specify the series they belong to. Series are subcategories in which general advice is classified. The following six generic series for general advice are proposed: “Clinical”, “Curiosities”, “Lifestyle”, “Technology”, “Special Dates”, and “Myths and Legends”. Each advice in this category can belong to only one series.

BPA descriptions specify the values of the targeted baseline characteristics. Examples of these values can include “Teenager” or “Adult” for the “Age” characteristic, or “Male” or “Female” for the “Sex” characteristic. A baseline personalized advice can target multiple baseline characteristics but only one specific value for each.

ESPA descriptions specify the topics they inform about. These topics group information related to different symptoms or triggers of episodes that share something in common. For example, advice belonging to the “Mood” topic may provide information about symptoms like “Irritability”, “Anxiety”, or “Sadness”. A piece of advice in this category can inform about more than one topic. The advice in this category are modeled with a 16-dimensional vector, in which each one corresponds to a topic which are represented in Table 3.2.

As mentioned in Chapters from 3.1.1 to 3.1.4 in total there were recorded 344 crises from all patients. To work with these data, each crisis is represented as a vector. Each dimension of the vector corresponds to a trigger, symptom, or type of aura recorded. This resulted in

Food	Drink	Stress	Sleep
Weather	Activity	Mood	Cognition
Temperature	Gastrointestinal	Fatigue	Skin
Sight	Hearing	Smell	Aura

Table 3.2: Topics of the ESPA

id	symptom	id	symptom	id	symptom
01	Trigger food	19	Concentration difficulties	37	Blurred vision
02	Trigger drink	20	Speaking difficulties	38	Dizziness
03	Trigger stress	21	Writing difficulties	39	Dermal/vascular changes
04	Trigger sleep	22	Feeling colder	40	Simultaneous pain nauseas
05	Trigger Weather	23	Feeling warmer	41	Simultaneous pain vomits
06	Trigger sports	24	Nausea and/or vomiting	42	Simultaneous pain Photophobia
07	Sadness	25	Flatulence	43	Simultaneous pain Sonophobia
08	Euphoria	26	Constipation	44	Simultaneous pain Osmophobia
09	Hyperactivity	27	Fluid retention	45	Post pain mood
10	Anxiety	28	Increased urinary vascular frequency	46	Post pain appetite
11	Irritability	29	Nuchal rigidity	47	Post pain specific appetite
12	Apathy	30	Yawning	48	Post pain cognitive
13	More appetite	31	Exhaustion	49	Post pain sleep
14	Less appetite	32	Photophobia	50	Post pain temperature
15	Food cravings	33	Sonophobia	51	Post pain yawn
16	Thirsty	34	Osmophobia	52	Aura visual
17	Insomnia	35	Alodynia	53	Aura language
18	Drowsiness	36	Tinnitus	54	Aura sensitive

Table 3.3: These are all the dimension codes for the migraine features. Enumerated from 1 to 54.

a vector of 54 dimensions. Table 3.3 represents these dimensions. The values in the vector were set to 0 if the associated characteristic was not experienced during the crisis, or 1 if it was present. By concatenating these vectors, the two inputs related to migraine crises—the patient and the cohort migraine vector—were modeled for being part of the Neural Network input.

The first functional block of the recommendation system, as shown in Figure 3.10, involves assigning patients to cohorts based on the characteristics of their migraine pain evolution, as explained in Chapter 3.1.11. It is important to note that the assignment of a patient to a cohort may change over time as new episodes occur, leading to dynamic assignment.

3.2.3 Advice selection

This section explains the process of how the information to be recommended is selected, taking into account the different categories of advice in the system. Advice is categorized into different series, with similar advice belonging to the same series. This approach ensures that recommendations meet the specific expectations.

GA selection: The recommendation of GA is done on a random basis due to the lack of an objective measure of relevance for each patient. Remember the reader that this is generalist advice on digital health and migraine.

However, to improve the recommendation process, it is included a filtering system based on feedback given by the patients. Negative feedback reduces the frequency of recommendations of advice. The way the GA are filtered is as follows:

1. **Novelty Filtering:** this filter separates the GA into two lists, one for those that have already been recommended to the patient, and one for those that have not. A GA that has been already recommended is given a score of $GA_{i_{score}} = 0$. Otherwise, it is given a score of $GA_{i_{score}} = 1$.
2. **Collaborative Filtering:** the score for each not recommended GA is then corrected by the number of negative feedback given to it by all patients. In order to leave a chance of being recommended, the advice will always have a minimum score of ρ_{GA} . This is shown in Equation 3.19, where δ_{GA} is a coefficient that weights the number of dislikes $GA_{i_{dislikes}}$ given to it.

$$GA_{i_{score}} = \max\{\rho_{GA}, 1 - \delta_{GA} * GA_{i_{dislikes}}\} \quad (3.19)$$

The absence of feedback or positive feedback has no impact on future recommendations, but the feedback is retrieved for statistical and population analysis purposes, allowing professionals to identify relevant advice among patients.

BPA selection: In a similar manner as for GA, the BPA are also split into lists and then filtered. In this occasion, the way to create the lists is slightly different.

1. **Patient-targeted advice:** first two groups are created: (i) patient-targeted and (ii) no-patient-targeted advice. Patient-targeted BPA is and advice that matches in characteristics with the patient. These are the 10 characteristics shown in Tables 3.4 and 3.5. If any of the characteristics of the advice is different from that of the patient, the advice is included in the no patient-targeted list.
2. **Novelty Filtering:** the list of patient-directed BPA is then divided into two lists, one for advice that has already been recommended to the patient and one for advice that has not. The tips are scored in the same way as is done with the GA Novelty Filter.
3. **Collaborative Filtering:** analogous to that for the GA, the BPA scores are corrected according to equation 3.20.

$$BPA_{i_{score}} = \max\{\rho_{BPA}, 1 - \delta_{BPA} * BPA_{i_{dislikes}}\} \quad (3.20)$$

The no-patient-targeted list of BPA is only used if all the advice of the patient-targeted list have been already recommended.

ESPA selection: The recommendation of ESPA represents the most important part of the system as they help the patient to better understand the crises of their disease at any given

Sex	Migraine	Family History	Treatment	Hypertension	Dyslipemia
0. Man	0. Without aura	0. No	0. No intake	0. No	0. No
1. Woman	1. With aura	1. Yes	1. Take	1. Yes	1. Yes

Table 3.4: Characteristics of interest of the BPA and the values of their categories [1/2].

Age	BMI	Crisis	HIT-6
0. Child	0. Low weight	0. Infrequent episodic	0. Low impact
1. Teenager	1. Healthy weight	1. Episodic	1. Some impact
2. Adult	2. Overweight	2. Frequent episodic	2. High impact
3. Elderly	3. Obesity	3. Chronic	3. Severe impact

Table 3.5: Characteristics of interest of the BPA and the values of their categories [2/2].

moment. This part of the system incorporates information from the migraine pain type cohort to which the patient belongs to make the advice recommendation.

The ESPA recommender system is a NN that receives as an input a vector with information of (i) the characteristics of the patient’s recent pain episodes, where the most recent episodes have a larger importance than the former ones, (ii) the cohort the patient belongs to, and (iii) the topics of Table 3.2 addressed by each advice. The output of the neural network for each advice is a score. The scores are sorted in decreasing order. Therefore, the scores change with each new episode the patient experiences.

This methodology is based on the concept of Neural Collaborative Filtering (NCF) described in the work of He *et al.* [81], where a MLP is used to obtain binary interest scores for a user and each item in a set. For the selection of ESPA, The scores are intended not to be dependent on the preferences of a patient but rather on their migraine episodes and corresponding cohort. For this reason, it is proposed an adaptation of this model to create a Hybrid Neural System (HNS) that replaces the content of patient feedback by the crises information of their cohort. The final architecture implementation has been done with an RNN. However, for the sake of completeness, the results section compares this implementation with that of an MLP.

Under this premise, it is presented the method for selecting ESPA, whose architecture is depicted in Figure 3.11. In the figure, E_p denotes the set of patient episodes, and e_p represents the information of the entire compressed set. Similarly, E_c represents the set of episodes from the patient’s cohort, and e_c symbolizes its compressed information. The set of available advice is denoted as C , and c_i represents an element of that set. Lastly, y_i represents the score assigned to advice c_i . The elements that constitute this method will now be described.

Inputs to the NN: the NN takes as input a vector comprising a total of 124 elements:

1. **Patient episodes:** 54 elements representing a weighted information of all the pain episodes of the patient for whom the advice is intended. Each vector contains information of each trigger, symptom, and type of aura.

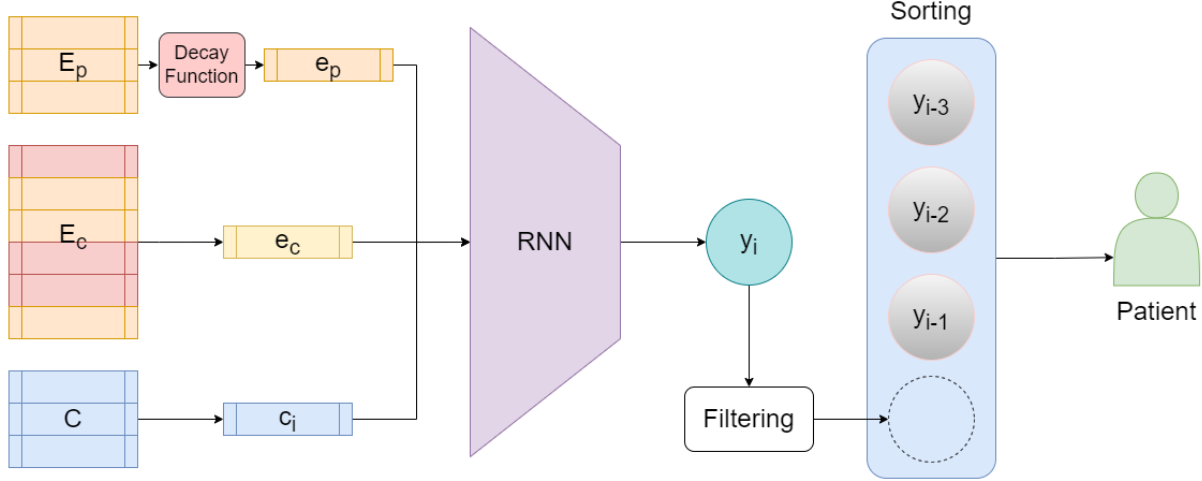


Figure 3.11: Architecture for ESPA selector. The input layer of the RNN is partitioned into three distinct blocks. The first block e_p representing the combination of symptoms from all the patient’s migraine occurrences, subject to a decay function. The second block e_c is the mean of all the cohort migraines, including those of the patient under consideration. In the figure, the red row are from the patient E_p and the orange are the rest of the cohort patients. The third block c_i is defining the advice to be assessed and scored.

To create these 54 elements of the vector, the weighted sum of all previous crises is used. Each crisis is weighted according to the expression in Equation 3.21, where t represents the index of the crisis in inverse order (0 represents the most recent migraine). The longer ago the crisis occurred, the less weight it will have. Thanks to it, the symptoms and triggers of the most recent crises hardly lose any contribution. This function has been modeled heuristically. After conducting several tests with different functions, this one was selected as it demonstrated better performance in the models.

$$f(t) = \frac{1}{2}(\tanh(2.5 - 0.7t) + 1), \quad (3.21)$$

2. **Standard cohort crisis:** 54 elements representing the pain episodes of the patients belonging to the same cohort as the target patient.

To create these 54 elements of the vector, the arithmetic mean of migraine episodes of all patients of the cohort is calculated, which is called as the “standard cohort crisis”. It is necessary to recalculate this standard crisis both when a patient of the cohort suffers a new migraine and when a patient enters or leaves the cohort.

3. **Topics of an advice:** a binary representation of the 16 topics of the advice to be scored.

In a recommendation for a specific patient, the first two representations e_p and e_c remain constant, and only the 16 elements representing the topics addressed by the advice being

scored (c_i) change. This would entail the neural network running once for each individual advice. While this approach may initially appear to lack scalability —particularly in cases where the number of advice increases—, it actually represents a more efficient way compared to one that processes all advice simultaneously. The justification for this explanation lies in the way topics are assigned to the advice. The encoding of advice facilitates the dynamic addition of new ones in many instances without necessitating a full model retraining. If a newly introduced advice shares the same encoding as an existing one, the model’s output remains consistent, as it already possesses knowledge of the information encoded by that specific advice.

At times, achieving patient classification based on the nature of their pain proves to be unfeasible due to the inherent difficulty in obtaining the requisite data for cohort formation. Consequently, an analytical endeavor is undertaken to assess the system performance in the absence of this information and to quantify the ensuing impact on the system. When cohort information is omitted, the input size diminishes from 124 inputs to 70. That implies the training of a different Neural Network for the ESPA recommendation system.

RNN: Its function is to reconstruct the suitability score of each advice for each pain crisis, as it would be assigned by a medical professional. Since the network’s functionality is regression, to evaluate the results of the network, the Mean Absolute Error (MAE), the coefficient of determination R^2 , and the MSE have been used .

RNN Outputs: The output of the network is a real number bounded between 0 and 1, corresponding to the score for the input advice. The closer the advice aligns with the patient’s condition, the closer the score is to 1. The appropriateness of recommending each piece of advice to each patient has been determined on expert opinions from the Neurology Unit of La Princesa University Hospital in Madrid. With their help, the conclusion was reached that the selection of the advice topic to recommend is primarily based on the symptoms and triggers of the current and previous crises.

Filtering: After obtaining the RNN result a novelty filter is applied in order to reduce repetition. Because an advice can be recommended twice or more times, the novelty filtering decreases the score by 2 if the advice has been recommended before. This prevents recommendations from getting stuck on the same advice and promotes variety.

Sorting and Delivery: Finally, once the scores for each advice have been adjusted, they are sorted in descending order. Each time an ESPA is needed, the selection is made sequentially based on the scores. In case of tie between two or more pieces of advice, they are considered equally relevant, and the order in which they are presented is not significant. If the patient experiences a migraine episode after creating this sorted list, the entire process needs to be repeated to generate a new list that better matches the patient’s new condition and possible cohort reassignment.

3.3 Sleep model

In this subchapter, the experiments and methodology concerning sleep are delved into. As highlighted in Chapter 2.3, the significance of sleep quality for overall health, particularly its

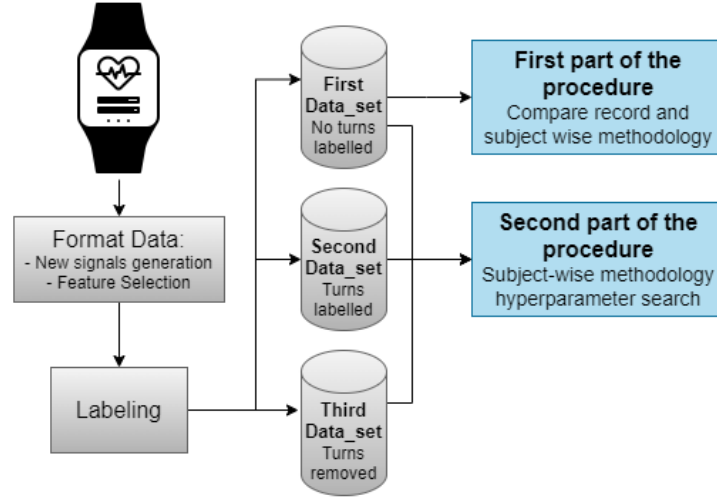


Figure 3.12: Methodology diagram.

association with headaches, notably migraines, cannot be overstated. The primary aim of this section is to leverage the same tools—wearable devices and AI technology employed throughout this thesis—to integrate insights into sleep quality into the system. This integration aims to refine predictions and offer additional valuable information to the patient.

This subchapter unfolds in two distinct parts: firstly, an exploration of the sleep state detection model, pivotal in calculating sleep quality, and secondly, an emphasis on posture detection. The latter is designed to augment and support the functionality of the primary system.

3.3.1 Posture detector model

The market size of wearable devices is growing year by year. The use in any population segment is generalized, which makes them an enabling tool to bring health to home through ambulatory monitoring. Generally all commercial wearable devices include an accelerometer; an inexpensive and energy efficient technology. Therefore, in this section, the focus will be on the use of wrist wearable devices, specifically the Empatica E4 already mention in the Chapter 2. This part of the work is based on the need to implement a torso posture detection algorithm for real life. After searching the state of the art, and seeing that it is feasible to do it using wearable devices, it was realized that the results were not able to be replicated. The main problem encountered was that testing with new subjects on machine learning models created by training-test split data on a randomization over all the data did not work.

Consequently, the decision was made to create an experimental dataset and investigate the reasons why the results of the state-of-the-art could not be replicated. In this work we: (i) evaluate classical machine learning algorithms to detect 4 torso postures during sleep, (ii) analyze the best features for these models based on clinical literature; (iii) show that the best solution for a real implementation is to create trained models with a subject-wise data partitioning; and (iv) demonstrate that it is essential to detect and isolate turns and sudden wrist movements to achieve a satisfactory classification.

Our experimental methodology is based on two hypotheses about why the models of the SoA are not replicable:

1. The system fails when new subjects are presented to models trained with data splitting techniques based on stratified random data splitting, or cross-validation (CV) over the entire dataset. This was demonstrated by Saeb *et al.* in [174]. The authors demonstrated that in order to bring machine learning models closer to real—clinical—practice, a subject-wise data splitting is needed. A subject-wise CV method does not overestimate the ability of machine learning models to predict clinical outcomes as a record-wise CV method does.
2. Motion signals are very stable during sleep. The evaluation of posture during movement of the subjects generates much noise that does not allow a good generalization of the models. Therefore, it is suggested labeling these movements which, in general, are due to body turns during sleep. Once identified, the feasibility and worth of classifying the turns will be evaluated, or on the contrary it would be better to detect them and override the posture classification during the time when the subject moves.

It is considered this proposal to identify and isolate movement periods constitutes a key point of differentiation in the work. This, combined with a subject-wise training strategy, has led to improvements in the metrics compared to the state of the art and has made them reproducible. The results are being compared with those referenced in Table 3.6, which have been deemed the most relevant.

Figure 3.12 shows the experimental scheme followed along this work. This section will go through each of the following blocks: (i) data acquisition, (ii) data labeling (with and without torso turns), (iii) feature generation and feature selection, and (iv) modeling, where it is compared record-wise and subject-wise methodology with and without labeling of the turns.

Data acquisition

It is conducted a controlled monitoring experiment with 13 individuals: 7 women (53.8%) and 6 men (46.2%), on average 32.1 ± 14.6 years old. The monitoring was performed using the Empatica E4 wristband shown in Figure 2.1. This wristband has a 3-axis accelerometer sensor in the range $[-2, 2] \times g_n$ of 8 bit resolution and 32 Hz sampling frequency. The wristband has an Event Mark Button to indicate timestamps.

As mentioned above, the four postures monitored have been: prone, supine, left lateral and right lateral. The posture will be estimated with respect to the torso, *i.e.*, with the torso at rest, no matter the movement of the extremities and/or head.

During the experiment, the participants wear the bracelet on the non-dominant wrist, always leaving the button on the thumb side. The experiment lasted about 34 minutes and was divided into three rounds of 360° turns. Starting from the supine position, the subject rotated four times to the right, passing through the other three postures and concluding in the starting position. Each posture was held for 5 minutes. The next round was analogous to the former, changing the direction of the turn to the left and holding each posture for 3 minutes. Finally, the last round, repeats the first one holding each posture for 30 seconds. The Event Mark

Table 3.6: Relevant works sleep position detection for comparison

Authors	Devices Used	# postures	# subjects	Hours monitored
[88]	FSR sensors and one infrared sensor array	6	?	?
[150]	Pressure sensor array	3 (13)	9	— ^a
[118]	Depth sensor and RGB camera	10	22	— ^a
[206]	Porcupine wearable wristband	3-5	4	63
[96]	1 wristband	4	2	0.3
[6]	9 (2) accelerometers	4	22+8	0.2+1.3
This work	1 wristband	4(+2)	13	7.4

Button was pressed at the end and the beginning of each posture.

Data labeling

As discussed at the beginning of this Chapter 3.3.1, the results of the models of the state of the art were not able to be reproduced. One of the hypotheses for which the results have not been reproducible is: the signal values during the turns in the bed add noise preventing a good performance of the model. Thus, using the timestamps, the time intervals in which the turns occurred were able to be delimited. An example is shown in Figure 3.13, where the 3 axes of accelerometry are shown for the four postures and turns.

Three different datasets have been used for experimentation, as shown in Figure 3.12: (i) the first dataset does not consider the labeling of the turns but only the four postures of the torso. The turns are labeled with the value of the previous posture; (ii) the second dataset contemplates the labeling of the right and left turns, resulting in a total of 6 classification labels; and, (iii) the third dataset marks and eliminates the time periods of the turns. Table 3.7 shows the percentage distribution of the data for each of the postures and turns. Note that, in the experiment, turns constitute approximately 3% of the total data.

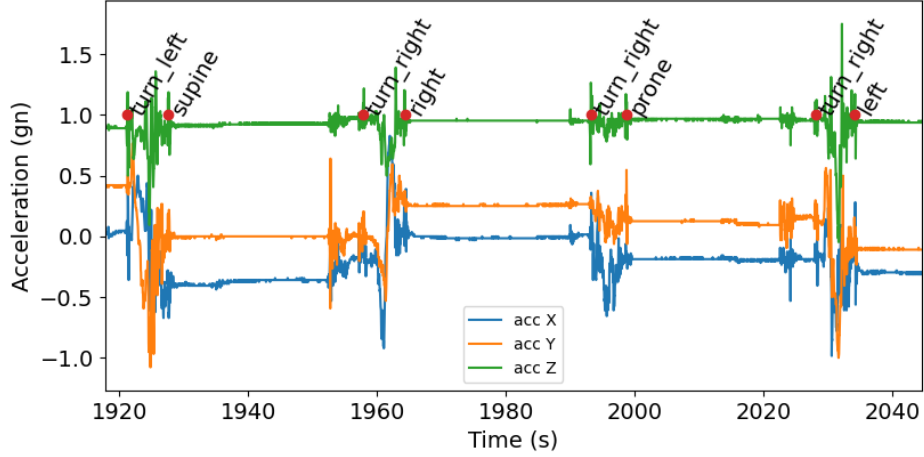


Figure 3.13: Example with the accelerometry values of 4 postures and turns.

Table 3.7: Data labeling distribution

Label (position)	Number of 5-second windows	Label distribution (%)
up (supine)	6663	25.32%
left	6472	24.60%
down (prone)	6443	24.49%
right	5947	22.60%
turn right	525	1.99%
turn left	260	1.00%

Feature generation and feature selection

The first experiments were based on the features used by [96], which were the mean value of every axis of accelerometry. As the results were not satisfactory, it was also decided to add other features considered of interest, as well as other features commonly used in the state of the art of clinical sleep studies [197].

After some preliminary tests, it was decided to use a 5-second window with 4-second time overlap, thus yielding results every 1 second. These values allow us (i) to cover the time duration of torso turns as measured in 6.3 ± 2.0 seconds, and (ii) to have a richer dataset for training. A final market implementation would not have such a high resolution according to the time standards of clinical studies which are 30 seconds [90].

Thus, 42 features have been calculated in each window. For each one of the 3 accelerometry axes, the Euclidean Norm Minus One (ENMO), the Z-angle, and the Energy there were computed as defined in [197]: the maximum, minimum, mean (μ), standard deviation (σ), mean value of the first derivative (μ_δ), the area of the absolute value of the signal ($A_{|si|}$), and the counts ($c_{|si|}$). The counts are defined as the number of crossings per zero crossover, or number of crossings for a threshold delta for positive signals only.

Previously to the modeling step, a feature selection has been performed based on Pearson's

correlation. The features with a correlation higher than 0.8 have been removed.

Modeling

The modeling methodology will present several experiments named as *procedures* in Figure 3.12.

The first procedure aim to show that the state-of-the-art models are not reproducible when used on data gathered from new individuals. The models are trained using a record-wise cross-validation and then tested on new subjects. This procedure corroborate the results of [174] and proceed to continue with a subject-wise split methodology on the training data.

The second part of the procedure intend to demonstrate that delimiting and removing the time intervals in which torso turns—or abrupt wrist movements occur—improves the performance of the models. To do so, subject-wise cross-validations on the three different labeled datasets were performed: (i) without marking the turns, (ii) labeling the turns, and (iii) removing the data acquired during turns.

During the whole modeling process we:

- Compare the results of three different machine learning algorithms: a RF, a SVM, and a KNN. 5-fold cross-validations were performed, and the evaluation metric is the F-value;
- Keep aside data from 2 randomly chosen subjects as if they were real life test data;
- Use a Bayesian hyperparameter optimization in order to search the best combination of hyperparameters for each model [57]. For that purpose, The Python libraries Scikit-learn and Tune (Scalable Hyperparameter Tuning and TuneSearchCV) were used.

The list of hyperparameters is as follows:

- RF: *# of trees*: [2000, 1000, 500, 250]; *maximum depth*: [500, 300, 100, 50]; *minimum samples per split*: [50, 20, 10, 5]; *minimum samples per leaf*: [50, 20, 10, 5]
- KNN: *weights*: [*uniform*, *distance*]; *algorithm*: [*auto*, *ball – tree*, *kd – tree*, *brute*]; *p*: [1, 2] being 1 the Manhattan distance, and 2 the Euclidean distance.
- SVM: *regularization parameter (C)*: [1, 2, 3, 4]; *kernel*: [*linear*, *poly*, *rbf*, *sigmoid*]; *gamma*: [*scale*, *auto*]; *maximum number of iterations*: without limit.

For the sake of repeatability, all training processes have been performed by setting a seed. The results chapter discuss an analyze the results obtained during the experiments.

3.3.2 Sleep model implementation

As observed in Chapter 2.3.1, the best results achieved in the attempt to detect sleep states using biomedical signals collected by a wrist-worn wearable device were obtained by Pedro Fonseca *et al.* The intention is to enhance these results by incorporating EDA signal into the system to provide additional relevant information. One aspect to consider in developing this methodology is the need to apply TL to achieve improved results, as it aided Fonseca *et al.* in enhancing their system.

Dataset

For this module, collaboration was undertaken a prospective investigation carried out at the Sleep Unit of "La Princesa" University Hospital. The main objective of this research was to evaluate individuals who underwent PSG for diagnostic reasons in 2021, while simultaneously collecting data from the Empatica E4 wristband wearable device, which was specifically selected for this study. Approval for the study protocol was obtained from the local Biomedical Research Ethics Committee, with approval number 4327, the documentation is in Annex 6.7. Before participating, all subjects provided informed consent by signing necessary documents, indicating their voluntary and informed participation in the study.

Throughout this investigation, each subject underwent a single night of in-laboratory PSG to facilitate screening. This process involved the recording of various critical biomedical parameters. Participants were instructed to retire to bed between 10:00 pm and 11:00 pm, with the duration of sleep extending until 7:00 a.m. to 8:00 a.m. Furthermore, the sleep laboratory meticulously maintained a stable room temperature of 23°C, creating an optimal setting for the sleep assessment.

Signal preprocessing and feature extraction

In sleep studies using PSG technique, a standardized temporal segment of half a minute has been established for sleep state detection. These sleep states are: awake, N1 and N2 for light sleep, N3 for deep sleep and Rapid Eye Movement (REM) phase. Therefore, both this approach and Fonseca *et al.* utilized this temporal step to generate network outputs. Fonseca *et al.* employed four-and-a-half-minute slicing windows for signal feature calculation, with the half minute steps, while here this model is adopted, but also it is decided to explore window sizes ranging from five minutes to half a minute.

An expert-reviewed signal analysis was conducted to classify the quality of the signals collected from each patient, thereby enabling the grouping of patients based on signal quality.

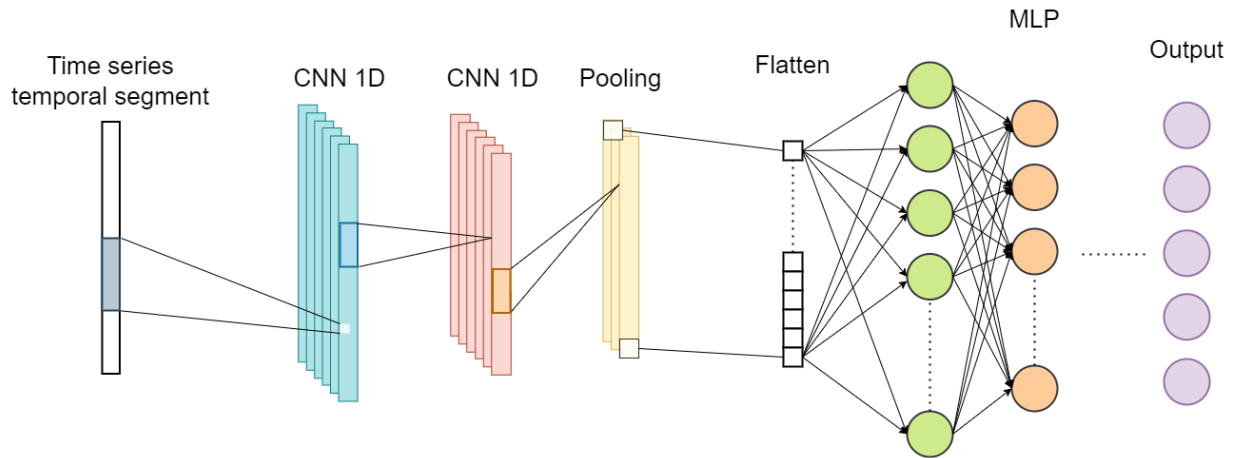


Figure 3.14: Network architecture with the automatic feature extraction model.

As is common in most state-of-the-art approaches, perform manual feature extraction were performed. 165 features extracted from various signals were obtained, including PPG, EDA, TEMP, ACC and HR signals from Empatica.

A trial would also be conducted involving automatic feature extraction using CNN. While it does not always guarantee performance improvement, as observed in [184], given the scarcity of studies on sleep detection in biomedical signals obtained from wrist-worn wearables and the significance of FS in these models, it could be verified whether this selection is appropriate through this procedure. This is the rationale behind the employability of this technique for feature extraction. A scheme of the configuration for automatic feature extraction can be observed in Figure 3.14.

Different experiments conducted

Several experiments were conducted on the data, including:

- Classification of patients by type of disease and training of models associated with each group.
- An automatic classification of patients based on the statistics of different biomedical signals using a ML model employing the K-means algorithm. Using the elbow method and heuristic analysis of the distributions as explained in Chapter 3.1.
- Cross-validation by patient for an analysis of their behavior.
- Models were also compared based on the classification of patients by signal quality. Additionally, a model employing TL techniques was developed, which is detailed in the following section.

Transfer Learning

After the model was finalized, in line with the approach described by Fonseca *et al.* [58], this study sought to improve the model using TL techniques. It was decided to apply these techniques directly to the PPG signal.

To address this, a limitation was the lacked access to a dataset as extensive as Fonseca *et al.* , so open-access models based on images [54] were used. To achieve this, the one-dimensional signals were transformed into images, a widely practiced approach achieved through spectral representation [95], [159], [185]. The most common method for this transformation is through the Discrete Wavelet Transform (DWT). However, it suffers from the drawback of losing intricate time-frequency details in its analysis [43]. This limitation is addressed by the Continuous Wavelet Transformation (CWT), which is widely used in biomedical signals [3], [189], [202], [214], but its computational cost is high [67], [190]. Nevertheless, for this study, the CWT would suffice; however, in a production scenario, the fast Continuous Wavelet Transformation (fCWT) could be employed [13].

Therefore, two types of image transformations were performed: CWT and Spectrogram. With these inputs, a fine-tuning of pre-trained models on images such as EfficientNet and ResNet would be conducted. In [199], they demonstrate better efficiency of the EfficientNet

architecture over ResNet, thus it was decided to employ the former in the experiment. In addition to testing combinations in the architecture by adding the remaining signals in tabular format with extracted features as inputs or even performing a patient-wise check, whether adding a mean image of the CWT of the PPG as a reference to the model improved its prediction.

3.4 Migraine pain predictor

This chapter will elucidate the methodology employed for the simplified model of pain prediction. Due to the scope constraints of the thesis, there was insufficient time to integrate the results obtained from the other modules into this one. Therefore, the discussion on this implementation will be elaborated in the section dedicated to future research.

3.4.1 Study population

For this module, the same population as in the pain classification chapter (Chapter 3.1) was utilized. However, patient filtering differed as it necessitated the recording of biometric signals using Empatica for at least one hour before the onset of pain and one hour after the commencement of pain. In contrast, the notation of pain intensity was not as crucial for this context.

Subjects with migraine recruited at a specialized Spanish Headache Unit were included, after signed informed consent. The inclusion criteria were: 1) diagnosis of episodic migraine with or without aura made by a neurologist specialized in headaches, according to the criteria established in ICHD-3 [91] with a range between two and eight attacks per month in the 3 months prior to the inclusion (to record a minimum of two episodes and prevent overlap between attacks); 2) age between 15 and 69 years; 3) minimum of 3 weeks of follow-up period or at least 2 attacks of migraine during the monitoring period; 4) history of migraine during at least 1 year; 5) proficiency user-level with smartphone-type electronic devices; 6) normal neurological examination; 7) given informed consent to be included in the study. Patients undertaking preventive treatments for migraine at a stable dose for at least three months at the time of inclusion in the study were allowed to participate. Symptomatic pain-relief treatment was allowed in moderate-intense headache episodes. Exclusion criteria were: 1) non-migraine headaches that could not be distinguished from migraine attacks by the patients; 2) highly incomplete (less than one hour period of recording) or incorrect information regarding attacks recording after visual inspection of the monitoring record and information included in the app; 3) cognitive impairment or any other disorder that may prevent them from correctly complying with study conditions, 4) any of the following comorbidities: heart disease, bronchopathy, Sjogren's syndrome, Diabetes Mellitus, uncontrolled blood pressure, hyper/hypothyroidism as well as adverse events; 5) presence of pacemakers or any electronic device that may interfere with the wireless device; 6) pregnancy; 7) active infections or acute diseases during monitoring.

Then a total of 8 patients with episodic migraine were included in the study. Most patients were women 7/8 (88%) and median age was 46 (IQR: 34-48) years. Among patients recorded,

Table 3.8: Baseline characteristics of the patients used in this study. P:Patent; F: Female; M:Male; HIT-6: Headache Impact Test. Level of education according to the International Standard Classification of Education (ISCED); 4: Post-secondary non-tertiary education; 5: short-tertiary education. - Part 1

N	Gender	Age (years)	Level of education	Aura	Duration of migraine disease (years)	HIT-6
P1	F	36	5	No	22	70
P2	F	45	5	No	20	59
P3	F	54	5	No	48	55
P4	M	48	5	No	32	63
P5	F	27	5	Yes	3	50
P6	F	22	4	No	12	59
P7	F	47	5	Yes	33	61
P8	F	48	5	Yes	44	60

Table 3.9: Baseline characteristics of the patients used in this study. P:Patent; BMI: Body Mass Index; CDS: Candesartan; Mg:Magnesium; OnabotA: OnabotulinumtoxinA; TPM:Topiramate; AMT:Amitriptyline. - Part 2

N	Family history	Vascular risk factors	Anxiety/Depression	BMI	Preventive treatment
P1	Yes	No	Anxiety	28.1	No
P2	No	No	No	22.3	AMT
P3	No	No	Depression	31.5	AMT CDS
P4	Yes	No	No	24.4	AMT
P5	No	No	No	20	No
P6	No	No	Anxiety	21.5	Mg
P7	Yes	No	No	34.6	OnabotA CDS Mg
P8	Yes	No	Anxiety	26.6	OnabotA AMT TPM

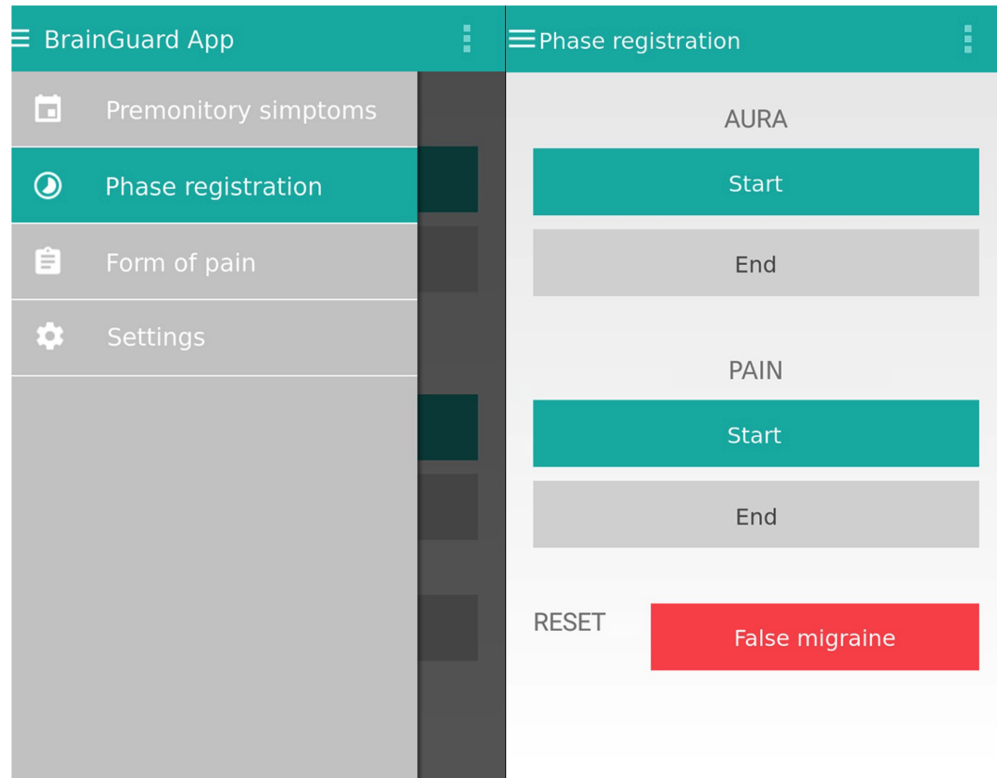


Figure 3.15: Screenshot of the BrainGuard application specifically designed for migraine studies.

3/8 (38%) had a diagnosis of migraine with aura. Half of the sample had family history of migraine. Median duration of migraine diagnosis was 27 (IQR: 18-35) years. Severity of migraine measured by HIT-6 scale revealed that migraine had an important impact with a median score of 60 (IQR: 58-62). No patient had vascular risk factors. Comorbidity of anxiety or depression was present in half of the sample. All the baseline characteristics of the 8 patients included in the current study are included in Tables 3.8 and 3.9.

3.4.2 Study protocol

In the initial visit, individualized instructional sessions were conducted with patients to elucidate the study's purpose and methodology, and to seek their participation. Following the signing of informed consent, baseline socio-demographic and clinical information was collected, encompassing variables such as gender, age, age of migraine onset, family history of migraine among first-degree relatives, educational level, vascular and mood-related comorbidities (including any vascular risk factors, anxiety or depression diagnoses), Body Mass Index (BMI) score, Headache Impact Test (HIT-6) scale score, and details of regular preventive treatments.

Simultaneously during this visit, patients were requested to install the BrainGuard App on their mobile devices. This application, developed by the research team specifically for Android operating systems Figure 3.15, was obtained from the research team's servers. It allowed patients to securely transmit data anonymously to the team's cloud servers.

Subsequently, the research team assisted patients in setting up the device and provided a comprehensive manual for downloading data. Throughout the study, patients were instructed to wear the device as consistently as possible, with allowances for its removal during sleep or showering, particularly during routine activities that reflected their typical environment. In the context of migraine attacks, patients were required to real-time mark the onset and conclusion of the pain phase within the app.

To confirm the patient’s adherence, the BrainGuard app incorporated a recurring alarm every three days to gently remind patients of the scientific community’s gratitude for their ongoing participation in the study, unless any warning signs, auras, or pain had occurred. The notifications received by the patient’s app during the study are detailed in Table 3.10. Additionally, patients received weekly phone calls from the neurologist and correspondence from technicians to ensure compliance with the study and address any concerns about using the wearable device or the BrainGuard app. Furthermore, patients were provided with an email address and phone number for inquiries related to the study. They had the option to withdraw from the study at any time based on their discretion.

Table 3.10: The table documents the notifications sent to patients via the BrainGuard app throughout their involvement in the research.

Event	Notification
Onset of aura indicated more than 3 hours ago and no onset of pain.	“An aura was indicated three hours ago but a migraine has not onset. The episode has been restarted.”
When the headache pain ends and before the patients can initiate symptoms, aura or another crisis.	“A migraine has been recorded but the associated questionnaire has not been fulfilled. Please fill it out and press “Send” in the section “Phase Registration””
The patient indicates one or more premonitory symptoms, but the pain has not started within 24 hours.	“Premonitory symptoms have been indicated 24 hours ago but the pain has not been initiated. If a headache does not start in the next 24 hours, the migraine episode will be restarted.”
The patient indicates one or more premonitory symptoms, but the pain has not started within 48 hours.	“Premonitory symptoms have been indicated 48 hours ago but no pain has been initiated. The migraine episode has been restarted.”
Every 3 days if no activity was reported in the app.	“Thank you for keeping participating in the study.”

3.4.3 Statistical analysis

Categorical variables as frequencies and percentages and continuous variables as means and SD or medians and Inter-Quartile Ranges (IQR) are presented. Due to the novelty of the

model, pre-study statistical power calculations were not feasible. Sample size was based on available data. A complete-case analysis method were performed. This analysis is the primary analysis of these data.

Data processing, feature generation and selection

The time records from both the wristband and the BrainGuard app were aligned. Biometric signals underwent preliminary processing through an in-house tool known as EDWAR (EDWAR library: "signal Error Detection for WBSN And data Recovery," accessible at [LSI edwar project](#). This tool enables the processing of data to identify and rectify anomalies, paving the way for the generation of features from data sourced across diverse devices.

Table 3.11: Algorithm 1 and 2 based on the different state-of-the-art methodologies.

Algorithm 1:

```
// Inputs:
// Electrodermal activity from the
// Empatica E4 wristband (fs=4 Hz)
1:      EDA;

// EDA artifact detection:
// Artifact detection based on SVM algorithm
// (Taylor et al., 2015)

// Oversampling from 4 to 8 Hz
2:      EDA_ovs = oversampling(EDA, 4, 8);
// Windowing: 5 sec, no oversampling
3:      EDA_win = create_windows(EDA_ovs, 5, 0);
// SVM based artifact detection
4:      EDA_artifacts = artifact_detection(EDA_win);

// EDA artifact correction:
// Delete detail coefficients of the
// Wavelet transform of segments
// with artifacts detected and recover the data by the
// inverse Wavelet transform
// (Can et al., 2019; Chen et al., 2015)
5:      EDA_corrected = correct_artifacts(EDA, EDA_artifacts);

// EDA features:
// Compute SCL (tonic),
// SCR (phasic),
// SMNA (sudo-motor nerve activity)
// applying the cvxEDA algorithm (Koskimaki et al., 2017)
6:      SCL, SCR, SMNA = compute_eda_features(EDA_corrected);
```

Algorithm 2:

```

// Inputs:

// Inter-beat intervals (IBI) from the Empatica E4 wristband
1:      IBI;
// Array of indexes of IBIs to remove
2:      idx_to_remove[];
// Array of IBIs to add, and where
3:      ibis_to_add[][];

// IBI artifact detection based on a modification of
// the methodology of (Fox & Haskell, 1968).
// Detect inter-beat intervals in the upper limit:
// noise, extra QRS peaks detected (220 bpm → 273 ms ()).
4:      extra_beats_idx = detect_extra_beats(IBI, 273);

// Detect inter-beat intervals in the lower limit:
// missed QRS peaks
// (40 bpm → 1500 ms (US Department of Health:
// https://medlineplus.gov/ency/article/003399.htm)).

// Indexes of the next useful beats
5:      missed_beats_idx = detect_missed_beats(IBI, 1500);

// Remove extra beats and join missed indexes
// if they are consecutive:
6:      IBI_clean = clean_IBI(IBI,
                             extra_beats_idx,
                             missed_beats_idx);

// For each beat compute N, the parameter of correction
// that indicates the number of beats that should exist
// between consecutive intervals:
    - If N equals 0, the faulty IBI is ignored and
      added to the adjacent IBI where it produces
      the least difference with the remaining signal.
    - If N equals 1, do not modify the interval
    - If N greater than 1, divides into N intervals
      with values linearly interpolated with respect
      to neighboring IBIs.

```

```
7:      for (i = 1, i < length( IBI ), i++){
8:          N = round( 2*( IBI[ i ]/( IBI[ i-1]+IBI[ i+1] ) ) );
9:          if (N == 0){
// Update the array
10:             idx_to_remove = append( idx_to_remove, i );
11:          }
12:          else if (N > 1){
// Update the array
13:             ibis_to_add = divide_IBI( ibis_to_add, IBI, i, N );
14:          }
15:      }

// Correct the IBI:
16:      IBI_corrected = correct_IBI( IBI, idx_to_remove, ibis_to_add );

// Compute the heart rate (HR):
17:      HR = IBI_corrected / 60;
```

The HR signal was derived from the Interbeat Interval (IBI) signal obtained from the Empatica E4 bracelet, extracted from the PPG signal. Error detection and data correction were applied to the IBI and EDA signals using modified versions of various cutting-edge methodologies shown in Table 3.11 [32], [38], [103], [175], [200]. In the case of EDA, artifact detection relied on machine learning using SVM Support Vector Machines (SVM) [200]. This method was applied to 5-second non-overlapping windows, and since the tool requires a signal at 8 Hz, the signal was oversampled and a low-pass filter at 1 Hz was applied beforehand. Following artifact detection in both the original and filtered signals, signal correction was performed using Wavelet transform filtering on the affected segments. It's important to note that the implemented solution [32], [38] does not eliminate detected artifacts but rather removes peaks by suppressing the detail coefficients of the Wavelet transform. As for the temperature signal, it underwent only a low-pass filtering process.

Following signal processing, the phasic (SCR) and tonic (SCL) components of skin conductance responses (EDA), along with SMNA, were calculated from the EDA signal using established methodologies [32], [38]. Additionally, three stress levels—amusement, baseline, and stress—were determined through a RF-based machine learning model trained on the WESAD dataset [103]. The ACC was solely employed in generating the stress signal and was not utilized in the prediction model.

In 1-minute displacement sliding windows, a total of 90 parameters were computed for the signals: TEMP, EDA, SCL, SCR, SMNA, HR, and stress, considering windows of 5, 15, and 30 minutes of past signal. Statistical features included (i) mean, maximum, minimum, mean of the derivative, and standard deviation for TEMP, EDA, SCL, SCR, HR; (ii) maximum and count for SMNA; and (iii) mode for each of the three stress levels. All features underwent normalization using the *Z - score*. Figure 3.16 shows the flow chart of the methodology of data collection and character extraction.

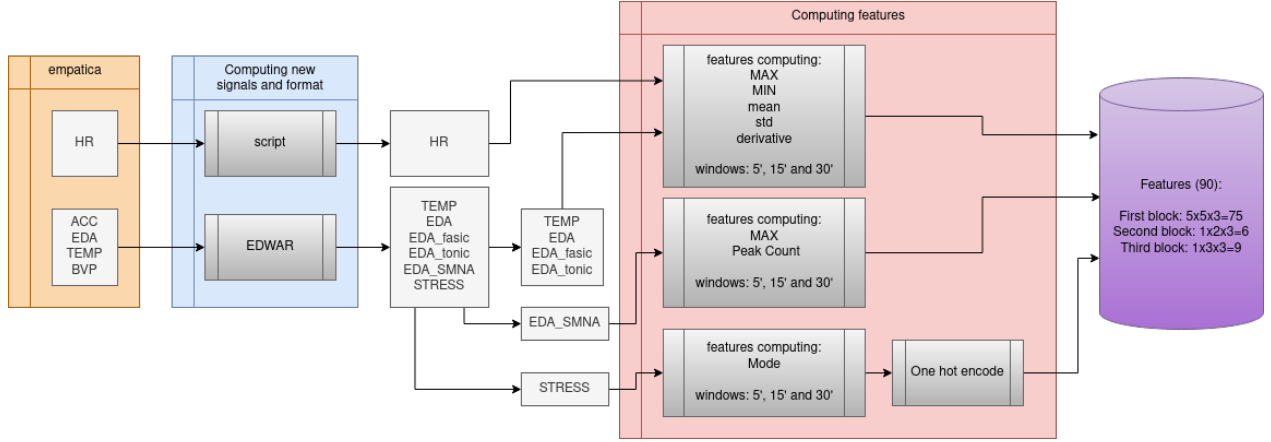


Figure 3.16: Data pre-processing methodology framework: Methodology and data flow from data collection through edwar to feature generation.

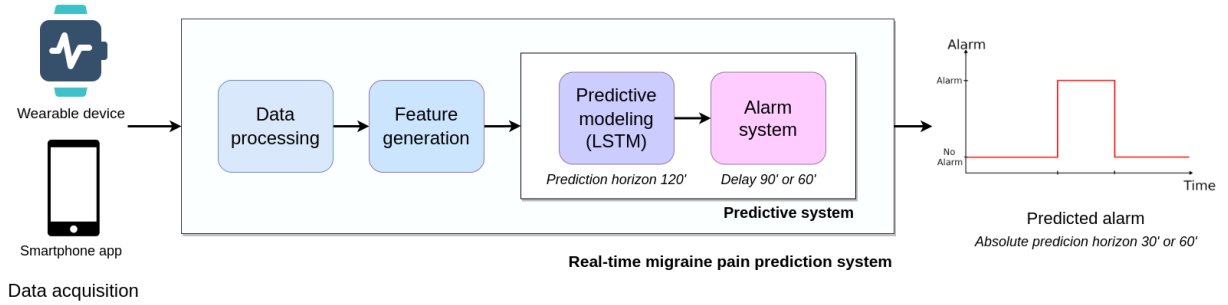


Figure 3.17: Methodology framework: The methodology developed based on the steps.

Finally, a sequential forward feature selection was executed using a wrapper based on a SVM algorithm with a polynomial kernel. The selection process identified the top 44 features with a merit score of 0.71. Out of the 44 features, 6 are related to stress, 8 to temperature, 7 to HR, and 23 to EDA. This represents a similar proportion of features for most signals, except for EDA, which accounts for half of the features considered relevant for prediction.

3.4.4 Predictive system

The processing and modeling methodology developed is shown in Figure 3.17. It shows the data acquisition, and data processing and feature generation steps already mentioned. The predictive system of the developed methodology consists of two blocks: (i) a predictive modeling and (ii) an alarm system.

The utilized prediction models were LSTM, a type of RNN commonly employed in DL. This model has demonstrated its efficacy in predictive techniques, particularly in the context of predicting epileptic seizures [145]. The entire patient dataset was employed to train the model, enabling pain/no pain predictions on a one-minute basis with a 120-minute prediction horizon. To ensure balanced training, the data were segmented into time periods of pain and no pain, maintaining a ratio of 33% to 66%, respectively. A 5-fold CV was conducted to

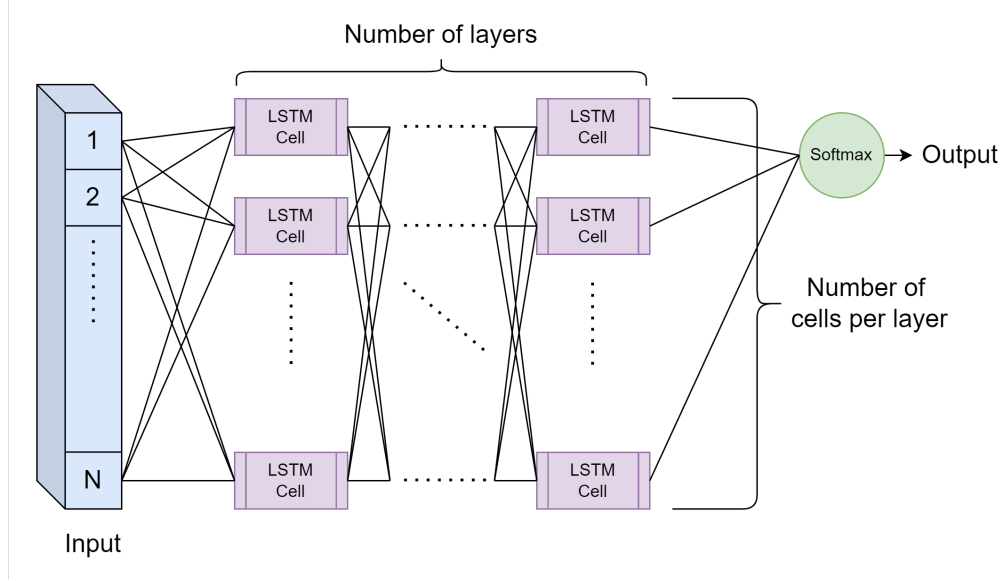


Figure 3.18: Generic architecture of the structure for LSTM models.

ensure the representativeness of both pain and no pain periods, employing a methodology consistent with a previous study on epilepsy prediction [145]. All of this combined with a heuristic grid search for optimizing the hyperparameters of the network and the comparison of two loss functions during training. The traditional Binary Cross Entropy (BCE) and an inverse f-score calculation [217], [223], [229]. The hyperparameters search was conducted over a search space with architectures of 2 and 3 layers, each containing 16, 32, and 64 neurons. Additionally, training batch sizes of 32, 64, 128, and 256 samples per batch were explored. To prevent overfitting, experiments incorporating dropout were included, analyzing a dropout rate of 20%. The graphical representation of the possible architectures explored is shown in Figure 3.18.

The metric for selecting the optimal model was the F-score, a metric defined as the harmonic mean of a model's precision and recall. The stress levels were one-hot-encoded before being input into the prediction model. Given the limited occurrence of crises, the model's performance was evaluated based on the attacks used for training to assess the effective prediction time and observe the number of False Negatives (FN). Subsequently, the model underwent an "individualized" evaluation on new baselines for each patient, focusing on identifying False Positives (FP). The prediction model's output was integrated into the alarm system.

The LSTM's output could generate numerous short- or long-lasting spurious events, which were considered as potential FP. To enhance prediction robustness, a delay filter was implemented in the alarm system. This filter only classified predictions as positive (indicating pain) if two conditions were met: (i) the prediction probability of "pain" remained consistently at 90% or higher during the last time window (ranging from 60 to 90 minutes), and (ii) there had been no migraine pain within the last 24 hours. A 60-minute time delay resulted in an Absolute Prediction Horizon (APH) of $120 - 60 = 60$ minutes, potentially missing some FP. On the other hand, a 90-minute delay was more conservative, minimizing FP but risking the

classification of True Positive (TP) predictions as FN if they did not persist for that duration. The APH in this case would be $120 - 90 = 30$ minutes.

The system ceased generating alarms 24 hours after an alert was triggered, assuming it was a genuine migraine attack. Consequently, in non-pain periods, only one FP could occur every 24 hours.

To analyze the importance of the features for the models developed, the Shapley values have been computed. Shapley values are used to explain individual predictions and represent the relevance of each feature on the model [30], [110], [129], [196].

Once this experiment was developed, the same methodology was implemented, aiming to integrate the last two chapters: the study of sleep (Chapter 3.3) and migraine prediction (Chapter 3.4). The objective was to ascertain the correlation between sleep quality and the occurrence of a migraine attack the following day, aiming to close the loop described at the beginning of the chapter and represented in Figure 3.1. During the course of this study, a preliminary investigation was conducted; however, further comprehensive research on this topic is warranted [68].

As mentioned, the idea was to reuse the information extracted from the sleep study and apply it to migraine patients. To this end, a study was conducted at the same hospital with reference number 3513 Annex 6.7, which included 8 migraine patients and 5 healthy controls aged between 21 and 60 years. The study participants were monitored with the Empatica wristband over several nights (2 weeks for controls and approximately 2 months for patients). Additionally, parallel to nocturnal monitoring, related information was collected from the patients.

To participate in the study, patients had to meet several criteria, including being between 18 and 65 years old, having at least a one-year history of migraines, and experiencing 4 to 9 migraine attacks per month (patients with episodic migraine, excluding those with chronic migraine). For the controls, the most important criteria were never having experienced a headache with migraine-like characteristics and having no family history of migraines (first and second degree).

Chapter 4

Results

In this chapter, we present the results of all the different modules investigated throughout this work. We will analyze the various outcomes, considering collectively whether they could be applicable to the migraine pain prediction model. In this chapter, the results of the different modules will be presented: (i) the characterization of the pain curve in Chapter 4.1, (ii) the results of patient clustering by pain curve type in Chapter 4.2, (iii) the results of the sleep model along with postural detection in Chapter 4.3, and (iv) the results of the migraine crisis prediction model in Chapter 4.4.

The sizes of the created models were as follows: For the recommender system, with weights under half a megabyte and around 33,500 parameters. For migraine pain prediction models, with weights less than one megabyte and approximately 40,500 parameters.

4.1 Pain curve characterization

In this section, we focus on the analysis of the clustering results using the k -means algorithm for values of k varying from 2 to 9. Recall that the objective of this part was to find a classification of migraines based on the type of pain they generate. To this end, this section divides the analysis into the following aspects: (i) analysis of the clustering cohesion index, (ii) analysis of the distribution of samples (migraine pain curves) per clustering result, (iii) analysis of the distribution of features per cluster for the best clustering results, and (iv) characterization and naming of the migraine pain curves per cluster for the best clustering result. Later, we discuss the result of the aforementioned cluster-then-classify strategy for supervised feature selection and classification modeling in Chapter 5. The results of all these experiments were published in [66],

4.1.1 Results of the unsupervised pain clustering model

Clustering cohesion index

Figure 4.1 depicts the results of the cohesion index using the k -means score. As it can be seen, the higher the number of clusters, the better the k -means score. The improvement of adding

a cluster from 4 to 5 is 3 points (from 20.5 to 17.5), but it is not as much as the improvement achieved by adding a cluster from 3 to 4, where the metric falls 4 points (from 24.5 to 20.5). To find the elbow of the curve we must find the farthest point from an imaginary line—red line in Figure 4.1—between the solutions for $k = 2$ and $k = 9$. The clustering model C_k for $k = 4$ clusters depicts the highest WCSS [47] and thus the the best solution for our problem.

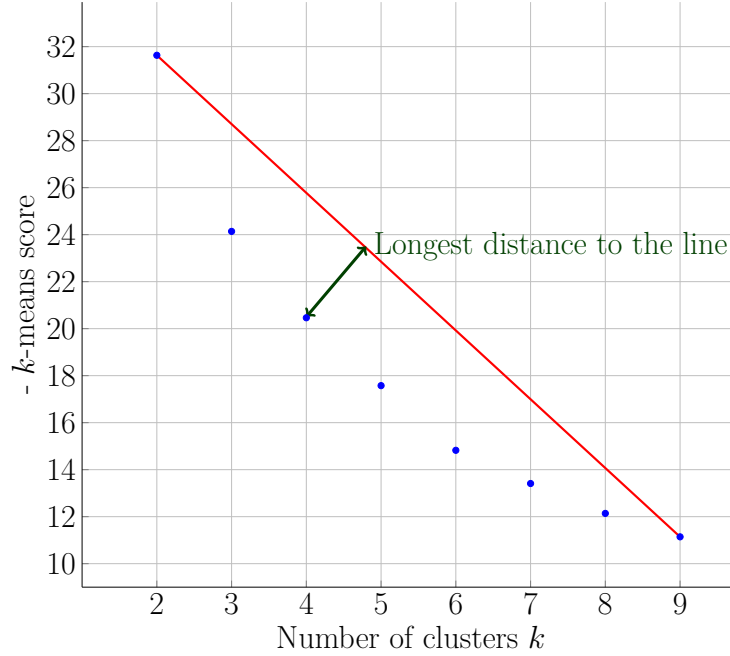


Figure 4.1: (*minus*) k -means score for the k -means clustering algorithm with k varying from 2 to 9 clusters.

Distribution of pain curves per clustering model

To understand how clustering behaves, Table 4.1 shows the percentage of data —the relative number of pain curves— that belong to each cluster $c_{k,i}$ for each clustering model C_k . The columns of the table represent each one of the clustering models. The rows represent the cluster $c_{k,i}$ of each clustering model for $i = 0, 1, \dots, 8$.

Notice the reader that the distribution of data change considerably from $k = 2, 3$ and 4. However, for $k = 6$ and higher, the new clusters that appear contain very few pain episodes and, in some of the cases, the new clusters host 1% episodes or less (1 episode represents $\frac{1}{344} = 0.3\%$), which objectively is an overclustering of the data and makes no sense for our purpose.

Looking closer to the numbers between the clusters of the models C_3 , C_4 and C_5 it can be noticed that $c_{3,1}$ and $c_{3,2}$ remain as $c_{4,1}$ and $c_{4,2}$, respectively, while the cluster $c_{3,0}$ is distributed in $c_{4,0}$ and $c_{4,3}$. Adding one cluster from C_4 to C_5 leads $c_{4,0}$ and $c_{4,3}$ remain as $c_{5,0}$ and $c_{5,3}$, respectively, and $c_{4,1}$ distributes in $c_{5,2}$ and the new one $c_{5,4}$.

Each clustering model has been analyzed sample by sample, and it has been checked how the samples have been distributed in successive subdivisions. Thus, 100% of the samples in

the cluster $c_{4,1}$ are in the cluster $c_{3,1}$, and 100% of the samples in the cluster $c_{4,2}$ are in the cluster $c_{3,2}$. So, it can be said that the cluster $c_{4,1}$ is the same cluster as $c_{3,1}$, and the cluster $c_{4,2}$ is the same cluster than $c_{3,2}$. In the cluster $c_{4,0}$, 8.33% (10/120) samples come from the cluster $c_{3,2}$ and the remaining **91.67%** come from the cluster $c_{3,0}$. In the cluster $c_{4,3}$ 2.35% (2/85) samples come from the cluster $c_{4,1}$ and the remaining **97.65%** come from the cluster $c_{3,0}$. So, it can be stated that most of the samples in the clusters $c_{4,0}$ and $c_{4,3}$ come from the subdivision of the cluster $c_{3,0}$.

Table 4.1: Data distribution for each cluster $c_{k,i}$ for each model C_k .

$c_{k,i} \ C_k$	2	3	4	5	6	7	8	9
$c_{k,0}$	68%	56%	34%	34%	15%	10%	15%	19%
$c_{k,1}$	32%	32%	31%	17%	31%	15%	15%	15%
$c_{k,2}$		12%	10%	15%	25%	21%	25%	20%
$c_{k,3}$			25%	25%	1%	1%	1%	9%
$c_{k,4}$				9%	11%	19%	11%	1%
$c_{k,5}$					17%	18%	16%	16%
$c_{k,6}$						16%	0%	0%
$c_{k,7}$							17%	12%
$c_{k,8}$								8%

Taking into account the distributions in Table 4.1 and the higher WCSS, the best solution is $k = 4$ clusters.

Distribution of features per cluster

Clustering the types of pain inevitably leads to thinking, from a clinical point of view, about what features describe a pain curve. Thus, Figure 4.2 shows the distributions of each feature for the clustering model $C_{k=4}$ to help interpret this.

Looking at the distributions in Figure 4.2 we can hypothesize that:

1. pain episodes of the cluster $c_{4,0}$ (in red) can be roughly defined by intermediate values of $\delta_{acme\uparrow}$ (speed of the onset of pain) with high $Acme$ (the maximum level of intensity) and $t_{acme\uparrow}$ (the time to the highest point);
2. pain episodes of the cluster $c_{4,1}$ (in orange) are characterized by the highest $\delta_{acme\uparrow}$ —consequence of having the lowest $t_{acme\uparrow}$ —and high $Acme$;
3. the pain episodes grouped by the cluster $c_{4,2}$ (in green) are the ones that last the longest (highest D , total duration) and they are the ones that reach the $Acme$ later (highest $t_{acme\uparrow}$). The pain of these pain episodes remains high for a longer time (high D_{90} , duration of the pain with an intensity higher than 90%), and all these characteristics combined lead the episodes of this cluster to be the ones with the highest energy E (energy) and areas A (total area) and $A_{acme\downarrow}$ (amounts expressed in the area of pain suffered after the $Acme$).
4. Pain episodes within the cluster $c_{4,3}$ (in blue) are events of low intensity (low $Acme$), and the times $\delta_{acme\uparrow}$ to reach the $acme$ are medium-high.

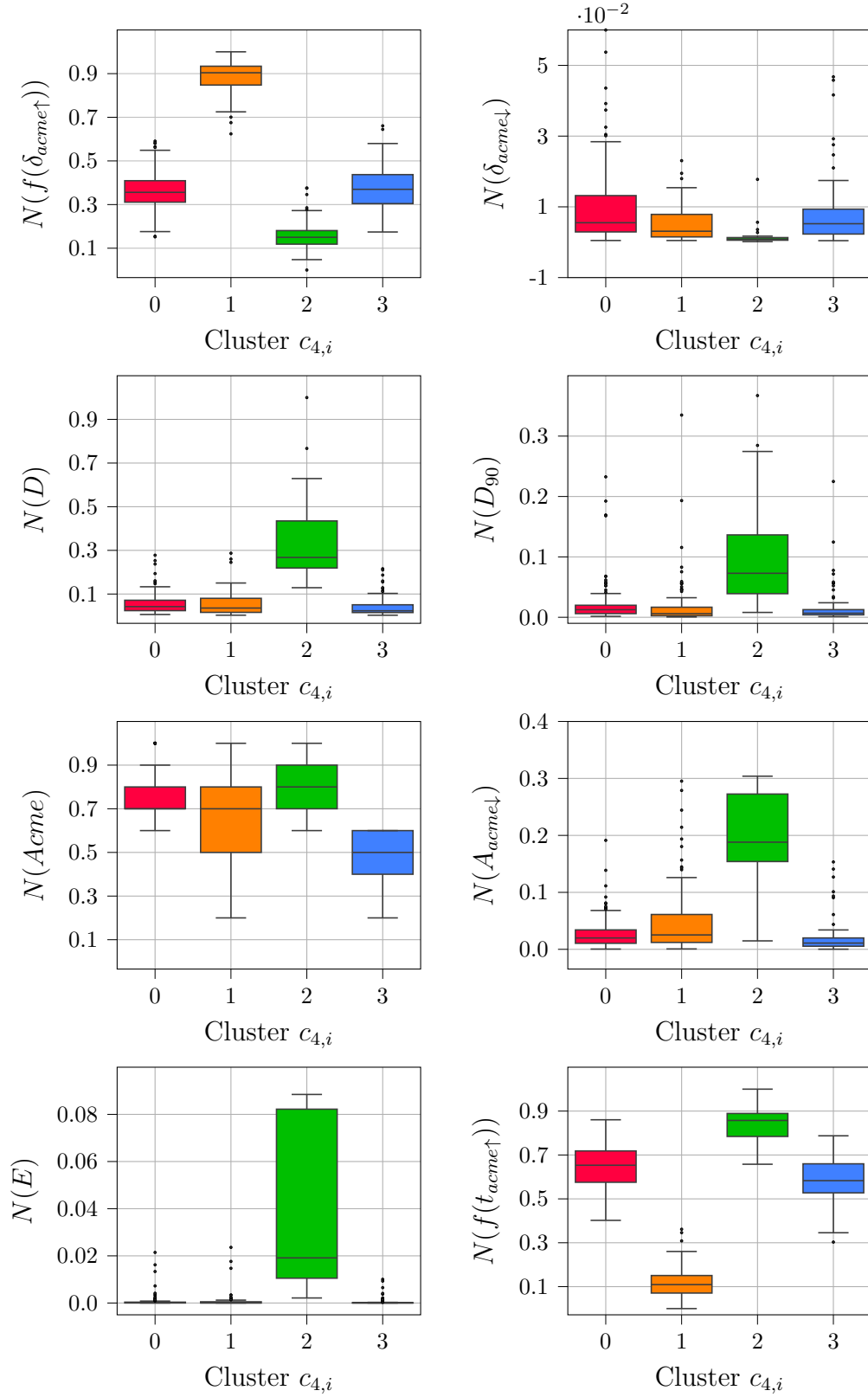


Figure 4.2: Distribution of features per class for $k = 4$ clusters.

As a reminder, we are referring to the features using the original names, while the ones that appears in the graphs are the transformed and normalized features used for modeling.

Characterization of the pain curves per cluster

In this section we study the shape of the migraine pain curves for each cluster for the chosen solution of $k = 4$ types of pain episodes, as well as average pain curve of each type of pain.

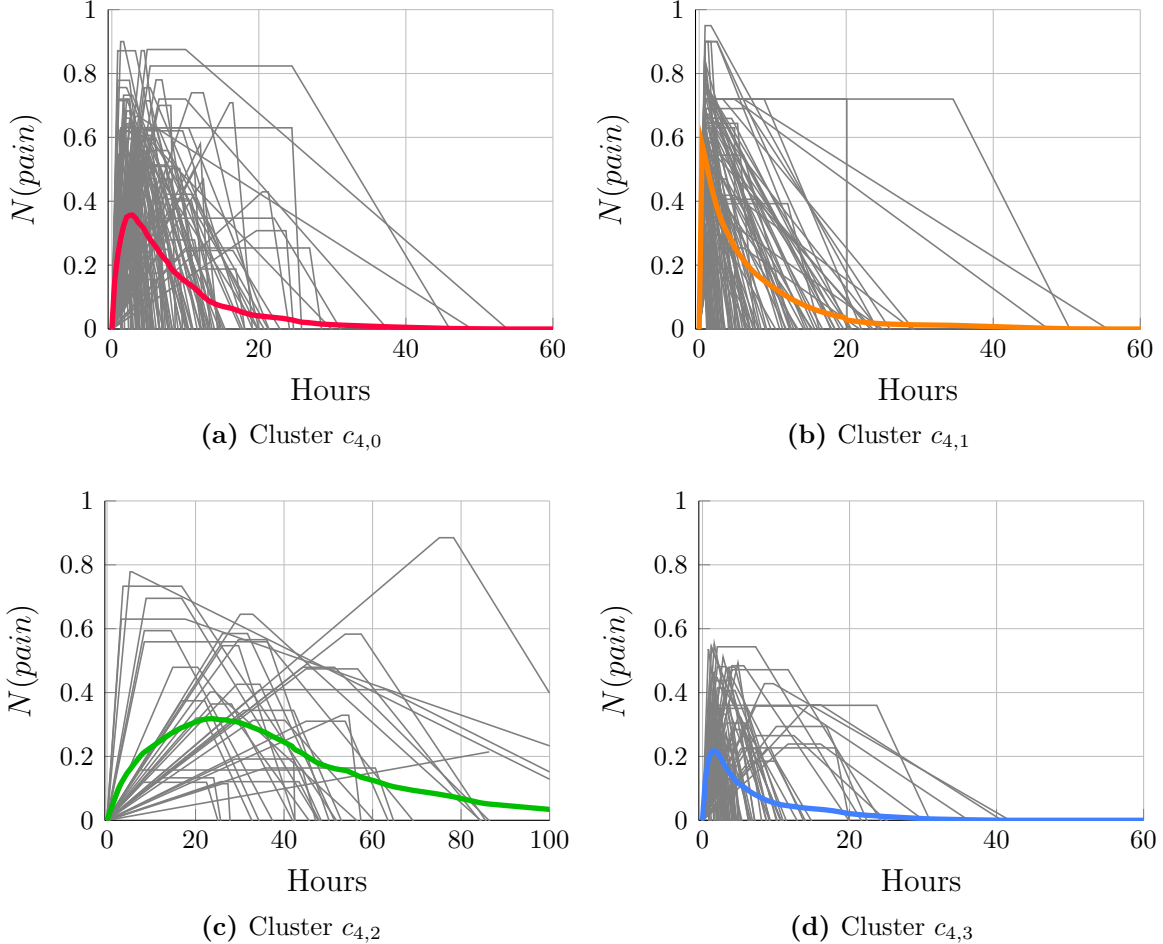


Figure 4.3: Schematic curves of each migraine in gray. The average curve of each cluster is drawn in color overlay.

Figure 4.3 displays schematic migraine curves for each cluster in model C_4 . Gray lines depict individual pain episodes, while colored lines show the average curve. Noticeable deviations exist within each group's representation, leading to quasi-Gaussian or semi-Gaussian average pain curves. This suggests that recent work proposing migraine pain can be modeled as Gaussian curves might not be unfounded [153], but further research is necessary to confirm this hypothesis.

Figure 4.3 remarks that whereas the features of these curves are suitable to run clustering models, the shape of the curve does not seem to be a good criterion and, therefore, the average

coloured curves should be considered only for qualitative and coarse-grained comparison. As an example, as it can be seen in the whiskers diagrams in Figure 4.2, while the pain episodes in cluster $c_{4,1}$ show high values of $\delta_{acme\uparrow}$ and short duration, there are some long lasting events ($D > 48h$) drawn in Figure 4.3. In the same way, while the pain episodes in cluster $c_{4,2}$ last a long time D , some of them last for less than $D < 24h$ as shown in Figure 4.3.

To reinforce the results, it was decided to verify the feasibility of an automated model for pain type detection and classification in the simplest form possible. The development and results are explained below.

4.1.2 Supervised feature selection

Using the 344 pain episodes available, the 8 aforementioned features and the labeling given by the unsupervised k -means have been computed by a correlation-based feature subset selection [78], with a search method of *BestFirst* using WEKA 3.8.3. platform.

39 subsets have been evaluated and the merit of the best one is 0.782. With this, 4 features are selected: $\delta_{acme\uparrow}$, D , $Acme$, and $t_{acme\uparrow}$. Unfortunately, with these features, it can be proved that we cannot classify the patient's type of pain in real time, and it must be waited for the pain to finish completely. However, it considerably reduces the amount of data collected by the patient in clinical studies. In the next section, these features are exploited to classify the four previously recognized types of migraine pain.

4.1.3 Supervised classification model

In order to support further research with an explanatory outcome on defining the characteristics that describe a type of pain, tree-based models have been trained. The trees allow for an easy interpretation of the threshold levels of the features that lead to the classification of episodes of pain.

The test results for the implementation of an SVM model, a NaiveBayes classifier, a single tree *J48* algorithm, a *J48 – graft* tree [210], and a *RandomForest* are shown in Table 4.2. WeKA 3.8.3 has been used in split 70 – 30% for training and test data sets. The table also shows the classification results of a LMT. A logistic model consists of a standard decision tree structure with logistic regression functions in the leaves [107]. The benefit of LMT is that an expression is obtained for each type of pain for the different cutoff thresholds of the variables. These algorithms have been commonly used because they allow physicians to interpret the impact of the features on the results.

The metrics in Table 4.2 demonstrate that the types of pain can be classified by a supervised model with 4 only features and a negligible error rate. To take into account the imbalance between classes in the analysis of the results, the F-measure is crucial. In any of the algorithms, the result is good. The best of the trees is the *J48-graft* with an F-measure = 98.1%. However, the logistic model returns better results, increasing the F-measure by one point up to 99.1%.

The most interesting result is that the LMT is made up of only one leaf, so a single expression is needed only for each type of pain, which means that there are no decision thresholds over the features.

Table 4.2: Accuracy, precision, recall and F-measure of the supervised models trained for migraine type classification.

Model	Acc. (%)	Prec. (%)	Recall (%)	F-meas. (%)
<i>SVM</i>	98.1	98.1	98.1	98.1
<i>NaiveBayes</i>	97.1	97.1	97.1	97.1
<i>J48</i>	97.1	97.4	97.1	97.2
<i>J48 – graft</i>	98.1	98.1	98.1	98.1
<i>RandomForest</i>	97.1	97.4	97.1	97.2
<i>LMT</i>	99.0	99.2	99.0	99.1

The coefficients of every feature for each linear function are shown in Table 4.3. For the sake of reproducibility, the hyperparameters of the models are described below. For SVM, a polynomial kernel has been proposed, setting the optimization tolerance to 10^{-3} . For the Naive Bayes model, a kernel estimator has been proposed. For the J48 tree, the minimum number of instances per tree has been set to 2, allowing pruning. The same hyperparameters have been used for the J48-graft. For the Random Forest, the maximum depth of the trees has not been limited. Finally, for the LMT, beta has been set to 0 for no weight trimming for LogitBoost, using cross-validation instead of Akaike’s Information Criteria. The batch size has been set at 100 for all experiments.

Table 4.3: Coefficients of the linear equations of the LMT model at the end leafs for each type of migraine.

	independent	$N(f(\delta_{acme\uparrow}))$	$N(D)$	$N(f(t_{acme\uparrow}))$	$N(Acme)$
f_0	-18.00	-9.47	-4.32	11.47	19.97
f_1	-31.80	160.92	0.00	-184.31	0.00
f_2	-24.62	-40.73	59.18	21.64	12.75
f_3	56.60	0.94	-27.56	-1.68	-89.24

The supervised feature selection shows that it is only needed a minimum of 4 points to identify the type of pain, These are: (i) when the pain starts, (ii) when the pain ends, (iii) when the pain reaches its maximum, and (iv) the level at maximum intensity of pain.

The coefficients in Table 4.3 show the importance of every feature in identifying each type of pain. It can be seen that: for cluster $c_{4,0}$ the most important feature is $Acme$. In the cluster $c_{4,1}$ the features $\delta_{acme\uparrow}$ and $t_{acme\uparrow}$ are the relevant ones, while the coefficients of D and $Acme$ are zero. This means that pains of type $c_{4,1}$ can be classified in real time before the pain ends. In the cluster $c_{4,2}$, D and $\delta_{acme\uparrow}$ are more important than the others. And for the cluster $c_{4,3}$ the most important feature is the $Acme$ and then the D .

The analysis partially answers the questions raised at the beginning of the discussion. In addition, we can venture a step forward for the convenience of future work. To do this, we name the types of pain based on the study of the coefficients and the representations shown in the Figure 4.3. Thus, the cluster $c_{4,0}$ can be named as the group of *high or medium-intense* pain episodes, the cluster $c_{4,1}$ can be named as the group of *sudden* pain episodes, the cluster $c_{4,2}$ can be named as the group of *long lasting* pain episodes, and the cluster $c_{4,3}$ can be named

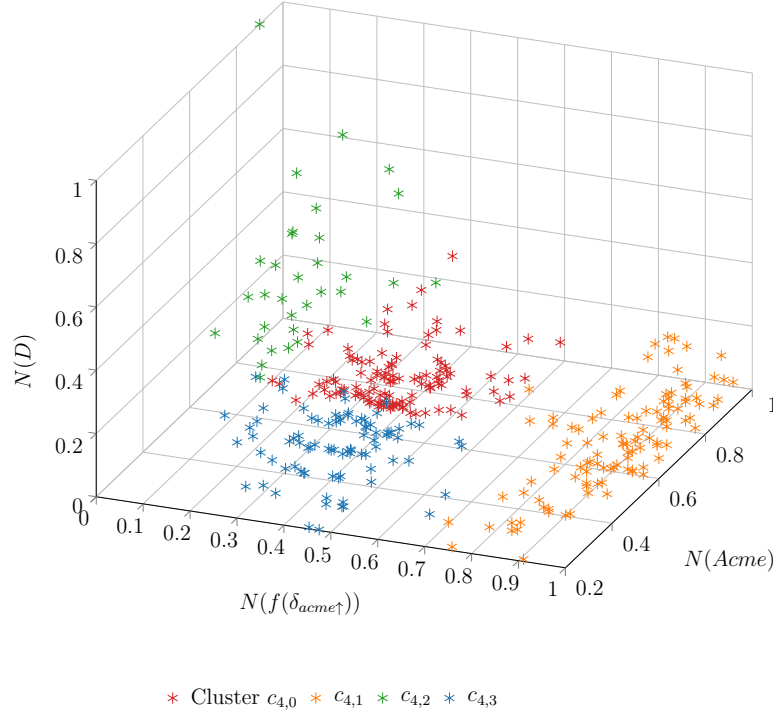


Figure 4.4: Classification of types of pain using the supervised LMT model with 99.0% accuracy. Plotting three of four features needed.

as the group of *mild or low-intense* pain episodes. These labels have been established based on the morphology of the average curves of the crises, and further clinical analysis will be conducted to validate them.

The justification of the naming is based on the following clinical evidence. According to the International Classification of Headache Disorders [148], the migraine intensity can be classified as: *mild*, *moderate* and *severe*. This classification only considers the intensity of the pain of the crisis, and it does not consider its evolution—which it is not considered as a clinical criterion for classification yet—. However, we can relate *moderate* migraines to our so-called *high or medium-intense* ($c_{4,0}$), and *mild* migraines can be related to our so-called *mild or low-intense* ($c_{4,3}$). Both clusters named as *sudden* and *long lasting* are related to the intensity of pain. However, looking at their most representative features in Figure 4.2 it can be seen that there are more relevant features than the intensity. Thus, the so-called *sudden* migraines ($c_{4,1}$) can be related to the *thunderclap* migraines. Thunderclap migraines reach their acme in less than one minute [182]. On the other hand, the so-called *long lasting* migraines ($c_{4,2}$) are migraines whose duration is above 48 hours on average. These migraines can be related to those highlighted by Kelman in [100].

Finally, in order to plot the classification of the pain curves using the LMT model, it is necessary to use the functions in Table 4.3 and the class membership probabilities given by Equation 4.1:

$$P(c_{4,k} = k | M = m) = \frac{e^{f_k(m)}}{\sum_{i=0}^4 e^{f_i(m)}}, \quad (4.1)$$

Table 4.4: The distribution of migraine pain types within each generated cohort as a percentage.

	A	B	C	D	E
medium-intense	87.1%	45.4%	17.4%	21.4%	3.8%
sudden	3.2%	23.0%	71.3%	14.3%	18.9%
long-lasting	0.0%	5.9%	0.0%	64.3%	11.3%
low-intense	9.7%	25.7%	11.3%	0.0%	66.0%

where $k \in \{0, 1, 2, 3\}$ represents each cluster, M is the dataset of pain curves and m represents each pain episode, and $f_k(m)$ are the values of the functions for classification at each leaf of the LMT for each episode. The result of the classification is shown in a 3D graph, in Figure 4.4 for three of the four features.

4.2 Group clustering by pain type

In this section, we present the results obtained from applying the described and implemented methodology of the recommender as specified, but first we present the results achieved in (i) formation of patient cohorts; we have carried out a novel classification of patients based on the evolution of pain during disease crises. Then we analyze (ii) RNN; the RNN is a crucial part of the ESPA selection. We will provide a detailed analysis of the RNN’s performance in reconstructing the suitability score assigned to each advice for every crisis. Additionally, we compare its performance with that of a MLP and analyze the impact of excluding cohort classification information from the RNN structure. (iii) Recommendation system; the patient cohorts and the RNN are essential components of a larger system—the clinical indication recommender as a whole—. We will demonstrate the effectiveness and performance of the entire recommendation system. The results of all these experiments were published and used in [133],

4.2.1 Patient cohorts

The k-means algorithm, an unsupervised clustering method, requires the number of clusters to be specified. As the optimal number of cohorts for patient distribution was uncertain, the results for different numbers of clusters were explored. To make the final decision, the “elbow method” mentioned in Chapter 3.1.10 was employed. Based on the curve, 5 cohorts were identified as corresponding to the coordinate on the horizontal axis nearest to the elbow of the curve. Subsequently, the data distribution within these cohorts will be examined.

Analyzing the distribution of migraine types across cohorts, as illustrated in Table 4.4, it can be inferred that four cohorts primarily exhibit one type of migraine attack (represented by columns A, C, D, and E in Table 4.4), while a fifth cohort comprises patients who typically experience multiple types of migraine attacks (Column B in Table 4.4).

Table 4.5 shows the statistical significance of the distribution of sociodemographic variables

Table 4.5: Statistical significance of cohorts regarding sociodemographic variables. The numbering in the columns of each group A, B, C, D, and E refers to how they should be regrouped. All those with a value of 0 will be in the same group, those with a value of 1 in another, and so on.

		Cohort grouping					
basals	groups	A	B	C	D	E	p-value
frequency of crises	[Episodic vs Others]	0	1	2	3	4	0.030
studies	[higher vs lower]	0	1	2	3	4	0.010
cognitive reserve	[≥ 14 vs < 14]	0	1	1	2	1	0.035
smoker		0	0	1	1	0	0.063
age of onset	[child vs others]	0	0	0	0	1	0.064
diagnosis	[No aura vs others]	0	1	1	1	1	0.082
sex		0	1	2	3	4	0.586
family history		0	1	2	2	0	0.125
dyslipidemia		0	1	2	0	0	0.239
age	[young vs mature]	0	1	0	0	0	0.160
cardiovascular risk factor		0	1	0	0	0	0.248

between the different cohorts of the patient sample. For an α of 0.05, the distribution can be considered representative. Groupings have been necessary to achieve adequacy in *cognitive reserve*, *smoking status*, *age of onset*, and *diagnosis*. In some other cases, although groupings have been made, it has not been possible to lower α to 0.05, but the best p-value achieved is shown. In Table 4.5, we show the groupings that have resulted in the best values, leading to the conclusion:

- Regarding with frequency of crises: This grouping of “episodic” versus “other” is representative, as we can assert that all type A patients have episodic frequency.
- Regarding with studies: In contrast to the other groups, group A exhibits a notable prevalence of members with primary education. This observation hints at a potential link between high-intensity migraines and educational discontinuity.
- Regarding with cognitive reserve: We can mention that it is representative that groups B, C, and E exhibit a higher cognitive reserve.
- Regarding with smokers: groups C and D stand out due to the presence of smokers, which could indicate that smoking may cause migraine to be sudden or lasting. Although group B also has presence of smokers, I believe it may be due to their patients suffering from a variety of types of migraines, including sudden ones. Considering that the proportion of smokers is the lowest among the three groups, and assuming that there is indeed causality between smoking and having a higher probability of suffering from sudden migraines, it may be logical to think that this percentage is due to those who suffer from sudden migraines due to tobacco; if they did not smoke, they would not

Table 4.6: Statistical significance of patient types regarding triggers. The numbering in the columns of each group A, B, C, D, and E refers to how they should be regrouped. All those with a value of 0 will be in the same group, those with a value of 1 in another, and so on.

	Cohort grouping					
triggers	A	B	C	D	E	p-value
food	0	1	2	3	4	0.022
drink	0	1	2	3	4	0.039
stress	0	1	1	0	1	0.013
sleep	0	1	2	3	4	0.003
environmental	0	1	2	3	4	0.001
sport	0	1	1	1	1	0.015

suffer from those sudden migraines and would belong to group A or E.

- Regarding with age of onset: With a p – *value* of 0.064, which is close to the significance threshold, it could indicate that the disease onset in patients from group E occurred earlier than in the others.
- Regarding with diagnosis: Even though the p-value does not reach the threshold, it is very close, suggesting a disparity in group A compared to the rest of the groups in this aspect. This is because group A has a majority of patients with aura compared to the other groups.

In any case, a clinical analysis of these hypotheses would be necessary.

Table 4.6 shows the statistical significance of the distribution of migraine triggers between patients in different cohorts. For an α of 0.05, the distribution can be considered representative for all triggers. Groupings have been necessary to achieve adequacy in *stress* and *sports*.

- With respect to food, it is noteworthy to mention that 29% of type D patients' migraines have that trigger, while the rest have around 9%.
- In terms of drink consumption, groups A, D, and E have an 11% occurrence compared to 3% in the rest.
- The stress trigger prevalence among groups A and D is 53% compared to 35% in groups B and C.
- With respect to sleep, groups A and D have a 57% compared to 33% in the rest.
- In terms of environmental factors, group A exhibits a significant majority with a prevalence of 61%, contrasting starkly with the 30% and 20% prevalence in other groups.
- And finally, regardless of sports, in group A it appears 13% of the time compared with a 4% in the rest.

As in the previously mentioned conclusions, a clinical analysis of these hypotheses would be

Table 4.7: Comparison of test results according to the number of batches for a fixed architecture of 16 and 1 neurons. The best results are highlighted.

Batch size	RNN			MLP		
	MSE	MAE	R^2	MSE	MAE	R^2
32	0.00171	0.0219	0.952	0.00199	0.0235	0.944
43	0.00188	0.0223	0.947	0.00200	0.0238	0.943
86	0.00218	0.0257	0.938	0.00194	0.0247	0.945
224	0.00258	0.0298	0.926	0.00223	0.0279	0.936
344	0.00243	0.0287	0.932	0.00233	0.0289	0.932
688	0.00262	0.0301	0.928	0.00258	0.0312	0.927
1376	0.00275	0.0333	0.923	0.00269	0.0324	0.922

necessary.

4.2.2 ESPA recommendation system

For this development, we used the same dataset and patients as in Chapter 3.1.1, where we mentioned that a total of 344 valid migraines were collected. We generate 28 ESPA (Episode-Specific Personalized Advice) for training our system. These 28 ESPA combined with our 344 migraines result in 962 instances to train and test our model.

Now we compare the performance of a RNN with a MLP for modeling advice scoring. We aim to find the best model for each architecture while considering batch size and network structure parameters. Prior to the training phase, we divided the 9632 instances we had into two subsets, one for training the networks, comprising 80% of the instances, and another for testing, consisting of the remaining 20%. Furthermore, we subdivided the training set into two fragments: one for model validation during the training phase, which included 5% of the training samples. All metrics presented in this chapter are computed using test data.

For batch size optimization, the metrics on the test data are indicated in Table 4.7. The batch size options to test were chosen because they were divisors of the total number of instances available to us. We concluded that the optimal batch size for both models was 32, although in the case of the MLP 86 batches were also a viable number. Subsequently, focusing on the optimal architecture, as we could see in Table 4.8, we determined that a three-layer RNN with 32, 16, and 1 neurons and a two-layer MLP with 32 and 1 neurons each were most accurate.

4.2.3 Study of the impact of migraine pain type information

We then repeated the process without including patient cohort information to assess its significance in model quality. If we compare the test results of Tables 4.8 and 4.9, it can be seen how the models in which patient cohort information was excluded generally exhibit lower performance. It is noteworthy that the MLP —despite having less information— needs a more intricate network, whereas the RNN, with a reduced input dataset, diminishes its intermediate layer. In both networks, there is a notable enhancement of approximately 11

Table 4.8: Comparison of test results according to the architecture used for 32 batches including patient cohort information by pain type. The best results are highlighted.

Architecture	RNN			MLP		
	MSE	MAE	R^2	MSE	MAE	R^2
32,1	0.00188	0.0230	0.948	0,00182	0,0216	0,949
16,1	0.00191	0.0229	0.947	0.00194	0.0232	0.942
8,1	0.00199	0.0256	0.943	0.00258	0.0280	0.931
32,16,1	0.00176	0.0210	0.951	0.00187	0.0219	0.946
32,8,1	0.00192	0.0223	0.947	0.00188	0.0227	0.942
32,4,1	0.00197	0.0227	0.945	0.00193	0.0224	0.940
16,8,1	0.00192	0.0232	0.947	0.00193	0.0231	0.947
16,4,1	0.00196	0.0225	0.946	0.00199	0.0230	0.943

Table 4.9: Comparison of test results according to the architecture used for 32 batches without including patient cohort information by pain type. The best results are highlighted.

Architecture	RNN			MLP		
	MSE	MAE	R^2	MSE	MAE	R^2
32,1	0.0045	0.0262	0.823	0.0042	0.0257	0.834
16,1	0.0050	0.0295	0.810	0.0045	0.0274	0.819
8,1	0.0053	0.0330	0.792	0.0049	0.0274	0.815
32,16,1	0.0044	0.0228	0.830	0.0050	0.0246	0.805
32,8,1	0.0040	0.0216	0.848	0.0043	0.0218	0.836
32,4,1	0.0047	0.0259	0.825	0.0044	0.0237	0.828
16,8,1	0.0045	0.0259	0.825	0.0052	0.0278	0.797
16,4,1	0.0051	0.0271	0.803	0.0056	0.0283	0.790

points in the metric R^2 . This substantial impact can be attributed to the increased availability of information for computing the final score, which alleviates the network’s difficulty in approximating the chosen score.

Despite the marginal improvement, we selected the RNN for our system due to its potential adaptability to new data on the migraine crisis.

After thorough parameter testing, we concluded that the best combination for our RNN was a batch size of 32 and three layers with 32, 16, and 1 neurons, and 32, 8, and 1 neurons for the model with and without cohort information, respectively. The RNN achieved notable results with 0.00176 MSE, 0.0210 MAE, and 0.951 R^2 . In Figure 4.5, we observe the cost function graph during training for validation and training sets, where we can verify that 150 epochs were sufficient for the model to achieve its best possible results. It is important to highlight that the validation error significantly decreases by nearly one order of magnitude when transitioning from the model without cohort information to the model incorporating cohort information.

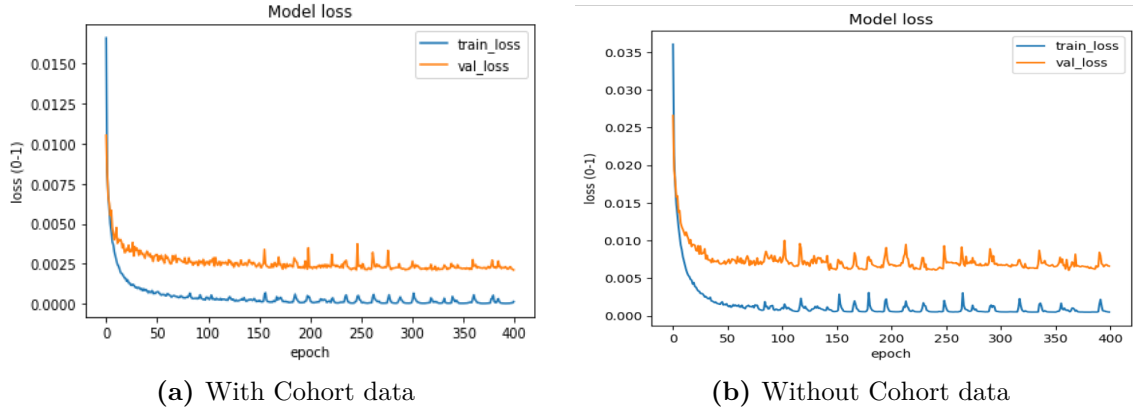


Figure 4.5: The graphs depicts the MSE or loss of the RNN using the best combination of parameters throughout the epochs for both the training (train) and validation (val) datasets. The x-axis represents the epochs, which denote the number of times the training algorithm goes through the entire dataset during training. The y-axis represents the MSE, which is a measure of the difference between the predicted and actual outputs.

4.2.4 Demonstration of a use case of the ESPA recommendation system

The system has been simulated and its functionality has been evaluated. We are going to explain a use case for one patient and the ESPA recommendation system. The demo patient was an adult woman, with aura, without family history, with treatment, without hypertension, without dyslipemia, with healthy weight based on BMI, with episodic crisis, and with a severe impact HIT-6 value. It is worth mentioning that the parameters used for the collaborative filters of the GA and BPA have been $\rho_{GA} = \rho_{BPA} = 0.25$ and $\delta_{GA} = \delta_{BPA} = 0.25$. These parameters can be freely and individually tuned by the experimenter.

Figure 4.6 depicts a timeline demo comparing the two distinct ESPA recommendation system developed—the one that incorporates cohort information, and the other one that does not—over a period of two months, delivering one piece of advice per week. On some occasions, both systems give aligned advice.

During the course of two months, the patient suffered from three episodes of migraine. The first pain occurs after the first recommendation, and this is the reason why neither of the two recommendation systems provides any ESPA-type advice at the beginning. This is a cold system start-up without patient-specific data.

Regarding **the system that does not use cohort information** (bottom side of the timeline), it can be observed that:

- The system advises the patient on *mood and sleep-related* aspects after the first migraine episode. This is due to the patient documented notable mood and sleep fluctuations subsequent to the migraine-related discomfort thus, the recommendation of such advice

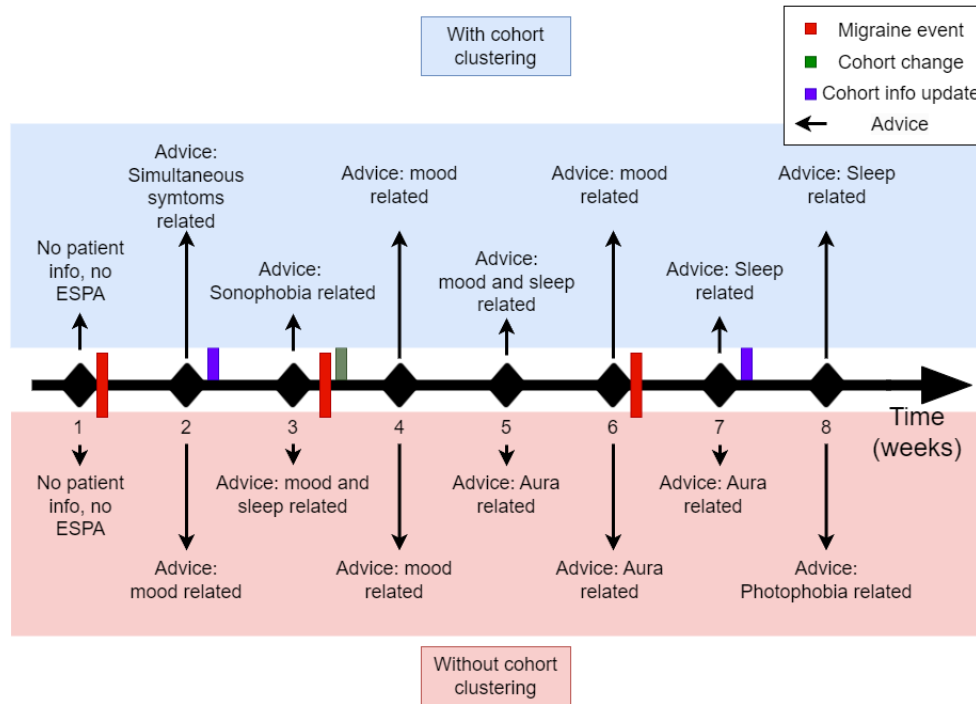


Figure 4.6: Timeline demo. Representation of ESPA advice for an example patient across an 8-week period, encompassing three migraine episodes, during which the patient’s cohort assignment dynamically changes.

aligns with logical expectations.

- In the fourth week, recommendations related to *mood* remain, while advice on auras is recommended in the fifth and sixth weeks. This observation aligns with the experience reported by the patient of visual aura during the second migraine.
- In the seventh week, the recommendations still focus on *aura-related* aspects. However, in the eighth week the emphasis shifts to advice concerning *photophobia*, as the patient consistently noted this symptom in all three migraines suffered.

It may be considered that, in view of our clinical desire to prioritize aura recommendations, it would have been more desirable to move the advice recommendation forward by one week. This would ensure that the three weeks following the migraine with aura would show advice adapted to the topic of auras.

Regarding **the system using cohort information** (top side of the timeline) it can be said that:

- In the second week, the recommendation provided is different from that shown by the recommendation system not using cohort data. However, it must be said that both pieces of advice are valid.
- At week three it can be seen that the system contextualizes the advice by relating them to the concomitant pain symptoms indicated by the patient—who reports experiencing

three out of five of these symptoms—. It is also noticed that the patient is classified to Cohort A after the first pain crisis. Therefore, the advice provided remains consistent and factually consistent.

- During the third week, after two migraine attacks, the patient is reassigned to Cohort D thereby resulting in a significant alteration of cohort-related information. Notably, the *sleep-related* topic within the cohort acquires the highest weighting during this period.
- Between the second week and the cohort change, the behavior of the model is not fully as expected. Although photophobia-related advice is the most important advice within Cohort D—and correlates with the patient’s symptoms—the system assigns a higher score to sonophobia-related advice. Photophobia is ranked third in the ranking of recommended advice, which delays the prioritization of advice closer related to the new cohort and the symptoms reported by the patient.
- During weeks four through eight, the system provides recommendations pertaining to *mood and sleep* topics, in a similar way as done by the aforementioned system.

4.2.5 Clinical analysis for patient cohorts

In this section a brief analysis is made on the possibility of grouping patients into cohorts according to the type of pain of their migraines and whether it is possible to find out how many types of patients there may be. This would lead, from a clinical point of view, to a new paradigm in the management of the pain episodes of migraine attacks.

To this end, we used a clustering model based on the k -means algorithm with Euclidean distance. There are 4 input variables, representing the percentage of migraines of each type experienced by the patient. Therefore, we have 51 rows, one per patient. To find the optimal number of cohorts, 10 iterations were performed of random initial centroid locations, for k varying from 2 to 14. Using the elbow method, the best number of cohorts for this unsupervised classification resulted in $k = 5$.

Table 4.10: Migraines of each type per patient cohort.

	medium-intense	sudden	long-lasting	low-intense
A	27 (87.1%)	1 (3.2%)	0 (0.0%)	3 (9.7%)
B	69 (45.4%)	35 (23.0%)	9 (5.9%)	39 (25.7%)
C	14 (17.5%)	57 (71.3%)	0 (0.0%)	9 (11.2%)
D	6 (21.4%)	4 (14.3%)	18 (64.3%)	0 (0.0%)
E	2 (3.8%)	10 (18.9%)	6 (11.3%)	35 (66.0%)

In parentheses the percentage of each type of pain over all migraines in the group.

The five cohorts of patients are labeled A through E. To study the relationship of groups found to be significant, we put the results in Table 4.10. This table shows the number and percentage of migraines by pain types in each cohort of patients. As can be seen, almost all cohorts—with the exception of B—are clearly represented by one type of pain mainly. This

clustering yields a $p - value = 10^{-49}$, so it is shown to be statistically significant, and we can hint that:

- Cohort A is represented in 87.1% by *medium – intense* crises.
- Patients in cohort B —which contains 39.2% of patients and 44.2% of migraines—, suffer from a mixture of types of pain, being the *medium-intense* crises the most prevalent (45.4), being the *long-lasting* ones uncommon.
- Cohort C is mainly represented by 71.3% of *sudden* migraines.
- Cohort D is represented in 64.3% by *long – lasting* migraines.
- Cohort E is represented in 66.0% by *low – intense* migraines.

To define a new clinical paradigm in the management of migraine pain episodes, it is necessary to analyze the characteristics of crises that could affect, for example, the choice of the type of drug recommended to treat each type of pain. To this end, we analyze statistically how the four features used to classify the pain curves behave in each cohort relative to the average total migraines. The average values of these features are: $\overline{\delta_{acme\uparrow}} = 54.3 \left[\frac{pain}{hour} \right]$, $\overline{Acme} = 6.6 [pain]$, $\overline{t_{acme\uparrow}} = 3.8 [hour]$ and $\overline{D} = 15.5 [hour]$. The results are significant in all cases: $p - value_{\delta_{acme\uparrow}} = 10^{-12}$, $p - value_{Acme} = 10^{-15}$, $p - value_{t_{acme\uparrow}} = 10^{-13}$, $p - value_D = 10^{-13}$. We can state that:

- Cohort A is clearly clustered, by three features: 96.8% of migraines have a $\delta_{acme\uparrow}$ below the mean, 87.1% have a duration D below the mean, and 87.1% have a level $Acme$ above the mean.
- Cohorts B and C distinguish by grouping 85.5% and 95.0% of migraines respectively below the mean of $t_{acme\uparrow}$.
- In cohort D , 89.3% of the migraines have a duration D above the mean.
- In cohort E , 92.5% of the migraines have a level $Acme$ below the mean.

These results confirm the possibility of grouping patients into cohorts according to their type of pain. This classification could be useful to educate them on better disease management according to the characteristics of the crises in their group. The next chapter discusses the results of this clustering.

4.3 Feasibility of sleep modeling for migraine patients

In this chapter, we present the results of two experiments aimed at predicting sleep postures and states using wearable devices and advanced ML techniques. In the first experiment, an actigraph sensor placed on the wrist was utilized to monitor and classify sleep postures, providing detailed insights into body movement patterns throughout the night. The results of this experiment is published in [65]. Subsequently, in the second experiment, DL and TL techniques were applied to analyze the biological signals collected by the same wearable device, with the goal of automatically and accurately detecting and classifying different sleep states. These innovative approaches promise to significantly contribute to the knowledge of postures

and sleep states, thus improving the diagnosis and treatment of sleep-related disorders such as migraine.

4.3.1 Posture detector model

Analysis of feature selection

The Pearson’s correlation-based feature selection has reduced the number of features for modeling from 42 to 16. The list of features selected is shown in Table 4.11. The reader should notice that the features related to the Z-angle and the Energy have not been chosen. It seems that the axes Y and Z are the most important signals, and the *mean value of the first derivative* (μ_δ) as well as the *counts* ($c_{|si|}$) are the most repeated features. On the other hand, the feature *maximum* has never been selected.

Table 4.11: List of feature selected.

feature\signal	X	Y	Z	ENMO
minimum				✓
μ	✓	✓	✓	
σ		✓	✓	
μ_δ	✓	✓	✓	✓
A_{s_i}		✓	✓	
c_{s_i}	✓	✓	✓	✓

Since there are not a large number of features, we will continue with these features while modeling. In the future, a more comprehensive analysis of the features should be done.

Record-wise methodology on new subjects data

Table 4.12 shows the results of the experiment in the *first procedure*, recalling that the samples were balanced as shown in Table 3.7. As suspected, the system fails when new cases of subjects are fed to a system trained on a record-wise fashion. As can be seen, the F-value decreases as much as 0.46, 0.08 and 0.47 for the RF, SVM and KNN algorithms respectively. Given such a difference in the results, and as mentioned above, the strategy of dividing data by records is not a valid solution and therefore we continue with a strategy of dividing data by subjects in training [174]. As a result, it is not even worth analyzing the architecture of these models and they are discarded.

Table 4.12: Mean F-value of record-wise method on new subjects data.

Tested on:	Model		
	<i>RF</i>	<i>SVM</i>	<i>KNN</i>
Record-wise folds	0.94	0.63	0.93
New subjects	0.48	0.55	0.46

Subject-wise models and detection of torso turns

In the second part of the procedure, based on a modeling methodology with subject-wise data splitting, we aim to demonstrate the need to detect and remove torso turns before classifying the posture in order to improve the performance of the models. The results of such experiments on the three datasets created are shown in Table 4.13. The standard deviation is small, which means that the performance of all the models is similar. One is chosen for each algorithm and their hyperparameters are shown in the Tables 4.14 through 4.16. Testing the models on new subjects data leads to average F-values of 0.966, 0.623 and 0.989 for RF, SVM and KNN algorithms respectively. It can be said that the KNN and RF algorithms perform better than SVM for all the experiments.

Table 4.13: Average f-values of subject-wise 5-cv over 10 runs of different test subjects.

F-value	Random Forest		SVM		KNN	
	<i>mean</i>	<i>std</i>	<i>mean</i>	<i>std</i>	<i>mean</i>	<i>std</i>
Non-labeled turns	0.946	0.029	0.564	0.175	0.985	0.017
Turns labeled	0.945	0.031	0.544	0.129	0.980	0.020
Turns removed	0.966	0.031	0.623	0.097	0.989	0.013

In view of the results in Table 4.13, it might appear that there is no substantial difference in the strategy of not labeling the data (first dataset), detecting and classifying them (second dataset), or detecting and eliminating them (third dataset). However, by looking at the confusion matrix of one of the models that detect torso turns (Figure 4.7 for an RF model), it can be seen that the detection of turns is not correct. The recall and precision of the *turn_left* and *turn_right* labels are 0.258, 1.000 and 0.777, 0.777, respectively. This shows the need to detect turns, yes, but to remove them and not to classify them. Removing the turns of the torso results in models that, in the worst case, show an F-value improvement of 0.009 with low deviation. The RF and KNN algorithms yield the best results with average F-values in test of 0.966 and 0.989, respectively.

In Tables 4.14, 4.15 and 4.16, the optimal hyperparameters selected using cross-validation is presented. These tables compare the configurations for the three ML models employed—RF, SVM, and KNN—all of them with and without labeled turns.

Table 4.14: Hyperparameters of RF algorithm.

experiment	Random Forest	
	<i>number of trees</i>	<i>max depth</i>
Non-labeled turns	500	100
Turns labeled	2000	50
Turns removed	1000	500
	<i>min samples per split</i>	<i>min samples per leaf</i>
Non-labeled turns	10	5
Turns labeled	5	5
Turns removed	10	5



Figure 4.7: Confusion matrix of the RF trained on the dataset with turns labeled.

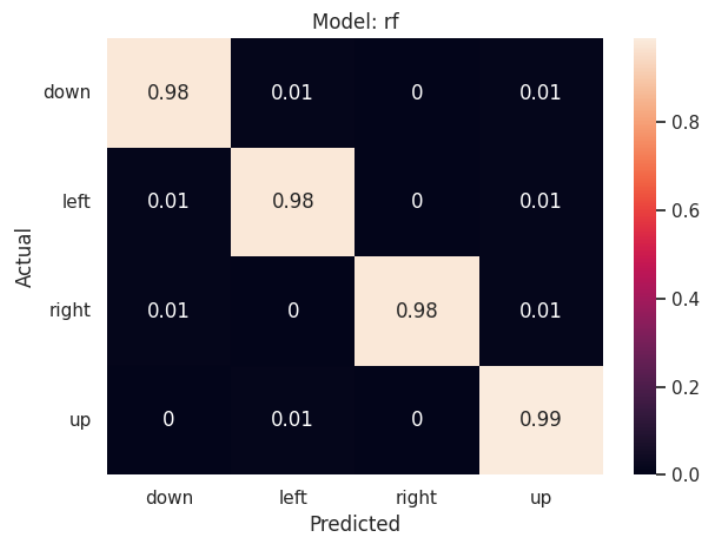


Figure 4.8: Confusion matrix of the RF trained on the dataset with turns removed.

For the RF model, it is observed that the chosen hyperparameters vary inversely based on the presence of turn labels. When turns are labeled, a greater number of trees with shallow depth are selected, whereas, without labeled turns, the model uses four times fewer trees but with double the depth. Additionally, the minimum number of samples per split is twice as high without turn labels compared to the model with turn labels, while the minimum number of samples per leaf remains consistent in both scenarios.

Table 4.15: Hyperparameters of SVM algorithm

experiment	SVM		
	C	<i>kernel</i>	<i>gamma</i>
Non-labeled turns	3	<i>rbf</i>	<i>scale</i>
Turns labeled	4	<i>rbf</i>	<i>scale</i>
Turns removed	4	<i>poly</i>	<i>scale</i>

For the SVM model, the differences are minimal, with the regularization parameter C being the only variant, differing by just one unit.

Table 4.16: Hyperparameters of KNN algorithm

experiment	KNN			
	<i>Neighbors</i>	<i>Weights</i>	<i>Algorithm</i>	p
Non-labeled turn	4	<i>distance</i>	<i>brute</i>	1
Turns labeled	6	<i>distance</i>	<i>ball – tree</i>	2
Turns removed	4	<i>distance</i>	<i>brute</i>	1

For the KNN model, an obvious difference is observed in the number of neighbors; when incorporating left and right turns, the model requires two additional neighbors for classification. The clustering algorithm also changes, but the most significant difference is the distance metric used. The model without turns utilizes the Manhattan distance, whereas the model with turns employs the Euclidean distance.

Notice the reader that comparing the results in Figure 4.7 and those in Figure 4.8, it seems clear that the inclusion of labeling and detection of torso turns leads the models to decrease the performance on the classification of the postures. This effect might be more accentuated in a real life scenario where more torso turns occur. Therefore it is recommended to implement a subject-wise classification strategy that implements a detection mechanism for turns or sudden movements in order to pause the posture classification until the accelerometry signals become stable. The issue of automatic detection will be addressed in future work.

Finally, Table 4.17 compares the performance of our results with the works that we consider most representative of the state of the art.

It is also worth highlighting that we achieve these results with a low-cost commercial ambulatory monitoring device placed on the wrist. Jeng *et al.* [96] already advanced that using only a wristband device improve the user experience with respect to solutions that require more sensors, cameras or more complex devices such as pressure blankets.

Table 4.17: Comparison of metrics from different posture detection studies with the study conducted, indicating the improvement index relative to each.

Author	Metric	Value	Improvement
P.-Y. Jeng. <i>et al.</i> [96]	accuracy	0.833	+0.156
K. Van Laerhoven <i>et al.</i> [206]	precision	0.808	+0.180
S. Ostadabbas <i>et al.</i> [150]	accuracy	0.916	+0.073
Hsiao <i>et al.</i> [88]	accuracy	0.881	+0.108
Y.-Y. Li, <i>et al.</i> [118]	accuracy	0.909	+0.080
Parastoo Alinia <i>et al.</i> [6]	F-value	0.978	+0.011
This work [65]	precision	0.988	+0.000
This work [65]	accuracy	0.989	+0.000
This work [65]	F-value	0.989	+0.000

4.3.2 Sleep model implementation

In this study, different references from patients were taken, mainly the primary pathology related to sleep. Information was also collected about the various pathologies of the patients who participated in the study, and although there were a few cases of patients diagnosed with migraines, they were very few and not relevant since the focus at this stage was to try to detect the different phases of sleep in general and not yet relate them to the condition of migraine.

During the study, 21 control patients, 8 patients with insomnia, 6 patients with PLM, 50 patients with SAHS, and 1 patient with Propriospinal myoclonus (MP) were analyzed. This influenced the classification of the patients when training the models. For trained sleep state classification models, the patient with MP was discarded, although they were analyzed with great interest from a clinical perspective, and the results can be observed in [212].

For the initial sleep phase models, the patients were separated by their primary pathology as previously mentioned, but the results were not satisfactory. Therefore, it was decided to test a model for detecting two states: awake and asleep. As documented in the state of the art, this is a viable model even considering only the ACC. Since no significant improvement was observed, it was hypothesized that the issue could be due to the quality of the EDA and PPG signals. Patients with better signal quality were filtered, resulting in Figure 4.9, which shows an improvement with an accuracy around 0.65. In general, it was observed that almost all models exhibited overfitting after a few epochs of training, another problems were found that will be explained in Chapter 5.2.2

A test was conducted to compare the manual feature selection carried out in Chapter 3.3.2 with automatic feature selection through a preceding layer of CNN with 8 neurons. The results can be seen in Figure 4.11, where an improvement in accuracy is observed, reaching a value of 0.88.

With this approach, the models were repeated with patients grouped by primary pathology, using automatic feature extraction as shown in Figure 4.11. Overfitting was observed after a few epochs of training, and the accuracy was similar in all models, around 0.4. It is true that

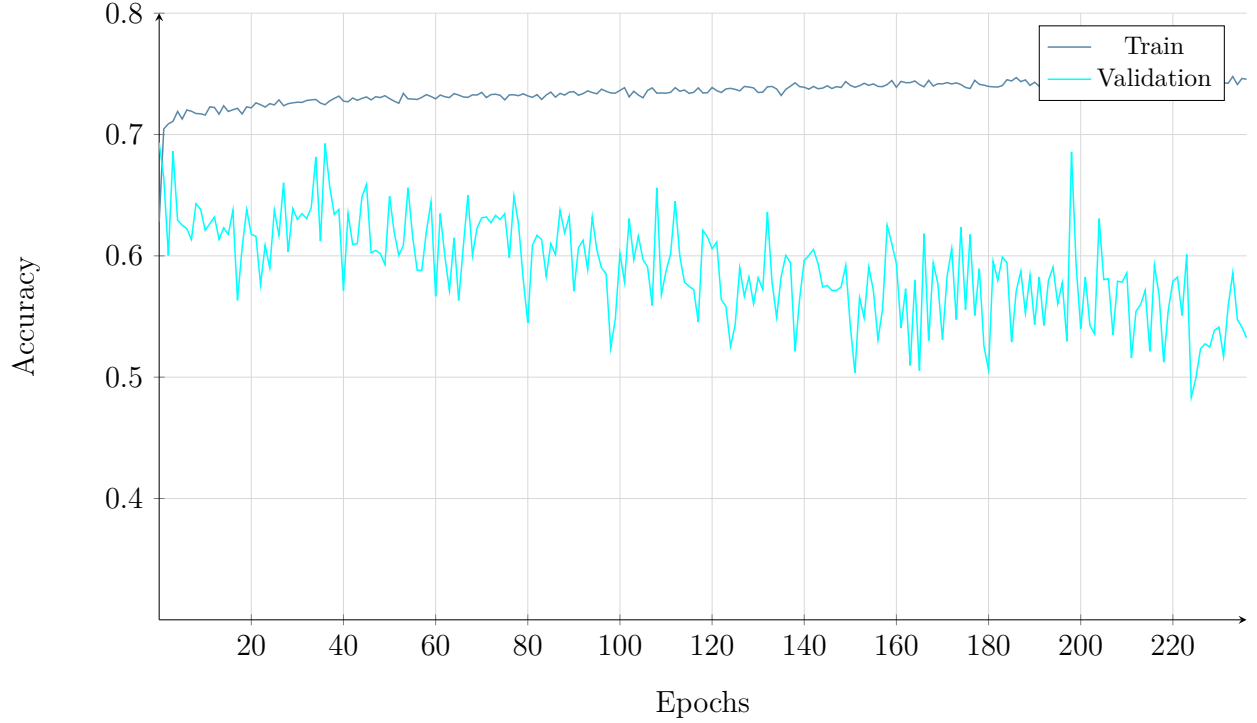


Figure 4.9: Manual FE plus MLP 50 neurons with softmax output. Blue color represent the train-accuracy and cyan color represent validation-accuracy. Awake-sleep model only with patients with EDA and PPG signals with the least amount of noise.

in the long run, the validation accuracy shows stability and is higher for the SAHS groups and the controls, although this may be due to the larger number of patients in these two groups.

Another improvement option was the variability of the patients, so it was proposed to perform a clustering based on the similarity of the characteristics of the signals collected from the different patients. This was carried out using a k-means procedure, resulting in two possible solutions: one classification into 4 groups and another into 7 groups. These analyses can be observed in Figures 4.12 through 4.13, where the training curves are classified by color based on the accuracy metric, with the dark colors representing training and the light colors representing validation. In these experiments, rapid overfitting is still observed, and there are barely any variations between the different groups, maintaining very similar metrics.

Based on the results, training using CV-Leave-One-Out (LOO) was proposed in the hope of identifying outlier patients for elimination. Following the same reasoning as patient clustering, the objective was to use only the common patients for each model.

In Figure 4.14, we can observe that the models with automatic FE and a MLP network of 50 neurons with softmax output, performing CV-LOO, perform better than the other models, achieving an accuracy of 0.6. However, they are somewhat unstable depending on the patient left out for the LOO development, and the improvement is not substantial.

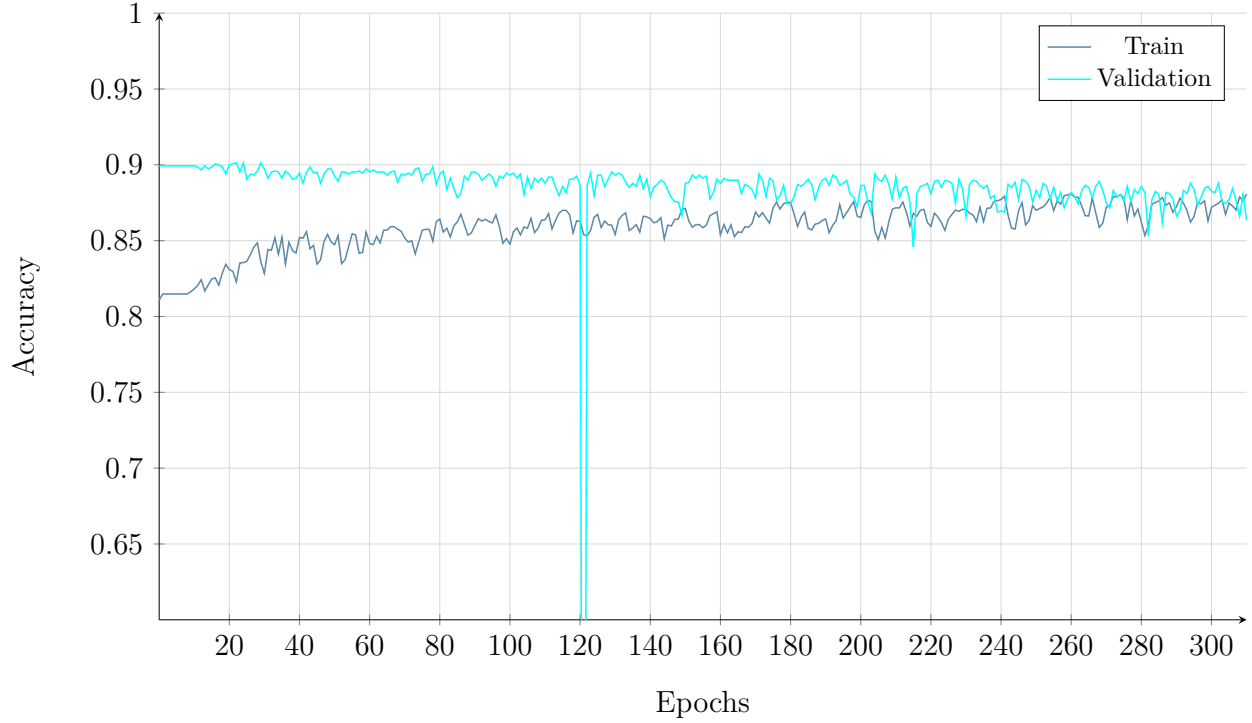


Figure 4.10: CNN 8 neurons plus MLP 50 neurons with softmax output. Intense color represent the train-accuracy and light color represent validation-accuracy. Awake-sleep model only with patients with EDA and PPG signals with the least amount of noise.

Transfer Learning

The next step was to try to improve the models by applying TL techniques. EfficientNet and ResNet models were applied to the images extracted from the PPG. The results were not sufficiently satisfactory to present. However, it is considered relevant to mention that this experiment was conducted.

4.4 Migraine pain alarm

In Chapter 3.4, we saw that, ultimately, only 8 patients were included for the migraine pain predictor, resulting in a total of 24 registered migraine attacks. The total registration period used for testing of 1072 hours (100%), ranging between 45 and 258 hours among patients. The dataset of the patients included in the study is represented in Figure 4.15. The number of migraine attacks, attacks detected, and false positive events during no pain monitoring periods as well as the total number of hours used for testing of each patient is shown in Tables 4.18 and 4.19 for each one of the time delays studied.

Regarding biometric records, median skin distal Temperature found in the sample was 33.3°C ranging from 21.41°C to 36.2°C and a SD of 1.8°C , EDA mean of $1.01\ \mu\text{S}$, ranged between $-0.2\ \mu\text{S}$ and $33.2\ \mu\text{S}$ and a SD of $2.26\ \mu\text{S}$, mean, mean HR computed from IBI was 79.4 ranging from 211.5 to 42.3 and a SD of 15.3. Regarding the feature selection performed, it

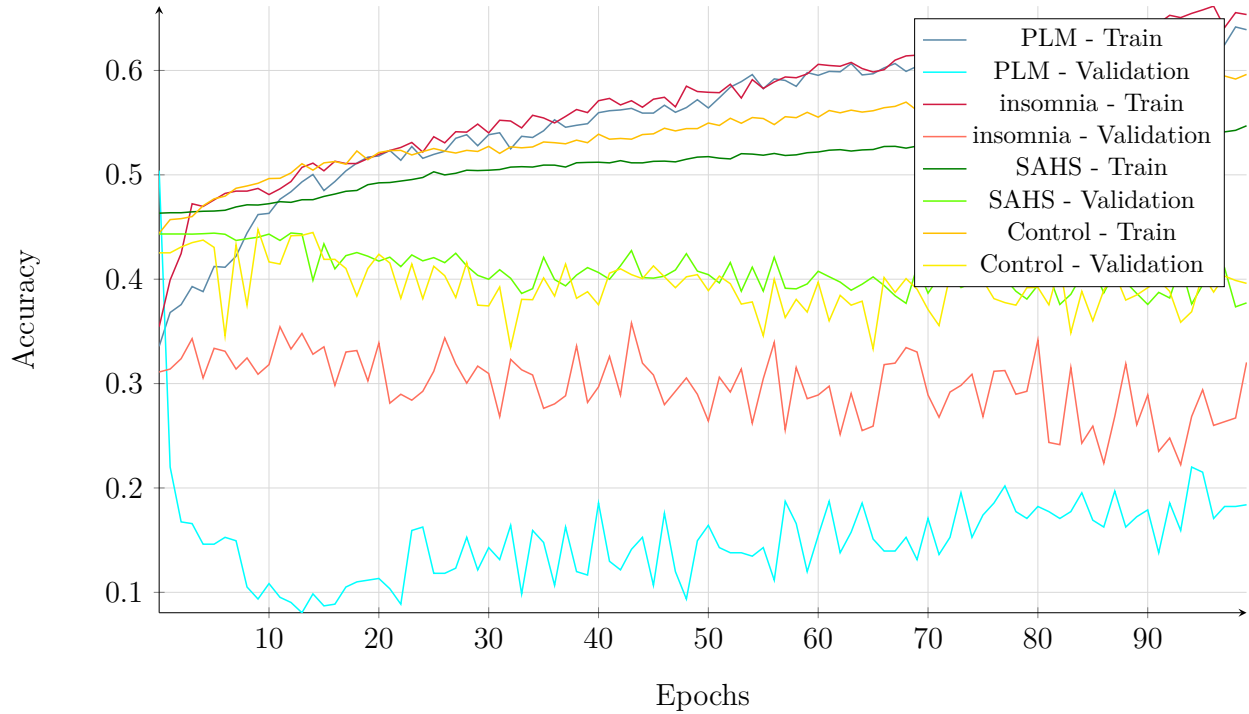


Figure 4.11: CNN 8 neurons plus MLP 50 neurons with softmax output. Intense colors represent the train-accuracy and light colors represent validation-accuracy. Blue - PLM patients, Red - insomnia patients, Green - SAHS patients and Yellow control patients

should be noted that 80% of the EDA features, 73% of the SCL features, and 60% of the HR features disappeared. Regarding the windows, 57% of the features calculated for the last 5 minutes were eliminated. Similarly, 67% of the features representing the maximum and 53% of those representing the minimum and the derivative of the mean were removed as well.

Table 4.18: Monitoring sessions with migraine attacks, TP (True Positive) detection per migraine.

Patient	# migraine attacks	Time (h)	# TP (%) APH 30'	# TP (%) APH 60'
P1	4	26.6	2 (50%)	3 (75%)
P2	5	56	0 (0%)	5 (100%)
P3	2	8.5	2 (100%)	2 (100%)
P4	4	11.9	0 (0%)	4 (100%)
P5	1	2.6	1 (100%)	1 (100%)
P6	1	14.4	1 (100%)	1 (100%)
P7	4	42.2	1 (100%)	4 (100%)
P8	3	27.1	1 (100%)	3 (100%)

The model generated with information on biometric variables successfully predicted pain using the LSTM network, as demonstrated in Figure 4.16. It captured the onset of pain at the beginning of the monitoring period and minor fluctuations at the end of the crisis, which were filtered out by the alarm system. Despite the delay caused by the alarm system, the prediction still anticipated the patient's pain by 30 minutes. In instances without pain crises, as depicted in Figure 4.16, the LSTM network predicted several pain episodes with multiple

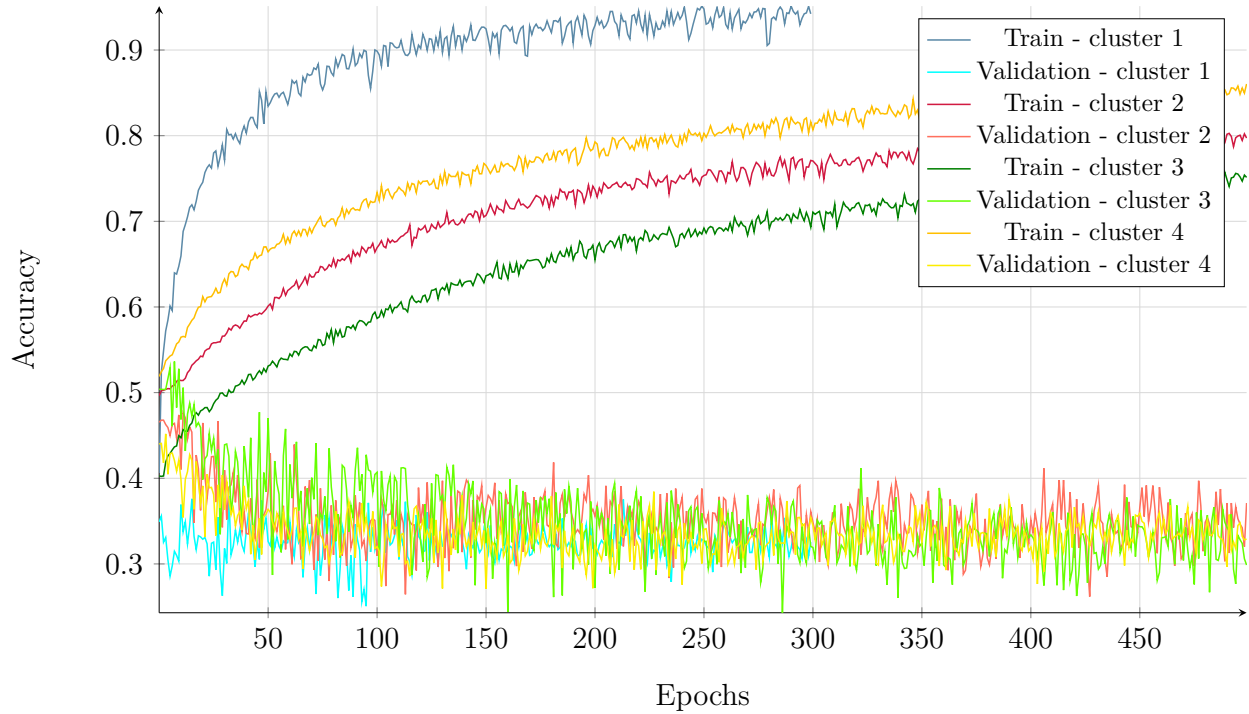


Figure 4.12: Manual FE plus MLP 50 neurons with softmax output. Intense color represent the train-accuracy and light color represent validation-accuracy. Awake-sleep models with the clusters obtained with the k-means 4 cluster classification.

fluctuations up to 60 minutes. However, the alarm detector corrected inconsistencies in the network’s output by not triggering any alerts to the patient. Although various changes may occur in the three hemodynamic signals utilized, some resembling precursors to a migraine pain crisis, the output is not sufficiently robust for consideration by the alarm system, thus minimizing potential false positives.

The predictive system implementing an alarm of 90’ delay was able to predict migraine attacks with 13/24 (54%) sensitivity. All migraine attacks were predicted in 5/8 (62%) of patients. The percentage of minutes with false positive alarms was under 0.4% in 5 patients and over 7.6% in 3 patients.

The predictive system implementing an alarm of 60’ delay was able to predict migraine attacks with 95% (23/24) sensitivity. All migraine attacks were predicted in 7/8 (86%) of patients. The percentage of minutes with false positive alarms was under 0.7% in 5 patients and over 11.2% in 3 patients. Shapley has been run for 100 samples on each model trained and the distribution of the Shapley values is shown in Figure 4.17 including the significance (p-values) of the distribution of the mean value of Shapley grouped by sensor, “past window” and “statistical feature”. The histograms have been divided into 3 levels of Shapley values based on visual inspection of Figure 4.17 (cuts: 0.01, 0.04). The contingency tables are the Tables 1, 2 and 3 shown in Annex 6.7.

In summary: i) There is no type of sensor that stands out from the rest, although the features related to the stress appear to be of higher importance (Shapley value ≥ 0.04)

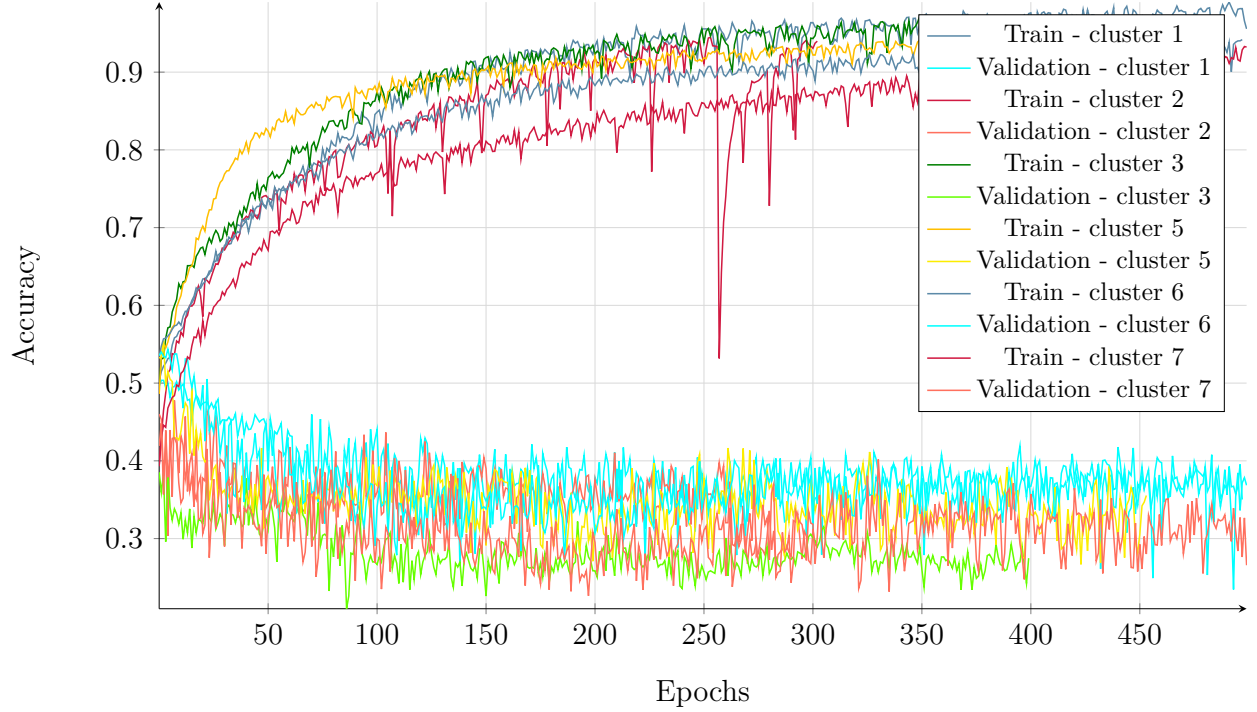


Figure 4.13: Manual FE plus MLP 50 neurons with softmax output. Intense color represent the train-accuracy and light color represent validation-accuracy. Awake-sleep models with the clusters obtained with the k-means 7 cluster classification.

and the contribution of the EDA seems to be of little importance (Shapley value < 0.01), ($p - value = 0.03$). According to this, no sensor of the wristband can be discarded so far; ii) The past window with the most important features is the window of the last 30 minutes. According to this, it is needed to process the data of, at least, the last 30 minutes ($p - value < 0.001$). iii) The statistics that contribute most are related to the *maximum* and *mean* values, and *mode of stress* ($p - value < 0.001$). And the *mean of the derivative*, and *minimum*, contribute little. The total number of registered migraine attacks was 24. Among them, algorithm was able to predict 23/24 (95%) of pain attacks. All migraine attacks were predicted in 7/8 (86%) of patients, with a sensitivity of 100% and no false negatives among them. Regarding global specificity, of the total registration period of 1072 hours (100%), ranging between 45 and 258 hours among patients, the percentage of minutes with false positive was under 2.5% in 4 patients and over 10% in 3 patients.

The number of migraine attacks, attacks detected, percentage of overall prediction, false negative and false positive per minute as well as total number of hours registered are included for each patient are included in Tables 4.18 and 4.19.

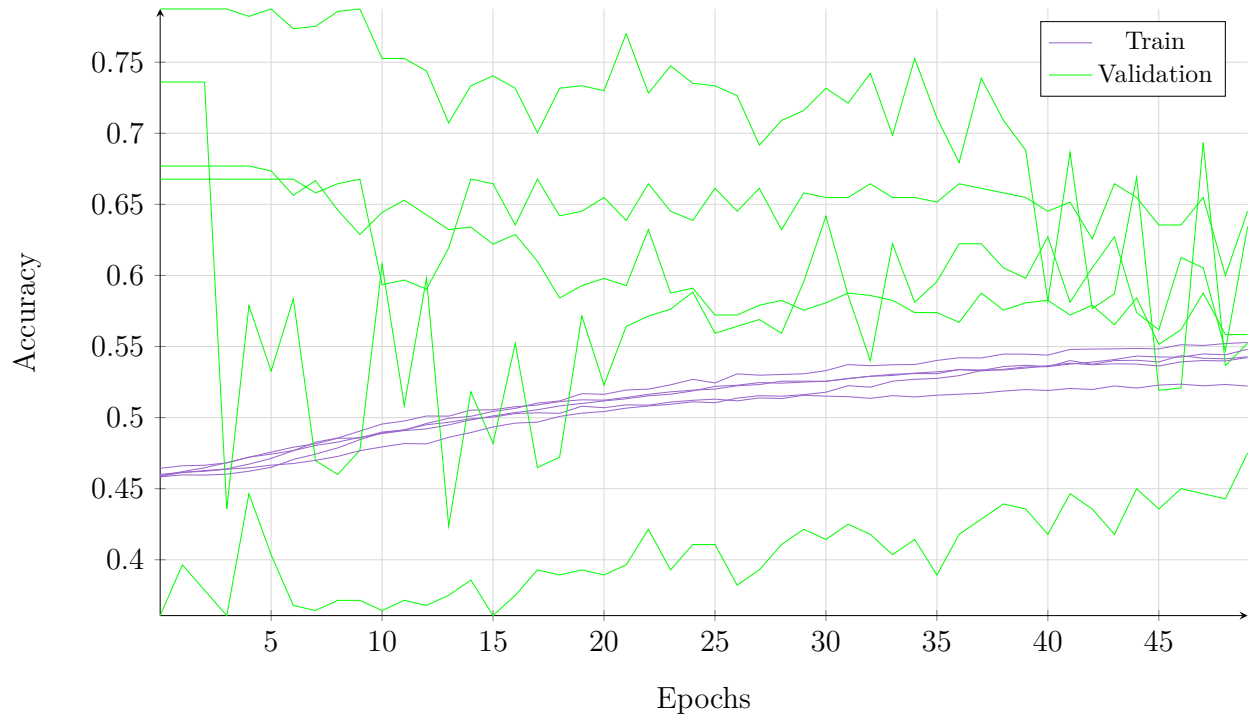


Figure 4.14: CNN 8 neurons plus MLP 50 neurons with softmax output. Purple color represent the train-accuracy and green color represent validation-accuracy. Sleep state detection models.

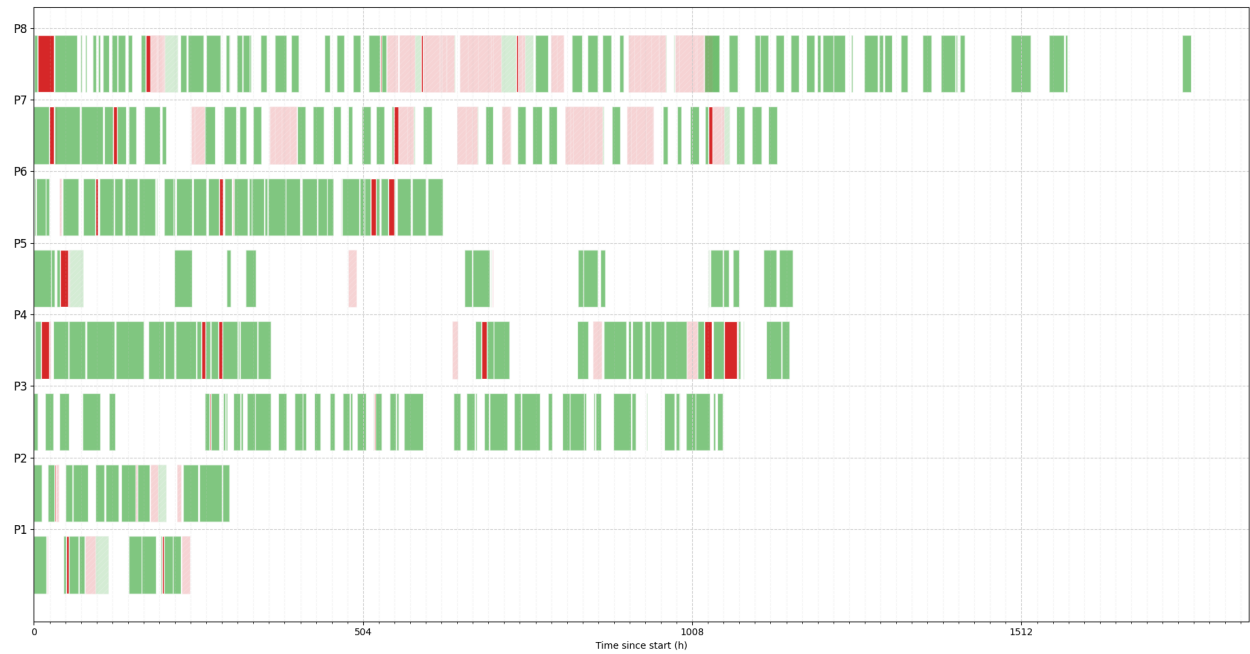


Figure 4.15: Dataset of the patients included in the study. Red: Pain period. Green: no pain period. Dark green: noisy data.

Table 4.19: Monitoring sessions in basal periods (± 24 h from migraine), FP (False Positives) detection per minute.

Patient	# basal periods	Time (h)	# FP (%) APH 30'	# FP (%) APH 60'
P1	4	142.4	1 (0.4 %)	2 (0.7 %)
P2	5	202.2	0 (0 %)	0 (0 %)
P3	2	81.5	0 (0 %)	1 (0.5 %)
P4	5	138.3	0 (0 %)	0 (0 %)
P5	2	69.3	0 (0 %)	1 (0.3 %)
P6	2	30.5	1 (17.6 %)	1 (22.4 %)
P7	6	122.3	6 (13.9 %)	6 (20.3 %)
P8	5	96.4	5 (7.6 %)	5 (11.2 %)

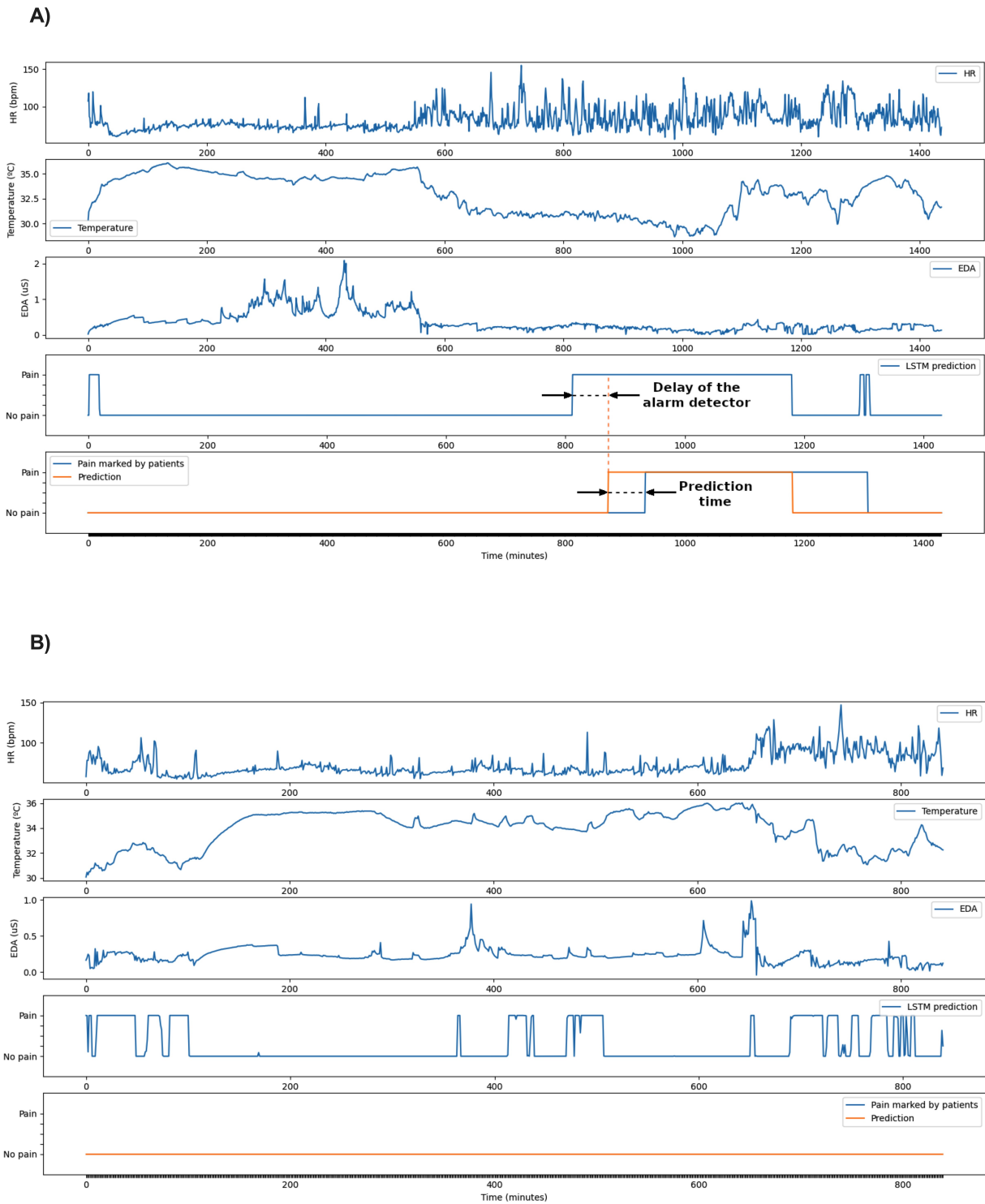


Figure 4.16: Variables registered with the Empatica E4 the wearable devices. A) Example of a monitoring period lasting 10 hours with a migraine crisis lasting roughly 2.5 hours. B) Example of the monitoring of a no pain period. The examples show the prediction of the LSTM and the output of the alarm filter. HR: Heart rate. Temperature ($^{\circ}C$).

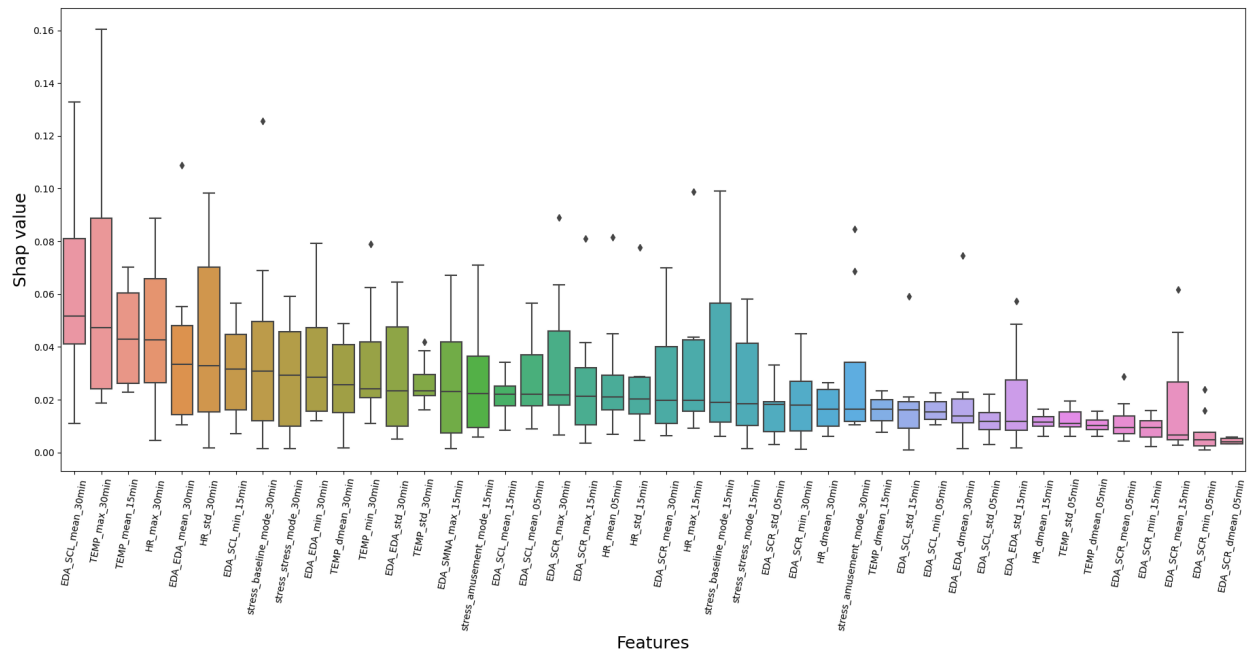


Figure 4.17: Boxplot of the distribution of the Shapley values of the features for all the models trained for all patients.

Chapter 5

Discussion

In this chapter, the results will be discussed, providing partial conclusions for each experiment conducted.

Finally, Chapter 6 presents the culmination of this holistic system for managing crises in chronic diseases, with migraine serving as a key example. The convergence of the various modules explored in this work and the potential combinations in both this case and in other related diseases provide the system with sufficient modularity for the personalization and particularization of each case, with the intention of improving its functionality.

5.1 Pain curve characterization

In this study, a novel methodology was presented for the clustering of the types of pain curves studied with migraine patients, and the results have shown that migraines can be clustered into 4 different types based on 8 features of the pain curves.

These results open a new research topic on the study of migraine and other chronic diseases cursing with symptomatic crises that must be addressed from a clinical point of view. For example: are all the pain episodes from one patient of the same type? And therefore, should we tackle these crises in the same way?

Questions about the patient's phenotype related to pain types should be addressed in further research studies. However, from a technical point of view, the results open many questions that refer to the motivation of this work and that must be considered from now on: How can these curves be useful for the prediction of the pain type in real time? For the development of personalized models or for the prediction of pain episodes, should patients continue to provide continuous information on the evolution of their pain? Is it possible to know in real time the type of pain that will occur?

To address these questions, we should look closer to the 8 used features: the slopes of the acme $\delta_{acme\uparrow}$ and $\delta_{acme\downarrow}$, the duration D , the energy E , the maximum of the pain $Acme$, the area at the end of the curve $A_{acme\downarrow}$, and the time of pain over 90% of acme D_{90} . It seems that it is not possible to know the type of developing pain in advance because there are

features that can only be computed at the end of the pain episode, such as energy E . Now that we have a label for each pain episode, a supervised feature selection can be run to detect redundancy and reduce the number of features. Similar to the multimodal pain classification systems highlighted in [149], which consider various factors including socioeconomic burdens and psychological aspects, understanding the nuances of these pain features may enhance personalized treatment approaches and address barriers in pain management effectively.

5.1.1 Group clustering by pain type

Despite the fact that, in general, high correlation metrics between the various variables and groupings of patients were not achieved, it is very interesting to highlight the different relationships found and the possibility of a new classification model for migraine patients. In addition, this methodology can be extrapolated to any chronic disease with pain crises.

Furthermore, it has been shown that this classification provides relevant information for the design and implementation of a clinical advice recommendation system, which helps educate patients with chronic diseases in as many aspects as possible. The result of this implementation has culminated in an algorithm that provides the most relevant information for each user based on their feedback, sociodemographic data, and the progression of their pain episodes.

During the development of the recommendation system, three types of information were identified based on their content, each requiring a different recommendation approach. Among these types of advice, the most important one, not only in our system but also for the patient, is ESPA, as it contains guidelines relevant to the specific symptoms of the disease. To model the selection of the most appropriate advice in this category for a migraine case, we collaborated with the Department of Neurology of the Hospital Universitario de La Princesa in Madrid. Their supervision enabled our system to incorporate criteria that closely align with those of healthcare professionals.

To attempt to reconstruct this medical reasoning, a RNN was chosen. Based on the patient's symptomatological and pain data, the RNN determines the relevance of the various available ESPA, prioritizing the most relevant advice in the recommendations. We evaluated the performance of this network against a MLP, obtaining quantitatively similar results in both cases. It is concluded that the properties of a recurrent network do not provide a significantly better option for modeling the evolution of the patient's crises.

One of the data points this RNN uses to determine the relevance of each piece of advice for a patient comes from the group or cohort to which they have been assigned. After an extensive research process, five distinct cohorts have been identified, differentiated by the proportion of types of pain progression experienced by the patients. We also considered the possibility of deducing the cohort of a patient based on factors other than their pain. While this seems feasible, this aspect has been left open for future lines of research.

This approach aligns with the evolving landscape of healthcare systems as outlined in recent studies [186], which emphasize the transition toward patient-centric models enabled by advanced technologies like the Internet of Things and machine learning. These innovations in

healthcare recommendation systems mirror the challenges and opportunities discussed in the broader context of Healthcare 4.0, underscoring the importance of integrating personal health data to enhance patient outcomes.

5.2 Feasibility of sleep modeling for migraine patients

Next, we will discuss the results of the sleep models, including both the posture model and the implementation of the sleep state detection model, along with the collateral results of the study. As mentioned in Chapter 4.3.2, an analysis of the different patients between the different diseases and the control group led to a publication [212].

5.2.1 Posture detector model

It has been verified that it is indeed possible to classify posture during sleep, however, to make it reproducible and accurate, it is necessary (i) to detect and remove torso turns and (ii) to perform a subject-wise training. The performance of our solution improves the results of the state-of-the-art [6], [88], [96], [118], [150], [206]. The test results of our models show F-values of 0.966 and 0.989 for the RF and KNN algorithms respectively.

The difference in metrics may seem meaningless between (i) not labeling, (ii) labeling and classifying, and (iii) labeling and removing torso turns. However, it has been shown that the low recall of the *turn_left* and *turn_right* labels when labeling the turns—0.258 and 0.777, respectively—claims to use the methodology proposed. In a real case of monitoring subjects, the number of turns and noisy data will be larger, and the difference between methodologies will be greater.

The solution here outperforms the results of the state-of-the-art from 0.011 to 0.180 points. In future work, we will tackle the problem of automating detection of turns to pause the classification in real time. These results are a step forward towards making ambulatory sleep monitoring with a clinical perspective feasible.

5.2.2 Sleep model implementation

The results of the clinical analysis from [212], are that polysomnography indicated that individuals with insomnia experienced a notable reduction in sleep efficiency and total sleep duration, extended sleep latency, and increased wakefulness after the onset of sleep. Accelerometry data showed significant differences exclusively during stage N3; no differences were observed between the groups during REM sleep. Both groups reached their lowest temperature during REM sleep. In the insomnia group, peripheral temperature did not decrease during sleep phases compared to wakefulness, in contrast to the control group. Insomniac patients had a higher heart rate than controls during both the wakefulness and sleep stages. Heart rate variability was reduced in stage N3 and increased in REM sleep compared to wakefulness in both groups. Sweating, as indicated by Skin Conductance Variability values and SMNA, was significantly higher in insomniac patients compared to the control group. These findings indicate that individuals with insomnia exhibit heightened

sympathetic activity during sleep, demonstrated by an elevated heart rate. In addition, temperature regulation and sweating have a significant impact on the various stages of sleep. These identified characteristics could aid in the creation of sleep disorder detection models for the subsequent application of a specific model to measure sleep quality. This could potentially improve current models and thus improve our holistic system for managing migraines.

In sleep state detection models, the results demonstrate a lack of complexity necessary to achieve outcomes comparable to the group of P. Fonseca *et al.* [58], [59], [165], [166], [216]. Additionally, the absence of labeled awake events contributes to an unbalanced data set. Moreover, our awake labels correspond to micro-awakenings during sleep, which likely differ significantly in features from regular awake events. These awake events have been used in the state of the art to balance the classes, which, in light of the results, might be considered a methodological error.

It is evident that there is an upper limit to the accuracy of the detection of the sleep state without the use of EOG or EEG. However, incorporating the EDA signal as a feature aids in the event classification, and automatic feature extraction also proves beneficial. This latter point is corroborated by the application of TL (Transfer Learning) by P. Fonseca *et al.*

Taking into account these findings, the use of a much simpler system that simply measures sleep quality at night, such as those already available on the market, would suffice to contribute valuable information to our holistic system to manage and predict migraine events. This integration would improve the overall functionality and effectiveness of the system in addressing chronic disease crises, exemplified by migraine.

5.3 Migraine pain alarm

The feasibility of predicting headache during a migraine attack has been demonstrated by the pilot study using a wearable device that monitors hemodynamic variables such as temperature, HR, and EDA. The onset of pain up to an hour in advance can be forecasted by ML models, opening new avenues to anticipate migraine episodes.

In this study, 24 migraine attacks and pain-free periods from 8 patients with episodic migraine were evaluated in a headache unit. A wearable device was used to measure hemodynamic variables, focusing on temperature, HR, and EDA signals. It was found that analyzing pain and non-pain records can generate accurate ML models to predict the headache phase of a migraine attack.

The potential utility of hemodynamic variables in the prediction of migraines has been supported by previous research [153], [156]. An earlier study using a monitoring kit before the availability of wearable devices demonstrated that migraine pain could be modeled as a curve using the main characteristics of the pain. It was also suggested that the pain phase of a migraine attack could be predicted on the basis of these models.

This study is among the first to show that ML models using only hemodynamic variables collected using a wearable device. The models per patient achieved high sensitivity (95%) and accuracy were achieved by the per-patient models. Using an alarm system, different

predictive probabilities are provided depending on the quality of the data and the accuracy of the model. High sensitivity and a low FP rate compared to other studies using hemodynamic variables were shown.

Future research should evaluate the average and maximum values of the signals. Although data from all patients for training and testing were used, similar to predicting seizures in epilepsy, it is necessary to validate this algorithm in a larger independent cohort.

Other studies have shown that identifying premonitory symptoms and administering treatments such as domperidone or naratriptan during this phase can prevent the pain phase in many patients [99]. This evidence suggests that identifying early stages of a migraine attack could allow timely treatment, potentially aborting the pain phase.

The study's main limitations include a small sample size and the complete-case analysis methodology. Although individual models were created, which limits generalizability, these models could inform a general model that learns from each patient as new data is added. The preventive treatments allowed in the study could have influenced the frequency, intensity, and duration of migraine, although symptomatic pain relief treatment did not affect the onset of pain.

The possibility of predicting migraine attack pain phases using hemodynamic variables via a wrist-worn device has been demonstrated by this proof-of-concept study. Future studies with larger patient samples and combining patient baseline characteristics with migraine features in both premonitory and pain phases could improve predictability. In future research, real-time alarms for oncoming migraines will also be considered.

5.3.1 LSTM and alarm detector

All the points mentioned in the previous section clearly indicate the potential for developing reliable models to predict migraine crises. This approach could also be extrapolated to other types of disease. Given the exponential advancements in machine learning models, their development could enhance our model, potentially eliminating the need for a secondary alarm module or achieving more sophisticated models with higher accuracy and longer prediction horizons.

Although the metrics analyzed solely from the model's perspective are insufficient, considering that the detection and error rate per minute, with the model generating a result every minute, shows a significant error percentage, this is mitigated when considering the entire model. The prediction window is sufficiently large to allow time to verify and check if a false positive is indeed such by comparing subsequent predictions. The model can be improved, for example, by incorporating the automated feature extraction strategy seen in the sleep model. However, it is adequate for inclusion in our holistic system, providing highly useful information to manage pain medications.

Chapter 6

Conclusions

This work presents a comprehensive investigation of the development and implementation of a holistic system for the management and prediction of symptomatic crises in chronic diseases using advanced AI techniques. The following key conclusions can be drawn from this research:

6.1 Feasibility and effectiveness of AI models

The research demonstrates the feasibility of using AI models for predicting symptomatic crises in chronic diseases that focus on migraine with a sensitivity of 95%, 30 ‘ before pain. The predictive models developed in this study, specifically for migraine crises, have shown promising results in terms of the percentage of false positives per minute. These models can be potentially adapted to other chronic conditions, leveraging the flexibility and adaptability of ML algorithms.

An automatic sleep position detection model that improves state-of-the-art results by at least 0.011 points. A sleep state detection model with an accuracy of 0.5 when detecting all sleep states and 0.9 when detecting two of them.

In the case of classifying pain curves using four features, we obtain an F-score of 99.1%, and in the recommender system, the neural network achieves an R^2 of 0.951.

6.2 Analysis and confrontation from different fronts

The study and analysis of the disease led to the application of AI techniques from various perspectives, (i) including a new classification of migraine pain, (ii) the potential to classify patients according to this new method of grouping of pain, and (iii) the use of this information to develop an informational application to improve patient management and adherence to treatment. In addition, (iv) the detection of different sleep states and their quality, due to the close relationship between sleep and migraine crises, and finally (v) the crisis prediction model were considered. These approaches have been studied and analyzed throughout this work. A significant advantage of a holistic and modular model is the ability to expand and improve the system through additional modules that address different aspects.

The importance of the type and classification of pain has been observed and the enhancement of this information in patient engagement with the disease and the information required by the patient. The best current method of incorporating sleep quality to provide additional information to the patient has been determined to be through commercial models. Furthermore, reliable support for pain prediction would be the pinnacle of the system.

6.3 Improvement in model performance

Incorporation of advanced techniques, such as automatic FE (feature extraction), has demonstrated an improvement in the model compared to manual extraction. However, the application of other techniques, such as TL (Transfer learning), has not shown such improvement, at least in the manner employed. This involved applying the knowledge gained from the image analysis models to the PPG spectrogram for image detection using the wavelet transformation. Nevertheless, despite these results, it is premature to completely discard this technique due to potential issues such as label imbalance and lack of eye movement information in the system.

The mathematical analysis and abstraction used to define pain curves have been crucial in classifying different types of migraine pain. Similarly, an effective alarm system has facilitated the successful performance of the migraine crisis prediction system.

6.4 Practical implications for clinical use

The practical application of the developed system in a clinical setting has been evaluated. The system's ability to provide timely and accurate predictions of symptomatic crises can significantly aid healthcare providers in decision-making processes, ultimately improving patient outcomes. The system's modularity also allows for customization based on individual patient needs, further enhancing its clinical utility.

These improvements will enable patients to better identify their symptoms and improve the prediction of their crises, allowing for the earlier administration of analgesics to prevent pain, which has been shown to increase their efficiency. Additionally, it will enhance the patient's understanding of the disease, helping them to adopt a healthier lifestyle and thereby improving the frequency of their crises.

These improvements will enable patients to better identify their symptoms and improve the prediction of their crises, allowing for the earlier administration of analgesics to prevent pain, which has been shown to increase their efficiency. Additionally, it will enhance the patient's understanding of the disease, helping them to adopt a healthier lifestyle and thereby improving the frequency of their crises. Thanks to the improvement in the patient's understanding of the disease, triggering factors will be reduced, leading to a considerable decrease in the number of migraine crises.

It demonstrates that a tool with these characteristics would be a great support for the healthcare team, enhancing their intervention and contact with the patient for treatment follow-up.

This system can also be effectively integrated into the industrial context of the sponsoring company, Brainguard, leveraging its expertise and leadership in developing innovative solutions for the control and treatment of migraines. The platform will enable Brainguard to enhance its technological offerings by providing a tool that not only facilitates early symptom identification and analgesic administration but also educates patients about their condition, thereby reducing triggering factors. In turn, this will strengthen Brainguard’s market position by offering a comprehensive approach that combines prevention, personalized treatment, and improvement in patients’ quality of life.

6.5 Problems and complications encountered during the course of the work

Time constraints, lack of technological proficiency, or lack of adherence to the study by patients have led to a reduction in the amount of data available to train models, which likely resulted in a decrease in their performance.

In some aspects, the lack of state-of-the-art guidance due to the novelty of the research has implied a heuristic development of the models, creating models never used before for these purposes. This represents a first step in these investigations and demonstrates that there is still much to explore and improve in this area.

Challenges such as data imbalance and false positives have been addressed through the development of robust validation frameworks and the implementation of sophisticated data preprocessing techniques. The use of larger prediction windows and subsequent verification of predictions help mitigate the impact of false positives, ensuring more reliable output. In the case of data balancing, the approach has not been sufficient. Nevertheless, the state-of-the-art approach using awake events in sleep state detection models, outside the sleep period, might be a poor approach for detection, as the number of micro-awakenings relative to the obtained metric does not present a significant difference.

6.6 Future directions and potential improvements

The research identifies several areas for future work and potential improvements. These include the incorporation of additional data types, such as genetic information, and the exploration of new machine learning algorithms and architectures. Continuous refinement and validation of the system with larger and more diverse datasets will further enhance its robustness and applicability.

The development of new related modules, such as those for promoting a healthy lifestyle by incorporating nutritional and fitness information into the holistic system, would enhance the effectiveness of the existing modules.

The characteristics of the models have been discussed, but a detailed study of their power consumption and efficiency falls outside the scope of this work. This could be addressed through the application of model quantization methodologies for their implementation in on

mobile devices or wearable devices.

Regarding the posture detection model, an analysis can be conducted to improve the model by using spatial representation through quaternions, as they could provide greater accuracy to the model.

The attempt at TL using models trained on images for feature extraction from PPG spectrogram images was not satisfactory. However, this methodology, with another model specifically trained for this type of images, could likely yield good results.

As another potential future line of research, the holistic system could be integrated as the foundation of a simulated system, functioning as a digital twin. The goal would be to closely replicate the digital system to the physical one.

For the migraine crisis prediction system, obtaining a larger dataset would enable the development of better personalized models for patients. Furthermore, applying TL techniques to create a general model and transferring its knowledge to individualized patient models could further improve outcomes.

Most importantly, integrating information from other modules into the migraine pain crisis prediction system is crucial. For example, incorporate a sleep quality module using commercial equipment into the model. This methodology could also be applied to other chronic conditions, such as epilepsy or oncological pain.

6.7 Application of AI methodologies to the field of healthcare

This thesis makes a significant contribution to the field of healthcare by demonstrating the potential of holistic AI systems in the management of chronic diseases. The findings underscore the importance of interdisciplinary research and the integration of AI with medical expertise to develop innovative solutions to complex healthcare challenges.

In conclusion, the holistic system developed represents a significant advance in the use of AI for the management and prediction of symptomatic crises in chronic diseases. The results of this research pave the way for future innovations and improvements in AI-driven healthcare solutions, ultimately aiming to enhance the quality of life of patients suffering from chronic diseases.

Bibliography

- [1] T. I. T. U. (ITU), “Itu-t recommendation y.2060 : Overview of the internet of things”, International Telecommunication Union, Tech. Rep., 2012.
- [2] M. Abou Jaoude, H. Sun, K. R. Pellerin, *et al.*, “Expert-level automated sleep staging of long-term scalp electroencephalography recordings using deep learning”, *Sleep*, vol. 43, no. 11, zsaa112, 2020.
- [3] H. Adeli, Z. Zhou, and N. Dadmehr, “Analysis of eeg records in an epileptic patient using wavelet transform”, *Journal of neuroscience methods*, vol. 123, no. 1, pp. 69–87, 2003.
- [4] A. Akbari, P. Alinia, H. Ghasemzadeh, and R. Jafari, “Transfer learning for wearable computers”, in *Wearable Sensors*, Elsevier, 2021, pp. 435–459.
- [5] M. Alazab, L. U. Khan, S. Koppu, *et al.*, “Digital twins for healthcare 4.0—recent advances, architecture, and open challenges”, *IEEE Consumer Electronics Magazine*, vol. 12, no. 6, pp. 29–37, 2022.
- [6] P. Alinia, A. Samadani, M. Milosevic, H. Ghasemzadeh, and S. Parvaneh, “Pervasive lying posture tracking”, *Sensors*, vol. 20, no. 20, p. 5953, 2020.
- [7] J. Allen, “Photoplethysmography and its application in clinical physiological measurement”, *Physiological measurement*, vol. 28, no. 3, R1, 2007.
- [8] S. U. Amin, M. S. Hossain, G. Muhammad, M. Alhussein, and M. A. Rahman, “Cognitive smart healthcare for pathology detection and monitoring”, *IEEE Access*, vol. 7, pp. 10 745–10 753, 2019. DOI: [10.1109/ACCESS.2019.2891390](https://doi.org/10.1109/ACCESS.2019.2891390).
- [9] A. Angelopoulos, E. T. Michailidis, N. Nomikos, *et al.*, “Tackling faults in the industry 4.0 era—a survey of machine-learning solutions and key aspects”, *Sensors*, vol. 20, no. 1, 2020. DOI: [10.3390/s20010109](https://doi.org/10.3390/s20010109).
- [10] S. A. H. Aqajari, R. Cao, E. K. Naeini, *et al.*, “Pain assessment tool with electrodermal activity for postoperative patients: Method validation study”, *JMIR mHealth and uHealth*, vol. 9, no. 5, pp. 1–11, 2021. DOI: [10.2196/25258](https://doi.org/10.2196/25258).
- [11] S. A. H. Aqajari, E. K. Naeini, M. A. Mehrabadi, S. Labbaf, A. M. Rahmani, and N. Dutt, “GSR Analysis for Stress: Development and Validation of an Open Source Tool for Noisy Naturalistic GSR Data”, *arXiv*, 2020. arXiv: [2005.01834](https://arxiv.org/abs/2005.01834).
- [12] M. Arnold, “Headache classification committee of the international headache society (ihs) the international classification of headache disorders”, *Cephalalgia*, vol. 38, no. 1, pp. 1–211, 2018.

- [13] L. P. Arts and E. L. van den Broek, “The fast continuous wavelet transformation (fcwt) for real-time, high-quality, noise-resistant time–frequency analysis”, *Nature Computational Science*, vol. 2, no. 1, pp. 47–58, 2022.
- [14] H. F. Atlam, R. J. Walters, and G. B. Wills, “Fog computing and the internet of things: A review”, *Big Data and Cognitive Computing*, vol. 2, no. 2, pp. 1–18, 2018. DOI: [10.3390/bdcc2020010](https://doi.org/10.3390/bdcc2020010).
- [15] B. Bag and N. Karabulut, “Pain-relieving factors in migraine and tension-type headache”, *International journal of clinical practice*, vol. 59, no. 7, pp. 760–763, 2005.
- [16] G. Bannerjee, U. Sarkar, S. Das, and I. Ghosh, “Artificial intelligence in agriculture: A literature survey”, *International Journal of Scientific Research in Computer Science Applications and Management Studies*, vol. 7, no. 3, pp. 1–6, 2018.
- [17] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives”, *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.
- [18] M. E. Bigal, D. Serrano, D. Buse, A. Scher, W. F. Stewart, and R. B. Lipton, “Acute migraine medications and evolution from episodic to chronic migraine: A longitudinal population-based study”, *Headache: The Journal of Head and Face Pain*, vol. 48, no. 8, pp. 1157–1168, 2008.
- [19] M. Bigal, A. Rapoport, F. Sheftell, S. Tepper, and R. Lipton, “Transformed migraine and medication overuse in a tertiary headache centre—clinical characteristics and treatment outcomes”, *Cephalalgia*, vol. 24, no. 6, pp. 483–490, 2004.
- [20] J. Blau, “Migraine: Theories of pathogenesis.”, *The Lancet*, vol. 339, no. 8803, pp. 1202–1207, 1992.
- [21] A. Boiardi, L. Munari, I. Milanese, C. Paggetta, E. Lamperti, and G. Bussone, “Impaired cardiovascular reflexes in cluster headache and migraine patients: Evidence for an autonomic dysfunction”, *Headache: The Journal of Head and Face Pain*, vol. 28, no. 6, pp. 417–422, 1988.
- [22] H. Bolhasani, M. Mohseni, and A. M. Rahmani, “Deep learning applications for IoT in health care: A systematic review”, *Informatics in Medicine Unlocked*, vol. 23, no. March, p. 100550, 2021. DOI: [10.1016/j.imu.2021.100550](https://doi.org/10.1016/j.imu.2021.100550).
- [23] M. Bonafede, S. Sapra, N. Shah, S. Tepper, K. Cappell, and P. Desai, “Direct and indirect healthcare resource utilization and costs among migraine patients in the united states”, *Headache: The Journal of Head and Face Pain*, vol. 58, no. 5, pp. 700–714, 2018.
- [24] M. Borazio, U. Blanke, and K. Van Laerhoven, “Characterizing sleeping trends from postures”, in *International Symposium on Wearable Computers (ISWC) 2010*, 2010, pp. 1–2. DOI: [10.1109/ISWC.2010.5665853](https://doi.org/10.1109/ISWC.2010.5665853).
- [25] A. Borrego, J. Latorre, M. Alcaniz, and R. Llorens, “Reliability of the Empatica E4 wristband to measure electrodermal activity to emotional stimuli”, *International Conference on Virtual Rehabilitation, ICVR*, vol. 2019-July, pp. 11–12, 2019. DOI: [10.1109/ICVR46560.2019.8994546](https://doi.org/10.1109/ICVR46560.2019.8994546).
- [26] E. Bose and K. Radhakrishnan, “Using unsupervised machine learning to identify subgroups among home health patients with heart failure using telehealth”, *CIN: Computers, Informatics, Nursing*, vol. 36, no. 5, pp. 242–248, 2018.

- [27] N. L. Bragazzi, “From p0 to p6 medicine, a model of highly participatory, narrative, interactive, and “augmented” medicine: Some considerations on salvatore iaconesi’s clinical story”, *Patient preference and adherence*, vol. 7, p. 353, 2013.
- [28] F. P. Bravo, A. A. D. B. García, M. M. Gallego, *et al.*, “Prediction of patient’s response to onabotulinumtoxin treatment for migraine”, *Heliyon*, vol. 5, no. 2, 2019.
- [29] F. P. Bravo, A. A. D. B. García, A. B. G. Veiga, *et al.*, “Smurf: Systematic methodology for unveiling relevant factors in retrospective data on chronic disease treatments”, *IEEE Access*, vol. 7, pp. 92 598–92 614, 2019.
- [30] J. Brown, “Classifiers and their metrics quantified”, *Molecular informatics*, vol. 37, no. 1-2, p. 1 700 127, 2018.
- [31] L. Brueggeman, T. Koomar, and J. J. Michaelson, “Forecasting risk gene discovery in autism with machine learning and genome-scale data”, *Scientific Reports*, vol. 10, no. 1, pp. 1–11, 2020. DOI: [10.1038/s41598-020-61288-5](https://doi.org/10.1038/s41598-020-61288-5).
- [32] Y. S. Can, N. Chalabianloo, D. Ekiz, and C. Ersoy, “Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study”, *Sensors*, vol. 19, no. 8, p. 1849, 2019.
- [33] D. Cary, K. Briffa, and L. McKenna, “Identifying relationships between sleep posture and non-specific spinal symptoms in adults: A scoping review”, *BMJ open*, vol. 9, no. 6, e027633, 2019.
- [34] F. Changyong, W. Hongyue, L. Naiji, C. Tian, H. Hua, L. Ying, *et al.*, “Log-transformation and its implications for data analysis”, *Shanghai archives of psychiatry*, vol. 26, no. 2, pp. 105–109, 2014.
- [35] M. D. Cheatle, “Depression, chronic pain, and suicide by overdose: On the edge”, *Pain medicine*, vol. 12, no. suppl_2, S43–S48, 2011.
- [36] J. Chen, J. Zheng, H. Yu, *et al.*, “Finding important terms for patients in their electronic health records: A learning-to-rank approach using expert annotations”, *JMIR medical informatics*, vol. 4, no. 4, e6373, 2016.
- [37] L. Chen, P. Chen, and Z. Lin, “Artificial intelligence in education: A review”, *IEEE Access*, vol. 8, pp. 75 264–75 278, 2020. DOI: [10.1109/ACCESS.2020.2988510](https://doi.org/10.1109/ACCESS.2020.2988510).
- [38] W. Chen, N. Jaques, S. Taylor, A. Sano, S. Fedor, and R. W. Picard, “Wavelet-based motion artifact removal for electrodermal activity”, in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2015, pp. 6223–6226.
- [39] Y. Chen, X. Qin, J. Wang, C. Yu, and W. Gao, “Fedhealth: A federated transfer learning framework for wearable healthcare”, *IEEE Intelligent Systems*, vol. 35, no. 4, pp. 83–93, 2020.
- [40] E. D. Chinoy, J. A. Cuellar, K. E. Huwa, *et al.*, “Performance of seven consumer sleep-tracking devices compared with polysomnography”, *Sleep*, vol. 44, no. 5, zsaa291, 2021.
- [41] J. Cho, “Current Status and Prospects of Health-Related Sensing Technology in Wearable Devices”, *Journal of Healthcare Engineering*, vol. 2019, 2019. DOI: [10.1155/2019/3924508](https://doi.org/10.1155/2019/3924508).
- [42] J. Cho, “Current status and prospects of health-related sensing technology in wearable devices”, *Journal of Healthcare Engineering*, vol. 2019, 2019.
- [43] L. Cohen, *Time-frequency analysis*. Prentice hall New Jersey, 1995, vol. 778.

- [44] E. R. Colson, N. L. Geller, T. Heeren, and M. J. Corwin, “Factors associated with choice of infant sleep position”, *Pediatrics*, vol. 140, no. 3, 2017.
- [45] Council and parliament of European Union, *Council regulation (EU) no 679/2016 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)*.
<https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN>, 2016.
- [46] E. Davcheva, “Classifying mental health conditions via symptom identification: A novel deep learning approach.”, in *ICIS*, 2019.
- [47] H. Delgado, X. Anguera, C. Fredouille, and J. Serrano, “Novel clustering selection criterion for fast binary key speaker diarization”, in *Interspeech 2015*, 2015. DOI: [10.13140/RG.2.1.3073.2886](https://doi.org/10.13140/RG.2.1.3073.2886).
- [48] A. Dhillon and A. Singh, “Machine learning in healthcare data analysis: A survey”, *Journal of Biology and Today’s World*, vol. 8, no. 6, pp. 1–10, 2019.
- [49] C. Ding and X. He, “Cluster merging and splitting in hierarchical clustering algorithms”, in *2002 IEEE International Conference on Data Mining, 2002. Proceedings.*, IEEE, 2002, pp. 139–146.
- [50] A. Ebbelohj, M. Ø. Thunbo, O. E. Andersen, M. V. Glindtvad, and A. Hulman, “Transfer learning for non-image data in clinical research: A scoping review”, *PLOS Digital Health*, vol. 1, no. 2, e0000014, 2022. DOI: [10.1371/journal.pdig.0000014](https://doi.org/10.1371/journal.pdig.0000014).
- [51] F. Emmert-Streib and M. Dehmer, “A machine learning perspective on personalized medicine: An automated, comprehensive knowledge base with ontology for pattern recognition”, *Machine Learning and Knowledge Extraction*, vol. 1, no. 1, pp. 149–156, 2019.
- [52] M. Enayati, M. Skubic, J. M. Keller, M. Popescu, and N. Z. Farahani, “Sleep posture classification using bed sensor data and neural networks”, in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2018, pp. 461–465.
- [53] O. Faust, A. Shenfield, M. Kareem, T. R. San, H. Fujita, and U. R. Acharya, “Automated detection of atrial fibrillation using long short-term memory network with RR interval signals”, *Computers in Biology and Medicine*, vol. 102, no. July, pp. 327–335, 2018. DOI: [10.1016/j.compbiomed.2018.07.001](https://doi.org/10.1016/j.compbiomed.2018.07.001).
- [54] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, “Transfer learning for time series classification”, in *2018 IEEE international conference on big data (Big Data)*, IEEE, 2018, pp. 1367–1376.
- [55] M. Ferdous, J. Debnath, and N. R. Chakraborty, “Machine Learning Algorithms in Healthcare: A Literature Survey”, *2020 11th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2020*, 2020. DOI: [10.1109/ICCCNT49239.2020.9225642](https://doi.org/10.1109/ICCCNT49239.2020.9225642).
- [56] L. Ferini-Strambi, A. Galbiati, and R. Combi, “Sleep disorder-related headaches”, *Neurological Sciences*, vol. 40, pp. 107–113, 2019.
- [57] M. Feurer, J. Springenberg, and F. Hutter, “Initializing bayesian hyperparameter optimization via meta-learning”, in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, 2015.

- [58] P. Fonseca, N. Den Teuling, X. Long, and R. M. Aarts, “Cardiorespiratory sleep stage detection using conditional random fields”, *IEEE journal of biomedical and health informatics*, vol. 21, no. 4, pp. 956–966, 2016.
- [59] P. Fonseca, M. M. van Gilst, M. Radha, *et al.*, “Automatic sleep staging using heart rate variability, body movements, and recurrent neural networks in a sleep disordered population”, *Sleep*, vol. 43, no. 9, zsaa048, 2020.
- [60] F. Forooghifar, A. Aminifar, L. Cammoun, *et al.*, “A self-aware epilepsy monitoring system for real-time epileptic seizure detection”, *Mobile Networks and Applications*, pp. 1–14, 2022.
- [61] A. Gaddam, T. Wilkin, M. Angelova, and J. Gaddam, “Detecting sensor faults, anomalies and outliers in the internet of things: A survey on the challenges and solutions”, *Electronics (Switzerland)*, vol. 9, no. 3, pp. 1–15, 2020. DOI: [10.3390/electronics9030511](https://doi.org/10.3390/electronics9030511).
- [62] A. Gago-Veiga, A. Gonzalez-Martinez, J. Galvez-Goicuria, *et al.*, “¿es posible predecir una crisis de migraña? estudio prospectivo para valorar la capacidad de predicción en tiempo real, a través de variables hemodinámicas, mediante un dispositivo portátil de muñeca.”, in *Libro de abstracts, Reunión anual asociación madrileña de neurología.*, 2022.
- [63] A. Gago-Veiga, A. Gonzalez-Martinez, J. Pagan, *et al.*, “¿cómo es la curva de dolor en una crisis de migraña?”, 2022, Poster presented at Reunión anual de la AMN, Madrid, Spain.
- [64] A. B. Gago-Veiga, J. Pagán, K. Henares, *et al.*, “To what extent are patients with migraine able to predict attacks?”, *Journal of pain research*, pp. 2083–2094, 2018.
- [65] J. Galvez-Goicuria, J. Pagan, L. Perez, J. Catalina-Gomez, J. M. Moya, and J. L. Ayala, “Reproducible and accurate subject-wise sleep posture detection by detecting and removing turns”, in *2022 IEEE International Conference on Omni-layer Intelligent Systems (COINS)*, IEEE, 2022, pp. 1–6.
- [66] J. Gálvez-Goicuría, J. Pagán, A. B. Gago-Veiga, J. M. Moya, and J. L. Ayala, “Cluster-then-classify methodology for the identification of pain episodes in chronic diseases”, *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 5, pp. 2339–2350, 2021.
- [67] R. X. Gao, R. Yan, R. X. Gao, and R. Yan, “Continuous wavelet transform”, *Wavelets: Theory and Applications for Manufacturing*, pp. 33–48, 2011.
- [68] B. García Heras, *Diseño e implementación de técnicas de aprendizaje automático para el estudio de calidad de sueño en pacientes con migraña*. 2023.
- [69] M. Garrido Cumbreña, D. Gálvez Ruiz, O. Braçe, M. I. Nieblas Rosado, and C. J. Delgado Domínguez, *Impacto y situación de la Migraña en España: Atlas 2018*. Editorial Universidad de Sevilla, 2018.
- [70] L. Gasser, K. Lakkaraju, S. Ray, and S. Swarup, “Darpa baa 05-29: Transfer learning issues and potential contributions”, *University of Illinois at Urbana-Champaign*, 2005.
- [71] M. Gil-Martín, J. M. Montero, and R. San-Segundo, “Parkinson’s disease detection from drawing movements using convolutional neural networks”, *Electronics*, vol. 8, no. 8, p. 907, 2019.
- [72] L. D. Goldberg, “The cost of migraine and its treatment.”, *The American journal of managed care*, vol. 11, no. 2 Suppl, S62–7, 2005.

- [73] A. Gonzalez-Martinez, J. Galvez-Goicuria, J. Pagan, *et al.*, “Classification of the pain curve and semiology of a migraine crisis”, in *Journal Of Headache And Pain*, BMC CAMPUS, 4 CRINAN ST, LONDON N1 9XW, ENGLAND, vol. 23, 2022.
- [74] A. Gonzalez-Martinez, J. Galvez-Goicuria, J. Pagan, *et al.*, “Predicción de una crisis de migraña utilizando variables hemodinámicas registradas en tiempo real mediante dispositivos portátiles de muñeca.”, 2022.
- [75] A. Gonzalez-Martinez, J. Galvez-Goicuria, J. Pagan, *et al.*, “Prediction of a migraine crisis in real-time using a wearable device”, in *Journal Of Headache And Pain*, BMC CAMPUS, 4 CRINAN ST, LONDON N1 9XW, ENGLAND, vol. 23, 2022.
- [76] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi, “Cvxeda: A convex optimization approach to electrodermal activity processing”, *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 797–804, 2015.
- [77] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, *Feature extraction: foundations and applications*. Springer, 2008, vol. 207.
- [78] M. A. Hall, “Correlation-based feature subset selection for machine learning”, *Thesis submitted in partial fulfillment of the requirements of the degree of Doctor of Philosophy at the University of Waikato, Department of Computer Science*, 1998.
- [79] J. Hannink, T. Kautz, C. F. Pasluosta, K.-G. Gaßmann, J. Klucken, and B. M. Eskofier, “Sensor-based gait parameter extraction with deep convolutional neural networks”, *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 85–93, 2016.
- [80] A. R. Hassan and A. Subasi, “A decision support system for automated identification of sleep stages from single-channel eeg signals”, *Knowledge-Based Systems*, vol. 128, pp. 115–124, 2017.
- [81] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, “Neural collaborative filtering”, in *Proceedings of the 26th international conference on world wide web*, 2017, pp. 173–182.
- [82] A. Hemingway and W. M. Price, “The autonomic nervous system and regulation of body temperature”, *The Journal of the American Society of Anesthesiologists*, vol. 29, no. 4, pp. 693–701, 1968.
- [83] A. Herlan, J. Ottenbacher, J. Schneider, D. Riemann, and B. Feige, “Electrodermal activity patterns in sleep stages and their utility for sleep versus wake classification”, *Journal of sleep research*, vol. 28, no. 2, e12694, 2019.
- [84] K. K. Holsteen, M. Hittle, M. Barad, and L. M. Nelson, “Development and internal validation of a multivariable prediction model for individual episodic migraine attacks based on daily trigger exposures”, *Headache: The Journal of Head and Face Pain*, vol. 60, no. 10, pp. 2364–2379, 2020.
- [85] A. Holzinger, M. Dehmer, and I. Jurisica, “Knowledge discovery and interactive data mining in bioinformatics-state-of-the-art, future challenges and research directions”, *BMC bioinformatics*, vol. 15, no. 6, pp. 1–9, 2014.
- [86] L. Hood and M. Flores, “A personal view on systems medicine and the emergence of proactive p4 medicine: Predictive, preventive, personalized and participatory”, *New biotechnology*, vol. 29, no. 6, pp. 613–624, 2012.
- [87] R. I. Horwitz, M. R. Cullen, J. Abell, and J. B. Christian, “(de) personalized medicine”, *Science*, vol. 339, no. 6124, pp. 1155–1156, 2013.

- [88] R.-S. Hsiao, T.-X. Chen, M. A. Bitew, C.-H. Kao, and T.-Y. Li, “Sleeping posture recognition using fuzzy c-means algorithm”, *Biomedical engineering online*, vol. 17, no. 2, pp. 1–19, 2018.
- [89] H.-L. Huttunen and R. Halonen, “Preferred biosignals to predict migraine attack”, in *International Conference on Well-Being in the Information Society*, Springer, 2018, pp. 200–210.
- [90] C. Iber, “The aasm manual for the scoring of sleep and associated events: Rules”, *Terminology and Technical Specification*, 2007.
- [91] IHS, *Headache Classification Committee of the International Headache Society (IHS) The International Classification of Headache Disorders, 3rd edition*, en. England, Jan. 2018, vol. 38, pp. 1–211.
- [92] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, “Real-time motor fault detection by 1-d convolutional neural networks”, *IEEE Transactions on Industrial Electronics*, vol. 63, no. 11, pp. 7067–7075, 2016.
- [93] L. A. Irish, C. E. Kline, H. E. Gunn, D. J. Buysse, and M. H. Hall, “The role of sleep hygiene in promoting public health: A review of empirical evidence”, *Sleep medicine reviews*, vol. 22, pp. 23–36, 2015.
- [94] Y. Itasaka, S. Miyazaki, K. Ishikawa, and K. Togawa, “The influence of sleep position and obesity on sleep apnea”, *Psychiatry and clinical neurosciences*, vol. 54, no. 3, pp. 340–341, 2000.
- [95] P. Jadhav, G. Rajguru, D. Datta, and S. Mukhopadhyay, “Automatic sleep stage classification using time–frequency images of cwt and transfer learning using convolution neural network”, *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 494–504, 2020.
- [96] P.-Y. Jeng, L.-C. Wang, C.-J. Hu, and D. Wu, “A wrist sensor sleep posture monitoring system: An automatic labeling approach”, *Sensors*, vol. 21, no. 1, 2021. DOI: [10.3390/s21010258](https://doi.org/10.3390/s21010258).
- [97] D. Johnson, S. Deterding, K.-A. Kuhn, A. Staneva, S. Stoyanov, and L. Hides, “Gamification for health and wellbeing: A systematic review of the literature”, *Internet interventions*, vol. 6, pp. 89–106, 2016.
- [98] C. Kappeler-Setz, F. Gravenhorst, J. Schumm, B. Arnrich, and G. Tröster, “Towards long term monitoring of electrodermal activity in daily life”, *Personal and Ubiquitous Computing*, vol. 17, no. 2, pp. 261–271, 2013. DOI: [10.1007/s00779-011-0463-4](https://doi.org/10.1007/s00779-011-0463-4).
- [99] N. Karsan and P. J. Goadsby, “Intervening in the premonitory phase to prevent migraine: Prospects for pharmacotherapy”, *CNS drugs*, pp. 1–14, 2024.
- [100] L. Kelman, “Pain characteristics of the acute migraine attack”, *Headache: The Journal of Head and Face Pain*, vol. 46, no. 6, pp. 942–953, 2006.
- [101] M. F. Kilkenny and K. M. Robinson, *Data quality: “garbage in–garbage out”*, 2018.
- [102] I. Kononenko, “Machine learning for medical diagnosis: History, state of the art and perspective”, *Artificial Intelligence in Medicine*, vol. 23, no. 1, pp. 89–109, 2001. DOI: [10.1016/S0933-3657\(01\)00077-X](https://doi.org/10.1016/S0933-3657(01)00077-X).
- [103] H. Koskimäki, H. Mönttinen, P. Siirtola, H.-L. Huttunen, R. Halonen, and J. Rönning, “Early detection of migraine attacks based on wearable sensors: Experiences of data collection using empatica e4”, in *Proceedings of the 2017 ACM International Joint*

- Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, 2017, pp. 506–511.
- [104] A. Krishnan, R. Zhang, V. Yao, *et al.*, “Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder”, *Nature neuroscience*, vol. 19, no. 11, pp. 1454–1462, 2016.
 - [105] L. Kuula, A.-K. Pesonen, *et al.*, “Heart rate variability and firstbeat method for detecting sleep stages in healthy young adults: Feasibility study”, *JMIR mHealth and uHealth*, vol. 9, no. 2, e24704, 2021.
 - [106] W. Kwong and D. Pathak, “Validation of the eleven-point pain scale in the measurement of migraine headache pain”, *Cephalalgia*, vol. 27, no. 4, pp. 336–342, 2007.
 - [107] N. Landwehr, M. Hall, and E. Frank, “Logistic model trees”, *Machine Learning*, vol. 95, no. 1-2, pp. 161–205, 2005.
 - [108] M. Långkvist, L. Karlsson, and A. Loutfi, “Sleep Stage Classification Using Unsupervised Feature Learning”, *Advances in Artificial Neural Systems*, vol. 2012, pp. 1–9, 2012. DOI: [10.1155/2012/107046](https://doi.org/10.1155/2012/107046).
 - [109] M. Lantéri-Minet, G. Duru, M. Mudge, and S. Cottrell, “Quality of life impairment, disability and economic burden associated with chronic daily headache, focusing on chronic migraine with or without medication overuse: A systematic review”, *Cephalalgia*, vol. 31, no. 7, pp. 837–850, 2011.
 - [110] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning”, *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
 - [111] J. B. Lee, Y. H. Park, J. H. Hong, *et al.*, “Determining optimal sleep position in patients with positional sleep-disordered breathing using response surface analysis”, *Journal of sleep research*, vol. 18, no. 1, pp. 26–35, 2009.
 - [112] R. S. Lee, *Artificial intelligence in daily life*. Springer, 2020.
 - [113] W.-H. Lee and M.-S. Ko, “Effect of sleep posture on neck muscle activity”, *Journal of physical therapy science*, vol. 29, no. 6, pp. 1021–1024, 2017.
 - [114] S. P. Leighton, R. Krishnadas, K. Chung, *et al.*, “Predicting one-year outcome in first episode psychosis using machine learning”, *Plos one*, vol. 14, no. 3, e0212846, 2019.
 - [115] A. Leino, H. Korkalainen, L. Kalevo, *et al.*, “Deep learning enables accurate automatic sleep staging based on ambulatory forehead eeg”, *IEEE Access*, vol. 10, pp. 26 554–26 566, 2022.
 - [116] J. Li and P. Carayon, “Health care 4.0: A vision for smart and connected health care”, *IIEE transactions on healthcare systems engineering*, vol. 11, no. 3, pp. 171–180, 2021.
 - [117] Q. Li, Q. Li, A. S. Cakmak, *et al.*, “Transfer learning from ecg to ppg for improved sleep staging from wrist-worn wearables”, *Physiological Measurement*, vol. 42, no. 4, p. 044 004, 2021.
 - [118] Y.-Y. Li, Y.-J. Lei, L. C.-L. Chen, and Y.-P. Hung, “Sleep posture classification with multi-stream cnn using vertical distance map”, in *2018 International Workshop on Advanced Image Technology (IWAIT)*, IEEE, 2018, pp. 1–4.
 - [119] C. Liden, M. Wolowicz, and J. Stivoric, “Characterization and implications of the sensors incorporated into the SenseWear armband for energy expenditure and activity detection”, *Bodymedia Inc.*, pp. 1–7, 2002.
 - [120] R. O. Lima, “Autonomous nervous system biosignal processing via eda and hrv from a wearable device”, Ph.D. dissertation, Universidade Nova, 2018.

- [121] Y.-K. Lin, G.-Y. Lin, J.-T. Lee, *et al.*, “Associations between sleep quality and migraine frequency: A cross-sectional case-control study”, *Medicine*, vol. 95, no. 17, 2016.
- [122] W. C. Lin and C. F. Tsai, “Missing value imputation: a review and analysis of the literature (2006–2017)”, *Artificial Intelligence Review*, vol. 53, no. 2, pp. 1487–1509, 2020. DOI: [10.1007/s10462-019-09709-4](https://doi.org/10.1007/s10462-019-09709-4).
- [123] Y.-P. Lin and T.-P. Jung, “Improving eeg-based emotion classification using conditional transfer learning”, *Frontiers in human neuroscience*, vol. 11, p. 334, 2017.
- [124] M. Linde, A. Mellberg, and C. Dahlöf, “The natural course of migraine attacks. a prospective analysis of untreated attacks compared with attacks treated with a triptan”, *Cephalalgia*, vol. 26, no. 6, pp. 712–721, 2006.
- [125] R. B. Lipton, W. F. Stewart, and A. I. Scher, “Epidemiology and economic impact of migraine”, *Current medical research and opinion*, vol. 17, no. sup1, s4–12, 2001.
- [126] R. Long, Y. Zhu, and S. Zhou, “Therapeutic role of melatonin in migraine prophylaxis: A systematic review”, *Medicine*, vol. 98, no. 3, 2019.
- [127] X. Lun, Z. Yu, T. Chen, F. Wang, and Y. Hou, “A simplified cnn classification method for mi-eeg via the electrode pairs signals”, *Frontiers in Human Neuroscience*, vol. 14, p. 338, 2020.
- [128] N. L. Lund, A. H. Snoer, P. J. Jennum, R. H. Jensen, and M. C. J. Barloese, “Sleep in cluster headache revisited: Results from a controlled actigraphic study”, *Cephalalgia*, vol. 39, no. 6, pp. 742–749, 2019.
- [129] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions”, *Advances in neural information processing systems*, vol. 30, 2017.
- [130] O. Maimon and L. Rokach, “Introduction to knowledge discovery and data mining”, in *Data mining and knowledge discovery handbook*, Springer, 2009, pp. 1–15.
- [131] J. A. Marcum, “P7 medicine: Humanizing systems medicine”, *Teorema*, vol. 40, no. 1, pp. 215–231, 2021.
- [132] A. B. Martin, M. Hartman, D. Lassman, A. Catlin, and N. H. E. A. Team, “National health care spending in 2019: Steady growth for the fourth consecutive year: Study examines national health care spending for 2019.”, *Health Affairs*, vol. 40, no. 1, pp. 14–24, 2021.
- [133] Á. Martínez Petit, *Diseño e implementación de un sistema de agrupamiento de pacientes y algoritmo de recomendación de indicaciones clínicas para enfermedades crónicas*, 2021.
- [134] A. Martínez-Fernández, M. Rueda Vega, S. Quintas, *et al.*, “Psychosocial repercussion of migraine: Is it a stigmatized disease?”, *Neurological Sciences*, vol. 41, no. 8, pp. 2207–2213, 2020.
- [135] N. Maslej, L. Fattorini, E. Brynjolfsson, *et al.*, “The ai index 2023 annual report”, *arXiv preprint arXiv:2310.03715*, Apr. 2023.
- [136] H. Masoumi, A. Behrad, M. A. Pourmina, and A. Roosta, “Automatic liver segmentation in mri images using an iterative watershed algorithm and artificial neural network”, *Biomedical signal processing and control*, vol. 7, no. 5, pp. 429–437, 2012.
- [137] N. T. Mathew, “Chronic refractory headache.”, *Neurology*, vol. 43, no. 6 Suppl 3, S26, 1993.
- [138] C. McCarthy, N. Pradhan, C. Redpath, and A. Adler, “Validation of the Empatica E4 wristband”, *2016 IEEE EMBS International Student Conference: Expanding the*

- Boundaries of Biomedical Engineering and Healthcare, ISC 2016 - Proceedings*, pp. 4–7, 2016. DOI: [10.1109/EMBSISC.2016.7508621](https://doi.org/10.1109/EMBSISC.2016.7508621).
- [139] J. McCarthy, “What is artificial intelligence”, URL: <http://www-formal.stanford.edu/jmc/whatisai.html>, 2004.
 - [140] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-efficient learning of deep networks from decentralized data”, in *Artificial intelligence and statistics*, PMLR, 2017, pp. 1273–1282.
 - [141] H. B. McMahan, E. Moore, D. Ramage, and B. A. y Arcas, “Federated learning of deep networks using model averaging”, *CoRR*, vol. abs/1602.05629, 2016. arXiv: [1602.05629](https://arxiv.org/abs/1602.05629).
 - [142] T. M. Mitchell and T. M. Mitchell, *Machine learning*. McGraw-hill New York, 1997.
 - [143] M. Muehlhan, M. Marxen, J. Landsiedel, H. Malberg, and S. Zaunseder, “The effect of body posture on cognitive performance: A question of sleep quality”, *Frontiers in human neuroscience*, vol. 8, p. 171, 2014.
 - [144] B. Muthu, C. Sivaparthipan, G. Manogaran, *et al.*, “Iot based wearable sensor for diseases prediction and symptom analysis in healthcare sector”, *Peer-to-peer networking and applications*, vol. 13, no. 6, pp. 2123–2134, 2020.
 - [145] M. Nasser, T. Pal Attia, B. Joseph, *et al.*, “Ambulatory seizure forecasting with a wrist-worn device using long-short term memory deep learning”, *Scientific reports*, vol. 11, no. 1, p. 21 935, 2021.
 - [146] A. Nguyen, R. Alqurashi, Z. Raghebi, *et al.*, “In-ear biosignal recording system: A wearable for automatic whole-night sleep staging”, *WearSys 2016 - Proceedings of the 2016 Workshop on Wearable Systems and Applications, co-located with MobiSys 2016*, pp. 19–24, 2016. DOI: [10.1145/2935643.2935649](https://doi.org/10.1145/2935643.2935649).
 - [147] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. V. Poor, “Federated learning for internet of things: A comprehensive survey”, *IEEE Communications Surveys & Tutorials*, vol. 23, no. 3, pp. 1622–1658, 2021.
 - [148] J. Olesen, “International classification of headache disorders”, *The Lancet Neurology*, vol. 17, no. 5, pp. 396–397, 2018.
 - [149] P. M. Orr, B. C. Shank, and A. C. Black, “The role of pain classification systems in pain management”, *Crit Care Nurs Clin North Am*, vol. 29, no. 4, pp. 407–18, 2017.
 - [150] S. Ostadabbas, M. B. Pouyan, M. Nourani, and N. Kehtarnavaz, “In-bed posture classification and limb identification”, in *2014 IEEE Biomedical Circuits and Systems Conference (BioCAS) Proceedings*, IEEE, 2014, pp. 133–136.
 - [151] B. Osumili, P. McCrone, S. Cousins, and L. Ridsdale, “The economic cost of patients with migraine headache referred to specialist clinics”, *Headache: The Journal of Head and Face Pain*, vol. 58, no. 2, pp. 287–294, 2018.
 - [152] J. Pagan, J. Fernandez-Blasco, J. Ayala, and J. García-Donas Jiménez, *Método y sistema para la detección precoz de episodios de dolor oncológico*. https://consultas2.oepm.es/pdf/ES/0000/000/02/90/54/ES-2905474_A1.pdf, 2022.
 - [153] J. Pagán, M. Irene De Orbe, A. Gago, *et al.*, “Robust and accurate modeling approaches for migraine per-patient prediction from ambulatory data”, *Sensors*, vol. 15, no. 7, pp. 15 419–15 442, 2015.
 - [154] J. Pagán, J. M. Moya, J. L. Risco-Martín, and J. L. Ayala, “Advanced migraine prediction simulation system”, in *Proceedings of the Summer Simulation Multi-Conference*,

- ser. SummerSim '17, Bellevue, Washington: Society for Computer Simulation International, 2017.
- [155] J. Pagán, J. L. Risco-Martín, J. M. Moya, and J. L. Ayala, “Modeling methodology for the accurate and prompt prediction of symptomatic events in chronic diseases”, *Journal of Biomedical Informatics*, vol. 62, pp. 136–147, 2016. DOI: <https://doi.org/10.1016/j.jbi.2016.05.008>.
 - [156] J. Pagán Ortiz, “Robust modeling for information acquisition in biophysical and critical scenarios”, Ph.D. dissertation, Universidad Complutense de Madrid, Departamento de Arquitectura de Computadores y Automática, 2018.
 - [157] M. Palacios-Ceña, K. Wang, M. Castaldo, *et al.*, “Variables associated with the use of prophylactic amitriptyline treatment in patients with tension-type headache”, *The Clinical Journal of Pain*, vol. 35, no. 4, pp. 315–320, 2019.
 - [158] S. J. Pan and Q. Yang, “A survey on transfer learning”, *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.
 - [159] O. Pavliuk, M. Mishchuk, and C. Strauss, “Transfer learning approach for human activity recognition based on continuous wavelet transform”, *Algorithms*, vol. 16, no. 2, p. 77, 2023.
 - [160] H. Phan, K. Mikkelsen, O. Y. Chén, *et al.*, “Personalized automatic sleep staging with single-night data: A pilot study with Kullback-Leibler divergence regularization”, *Physiological Measurement*, vol. 41, no. 6, 2020. DOI: [10.1088/1361-6579/ab921e](https://doi.org/10.1088/1361-6579/ab921e).
 - [161] L. Piwek, D. A. Ellis, S. Andrews, and A. Joinson, “The rise of consumer health wearables: Promises and barriers”, *PLoS medicine*, vol. 13, no. 2, e1001953, 2016.
 - [162] J. Plazak and M. Kersten-Oertel, “A survey on the affordances of “hearables””, *Inventions*, vol. 3, no. 3, pp. 1–16, 2018. DOI: [10.3390/inventions3030048](https://doi.org/10.3390/inventions3030048).
 - [163] P. Polysomnography, “A sleep analysis method based on heart rate variability”, 2019.
 - [164] H. F. Posada-Quintero and K. H. Chon, “Innovations in electrodermal activity data collection and signal processing: A systematic review”, *Sensors (Switzerland)*, vol. 20, no. 2, 2020. DOI: [10.3390/s20020479](https://doi.org/10.3390/s20020479).
 - [165] M. Radha, P. Fonseca, A. Moreau, *et al.*, “Sleep stage classification from heart-rate variability using long short-term memory neural networks”, *Scientific reports*, vol. 9, no. 1, p. 14149, 2019.
 - [166] M. Radha, P. Fonseca, A. Moreau, *et al.*, “A deep transfer learning approach for wearable sleep stage classification with photoplethysmography”, *NPJ digital medicine*, vol. 4, no. 1, p. 135, 2021.
 - [167] R. Rao, S. Choubey, G. Gutjahr, and P. Nedungadi, “Document recommendation for medical training using learning to rank”, in *Machine Learning for Predictive Analysis: Proceedings of ICTIS 2020*, Springer, 2021, pp. 353–362.
 - [168] N. B. Rappaport, D. P. McAnulty, C. D. Waggoner, and P. J. Brantley, “Cluster analysis of minnesota multiphasic personality inventory (mmpi) profiles in a chronic headache population”, *Journal of behavioral medicine*, vol. 10, no. 1, pp. 49–60, 1987.
 - [169] D. Ravi, C. Wong, F. Deligianni, *et al.*, “Deep Learning for Health Informatics”, *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 4–21, 2017. DOI: [10.1109/JBHI.2016.2636665](https://doi.org/10.1109/JBHI.2016.2636665).

- [170] P. P. Rebouças Filho, R. M. Sarmiento, G. B. Holanda, and D. de Alencar Lima, “New approach to detect and classify stroke in skull ct images via analysis of brain tissue densities”, *Computer methods and programs in biomedicine*, vol. 148, pp. 27–43, 2017.
- [171] M. Rijken, V. Struckmann, M. Dyakova, M. G. Melchiorre, S. Rissanen, E. van Ginneken, *et al.*, “Icare4eu: Improving care for people with multiple chronic conditions in europe”, *Eurohealth*, vol. 19, no. 3, pp. 29–31, 2013.
- [172] D. M. Roberts, M. M. Schade, G. M. Mathew, D. Gartenberg, and O. M. Buxton, “Detecting sleep using heart rate and motion data from multisensor consumer-grade wearables, relative to wrist actigraphy and polysomnography”, *Sleep*, vol. 43, no. 7, zsaa045, 2020.
- [173] W. L. Romine, N. L. Schroeder, J. Graft, *et al.*, “Using machine learning to train a wearable device for measuring students’ cognitive load during problem-solving activities based on electrodermal activity, body temperature, and heart rate: Development of a cognitive load tracker for both personal and classroom use”, *Sensors (Switzerland)*, vol. 20, no. 17, pp. 1–14, 2020. DOI: [10.3390/s20174833](https://doi.org/10.3390/s20174833).
- [174] S. Saeb, L. Lonini, A. Jayaraman, D. C. Mohr, and K. P. Kording, “The need to approximate the use-case in clinical machine learning”, *Gigascience*, vol. 6, no. 5, gix019, 2017.
- [175] S. Sami, M. Seppänen, and A. Kuusela, “Artefact correction for heart beat interval data”, in *1st Advanced Methods for Processing Bioelectrical Signals Meeting: University of Jyväskylä*, 2004.
- [176] R. San-Segundo, M. Gil-Martín, L. F. D’Haro-Enríquez, and J. M. Pardo, “Classification of epileptic eeg recordings using signal transforms and convolutional neural networks”, *Computers in biology and medicine*, vol. 109, pp. 148–158, 2019.
- [177] A. Sano, W. Chen, D. Lopez-Martinez, S. Taylor, and R. W. Picard, “Multimodal ambulatory sleep detection using lstm recurrent neural networks”, *IEEE journal of biomedical and health informatics*, vol. 23, no. 4, pp. 1607–1617, 2018.
- [178] A. Sathyanarayana, S. Joty, L. Fernandez-Luque, *et al.*, “Sleep quality prediction from wearable data using deep learning”, *JMIR mHealth and uHealth*, vol. 4, no. 4, pp. 1–13, 2016. DOI: [10.2196/mhealth.6562](https://doi.org/10.2196/mhealth.6562).
- [179] R. Savage, M. Messenger, R. D. Neal, *et al.*, “Development and validation of multivariable machine learning algorithms to predict risk of cancer in symptomatic patients referred urgently from primary care: A diagnostic accuracy study”, *BMJ open*, vol. 12, no. 4, e053590, 2022.
- [180] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, “Introducing wesad, a multimodal dataset for wearable stress and affect detection”, in *Proceedings of the 20th ACM international conference on multimodal interaction*, 2018, pp. 400–408.
- [181] A. A. Schuurmans, P. De Looft, K. S. Nijhof, *et al.*, “Validity of the empatica e4 wristband to measure heart rate variability (hrv) parameters: A comparison to electrocardiography (ecg)”, *Journal of medical systems*, vol. 44, pp. 1–11, 2020.
- [182] T. J. Schwedt, M. S. Matharu, and D. W. Dodick, “Thunderclap headache”, *The Lancet Neurology*, vol. 5, no. 7, pp. 621–631, 2006.
- [183] G. Selby, *Migraine and its variants*. Williams & Wilkins, 1983.
- [184] F. Shaheen, B. Verma, and M. Asafuddoula, “Impact of automatic feature extraction in deep learning architecture”, in *2016 International Conference on Digital Image*

- Computing: Techniques and Applications (DICTA)*, 2016, pp. 1–8. DOI: [10.1109/DICTA.2016.7797053](https://doi.org/10.1109/DICTA.2016.7797053).
- [185] S. Shao, S. McAleer, R. Yan, and P. Baldi, “Highly accurate machine fault diagnosis using deep transfer learning”, *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2446–2455, 2018.
 - [186] D. Sharma, G. Singh Aujla, and R. Bajaj, “Evolution from ancient medication to human-centered healthcare 4.0: A review on health care recommender systems”, *International Journal of Communication Systems*, vol. 36, no. 12, e4058, 2023.
 - [187] B. Shekarraz, H.-F. Lu, and M. L. Stoller, “Correlation of unilateral urolithiasis with sleep posture”, *The Journal of Urology*, vol. 165, no. 4, pp. 1085–1087, 2001.
 - [188] P. Siirtola, H. Koskimäki, H. Mönttinen, and J. Rönning, “Using sleep time data from wearable sensors for early detection of migraine attacks”, *Sensors*, vol. 18, no. 5, p. 1374, 2018.
 - [189] S. Singaram, K. Ramakrishnan, and S. Periyasamy, “Electrodermal signal analysis using continuous wavelet transform as a tool for quantification of sweat gland activity in diabetic kidney disease”, *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 237, no. 8, pp. 919–927, 2023.
 - [190] S. Sinha, P. S. Routh, P. D. Anno, and J. P. Castagna, “Spectral decomposition of seismic data with continuous-wavelet transform”, *Geophysics*, vol. 70, no. 6, P19–P25, 2005.
 - [191] M. Smuck, C. A. Odonkor, J. K. Wilt, N. Schmidt, and M. A. Swiernik, “The emerging clinical role of wearables: Factors for successful implementation in healthcare”, *NPJ Digital Medicine*, vol. 4, no. 1, pp. 1–8, 2021.
 - [192] Statista, *Statista: Wearable device unit sales worldwide by region from 2015 to 2021 (in millions)*. <https://www.statista.com/statistics/490231/wearable-devices-worldwide-by-region/>, 2022.
 - [193] E. J. Stepanski and J. K. Wyatt, “Use of sleep hygiene in the treatment of insomnia”, *Sleep medicine reviews*, vol. 7, no. 3, pp. 215–225, 2003.
 - [194] K. Stokes, R. Castaldo, M. Franzese, *et al.*, “A machine learning model for supporting symptom-based referral and diagnosis of bronchitis and pneumonia in limited resource settings”, *Biocybernetics and Biomedical Engineering*, vol. 41, no. 4, pp. 1288–1302, 2021. DOI: <https://doi.org/10.1016/j.bbe.2021.09.002>.
 - [195] L. J. Stovner and C. Andrée, “Impact of headache in europe: A review for the eurolight project”, *The journal of headache and pain*, vol. 9, no. 3, pp. 139–146, 2008.
 - [196] M. Subramanian, N. P. Lv, and S. VE, “Hyperparameter optimization for transfer learning of vgg16 for disease identification in corn leaves using bayesian optimization”, *Big Data*, vol. 10, no. 3, pp. 215–229, 2022.
 - [197] K. Sundararajan, S. Georgievska, B. H. Te Lindert, *et al.*, “Sleep classification from wrist-worn accelerometer data using random forests”, *Scientific Reports*, vol. 11, no. 1, pp. 1–10, 2021.
 - [198] Y. R. Tabar, K. B. Mikkelsen, N. Shenton, *et al.*, “At-home sleep monitoring using generic ear-eeg”, *Frontiers in Neuroscience*, vol. 17, p. 987578, 2023.
 - [199] M. Tan and Q. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks”, in *International conference on machine learning*, PMLR, 2019, pp. 6105–6114.

- [200] S. Taylor, N. Jaques, W. Chen, S. Fedor, A. Sano, and R. Picard, “Automatic identification of artifacts in electrodermal activity data”, in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2015, pp. 1934–1937.
- [201] S. Taylor, N. Jaques, W. Chen, S. Fedor, A. Sano, and R. Picard, “Automatic identification of artifacts in electrodermal activity data”, *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2015–November, pp. 1934–1937, 2015. DOI: [10.1109/EMBC.2015.7318762](https://doi.org/10.1109/EMBC.2015.7318762).
- [202] P. Thamarai and K. Adalarasu, “Denoising of eeg, ecg and ppg signals using wavelet transform”, *Journal of Pharmaceutical Sciences and Research*, vol. 10, no. 1, pp. 156–161, 2018.
- [203] R. Thirukovalluru, S. Dixit, R. K. Sevakula, N. K. Verma, and A. Salour, “Generating feature sets for fault diagnosis using denoising stacked auto-encoder”, *2016 IEEE International Conference on Prognostics and Health Management, ICPHM 2016*, 2016. DOI: [10.1109/ICPHM.2016.7542865](https://doi.org/10.1109/ICPHM.2016.7542865).
- [204] V. Toral, A. García, F. J. Romero, *et al.*, “Wearable system for biosignal acquisition and monitoring based on reconfigurable technologies”, *Sensors (Switzerland)*, vol. 19, no. 7, pp. 1–17, 2019. DOI: [10.3390/s19071590](https://doi.org/10.3390/s19071590).
- [205] M. Vailati-Riboni, V. Palombo, and J. J. Loor, “What are omics sciences?”, *Periparturient diseases of dairy cows: a systems biology approach*, pp. 1–7, 2017.
- [206] K. Van Laerhoven, M. Borazio, D. Kilian, and B. Schiele, “Sustained logging and discrimination of sleep postures with low-level, wrist-worn sensors”, in *2008 12th IEEE International Symposium on Wearable Computers*, 2008, pp. 69–76. DOI: [10.1109/ISWC.2008.4911588](https://doi.org/10.1109/ISWC.2008.4911588).
- [207] B. R. Wakerley, “Medication-overuse headache”, *Practical neurology*, vol. 19, no. 5, pp. 399–403, 2019.
- [208] S. Wan, L. Qi, X. Xu, C. Tong, and Z. Gu, “Deep learning models for real-time human activity recognition with smartphones”, *Mobile Networks and Applications*, vol. 25, no. 2, pp. 743–755, 2020.
- [209] Y. Wang, Y. Zhao, T. M. Therneau, *et al.*, “Unsupervised machine learning for the discovery of latent disease clusters and patient subgroups using electronic health records”, *Journal of biomedical informatics*, vol. 102, p. 103364, 2020.
- [210] G. Webb, “Decision tree grafting from the all-tests-but-one partition”, San Francisco, CA: Morgan Kaufmann, 1999, pp. 702–707.
- [211] J. Wiens and E. S. Shenoy, “Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology”, *Clinical Infectious Diseases*, vol. 66, no. 1, pp. 149–153, 2018. DOI: [10.1093/cid/cix731](https://doi.org/10.1093/cid/cix731).
- [212] R. Wix-Ramos, J. Gálvez-Goicuría, M. Verona-Almeida, *et al.*, “Monitoring differences in the function of the autonomic nervous system in patients with chronic insomnia using a wearable device”, *Sleep Medicine*, 2024.
- [213] H. Wolff, “Headache in the migraine syndrome”, *Headache and other head pain*, pp. 255–396, 1948.
- [214] J. Wu, H. Liang, C. Ding, X. Huang, J. Huang, and Q. Peng, “Improving the accuracy in classification of blood pressure from photoplethysmography using continuous wavelet transform and deep learning”, *International journal of hypertension*, vol. 2021, 2021.

- [215] Q. Wu, K. He, and X. Chen, “Personalized federated learning for intelligent iot applications: A cloud-edge based framework”, *IEEE Open Journal of the Computer Society*, vol. 1, pp. 35–44, 2020.
- [216] B. M. Wulterkens, P. Fonseca, L. W. Hermans, *et al.*, “It is all in the wrist: Wearable sleep staging in a clinical population versus reference polysomnography”, *Nature and Science of Sleep*, pp. 885–897, 2021.
- [217] H. Yang, C. Shan, A. F. Kolen, and P. H. de With, “Efficient catheter segmentation in 3d cardiac ultrasound using slice-based fcn with deep supervision and f-score loss”, in *2019 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2019, pp. 260–264.
- [218] Q. Yang, Y. Liu, T. Chen, and Y. Tong, “Federated Machine Learning”, *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, pp. 1–19, 2019. DOI: [10.1145/3298981](https://doi.org/10.1145/3298981).
- [219] W. Z. Yeh, L. Blizzard, and B. V. Taylor, “What is the actual prevalence of migraine?”, *Brain and behavior*, vol. 8, no. 6, e00950, 2018.
- [220] Y. Zang, Y. An, and X. T. Hu, “Automatically recommending healthy living programs to patients with chronic diseases through hybrid content-based and collaborative filtering”, in *2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, 2014, pp. 578–582.
- [221] M. Zanon, M. Mueller, P. Zakharov, *et al.*, “First experiences with a wearable multi-sensor device in a noninvasive continuous glucose monitoring study at home, part ii: The investigators’ view”, *Journal of Diabetes Science and Technology*, vol. 12, no. 3, pp. 554–561, 2018.
- [222] J. Zhang, R. L. Lafta, X. Tao, *et al.*, “Coupling a fast fourier transformation with a machine learning ensemble model to support recommendations for heart disease patients in a telehealth environment”, *IEEE Access*, vol. 5, pp. 10 674–10 685, 2017.
- [223] J. Zhang, E. H. Cain, A. Saha, Z. Zhu, and M. A. Mazurowski, “Breast mass detection in mammography and tomosynthesis via fully convolutional network-based heatmap regression”, in *Medical Imaging 2018: Computer-Aided Diagnosis*, SPIE, vol. 10575, 2018, pp. 474–479.
- [224] Q. Zhang, J. Ding, W. Kong, Y. Liu, Q. Wang, and T. Jiang, “Epilepsy prediction through optimized multidimensional sample entropy and bi-lstm”, *Biomedical Signal Processing and Control*, vol. 64, p. 102 293, 2021.
- [225] Y.-L. Zheng, X.-R. Ding, C. C. Y. Poon, *et al.*, “Unobtrusive sensing and wearable devices for health informatics”, *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 5, pp. 1538–1554, 2014.
- [226] C. Zhou, A. Fu, S. Yu, W. Yang, H. Wang, and Y. Zhang, “Privacy-Preserving Federated Learning in Fog Computing”, *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 10 782–10 793, 2020. DOI: [10.1109/JIOT.2020.2987958](https://doi.org/10.1109/JIOT.2020.2987958).
- [227] Z. Zhu, X. Fan, X. Li, G. Tan, L. Chen, and J. Zhou, “Prevalence and predictive factors for poor sleep quality among migraineurs in a tertiary hospital headache clinic”, *Acta Neurologica Belgica*, vol. 113, pp. 229–235, 2013.
- [228] S. A. Ziaee, A. Moradi, M. Fateh, and S. M. Hosseini Moghaddam, “Sleep posture and unilateral renal stone formation”, *Scandinavian journal of urology and nephrology*, vol. 42, no. 6, pp. 551–554, 2008.

- [229] S. Zwarts and M. Johnson, “The impact of language models and loss functions on repair disfluency detection”, in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 2011, pp. 703–711.

Appendix A: Tables

Table 1: Distribution of Shapley values by type of sensor

Sensor/Shapley value	$x < 0.01$	$0.01 \leq x < 0.04$	$0.04 \leq x$
TEMP	14.10%	64.10%	21.90%
EDA	31.50%	47.80%	20.70%
HR	21.40%	53.60%	25.00%
Stress	18.80%	45.80%	35.40%

Table 2: Distribution of Shapley values by past time window

Window/Shapley value	$x < 0.01$	$0.01 \leq x < 0.04$	$0.04 \leq x$
5'	32.70%	51.00%	16.30%
15'	26.60%	50.00%	23.40%
30'	13.90%	52.10%	34.00%

Table 3: Distribution of Shapley values by type of statistical feature

Stat/Shapley value	$x < 0.01$	$0.01 \leq x < 0.04$	$0.04 \leq x$
Maximum	16.70%	43.80%	39.60%
Minimum	25.00%	51.40%	23.60%
Mean	20.80%	50.00%	29.20%
Mean of the Derivative	37.50%	55.40%	7.10%
Standard Deviation	27.80%	55.60%	16.70%
Mode	18.80%	45.80%	35.40%

Appendix B: Ethics committee agreement

COMITÉ ÉTICO DE INVESTIGACIÓN CLÍNICA

Madrid, 23 de diciembre de 2015

El **Comité Ético de Investigación Clínica del Hospital Universitario de la Princesa** en su reunión del día 21-12-2015 (acta 23/15) después de evaluar la respuesta a las aclaraciones solicitadas del siguiente proyecto de investigación:

TITULO: ¿Existe un perfil de pacientes con mayor capacidad para predecir una crisis de migraña?

Nº de Registro: PI-879

Investigador principal: Ana Gago (Servicio de Neurología)

Decisión tomada: Aprobación (21-12-15)

Este Comité Ético de Investigación Clínica considera que el tanto el **proyecto de investigación como la Hoja de información y consentimiento informado son ética y metodológicamente aceptables**. Asimismo, considera que los investigadores son competentes para llevar a cabo este proyecto que está enmarcado dentro de las líneas de investigación prioritarias del Hospital Universitario de La Princesa.



Fdo: **Dra. Mª de Mar Ortega Gómez**
Secretaria del C.E.I.C.

COMITÉ DE ÉTICA DE LA INVESTIGACIÓN CON MEDICAMENTOS

Madrid, a 14 de Agosto de 2018

El **Comité de Ética de La Investigación con Medicamentos del Hospital Universitario de la Princesa** en su reunión del día **13-08-2018 (Acta CEIm 14/18)** después de evaluar la respuesta a las aclaraciones solicitadas del siguiente proyecto de investigación:

TITULO: Estudio de viabilidad para la predicción en tiempo real de una crisis de migraña mediante dispositivos “WEARABLE”.

TIPO: Proyecto de investigación

CODIGO: BRAINGUARD-MIGRAINE

Nº de Registro: 3513

Investigador Principal: José Vivancos Mora (Servicio de Neurología)

DECISIÓN TOMADA: Aprobación (13-08-18)

Este **Comité de Ética de La Investigación con Medicamentos** considera que tanto el **proyecto de investigación** como la **Hoja de información al paciente y Consentimiento informado para el paciente y para los controles sanos**, son **ética y metodológicamente aceptables**. Asimismo, considera que los investigadores son competentes para llevar a cabo este proyecto que está enmarcado dentro de las líneas de investigación prioritarias del Hospital Universitario de La Princesa.

SECRETARIA DEL CEIm

COMITÉ DE ÉTICA DE LA INVESTIGACIÓN CON MEDICAMENTOS

Madrid, a 28 de enero de 2020

El **Comité de Ética de La Investigación con Medicamentos del Hospital Universitario de la Princesa** en su reunión del día 28 de enero de 2021 tras evaluar la respuesta a las aclaraciones solicitadas del siguiente proyecto de investigación:

CODIGO: N/A

Nº EudraCT: N/A

Nº de Registro: 4327

Investigador Principal: Rybel Wix Ramos (Servicio de Neurofisiología clínica)

Investigadores Colaboradores: Servicio de Neurología

Promotor: N/A

CEIm: Hospital Universitario La Princesa

Centros Participantes: Hospital Universitario La Princesa, Universidad Politécnica de Madrid

DECISION TOMADA: Aprobación (28-01-21, acta CEIm 02/21)

Este **Comité de Ética de La Investigación con Medicamentos** considera que tanto el **proyecto de investigación** como la **Hoja de información al paciente y Consentimiento Informado** son **ética y metodológicamente aceptables**. Asimismo, considera que los investigadores son competentes para llevar a cabo este proyecto que está enmarcado dentro de las líneas de investigación prioritarias del Hospital Universitario de La Princesa.

Fdo.: Dra. M^a del Mar Ortega Gómez
SECRETARIA DEL CEIm