

*Clasificación del miedo usando
computación afectiva, información
fisiológica, y dispositivos portables e
inteligentes para ayudar a combatir
la violencia de género.*

TESIS DOCTORAL PRESENTADA POR
JOSE ÁNGEL MIRANDA CALERO

14. COLECCIÓN:

Premio de la Delegación del Gobierno
contra la Violencia de Género
a Tesis Doctorales sobre Violencia contra la Mujer

Primer premio - Año 2022





© Ministerio de Igualdad
Centro de Publicaciones
C/ Alcalá, 37 - 28071 Madrid

NIPO en línea: 048-23-037-2

Las opiniones expresadas en esta publicación son responsabilidad exclusiva de su autor y su publicación no significa que la Delegación del Gobierno contra la Violencia de Género se identifique con las mismas.

"En esta publicación se ha utilizado papel reciclado libre de cloro de acuerdo con los criterios medioambientales de la contratación pública".

Correo electrónico: dgviolenciagenero@igualdad.gob.es

<http://www.publicacionesoficiales.boe.es>

A mis gatos.

Agradecimientos

En el año 2015 decidí embarcarme de nuevo en el ámbito académico. Antes de hacerlo, la primera llamada que realicé para coger impulso fue a David Puerta, mi amigo, mi maestro, y mi mentor académico. Sin ese empujón, no estaría donde estoy hoy. En su día me enseñaste que el rigor y la pasión pueden ir de la mano. Gracias por darme esa y otras muchas lecciones tan valiosas que me han servido, me sirven y servirán.

Esta tesis no sería posible ni habría llegado a buen puerto sin la supervisión de Celia López y Marta Portela, mis supervisoras. La primera me ha enseñado a ser creativo a la vez que ingeniero, me ha guiado en momentos fáciles y en momentos no tan fáciles, me ha permitido equivocarme, y me ha regalado mucho de su tiempo. Gracias por ser mi mentora científica. La segunda, la cual siento como hermana mayor científica y ejemplo a seguir académicamente, me ha enseñado a no rendirme, me ha dado ánimos en los momentos difíciles, y también me ha regalado mucho de su tiempo. De ambas he aprendido lo mucho que aún me queda por aprender.

Agradecer también a todas las personas integrantes del Departamento de Tecnología Electrónica por su apoyo durante estos años. En especial a los Técnicos de Laboratorio con los que he tenido el placer de coincidir y trabajar. Sin vuestro trabajo y ayuda, estaríamos vendidos.

El equipo UC3M4Safety y el proyecto EMPATIA-CM. Dos cosas que han formado parte de esta tesis y viceversa. Gracias a todas las personas involucradas en ambas, habéis dado sentido a esta trabajo.

Gracias a toda mi familia. A mi padre y a mi madre, soy el reflejo de toda vuestra paciencia, perseverancia y educación que habéis mantenido a lo largo de vuestra vida y nos habéis sabido transmitir a mi hermana y a mi.

Finalmente, muchas gracias a mi pareja, Patricia. Este trabajo también es tuyo. Te ha tocado la peor parte, pero ya hemos llegado. Gracias por ser paciente y por enseñarme a ser paciente. Gracias por estar conmigo y aguantar carros y carretas. Espero poder ser capaz de devolverte algún día, aunque sea una pequeña parte, el tiempo invertido.

Jose Angel Miranda

Marzo 2022

*"iustum et tenacem propositi virum
non civium ardor prava iubentium,
non vultus instantis tyranni
mente quatit solida neque Auster,
dux inquieti turbidus Hadriae,
nec fulminantis magna manus Iovis:
si fractus illabatur orbis,
impavidum ferient ruinae"*

Horacio, Carmina III, 3, 1-8.

Contenido enviado y publicado

- **Artículos en Revistas**

- J. A. Miranda, M. F. Canabal, L. Gutiérrez-Martín, J. M. Lanza-Gutierrez, M. Portela-García, C. López-Ongil, "Fear Recognition for Women Using a Reduced Set of Physiological Signals," *Sensors*, 2021, 21(5), 1587.

This document is partially included in Chapter 4.

- J. A. Miranda, M. F. Canabal, L. Gutiérrez-Martín, J. M. Lanza-Gutierrez, C. López-Ongil, "Edge computing design space exploration for heart rate monitoring," *VLSI Integration*, 2022.

This document is partially included in Chapter 5.

- J. A. Miranda, E. Rituerto-González, M. F. Canabal, A. R. Bárcenas, J. M. Lanza-Gutiérrez, C. Pelaez-Moreno, and C. López-Ongil, "Bindi: Affective internet of things to combat gender-based violence," *IEEE Internet of Things*, 2022, manuscript submitted for publication.

This document is partially included in Chapter 6.

- J. A. Miranda, A. P. Montoro, C. López-Ongil, and J. Andreu-Pérez, "FT2F-SQA: Few-shot type-2 fuzzy-based subject-invariant ppg quality assessment for extreme edge physiological monitoring," *IEEE TIM*, 2022, manuscript submitted for publication.

This document is partially included in Chapter 5.

- M. F. Canabal, J. A. Miranda, A. P. Montoro, I. P. Garcilópez, S. P. Álvarez, E. G. Ares, and C. López-Ongil, "Design and validation of an efficient and adjustable GSR sensor for emotion monitoring," *IEEE Sensors*, 2022, manuscript in progress.

- **Artículos en conferencias**

- J. A. Miranda, A. Vaskova, M. Portela-García, M. García-Valderas and C. López-Ongil, "On-line testing of sensor networks: A case study," IEEE 23rd International Symposium on On-Line Testing and Robust System Design (IOLTS), 2017, pp. 201-202, doi: 10.1109/IOLTS.2017.8046218.
- J. A. Miranda, M. F. Canabal, J. M. Lanza-Gutiérrez, M. Portela-García, C. López-Ongil, Teresa Riesgo Alcaide, "Meaningful Data Treatment from Multiple Physiological Sensors in a Cyber-Physical System," DCIS 2017: XXXII Conference on Design of Circuits and Integrated Systems, 22nd-24th November 2017, Barcelona (Spain). pp. 100-104.
- J. A. Miranda, M. F. Canabal, M. Portela García, C. Lopez-Ongil, "Embedded emotion recognition: Autonomous multimodal affective internet of things," Proceedings of the cyber-physical systems workshop, 2018, (Vol. 2208, pp. 22-29).
- J. A. Miranda, R. Marino, J.M Lanza-Gutierrez, Teresa Riesgo, M. Garcia-Valderas, C. Lopez-Ongil, "Embedded emotion recognition within cyber-physical systems using physiological signals," In 2018 Conference on design of circuits and integrated systems (DCIS) (pp. 1-6). IEEE.
- J. A. Miranda, M. F. Canabal, J. M. Lanza-Gutiérrez, M. P. García and C. López-Ongil, "Toward Fear Detection using Affect Recognition," 2019 XXXIV Conference on Design of Circuits and Integrated Systems (DCIS), 2019, pp. 1-4, doi:10.1109/DCIS201949030.2019.8959852.
- E. Rituerto-González, J. A. Miranda, M. F. Canabal, J. M. Lanza-Gutiérrez, C. Peláez-Moreno, C. López-Ongil, "A hybrid data fusion architecture for bindi: A wearable solution to combat gender-based violence," In International Conference on Multimedia Communications, Services and Security (pp. 223-237). Springer, Cham., 2020.
- J. A. Miranda, M. F. Canabal, L. Gutiérrez-Martín, J. M. Lanza-Gutiérrez, C. López-Ongil, "A Design Space Exploration for Heart Rate Variability in a Wearable Smart Device," In 2020 XXXV Conference on Design of Circuits and Integrated Systems (DCIS) (pp. 1-6). IEEE.
- M. F. Canabal, J. A. Miranda, J. M. Lanza-Gutiérrez, A. P. Garcilópez, C. López-Ongil, "Electrodermal Activity Smart Sensor Integration in a

Wearable Affective Computing System," In 2020 XXXV Conference on Design of Circuits and Integrated Systems (DCIS) (pp. 1-6). IEEE.

- J. A. Miranda, A. P. Montoro, C. López-Ongil, and J. Andreu-Pérez, "Towards Type-2 Fuzzy-Based PPG Quality Assessment for Physiological Monitoring," IEEE WCCI 2022, manuscript submitted for publication.

- **Otras publicaciones**

- J. A. Miranda, E. Rituerto-González, L. Gutiérrez-Martín, C. Luis-Mingueza, M. F. Canabal, A. R. Bárcenas, C. López-Ongil, "WEMAC: Women and Emotion Multi-modal Affective Computing dataset," arXiv preprint arXiv:2203.00456, 2022.
- M. Á. Blanco Ruiz, L. Gutiérrez Martín, J. Á. Miranda Calero, M. F. Canabal Benito, E. Romero Perales, C. Sainz de Baranda Andujar, R. San Segundo Manuel, D. Larrabeiti López, C. Peláez-Moreno, and C. López Ongil, "UC3M4Safety Database - List of Audiovisual Stimuli Annotations," 2021, [Online]. Available: <https://doi.org/10.21950/CXAAHR>
- C. S. de Baranda Andújar, M. B. Ruiz, J. Á. Miranda, L. Gutierrez-Martín, M. F. Canabal, R. San Segundo, C. López-Ongil, "Perspectiva de género y social en las STEM: La construcción de sistemas inteligentes para detección de emociones," Sociología y tecnociencia: Revista digital de sociología del sistema tecnocientífico, 11(1), 83-115, 2021.
- E. Rituerto-González, J. A. Miranda, M. F. Canabal, J. M. Lanza-Gutiérrez, C. Peláez-Moreno, C. López-Ongil, "A Hybrid Data Fusion Architecture for BINDI: A Wearable Solution to Combat Gender-Based Violence," In: Dziech A., Mees W., Czyżewski A. (eds) Multimedia Communications, Services and Security. MCSS 2020. Communications in Computer and Information Science, vol 1284. Springer, Cham. https://doi.org/10.1007/978-3-030-59000-0_17.
- Andres Russu, Antonio J. de Castro, Francisco Cortes, Celia López-Ongil, Marta Portela, Ernesto Garcia, José A. Miranda, Manuel F. Canabal, Ignacio Arruego, Javier Martinez-Oter, Fernando López, "A light compact and rugged IR sensor for space applications," Proc. SPIE 11129, Infrared Sensors, Devices, and Applications IX, 1112907 (9 September

2019); <https://doi.org/10.1117/12.2529846>.

Otros méritos de investigación

- **Premios**

- Best poster award, XXXIV Conference on Design of Circuits and Integrated Systems (DCIS), 2019.
- YERUN Research Mobility Award 2021.

- **Patentes**

- Sistema y método para for determinar un estado emocional de un usuario (System and method for determining a user's emotional state), Modelo de Utilidad ES1269890 (Expedition date: 20/09/2021). Owned by UC3M / UPM.

- **Participación en proyectos de investigación**

- Ciencia y Tecnología para la caracterización in situ de la atmósfera de Marte. Desarrollo del Instrumento Dust Sensor para la misión EXO-MARS'18 de ESA/IKI. Fases A/B y C/D. ESP2015-67624-R, MINECO.
- Comunidad de Madrid, protección integral de las víctimas de violencia de género Mediante computación afectiva multimodal (EMPATÍA-CM), Y2018/TCS-5046.

- **Trabajos Fin de Grado supervisados**

- Ismael Granados Neira, "Desarrollo e implementación de drivers para aplicación IoT," Internship coordinator, Universidad Carlos III de Madrid, 2016.
- Adrian Neira Robledo, "Integración de cámaras IR en una red de sensores inalámbrica para recintos críticos," Trabajo Fin de Grado, Universidad Carlos III de Madrid, 2017.
- George Sebastian Roma, "Investigación, comparación e implementación

- de RTOSs para plataformas orientadas a IoT," Trabajo Fin de Grado, Universidad Carlos III de Madrid, 2017.
- Oscar Escobar Muñoz, "Diseño Hardware de un Sistema Eficiente para la Medida del Consumo de Energía en Nodos Inalámbricos de Redes de Sensores," Trabajo Fin de Grado, Universidad Carlos III de Madrid, 2017.
 - Javier Plaza Arenas, "Implementación de técnicas de Machine Learning para la detección de emociones en FPGA," Trabajo Fin de Grado, Universidad Carlos III de Madrid, 2018.
 - A. Aranzana Sánchez, "Implementación de Técnicas de Extracción de Características para el Reconocimiento de Emociones Usando Sensores Fotopleletismográficos," Trabajo Fin de Grado, Universidad Carlos III de Madrid, 2020.
 - W. Allay Bakhtaoui, "Implementación de Técnicas de Extracción de Características para el Reconocimiento de Emociones Usando Sensores para la Conductividad de la Piel," Trabajo Fin de Grado, Universidad Carlos III de Madrid, 2020.
 - Pedro Ruiz Perez, "Estudio e implementación embebida de filtros digitales con señales fotopleletismográficas," Trabajo Fin de Grado, Universidad Carlos III de Madrid, 2020.
 - L. Velasco Gonzalez, "Diseño e implementación de un sistema de recuperación fisiológica para experimentos de reconocimiento de emociones," Trabajo Fin de Grado, Universidad Carlos III de Madrid, 2021.
 - Alexandru Stoica Stoica, " Estudio e implementación de técnicas de procesamiento orientadas a la eliminación de artefactos con señales fotopleletismográficas para bindi," Trabajo Fin de Grado, Universidad Carlos III de Madrid, 2021.
 - Juan Marcos Torero, "Diseño y desarrollo de pendiente inteligente con micrófono y sensor de pulso para su integración en el sistema BINDI," Trabajo Fin de Grado, Universidad Carlos III de Madrid, 2021.
 - Mario Iañez Diaz, "Diseño e integración de Front-End para ECG de dos electrodos en Bindi," Trabajo Fin de Grado, Universidad Carlos III de Madrid, 2021.

- Andrés Lucas Rodríguez, "Diseño e implementación de un sensor textil inteligente para Bindi", Trabajo Fin de Grado, Universidad Carlos III de Madrid, 2021.
- **Trabajo Fin de Máster supervisados**
 - Jesus Jose Garcia de Cuerva Camacho, "Eliminación de artefactos de movimiento en señales fotoplestismográficas para sistemas portables orientados a la detección de emociones," Trabajo Fin de Máster. Universidad Carlos III de Madrid, 2019.
 - F. Adrián Hernández Gant, "Diseño de modelos de aprendizaje para detección de miedo en Bindi," Trabajo Fin de Máster. Universidad Carlos III de Madrid, 2021.
 - Marta Subirán Adrados, "Design Space Exploration for the Multi-modal Data Fusion Architectures of Bindi," Trabajo Fin de Máster. Universidad Carlos III de Madrid, 2021.

Biografía

José A. Miranda Calero se licenció en Ingeniería Electrónica Industrial y Automática en 2013, y en Ingeniería de Sistemas y Aplicaciones Electrónicas en 2016 con matrícula de honor, ambos por la Universidad Carlos III de Madrid. Actualmente es candidato a doctor en el grupo de investigación de Diseño y Aplicaciones Microelectrónicas (DMA), perteneciente a la Universidad Carlos III de Madrid. Desde 2013 hasta 2015, ha trabajado como Ingeniero de Software Embebido en diferentes países de Europa para el sector público y privado. Su campo de investigación comprende un sensor inalámbrico, redes, diseño, desarrollo e integración de wearables para aplicaciones de seguridad, implementación de computación afectiva en dispositivos de edge computing y aceleración de hardware. En lo que respecta al desarrollo de tecnología vestible para aplicaciones de seguridad, su principal contribución está relacionada con el diseño de Bindi, que es una nueva solución autónoma, inteligente, discreta, conectada, basada en la computación de borde y vestible, capaz de detectar y alertar cuando un usuario se encuentra en una situación violenta empleando el reconocimiento de emociones utilizando las señales fisiológicas y auditivas del usuario. Bindi se está desarrollando dentro del equipo UC3M4Safety. Además, también ha participado en otros proyectos relacionados con aplicaciones espaciales, como DS-EXOMARS20.

Índice

1	Introducción	1
1.1	Contexto y motivación	1
1.2	Alcance de esta disertación	6
1.3	Esquema del documento	8
I	Emociones humanas, señales fisiológicas y computación afectiva	11
2	Clasificación de las emociones y cuantificación fisiológica	13
2.1	Supuestos y definiciones	14
2.2	Teorías Emocionales	14
2.3	Clasificación de las emociones humanas	17
2.3.1	Clasificación discreta de las emociones	18
2.3.2	Clasificación dimensional de las emociones	20
2.3.3	Rasgos personales, procesos cognitivos, atención y sesgo de género	25
2.3.4	El mapa del miedo dentro de los métodos de clasificación de las emociones humanas	27
2.4	Herramientas para el análisis científico de las respuestas emocionales humanas	29
2.5	Indicadores fisiológicos de las respuestas de las emociones humanas	33
2.5.1	Actividad cardiovascular	35
2.5.2	Actividad Electrodermica	40
2.5.3	Temperatura de la piel	49
2.6	Conclusión	52

3	Bases de datos y aprendizaje automático para el reconocimiento de emociones	55
3.1	Metodología general de las bases de datos	58
3.1.1	Selección y análisis de estímulos	58
3.1.2	Procesamiento y adquisición de los sensores	60
3.1.3	Análisis exploratorio de los datos	62
3.1.4	Ingeniería de características	63
3.1.4.1	Extracción de características	64
3.1.4.2	Selección de características	67
3.1.4.3	Reducción de dimensionalidad	68
3.1.5	Optimización de Hyperparámetros	69
3.1.6	Fusion de datos	71
3.1.7	Clasificación emocional	73
3.1.7.1	Compensación del sesgo (<i>bias</i>) y varianza (<i>variance</i>)	73
3.1.7.2	Algoritmos de aprendizaje máquina	75
3.1.7.3	Técnicas de validación cruzada	81
3.2	Bases de datos abiertas	85
3.3	Conclusión	88
II	Clasificación del miedo usando el estado del arte	90
4	Prueba de concepto para clasificar miedo	91
4.1	Clasificación del miedo usando DEAP	95
4.1.1	Consideraciones sobre el equilibrio de los estímulos y las etiquetas	100
4.1.2	Análisis exploratorio de datos y procesamiento de filtros . . .	108
4.1.3	Extracción de características	111
4.1.3.1	Dominio Temporal	112
4.1.3.2	Dominio Frecuencial	115
4.1.3.3	Dominio No-lineal	117
4.1.4	Sistemas de clasificación del miedo	119
4.1.4.1	Sistema DEAP-b1	119
4.1.4.2	Sistema DEAP-b2	125

4.2	Clasificación del miedo mediante MAHNOB	132
4.2.1	Consideraciones sobre el equilibrio de los estímulos y las etiquetas	135
4.2.2	Análisis exploratorio de datos, segmentación y filtrado de datos	139
4.2.3	Extracción de características	141
4.2.3.1	Dominio del tiempo y de la frecuencia	145
4.2.3.2	Dominio No-lineal	146
4.2.4	Sistemas de clasificación del miedo	149
4.2.4.1	Modelos dependientes de la voluntaria	150
4.2.4.2	Modelos independientes de la voluntaria	155
4.3	Discusión y Conclusión	159

III Hacia un nuevo paradigma de detección del miedo en situaciones de violencia de género 166

5	Un nuevo sistema autónomo de reconocimiento de emociones:	
	Bindi	167
5.1	Tecnología actual para luchar contra la violencia de género	168
5.2	Bindi	174
5.2.1	Arquitectura del sistema	177
5.2.1.1	Diseño e integración de sensores fisiológicos	178
5.2.1.2	Diseño de procesamiento digital de señales	185
5.2.2	Evaluación del filtrado integrado	193
5.2.3	Signal Quality Assessment	195
5.2.3.1	Diseño, formación y validación de SQA	200
5.2.3.2	Implantación de SQA y autoajuste	208
5.2.3.3	Herramientas y métodos	211
5.2.3.4	Resultados	212
5.2.4	Extracción de características Exploración del espacio de diseño	220
5.2.4.1	Extracción de características: Detección de picos	221
5.2.4.2	Extracción de características: Información de la VFC	225
5.2.4.3	Implementación del caso de uso de VFC	228
5.2.5	Análisis del consumo de energía	231

5.3	Conclusión	232
6	Un nuevo conjunto de datos para el reconocimiento de emociones:	
	WEMAC	237
6.1	Métodos, herramientas y estímulos	240
6.2	Exploración de respuestas de etiquetado autodirigidas	244
6.3	Exploración de la respuesta fisiológica	250
6.3.1	Patrones fisiológicos y recuperaciones	251
6.3.1.1	Análisis de patrones	252
6.3.1.2	Análisis de recuperación	257
6.3.2	Resultados fisiológicos uni-modales	265
6.3.2.1	Extracción de características	267
6.3.2.2	Selección de características	269
6.3.2.3	Resultados de la validación y las pruebas	273
6.4	Marco de fusión de datos multimodales	279
6.4.1	Métodos de fusión de datos multimodales	284
6.4.2	Resultados de la fusión de datos multimodales	287
6.5	Discusión y conclusión	293
IV	Conclusión	297
7	Conclusión	299
7.1	Contribuciones	301
7.2	Trabajo futuro	303
A	Esquemáticos del brazalete	305

Lista de Figuras

1-1	Número total de víctimas de violencia de género asesinadas desde 2003 hasta octubre de 2021. Datos proporcionados por [1].	4
2-1	Orden de activación emocional para la teoría de James-Lange.	15
2-2	Orden de activación emocional para la teoría desarrollada por Magda Arnold y Richard Lazarus [2].	16
2-3	Orden de activación para la teoría emocional racional-emotiva por Albert Ellis [3].	17
2-4	Relación entre los modelos PAD y ABC. Este último es el modelo de actitudes por Bakker et al.. Esta relación trata de dar una visión más clara que la propuesta inicial del modelo PAD [4].	23
2-5	Los 24 términos emocionales se han mapeado en el esquema cuatridimensional propuesto por Fontaine et al. en [5].	24
2-6	De izquierda a derecha: concepto de miedo unidimensional (niveles de intensidad discretos), miedo contenido en un espacio bidimensional (modelo Pleasure-Arousal model (PA)), conceptos tridimensionales (modelo pleasure-arousal-dominance model (PAD)) y cuatridimensionales (modelo PAD más cualquier dimensión intrínseca individual).	27
2-7	Self-Assessment Manikins (SAM) original [6].	32
2-8	SAM modificado por el equipo UC3M4Safety.	32
2-9	Location of the two main parts, amygdala and hypothalamus, involved in the emotional processing and autonomous nervous system regulation.	34
2-10	Ilustración de las dos técnicas de medición del PPG, reflexión y transmisión. Obsérvese que la señal obtenida está invertida en un método con respecto al otro.	37

2-11	Ejemplificación de los diferentes puntos característicos a extraer dentro de la morfología de la señal PPG.	39
2-12	Ilustración del comportamiento de las glándulas merócrinas y del proceso de difusión a través de las diferentes capas de la piel.	41
2-13	Ejemplo de configuración de un amplificador operacional inversor para la adquisición de DC Electrodermal Activity (EDA) exosomática.	44
2-14	Un ejemplo ilustrativo de una Event-Related Skin Conductance Response (ERSCR) y algunas de las métricas que se pueden extraer de ella (<i>Stimulus - Estímulo, Latency - Latencia, Onset - Punto de inicio, Rise Time - Tiempo de subida, Response Peak - Pico de respuesta, Recovery time - Tiempo de recuperación, Half Recovery - Mitad de recuperación, Offset - Punto de parada</i>).	45
2-15	Diferencia entre los electrodos secos y húmedos que miden en la parte ventral sobre la parte derecha (húmeda) e izquierda (seca) de la muñeca. Obsérvese que las unidades están normalizadas μ S y que los electrodos húmedos contienen 0,5% de sal de cloruro.	48
3-1	Elementos, procesos y acciones comunes necesarios para la generación de una base de datos de reconocimiento de emociones (<i>Volunteers - Voluntarios, Sensors - Sensores, DB Experiment - Experimento de la BBDD, Raw Data - Datos crudos, Self-Reported Annotations - Anotaciones reportadas, Stimuli Pool - Conjunto de Estímulos, Stimuli Reduction/Selection - Selección/Reducción de estímulos, Physiological Data Processing - Procesamiento de los datos fisiológicos, Exploratory Data Analysis - Análisis exploratorio de los datos, Feature Engineering - Ingeniería de los datos, Classification - Clasificación, Performance assessment - Evaluación del desempeño, Data Fusion - Fusion de Datos, Hyperparameter Optimisation - Optimización de hiperparámetros, Fully trained operational model - modelo completamente entrenado</i>).	57
3-2	Procesos convencionales de ingeniería de rasgos para la selección supervisada de rasgos.	64

3-3	Técnicas de fusión temprana y tardía de datos para características fisiológicas y de audio/voz, con dimensiones N y M respectivamente.	73
3-4	Compensación entre el sesgo y la varianza con el infraajuste (<i>Underfitting Zone</i>), el sobreajuste (<i>Overfitting Zone</i>) y las zonas óptimas (<i>Optimal Zone</i>).	75
3-5	Ilustración del hiperplano para el clasificador Support Vector Machine (SVM) para la clasificación binaria (los puntos negros son la clase positiva, y los puntos grises son la clase negativa).	76
3-6	Ilustración del truco del núcleo para un problema binario.	78
3-7	Radial Basis Function (RBF) valores del núcleo en función de la distancia entre los dos puntos que se evalúan para diferentes σ .	79
3-8	Representación gráfica de las técnicas Leave-One-Subject-Out (LOSO), Leave-One-Trial-Out (LOTO) y Leave-hAlf-Subject-Out (LASO) [7].	84
4-1	Visión general del proceso de entrenamiento para el sistema de reconocimiento de miedo propuesto que emplea datos de sensores fisiológicos y etiquetado de emociones de enfoque dimensional. Este último se introduce en el procedimiento de mapeo binario del miedo. Obsérvese que $w \# n$ denota las diferentes ventanas obtenidas tras la segmentación de los datos, si procede.	94
4-2	Diagrama simplificado de la experimentación aplicada para cada voluntario y cada estímulo para la base de datos DEAP.	96
4-3	Diferencias de etiquetado para la base de datos DEAP y la numeración original de los videoclips seleccionados.	101
4-4	PAD para las etiquetas autodeclaradas por los voluntarios. El mapeo del miedo propuesto en la sección 2.3.4 está marcado con un cubo de color.	103
4-5	Balance de clases por voluntario después de haber aplicado el mapeo binario de miedo de un espacio PA.	105
4-6	Balance de clases por voluntario después de haber aplicado el mapeo binario de miedo de un espacio PAD.	106

4-7	Valores p promediados para todos los voluntarios considerados y sus etiquetas aplicando: a) la correlación de Spearman, y b) para la prueba Chi-cuadrado de independencia. En este caso, las etiquetas se binarizan utilizando el mapeo basado en el miedo binario.	107
4-8	Valores p promediados para todos los voluntarios considerados y sus etiquetas aplicando: a) la correlación de Spearman, y b) para la prueba Chi-cuadrado de independencia. En este caso, las etiquetas se binarizan utilizando el mapeo basado en el miedo binario.	108
4-9	Ejemplo de una de las representaciones gráficas de la evaluación visual fisiológica realizada.	110
4-10	Ejemplo de filtrado para la extracción y eliminación de la línea de base mediante el filtrado IIR, y la eliminación del ruido alto.	111
4-11	Ilustración de la resolución de la frecuencia y ubicación de los bins de frecuencia basados en una ventana de procesamiento de T segundos.	115
4-12	Representación ideal y relación entre las partes de baja frecuencia (LF) y de alta frecuencia (HF) de la Inter-Beat-Interval (IBI) Power Spectral Density (PSD) [8].	116
4-13	Cada serie temporal de grano grueso obtenida para cada nivel de la técnica o algoritmo de extracción de características MSE.	119
4-14	Exactitud frente al coste de la clasificación errónea para $p18$	121
4-15	Sensibilidad frente al coste de la clasificación errónea para $p18$	122
4-16	Especificidad frente al coste de la clasificación errónea para $p18$	122
4-17	Metodología seguida durante los experimentos de la base de datos MAHNOB.	133
4-18	Distribución de clases para el mapeo binario del miedo sobre los autoinformes subjetivos en MANHOB para todas las diferentes voluntarias consideradas, y la distribución de clases original prevista en el experimento.	137
4-19	Valores p promediados para todos los voluntarios MAHNOB considerados y sus etiquetas aplicando: a) la correlación de Spearman, y b) para la prueba Chi-cuadrado de independencia. En este caso, las etiquetas se binarizan utilizando el mapeo basado en el miedo binario.	138

4-20	Proceso típico de segmentación de datos en los sistemas de reconocimiento de emociones basados en el aprendizaje automático.	141
4-21	Esquema de la arquitectura del algoritmo de identificación de picos Electrocardiogram (ECG) aplicado en este trabajo.	145
4-22	Matrices de confusión para un modelo dependiente del sujeto en V11, detectado como un problema de asimetría.	153
4-23	Matrices de confusión para un modelo dependiente del sujeto en V7. .	154
4-24	Matrices de confusión para los clasificadores ENS y los voluntarios probados (datos no vistos) sobre sus respectivos modelos independientes del sujeto: (a) probado V4, (b) probado V7.	158
5-1	Arquitectura simplificada del sistema Bindibasado en las diferentes tecnologías del IoT.	168
5-2	Dispositivos considerados para el sistema de vigilancia electrónica dentro del "Protocolo de actuación del sistema de vigilancia por medios telemáticos de las medidas y penas de alejamiento en materia de violencia de género". DLI: Dispositivo que lleva el agresor; DLV: Dispositivo que lleva la víctima [9].	173
5-3	Evolución de la tecnología Bindi desde 2016 hasta 2022.	176
5-4	Arquitectura simplificada del brazalete.	178
5-5	Fotodiodo de eficiencia cuántica MAX30101 [10]	180
5-6	Implementación del frente analógico del sensor de GSR en el brazalete de Bindi.	181
5-7	Respuesta GSR de Bindi considerando diferentes resistencias de la piel.	182
5-8	Respuesta no lineal de la corriente de piel dada por el divisor de tensión entre R_{14} y R_{skin}	182
5-9	Señales filtradas normalizadas del sensor de Galvanic Skin Response (GSR) obtenidas por Bindi y el sensor de validación para un voluntario en dos estímulos. La línea vertical de guiones denota la separación de estímulos.	183
5-10	Integración de los sensores de temperatura de la piel (círculo amarillo) y de frecuencia cardíaca en la pulsera. La zona gris determina el plano de tierra.	186

5-11	Salida filtrada del MAX30205 tras colocar un dedo sobre el chip integrado en condiciones de temperatura ambiente controlada.	186
5-12	Modificación realizada al Brazalete para incluir el MAX30208 y comparación del experimento para ambos sensores de temperatura. A la derecha parte de la placa de evaluación del MAX30208.	187
5-13	Pila actual del firmware del brazalete de Bindi.	189
5-14	Sincronización fisiológica actual y tiempos de procesamiento de datos en el Brazalete.	190
5-15	Arquitectura actual del sistema para las principales tareas de procesamiento digital del Brazalete.	192
5-16	Arquitectura de formación SQA propuesta.	200
5-17	Representación de intervalo con tres (m) particiones (ν) y dos puntos finales encontrados (τ). Los valores τ_{min} y τ_{max} son el mínimo y el máximo de la secuencia entrante que se evalúa o los puntos finales izquierdo y derecho.	203
5-18	Funciones de pertenencia de tipo II generadas a partir de los datos de características del perfil de la matriz aplicando IA para todos los sujetos de entrenamiento. Tres variables lingüísticas: Baja (L), Media (M), Alta (H). El área sombreada en gris es la FOU obtenida.	204
5-19	Arquitectura integrada SQA implementada. SoA: Fuerza de Activación. sSoA: Fuerza de activación escalada.	209
5-20	Captura en tiempo real de la implementación de SQA integrada que muestra los diferentes valores de las características en cada ventana de procesamiento (3 segundos).	216
5-21	Parameters and processes involved in the BVP-based DSE.	221
5-22	Diferencias morfológicas del PPG entre tres grupos de edad. (a) Persona de 18-24 años. (b) Persona de 35-44 años. (c) Persona de 55-65 años. Las señales mostradas fueron adquiridas por el brazalete Bindi.	222
5-23	Análisis del impacto temporal de los algoritmos de detección de picos considerados.	225
5-24	Análisis del impacto temporal basado en diferentes métodos de interpolación y en la FFT implementada y considerada.	227

5-25	Cadena de datos completa para el procesamiento de una ventana de 4 segundos teniendo en cuenta las compensaciones comentadas.	229
5-26	Efectos de artefactos de movimiento mostrados en un segmento del estímulo audiovisual de estrés del voluntario 2.	231
5-27	Consumo medio de corriente en el Brazalete [11].	232
5-28	La ventaja competitiva de Bindi sobre sus principales y más directos competidores.	233
6-1	Metodología experimental seguida durante la elaboración del conjunto de datos WEMAC. Antes y durante la experimentación.	241
6-2	Distribución de clases para el mapeo de miedo binario sobre los autoinformes subjetivos discretos en WEMAC para todas las 47 voluntarias consideradas, y la distribución de clases original prevista en el experimento: G2 y G1 para el segundo y primer lote, respectivamente.	246
6-3	Distribución de clases para el mapeo binario del miedo sobre los autoinformes subjetivos dimensionales en WEMAC para todas las 47 voluntarias consideradas, y la distribución de clases original prevista en el experimento: G2 y G1 para el segundo y primer lote, respectivamente.	247
6-4	Intercorrelación de Spearman uno a uno entre los sujetos de los 47 voluntarias para ambas metodologías de etiquetado: a) discreta, y b) dimensional (PAD).	248
6-5	Valores P obtenidos a partir de la intercorrelación de sujetos de Spearman uno a uno a través de los 47 voluntarias para ambas metodologías de etiquetado: a) discreto, y b) dimensional (PAD).	249
6-6	Valores p promediados para todos los voluntarias considerados y sus etiquetas aplicando la correlación de Spearman para sus etiquetas de mapeo binario de miedo basadas en PAD.	250
6-7	Valores p promediados para todos los voluntarias considerados y sus etiquetas aplicando la correlación de Spearman para sus etiquetas de mapeo binario del miedo basado en la discreción.	250
6-8	Señales GSR extraídas de la visualización completa del sexto estímulo de la primera tanda (último estímulo) de los voluntarias 4, 15 y 27.	253

6-9	Matriz de distancia Dynamic Time Warping (DTW) promediada para los 32 voluntarias que visualizan los 6 estímulos de miedo del primer lote de estímulos relacionados con la emoción.	255
6-10	Resultados agregados obtenidos a partir de la matriz de distancia DTW promediada para los 32 voluntarias de la Figura 6-9.	256
6-11	Comparación de resultados promediados obtenidos del proceso de extracción de picos GSR mediante el algoritmo cvxEDA para los voluntarias de 47 y ambos lotes.	259
6-12	Comparación de los resultados promediados obtenidos del proceso de extracción de amplitudes relativas GSR mediante el algoritmo cvxEDA para los 47 voluntarias y ambos lotes.	260
6-13	Comparación de los resultados promediados obtenidos del proceso de extracción del tiempo de recuperación de picos GSR mediante el algoritmo cvxEDA para los voluntarias 47 y ambos lotes.	260
6-14	Ejemplificación de un gráfico de Poincaré recurrente y su métrica de desviación estándar a lo largo (SD_2) y perpendicular (SD_1) a la línea de identidad.	261
6-15	Diferentes perspectivas de Poincaré para todos los 47 voluntarias considerando los estímulos de miedo (rojo-abajo), los estímulos sin miedo (verde-medio) y las etapas de recuperación (azul-arriba). Vista frontal.	262
6-16	Diferentes perspectivas de Poincaré para todos los 47 voluntarias considerando los estímulos de miedo (rojo-abajo), los estímulos sin miedo (verde-medio) y las etapas de recuperación (azul-arriba). Vista longitudinal.	262
6-17	Diferentes perspectivas de Poincaré para todos los 47 voluntarias considerando los estímulos de miedo (rojo-abajo), los estímulos sin miedo (verde-medio) y las etapas de recuperación (azul-arriba). Vista 2D.	263
6-18	Arquitectura de procesamiento de datos fisiológicos para entrenar y probar los modelos de aprendizaje automático generados utilizando nuestro propio conjunto de datos.	267

6-19	Evaluación de la métrica de prueba MCC para todos los 42 modelos considerados dentro del caso de uso de detección de miedo discreto binarizado.	276
6-20	Evaluación de la métrica de prueba MCC para todos los 38 modelos considerados dentro del caso de uso de detección de miedo dimensional binarizado.	276
6-21	Distribución del diagrama de caja de la métrica de prueba MCC para todos los modelos de 42 y 38 considerados dentro de los casos de uso de detección de miedo discreto y dimensional binarizado.	277
6-22	Relación LF/HF extraída del voluntario 3 del conjunto de datos WEMAC. Obsérvese que en las abscisas están representadas las emociones dirigidas para el primer lote.	280
6-23	Esbozo de exploración del espacio de diseño para los diferentes arreglos de la modalidad a realizar con la arquitectura de Bindi.	282
6-24	Diagrama de bloques de fusión de datos para Bindi 2.0a y Bindi 2.0b.	284
6-25	Barrido de parámetros para a) th_{phy} y b) th_{sp} en los subsistemas fisiológico y unimodal del habla, respectivamente.	288
6-26	Análisis del rendimiento medio de las puntuaciones F1 prediciendo sobre los 42 voluntarias de prueba para las diferentes configuraciones de arquitectura.	289
6-27	Análisis del rendimiento de la puntuación media de exactitud que predice sobre los 42 voluntarias de prueba para las diferentes configuraciones de arquitectura.	290
6-28	Análisis del rendimiento individual de los dos subsistemas unimodales.	292

Lista de Tablas

2.1	Main categorical models of emotions developed since 19th century. . .	19
2.2	Revisión del tipo de estímulo utilizado en entornos de laboratorio controlados.	31
2.3	Principales diferencias entre las mediciones exosomáticas de CC y CA.	42
3.1	Las bases de datos de reconocimiento de emociones más comunes con una configuración de laboratorio utilizadas dentro de la comunidad científica de la computación afectiva.	86
4.1	vídeos que se encuentran en un cuadrante diferente con respecto al preetiquetado frente a las etiquetas autodeclaradas.	102
4.2	Ratios de desequilibrio reportados por las voluntarias y los voluntarios para la base de datos DEAP.	104
4.3	Ratios de balanceo PA y PAD para la base de datos DEAP.	105
4.4	Características extraídas del sistema DEAP-b2.	113
4.5	Impacto del tamaño del conjunto de entrenamiento en la memoria y el cálculo para $p18$. Enfoque dependiente del sujeto.	123
4.6	Métricas de precisión, sensibilidad, especificidad y media geométrica para cada voluntario asumiendo un coste de retención y de clasificación errónea de 0,99 y 8, respectivamente. Enfoque dependiente del sujeto.	124
4.7	Impacto del tamaño del conjunto de entrenamiento en la memoria y el cálculo. Enfoque independiente de la materia.	125
4.8	Exactitud, sensibilidad, especificidad y media geométrica para cada retención probada asumiendo un coste de clasificación errónea de 8. Enfoque independiente del sujeto.	125

4.9	Métricas de precisión, sensibilidad, especificidad y AUC para cada caso asumiendo las condiciones especificadas, respectivamente. Enfoque independiente del sujeto.	129
4.10	Exactitud, sensibilidad, especificidad y métrica AUC para la selección de características Maximum Relevance — Minimum Redundancy (mrMR) y SVM con kernel RBF. Enfoque independiente del sujeto. .	131
4.11	Mapeo de dimensiones discretas para el arousal y la valencia basado en [5] y adoptado por MAHNOB [12].	136
4.12	Características extraídas para la señal ECG y la propuesta de reconocimiento binario de emociones de miedo utilizando el conjunto de datos MAHNOB.	142
4.13	Características extraídas para la señal GSR y la propuesta de reconocimiento binario de emociones de miedo utilizando el conjunto de datos MAHNOB.	143
4.14	Características extraídas para la señal Skin Temperature (SKT) y la propuesta de reconocimiento binario de emociones de miedo utilizando el conjunto de datos MAHNOB.	143
4.15	Métricas de rendimiento para cada modelo generado en función del sujeto y métricas de rendimiento medio y dispersión para cada algoritmo de clasificación.	151
4.16	Métricas de rendimiento para cada modelo generado independiente del sujeto y métricas de rendimiento medio y dispersión para cada algoritmo de clasificación. La etapa de entrenamiento se realiza utilizando todos los voluntarios excepto el voluntario probado en cada modelo generado (datos de prueba no vistos).	156
4.17	Métricas de rendimiento para cada modelo generado independiente del sujeto y métricas de rendimiento medio y dispersión para Ensemble methods (ENS) tras la selección de características mrMR. La etapa de entrenamiento se realiza utilizando todos los voluntarios excepto el voluntario probado en cada modelo generado (datos de prueba no vistos).	160

4.18	Los mejores resultados obtenidos a lo largo del capítulo 4 para el reconocimiento de la emoción binaria del miedo cuando se trata de un modelo independiente del sujeto.	162
4.19	Los trabajos más recientes y principales del estado de la técnica que están directamente relacionados con esta investigación y que han influido en ella en lo que respecta a la computación afectiva utilizando información fisiológica.	165
5.1	Resultados obtenidos para las arquitecturas de filtrado integradas evaluadas.	195
5.2	Métricas de rendimiento de validación utilizando los métodos de razonamiento α y τ y nuestro propio conjunto de datos.	213
5.3	Métricas de rendimiento de las pruebas para los diferentes conjuntos de datos de prueba utilizando los métodos de razonamiento α y β , y el autoajuste (s-T).	213
5.4	Métricas de rendimiento promedio de las pruebas utilizando los métodos de razonamiento α y β , y el autoajuste.	215
5.5	Comparación entre las métricas de rendimiento registradas por MATLAB para los 33 segmentos de PPG evaluados en el SoC utilizando el método de razonamiento <i>beta</i>	216
5.6	Coefficiente de determinación (R^2) para los principales procesos realizados dentro del SoC. SoA: Fuerza de Activación. MP: Perfil de la matriz. $-y + AD$: Grados de Asociación de Clase Negativa.	217
5.7	Análisis de ahorro de energía en tiempo real con y sin el método SQA. EC_{SQA} : Consumo de energía para el sistema implementado con SQA. EC_{Sensor} : Consumo de energía del sensor PPG. EC_{TR} : Consumo de energía para la transmisión BLE. <i>NE</i> : No ejecutado.	218
5.8	Comparación con los trabajos publicados sobre SQA.	218
5.9	Resultado de la medición del caso de uso específico del detector de estrés HRV.	230
6.1	Lista de los estímulos audiovisuales utilizados en el conjunto de datos WEMAC.	243

6.2	Estudio de agrupación de un solo segmento para los sujetos dependientes e independientes. SPE: especificidad, SEN: sensibilidad, Gmean: media geométrica.	257
6.3	Evaluación de las características del diagrama de Poincaré para los estímulos con y sin miedo, y sus respectivas etapas de recuperación. Estas métricas son la media y la desviación estándar promediadas para los 47 voluntarias.	263
6.4	Características extraídas para la señal Blood Volume Pulse (BVP) y la propuesta de reconocimiento de emociones binarias de miedo utilizando nuestro conjunto de datos.	270
6.5	Características extraídas para la señal GSR y la propuesta de reconocimiento binario de emociones de miedo utilizando nuestro conjunto de datos.	271
6.6	Número total de instancias para nuestro conjunto de datos basado en etiquetas discretas binarizadas y dimensionales autodeclaradas. . . .	274
6.7	Resultados de validación y prueba de los diferentes sistemas de aprendizaje automático fisiológico utilizando la primera versión de WEMAC. Se muestran los resultados de ambos enfoques binarizados discretos (Disc) y dimensionales (Dim).	274
6.8	Resultados de validación y prueba para los sistemas de aprendizaje automático K-Nearest Neighbours (KNN) utilizando el etiquetado discreto binarizado y una técnica LOSO Cross-Validation (CV) para la partición entrenamiento-prueba.	278
6.9	Análisis del rendimiento medio que predice los 42 voluntarias de las pruebas. Media y desviaciones estándar (Std).	291

Lista de abreviaciones¹

SoC	System-On-Chip
PPG	Photoplethysmography
ECG	Electrocardiogram
EDA	Electrodermal Activity
GSR	Galvanic Skin Response
SKT	Skin Temperature
VR	Virtual Reality
PA	Pleasure-Arousal model
PAD	pleasure-arousal-dominance model
PTSD	Post-Traumatic Stress Disorder
VR	Virtual Reality
SAM	Self-Assessment Manikins
ANS	Autonomous Nervous System
SNS	Sympathetic Nervous System
PNS	Parasympathetic Nervous System
BVP	Blood Volume Pulse
LED	Light Emitting Diode
BPM	Beats Per Minute
MAR	Motion Artifact Removal
SQA	Signal Quality Assessment
SCL	Skin Conductance Level
SCR	Skin Conductance Response
ERSCR	Event-Related Skin Conductance Response
NSSCR	Nonspecific Skin Conductance Response
SMNA	Sudomotor Nerve Activity
RLSD	Regularized Least-Squares Detrending
PSD	Power Spectral Density
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
RP	Recurrence Plot
SFS	Sequential Forward Selection
SBE	Sequential Backward Elimination

¹(Nota: Por la naturaleza técnica e internacional de este trabajo, las abreviaciones se encuentran definidas en inglés)

SVM	Support Vector Machine
KNN	K-Nearest Neighbours
RF	Random Forests
ENS	Ensemble methods
PCA	Principal Component Analysis
SMBO	Sequential Model-Based Optimisation
CV	Cross-Validation
LOO	Leave-One-Out
LOTO	Leave-One-Trial-Out
LOSO	Leave-One-Subject-Out
LASO	Leave-hAlf-Subject-Out
TEAP	Toolbox for Emotional feAture extraction from Physiological signals
HRV	Heart Rate Variability
FIR	Finite Impulse Response
IIR	Infinite Impulse Response
AGC	Automatic Gain Control
ACC	Accuracy
DSE	Design Space Exploration
IBI	Inter-Beat-Interval
MSE	Multi-Scale Entropy
SMOTE	Synthetic Minority Over-sampling TEchnique
AUC	Area Under the Curve
RBF	Radial Basis Function
mrMR	Maximum Relevance — Minimum Redundancy
MDI	Mutual Information Difference
ANOVA	Analysis of Variance
DFA	Detrended Fluctuation Analysis
BLE	Bluetooth Low Energy
DTW	Dynamic Time Warping
IoT	Internet of Things
SQA	Signal Quality Assessment
GPS	Global Positioning System
SoC	System on Chip
PCB	Printed Circuit Board
COTS	Commercial-Off-The-Shell
SDK	Software-Development-Kit
HAL	Hardware Abstraction Layer

Abstract

Entre los 17 Objetivos de Desarrollo Sostenible propuestos dentro de la Agenda 2030 y adoptados por todos los Estados miembros de las Naciones Unidas, el quinto ODS es un llamamiento a la acción para convertir efectivamente la igualdad de género en un derecho humano fundamental y una base esencial para un mundo mejor. Incluye la erradicación de todo tipo de violencia contra las mujeres. Centrándonos en la perspectiva tecnológica, el abanico de soluciones disponibles para prevenir este problema social es muy limitado. Además, la mayoría de las soluciones se basan en un enfoque de botón de pánico, dejando de lado el uso y la integración de las tecnologías actuales de vanguardia, como el Internet de las cosas (IoT), la computación afectiva, los sistemas ciberfísicos y los sensores inteligentes. Así, el objetivo principal de esta investigación es aportar nuevos conocimientos en el diseño y desarrollo de herramientas para prevenir y combatir las situaciones de riesgo e, incluso, las agresiones de Violencia de Género, desde una perspectiva tecnológica, pero sin dejar de lado las diferentes consideraciones sociológicas directamente relacionadas con el problema. Para lograr tal objetivo, nos basamos en la aplicación de la computación afectiva desde un punto de vista realista, es decir, apuntando a la generación de sistemas y herramientas capaces de ser implementados y utilizados en la actualidad o en un plazo alcanzable. Esta visión pragmática se canaliza a través de: 1) un estudio exhaustivo de las herramientas y mecanismos tecnológicos existentes orientados a la lucha contra la Violencia de Género, 2) la propuesta de un nuevo sistema smart-wearable destinado a solventar algunas de las limitaciones tecnológicas encontradas en la actualidad, 3) un novedoso enfoque de clasificación de las emociones relacionadas con el miedo para desentrañar la relación entre las emociones y la fisi-

ología, y 4) la definición y publicación de un nuevo conjunto de datos multimodales para el reconocimiento de emociones en mujeres.

En primer lugar, se exploran y diseñan diferentes sistemas de clasificación del miedo utilizando un conjunto reducido de señales fisiológicas. Para ello se emplean conjuntos de datos abiertos junto con la combinación de técnicas de dominio temporal, frecuencial y no lineal. Este proceso de diseño está rodeado de compensaciones entre las consideraciones fisiológicas y las capacidades integradas. Esto último es de suma importancia debido al enfoque de esta investigación en la computación de borde. En esta primera tarea se destacan dos resultados: el sistema de clasificación del miedo diseñado que empleó los datos del conjunto de datos DEAP y logró un AUC de 81,60% y un Gmean de 81,55% de media para un enfoque independiente del sujeto, y sólo dos señales fisiológicas; y el sistema de clasificación del miedo diseñado que empleó los datos del conjunto de datos MAHNOB logrando un AUC de 86,00% y un Gmean de 73,78% de media para un enfoque independiente del sujeto, sólo tres señales fisiológicas, y una configuración Leave-One-Subject-Out. Se presenta una comparación detallada con otros sistemas de reconocimiento de emociones propuestos en la literatura, que demuestra que las métricas obtenidas están en línea con el estado del arte.

En segundo lugar, se presenta Bindi. Se trata de un sistema multimodal autónomo de extremo a extremo que aprovecha el IoT afectivo a través de sensores comerciales inteligentes auditivos y fisiológicos, la fusión multisensorial jerárquica y una arquitectura de servidor segura para combatir la violencia de género mediante la detección automática de situaciones de riesgo basada en un motor de inteligencia multimodal y la posterior activación de un protocolo de protección. En concreto, esta investigación se centra en el diseño de hardware y software de uno de los dos dispositivos informáticos de borde de Bindi. Se trata de un brazalete que integra tres sensores fisiológicos, actuadores, chips integrados de monitorización de energía y un System-On-Chip con capacidades inalámbricas. En este contexto, se presentan diferentes exploraciones del espacio de diseño embebido: evaluación del filtrado embebido, evaluación de la calidad de la señal fisiológica en línea, extracción de características y análisis del consumo de energía. Los resultados reportados en todos estos procesos se validan con éxito y, para algunos de ellos, incluso se comparan con

equipos de medición fisiológica estándar. Entre los diferentes resultados obtenidos respecto al diseño e implementación embebida dentro de la pulsera de Bindi, cabe destacar que su bajo consumo de energía proporciona una duración de la batería de aproximadamente 40 horas cuando se utiliza una batería de 500 mAh.

Finalmente, las particularidades de nuestro caso de uso y la escasez de conjuntos de datos multimodales abiertos que traten sobre tecnología inmersiva emocional, metodología de etiquetado considerando la perspectiva de género, distribución equilibrada de estímulos respecto a las emociones objetivo, y procesos de recuperación basados en las señales fisiológicas de las voluntarias para cuantificar y aislar la activación emocional entre estímulos, nos llevaron a la definición y elaboración del conjunto de datos Women and Emotion Multi-modal Affective Computing (WEMAC). Se trata de un conjunto de datos multimodal en el que 104 mujeres que nunca han experimentado la violencia de género que realizaron diferentes visualizaciones de estímulos relacionados con la emoción en un entorno de laboratorio. Los anteriores sistemas de clasificación binaria del miedo fueron mejorados y aplicados a este novedoso conjunto de datos multimodal. Por ejemplo, el sistema de reconocimiento de miedo multimodal propuesto utilizando este conjunto de datos reporta hasta un 60,20% y un 67,59% para el ACC y la puntuación F1, respectivamente. Estos valores representan un resultado competitivo en comparación con el estado del arte que se ocupa de casos de uso multimodal similares.

En general, esta tesis doctoral ha abierto una nueva línea de investigación dentro del grupo de investigación bajo el que se ha desarrollado. Además, este trabajo ha establecido una base sólida desde la que ampliar el conocimiento y continuar la investigación orientada a la generación tanto de mecanismos de ayuda a colectivos vulnerables como de tecnología de orientación social.

Introducción

1.1 Contexto y motivación

La violencia de género constituye una violación de los derechos humanos y de las libertades fundamentales reconocida por la Declaración de las Naciones Unidas sobre la Eliminación de la Violencia contra la Mujer de 1993 [13]. Esta declaración proporciona una definición clara y completa de lo que significa este tipo de violencia, que se recoge en su primer artículo al considerar cualquier acto de violencia, ya sea física, sexual o psicológica, basada en la pertenencia al género femenino. En 2020, la Comisión Europea amplió dicha definición y declaró que esta violencia incluye la ejercida contra las mujeres, los hombres y los niños [14]. En cuanto a las cifras concretas, desde el año 2000 hasta 2018, más de una de cada cuatro (27%) mujeres alguna vez emparejadas con edades comprendidas entre los 15 y los 49 años habían sufrido violencia física o sexual, o ambas, por parte de su pareja desde los 15 años [15]. Este problema no es nuevo, de hecho, en la Unión Europea, el primer principio de igualdad de trato entre hombres y mujeres se introdujo en 1975 en el Tratado de Roma [16]. Sin embargo, es en 2007 a través del Tratado de Lisboa [17] que la Comunidad Europea incluyó este principio entre los valores y objetivos de la Unión. Desde entonces, diferentes territorios de Europa han tomado estas medidas como base para sus leyes de violencia de género. A pesar de estos esfuerzos, seguía siendo necesario contar con un conjunto de normas o reglas comunitarias aplicables a este problema. Así, en 2011 se aprobó el Convenio del Consejo de Europa para prevenir y combatir la violencia contra las mujeres y la violencia doméstica, también conocido como Convenio de Estambul, que entró en vigor más tarde, en 2014 [18]. Este

convenio estableció un marco o instrumento común a partir del cual se establecen diferentes normas sobre prevención, protección, persecución y prestación de servicios para responder a las necesidades de las víctimas y de las personas en riesgo. Hasta la fecha, todos los miembros han firmado la convención y 35 de los 47 la han ratificado, aunque en julio de 2021, Turquía se convirtió oficialmente en el primer y único país en retirarse de ella. Hay que tener en cuenta que este país estuvo entre los precursores iniciales de este acuerdo. Además, el Convenio de Estambul creó un mecanismo de seguimiento encargado de controlar, informar y evaluar las medidas legislativas y de otro tipo adoptadas por los Estados ratificantes. Sin embargo, la aplicación de todas las recomendaciones del convenio no siempre es una tarea sencilla, ya que depende de los recursos de cada Estado. Por ello, se han puesto en marcha diferentes programas de financiación de la Unión Europea para facilitar la puesta en marcha de estas acciones (DAPHNE, PROGRESS, REC), pero siempre con un enfoque de aprendizaje mutuo aprovechando el mensaje dentro y fuera de la Comunidad Europea, ya que se concibe como un problema mundial [19]. Junto a estos acuerdos, convenios y programas de financiación, también se han creado diferentes pactos y organizaciones, como el Pacto Europeo por la Igualdad de Género (2011-2020) y el Instituto Europeo de la Igualdad de Género. Estas acciones han ido acompañadas de normativas europeas, que pretenden salvaguardar los derechos de las víctimas desde el punto de vista jurídico (UE 606/2013, 2012/29/UE).

Centrándonos en España, país en el que se ha desarrollado esta investigación, hay que destacar la aprobación por unanimidad en 2004 de la Ley Orgánica 1/2004 a través de la cual este país se convirtió en un referente fundamental en el mundo por la forma de afrontar este problema. En concreto, se trata de una ley integral contra la Violencia de Género, que considera también este tipo de violencia como la que se ejerce sobre las personas dependientes de la mujer cuando se abusa de ellas para causarles daño. Además, España fue uno de los primeros países en firmar el Convenio de Estambul en 2011, para posteriormente ratificarlo en 2014. Otra fecha clave en la hoja de ruta nacional fue la ratificación del Acuerdo Nacional contra la Violencia de Género por parte de los diferentes Grupos del Parlamento Nacional, los Gobiernos Autonómicos y las Entidades Locales en diciembre de 2017. Al igual que muchos otros países de la Comunidad Europea, España está dividida

en Comunidades Autónomas que, al margen de lo que se refiere a la aplicación legislativa nacional, tienen sus propias leyes y normas regionales. Por ejemplo, la Comunidad Autónoma de Madrid cuenta con una dilatada experiencia en materia de políticas y actuaciones legales en materia de Violencia de Género. Uno de los mayores logros de esta comunidad en relación con este tema fue la aprobación de la primera Ley de Ordenación Regional nº 5/2005, de 20 de diciembre, de medidas de protección integral contra la Violencia de Género. Además, esta comunidad también creó un órgano institucional regional específico para asesorar y evaluar la integración de dichas políticas, éste fue concebido como el observatorio regional de Violencia de Género en 2003 (decreto 256/2003, de 27 de noviembre). Por último, en 2016 presentaron una estrategia integral de prevención y lucha contra la Violencia de Género, que consistía en una serie de medidas y acciones a desarrollar desde ese año hasta 2021 y que reflejaba el compromiso del gobierno madrileño de luchar por la erradicación de este problema. A pesar de todas estas medidas nacionales y regionales, la oficina del Gobierno contra la Violencia de Género contabilizó un total de 1117 mujeres asesinadas desde 2003 hasta octubre de 2021, Fig. 1-1.

Basándonos en los hechos anteriores, podemos concluir que la violencia de género es un problema de emergencia que lleva a la sociedad a enfrentarse a él utilizando diferentes perspectivas y adoptando un enfoque multidisciplinar. Por ejemplo, desde el punto de vista sociológico, la educación y la concienciación informativa sobre la prevención y el combate de la violencia contra las mujeres es fundamental. Además, la perspectiva tecnológica es también un aspecto fundamental relacionado con el desarrollo de nuevas tecnologías emergentes que faciliten la creación de nuevas plataformas para prevenir y responder a la violencia de género [20]. De hecho, estas y otras perspectivas como la jurídica, la psicológica y la médica, entre otras, están vinculadas y colaboran en la búsqueda de soluciones cotidianas para combatir este problema. Esta reivindicación multidisciplinar está fuertemente respaldada por un amplio abanico de profesionales que trabajan estrecha y personalmente con las víctimas (agentes de la autoridad, jueces y psicólogos) [21]. Sin embargo, identifican dos grandes inconvenientes de los actuales instrumentos públicos coincidiendo en que se deberían invertir más esfuerzos y mejor organizados en los mecanismos o herramientas de prevención y en la formación de los profesionales que tratan directamente con

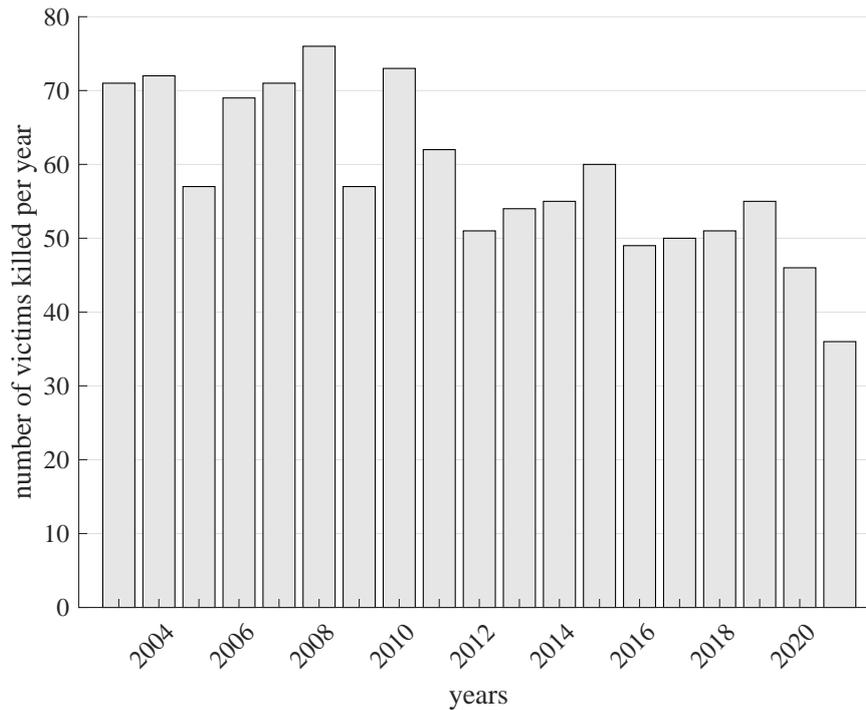


Figure 1-1: Número total de víctimas de violencia de género asesinadas desde 2003 hasta octubre de 2021. Datos proporcionados por [1].

las víctimas para evitar una revictimización más dañina.

A pesar del esfuerzo institucional, desarrollar soluciones aplicando un enfoque multidisciplinar para crear comunidades más seguras es una tarea difícil. Sin embargo, es necesario que todas las perspectivas mencionadas se comprometan y cooperen de forma más estrecha para combatir la violencia de género de forma eficaz. Debido a la transformación digital que están experimentando algunos países de Europa, están tratando de aprovechar el desarrollo de las nuevas tecnologías para proporcionar servicios a las comunidades, de los cuales, algunos están destinados a tratar este problema en cuestión. Por ejemplo, en España se han puesto en marcha diferentes servicios, como VioGen [22], ATENPRO [23], y COMETA [9]. En primer lugar, VioGen permite estimar el nivel de riesgo al que se enfrenta una víctima de violencia de género y determinar el tipo y grado de protección adecuados para ella. Este nivel de riesgo se actualiza continuamente en función de su situación jurídica y social. Esta herramienta es el resultado de una intensa investigación del Ministerio del Interior español con varios grupos de investigación universitarios españoles con expertos en psicología, criminología y sociología. En segundo lugar, ATENPRO es un servicio que proporciona una línea directa y 24 horas al día a las fuerzas de seguridad españolas a través de un botón del pánico. En concreto, se entrega a la

víctima un dispositivo móvil que permite una comunicación continua en cualquier momento y circunstancia. Dicha comunicación es atendida por un centro de asistencia telefónica especializado, donde asistentes específicamente formados dan una respuesta adecuada para atender este tipo de situaciones en tiempo real. Por último, COMETA es un sistema concebido como un conjunto de dispositivos de control telemático que se adopta cuando se dicta una orden de alejamiento del agresor. En este caso, tanto la víctima como el agresor reciben un dispositivo de geolocalización con capacidades básicas de telecomunicación de voz y datos para comunicarse con el centro de control. El agresor también debe llevar un ligero dispositivo de radiofrecuencia en forma de pulsera que se conecta a los dispositivos de geolocalización. Aunque COMETA ofrece una solución tecnológica para luchar contra la violencia de género, su limitada duración de la batería y su tecnología anticuada presentan una elevada tasa de falsos positivos [24, 25], además del riesgo de acoso para las víctimas.

Además de los esfuerzos de las organizaciones públicas gubernamentales, el sector privado también se anima en todo el mundo a aportar una solución al problema de la violencia de género. De hecho, diferentes iniciativas privadas salen continuamente con ideas para prevenir y evitar este problema. Por ejemplo, la Fundación XPrize lanzó en 2018 un concurso mundial de 1 millón de dólares para desafiar a equipos de todo el mundo a aprovechar la tecnología para empoderar a las mujeres a responder a las agresiones sexuales. El objetivo de este concurso era desarrollar una solución tecnológica capaz de activar alertas de emergencia de forma autónoma, transmitir información a una red de respondedores de la comunidad y ser lo más asequible posible, todo ello en 90 segundos. Los siete finalistas utilizaban tecnología vestible con los últimos protocolos de comunicación inalámbrica conectados con los diferentes intervinientes o incluso con las fuerzas del orden. Sólo uno de ellos incluyó en sus dispositivos capacidades de computación afectiva para hacer un seguimiento perfecto de los niveles de amenaza emocional mediante el uso de información fisiológica cardíaca. El seguimiento de dicha información se ha demostrado como un sólido indicador emocional [26].

Teniendo en cuenta toda la información revisada y apuntando a la generación de nuevos mecanismos de prevención y lucha, se podría aprovechar una nueva her-

ramienta autónoma, inteligente, discreta, conectada, con *edge-computing* (tecnología basada en la computación en el nodo, e.g. en dispositivos con muy pocos recursos) y preparada para ser usada como portable o vestible, capaz de detectar y alertar cuando una usuaria está bajo una situación de Violencia de Género. Sobre esta base, el trabajo de investigación descrito en este documento se centra en proporcionar una solución tecnológica inteligente que ayude a tratar el problema planteado. Este sistema se denominará en adelante BINDI a lo largo de todo el documento, y ha sido desarrollado por el grupo UC3M4Safety de la Universidad Carlos III de Madrid. En concreto, esta investigación se centra en el diseño, desarrollo e implementación de uno de los tres dispositivos que componen el sistema, que es una pulsera inteligente que utiliza computación afectiva embebida basada en la monitorización fisiológica para la detección de estados emocionales relacionados con el miedo. Los otros dos dispositivos son un colgante inteligente y una aplicación para teléfonos inteligentes. El primero capta el audio a la carta, mientras que el segundo realiza la fusión de datos fisiológicos y físicos y gestiona las alarmas de emergencia para enviarlas a una red de respondedores de confianza o incluso a las fuerzas del orden. La naturaleza del problema a tratar hizo que este trabajo se derivara por un enfoque multidisciplinar, reuniendo conocimientos de los estudios de género, la electrónica, la telemática, la fisiología, las tecnologías del habla y el audio, y la informática o computación afectiva.

1.2 Alcance de esta disertación

Esta investigación pretende aportar una nueva visión en el desarrollo de herramientas para prevenir y evitar situaciones de riesgo de Violencia de Género e, incluso, agresiones, desde una perspectiva tecnológica, pero sin dejar de lado las diferentes consideraciones sociológicas relacionadas con el problema.

Desde un punto de vista teórico, este trabajo propone una nueva forma de utilizar las señales fisiológicas y el reconocimiento de emociones para proporcionar soluciones autónomas, vestibles y discretas para proteger a las personas vulnerables. El objetivo en ese aspecto es desentrañar la relación entre las señales fisiológicas y las emociones relacionadas con el miedo, proporcionando alternativas a los sistemas de clasificación de reconocimiento de emociones ya propuestos en la literatura, nuevas arquitecturas

de sistemas de monitorización fisiológica listos para llevar, nueva integración de sensores e implementación embebida en dispositivos vestibles, nuevas técnicas para mitigar el ruido de los artefactos de movimiento fisiológico, y realizando un estudio analítico de toda la solución propuesta.

Desde un punto de vista práctico, se proporcionaron diferentes sistemas de reconocimiento de emociones binarias de miedo basados en bases de datos disponibles abiertamente que contienen emociones evocadas no actuadas para un conjunto de voluntarios. Además, se desarrolló e implementó una nueva solución hardware wearable para la detección de la violencia de género, basada en la familia de procesadores ARM Cortex-M[®], utilizando tres de los sensores fisiológicos más discretos, Photoplethysmography (PPG), EDA, y SKT, y comunicaciones inalámbricas de baja potencia. Este dispositivo forma parte del sistema BINDI, que ha sido desarrollado junto con el grupo UC3M4Safety. Los requisitos para el sistema completo son el menor consumo de energía posible, una integración discreta y vestible de todos los dispositivos y componentes, y el menor tiempo de cálculo para las diferentes arquitecturas de procesamiento digital para lograr el mayor tiempo de respuesta posible. Por último, se generó una nueva base de datos utilizando estímulos inmersivos y estímulos específicos orientados al caso de uso de la violencia de género. Esto último es especialmente relevante ya que la base de datos generada es única en la literatura.

En concreto, los objetivos de esta investigación son los siguientes:

- Proponer un nuevo enfoque para detectar las emociones relacionadas con el miedo haciendo uso de las diferentes teorías emocionales e indicadores fisiológicos afectivos.
- Derivar nuevos sistemas de detección de emociones relacionadas con el miedo, listos para ser llevados, utilizando bases de datos abiertas que han utilizado señales fisiológicas para el reconocimiento de emociones.
- Tratar el comportamiento fisiológico (cuasi-estacionario, no estacionario y no lineal) y proponer nuevas técnicas de procesamiento digital que lo tengan en cuenta para los sistemas relacionados con el miedo de inferencia rápida.
- Analizar y estudiar diferentes restricciones de integración que deben considerarse en los sistemas de computación afectiva vestibles.
- Diseñar una nueva solución de hardware vestible para desplegar las arqui-

tecturas de sistemas de detección relacionados con el miedo propuestas. En este punto, afrontaríamos la implementación de hardware y software embebido hacia una solución discreta, autónoma, de bajo consumo, inalámbrica y conectada.

- Comparar los resultados obtenidos con arquitecturas similares publicadas y soluciones comerciales para la prevención de la violencia de género.
- Generar una nueva base de datos enfocada al caso de uso específico de Violencia de Género que recoja las respuestas fisiológicas, físicas y emocionales a los estímulos inmersivos.

1.3 Esquema del documento

El documento se divide en tres partes. En la primera parte se revisa la relación entre las emociones y las señales fisiológicas, investigando las diferentes teorías emocionales y los indicadores fisiológicos afectivos. También se analiza el marco general de las bases de datos utilizadas en la literatura, que tratan del reconocimiento de emociones mediante el uso de señales fisiológicas. En la segunda parte, se presenta la aplicación de la teoría revisada en la primera parte a la propuesta y análisis de un nuevo sistema de reconocimiento de emociones relacionadas con el miedo. En la tercera parte se presentan los resultados tanto del hardware como del software embebido del sistema edge-computing desarrollado para el reconocimiento binario del miedo. Finalmente, en la última parte se presenta la nueva base de datos para el reconocimiento de emociones centrada en la detección del miedo. Además, a lo largo de todo el documento se exponen y analizan las limitaciones de la integración de los wearables y las dinámicas fisiológicas a tener en cuenta.

Así, el esquema del documento es el siguiente:

Parte I

El Capítulo 2 describe las nociones básicas y avanzadas necesarias para comprender bien de los temas de esta investigación. En concreto, se estudian las teorías emocionales, las metodologías de clasificación de las emociones humanas, las herramientas para la elicitación de las emociones y la cuantificación de los indicadores fisiológicos afectivos. Todos estos temas se apoyan en referencias del estado del arte que ayudarán a comprender el contenido original que se ofrece en los siguientes

capítulos.

El Capítulo 3 se centra en proporcionar un análisis en profundidad sobre la estructura y los procedimientos experimentales utilizados para la generación de bases de datos diseñadas para el reconocimiento de emociones. Además, también se detalla y explica cada parte de la cadena completa de procesamiento de datos para la tarea de diseño de sistemas de computación afectiva utilizando dichas bases de datos.

Parte II

El Capítulo 4 aborda uno de los principales objetivos de la tesis. Se trata del diseño y la validación de sistemas novedosos de reconocimiento del miedo basados en un conjunto reducido de señales fisiológicas. Se seleccionan diferentes bases de datos públicas disponibles para diseñar dos sistemas principales de reconocimiento de emociones binarias de miedo. Las limitaciones encontradas en dichas bases de datos se detectan y se tienen en cuenta para el trabajo presentado en el Capítulo 6. Además, los resultados presentados en este capítulo se comparan con el estado actual de la técnica.

Parte III

El Capítulo 5 detalla el proceso de diseño e integración de una nueva solución de hardware wearable para desplegar partes de las arquitecturas del sistema de detección de miedo propuesto en el Capítulo 4. Por un lado, se contextualiza esta nueva solución vestible analizando la tecnología actual que se aplica para la prevención y el combate de la violencia de género. Por otro lado, se detallan y explican exhaustivamente los retos de diseño e integración, tanto desde el punto de vista del hardware como del software.

Capítulo 6 desarrolla una de las principales contribuciones de esta investigación. Se trata de la generación de un novedoso conjunto de datos multimodales, Dataset-Name, consistente en experimentos realizados en un entorno de laboratorio con sólo mujeres voluntarias. Además, se emplean las diferentes arquitecturas de computación afectiva propuestas y presentadas en el Capítulo 4, utilizando los datos recogidos en este conjunto de datos. Por último, también se presenta una aproximación multimodal mediante la fusión de datos fisiológicos y del habla para proporcionar una primera línea de base a tener en cuenta para futuros trabajos.

Parte IV

Capítulo 7 concluye esta investigación y ofrece algunas sugerencias sobre la posible ampliación de este trabajo en un futuro próximo.

Parte I

Emociones humanas, señales fisiológicas y computación afectiva

Capítulo 2

Clasificación de las emociones y cuantificación fisiológica

Este capítulo se basa en cuatro temas esenciales, necesarios para el desarrollo de esta investigación: las teorías emocionales, las metodologías de clasificación de las emociones humanas, las herramientas para la elicitación de las emociones y la cuantificación de los indicadores fisiológicos afectivos. En primer lugar, se analizan cronológicamente las principales teorías emocionales y se evalúan con vistas al caso de uso específico de esta investigación. Para ello, se relaciona su orden de activación emocional con los efectos relativos a la elaboración de sistemas informáticos afectivos. También se presentan y analizan las diferentes metodologías de clasificación de las emociones humanas revisando sus ventajas e inconvenientes y explotando su relación con las teorías emocionales anteriores. A continuación, se realiza un análisis exhaustivo de los efectos en la modulación de las emociones por parte de factores intrapersonales como los rasgos de personalidad, la cognición, la atención y el sesgo de género, que profundiza en la aportación de la contextualización personal dentro de las metodologías de clasificación de las emociones humanas. Posteriormente, se presenta un nuevo enfoque pragmático para fusionar esos métodos de clasificación de las emociones humanas, que se aplicará posteriormente en los modelos generados con el fin de limitar la identificación o el reconocimiento de la emoción del miedo. En segundo lugar, se comparan las diferentes herramientas para evocar emociones. Por último, se presenta y analiza un conjunto reducido de señales fisiológicas y su relación con las emociones y los modelos emocionales. Aunque se ha demostrado que

muchas otras señales fisiológicas garantizan características específicas relacionadas con las emociones, no pueden ser adquiridas por sensores poco visibles para su uso diario. Por lo tanto, se estudia a fondo la información relacionada con las emociones que puede proporcionar cada uno de los sensores fisiológicos más discretos, así como su estado actual de desarrollo e integración en dispositivos portátiles y los retos actuales.

2.1 Supuestos y definiciones

Antes de entrar en detalles sobre los diferentes temas que se van a tratar en este capítulo, es necesario dar algunas premisas y definiciones.

- En primer lugar, las emociones son un compuesto de reacciones conductuales, procesos cognitivos subjetivos y cambios fisiológicos, en su mayoría desencadenados por estímulos emocionales [27].
- Se denomina estímulo emocional a cualquier tipo de material o proceso a través del cual se provoca una emoción específica a una persona. Derivan en respuestas emocionales específicas.
- Las respuestas emocionales pueden cuantificarse o medirse utilizando autoinformes subjetivos, información física y/o fisiológica, y cualquier tipo de datos procedentes de la persona sometida a la elicitación de la emoción y recogidos durante dicho proceso.
- Las bases de datos de reconocimiento de emociones son aquellas que utilizan estímulos emocionales bajo una presentación o método de interacción específico para recoger diferentes respuestas emocionales. Toda esa información puede utilizarse posteriormente para entrenar sistemas informáticos afectivos inteligentes.
- Los sistemas de computación afectiva utilizan todos los elementos anteriores para generar un sistema de reconocimiento de emociones entrenado.

2.2 Teorías Emocionales

A pesar de la teoría emocional considerada, se está de acuerdo en que las emociones intervienen directamente en el ajuste de nuestra respuesta a un estímulo externo. Sin embargo, no existe un acuerdo común respecto al orden en que se desencadenan los

compuestos de las emociones al recibir dicho estímulo. Diferentes teorías emocionales a lo largo de la historia se han postulado tratando de abordar este proceso, algunas basadas y otras refutando a las anteriores o predecesoras. Por ejemplo, las teorías emocionales de Darwin [28], James-Lange [29], y Cannon-Bard [30], que se preceden respectivamente, fueron las tres principales teorías de la emoción antes de 1950 y cada una de ellas sigue un orden de activación diferente. Darwin fue el primero en intentar averiguar el origen de las emociones y sus mecanismos de activación. Aunque su trabajo no tenía en cuenta la información fisiológica dentro del proceso de reacción emocional, propuso tres principios diferentes que estaban profundamente vinculados al proceso cognitivo de las emociones. Entre esos principios, destaca el primero (principio de los hábitos útiles), que se basa en la asociación entre diferentes acciones y estados mentales específicos, ya que sirvió de base para algunas de las siguientes teorías cognitivo-emocionales. La teoría de James-Lange afirma que las emociones se vuelven conscientes para la persona una vez que toda la información fisiológica ha sido procesada dentro del neocórtex, Fig. 2-1. Por otro lado, la teoría de Cannon-Bard refuta la afirmación de James-Lange de que la respuesta fisiológica se produce al mismo tiempo que la reacción de la emoción y de forma independiente.

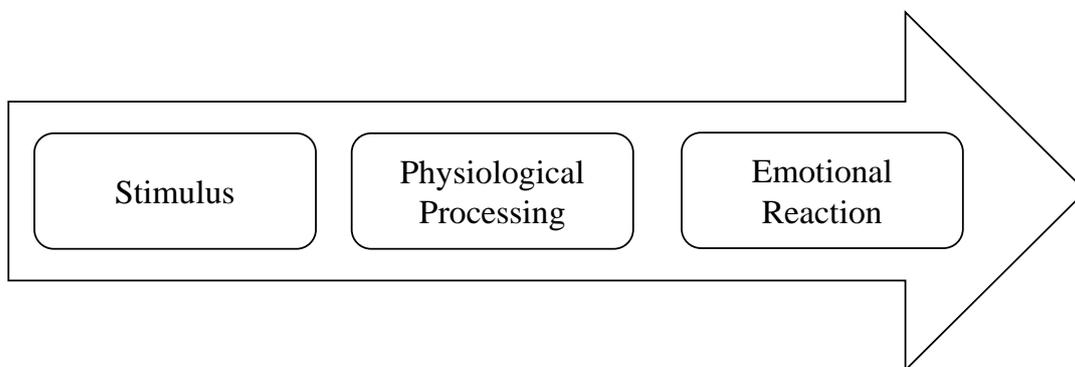


Figure 2-1: Orden de activación emocional para la teoría de James-Lange.

Estas tres teorías fueron muy criticadas, ya que las dos primeras carecen de evidencia empírica y la de Bard impone la total independencia de las reacciones fisiológicas y emocionales. Así, esto dio lugar al nacimiento de las teorías emocionales cognitivas, en las que el contexto de la situación y nuestra experiencia previa también afectan directamente a esa respuesta conductual. Por ejemplo, la teoría de la valoración de las emociones, desarrollada principalmente por Magda Arnold y Richard Lazarus [2], entre otros, fue una de las primeras teorías emocionales cognitivas. Se

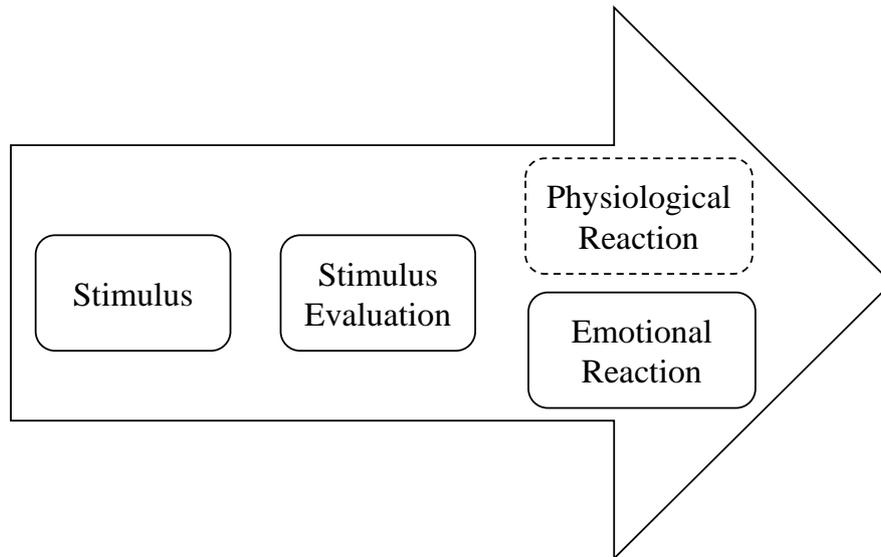


Figure 2-2: Orden de activación emocional para la teoría desarrollada por Magda Arnold y Richard Lazarus [2].

basa en el supuesto de que las emociones están directamente determinadas por nuestras apreciaciones o evaluaciones de los estímulos, que pueden provocar reacciones fisiológicas y emocionales específicas y simultáneas en diferentes personas, Fig.2-2. Una de las afirmaciones más controvertidas de esta teoría es que las emociones podrían originarse directamente a partir de nuestras propias valoraciones sin necesidad de una excitación fisiológica. Este hecho implica la posible falta de correlación de la respuesta fisiológica con respecto al estímulo externo específico. Junto con la aparición de la teoría de la valoración de las emociones, Albert Ellis introdujo la teoría racional-emotiva [3], que incluye el proceso de valoración o evaluación al afirmar que las emociones se ven directamente afectadas por nuestros pensamientos o creencias, pero no descuida la respuesta fisiológica a un estímulo. Además, esta última va precedida de la reacción emocional a diferencia de las otras teorías, Fig. 2-3. Aunque las teorías cognitivas son ampliamente aceptadas, se observan muchas variaciones dentro de ellas. En la actualidad, el papel completo de la cognición sobre las emociones sigue siendo una cuestión abierta [31].

A pesar del enorme esfuerzo por desentrañar el paradigma del origen de la emoción a lo largo de los años, todavía no existe una definición consensuada para la emoción ni un orden de activación de los diferentes elementos que intervienen en ella. En concreto, en nuestro caso, la teoría en la que se basa este trabajo de investigación es la teoría racional-emotiva. Esta teoría permite la cuantificación emocional a través

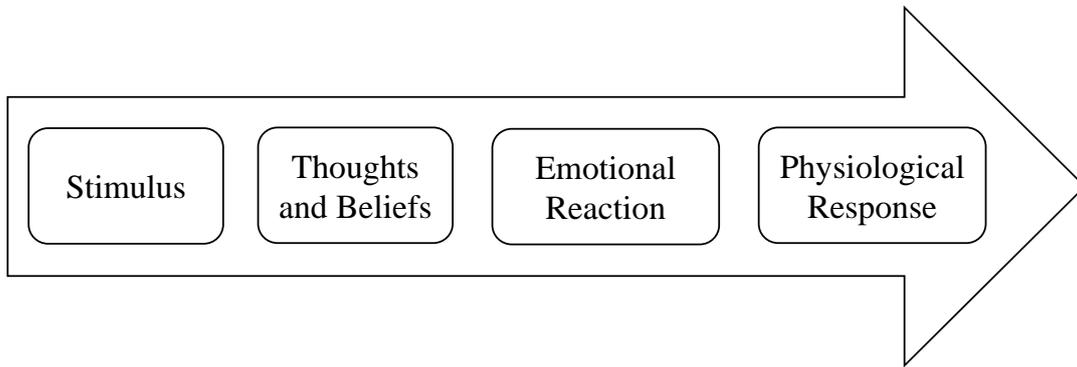


Figure 2-3: Orden de activación para la teoría emocional racional-emotiva por Albert Ellis [3].

de la monitorización fisiológica y admite la repercusión de los pensamientos y creencias sobre la emoción sentida. Esta última es, efectivamente, un factor clave a la hora de abordar el desarrollo de herramientas de reconocimiento de emociones para prevenir y evitar situaciones de violencia de género. Las experiencias vitales de cada víctima de violencia de género son diferentes, y la necesidad de no sólo considerar las interdiferencias reales entre las víctimas, sino también las intradiferencias individuales a lo largo del tiempo, es esencial para proporcionar una solución mejor y más inteligente, integrada social y tecnológicamente.

2.3 Clasificación de las emociones humanas

En el contexto de la categorización e identificación inequívoca de las emociones, la literatura presenta un amplio esfuerzo por llegar a un etiquetado o modelado estandarizado respecto a las mismas [32]. En la comunidad científica de la informática afectiva, esto se traduce en la identificación de la emoción experimentada por una persona a través de diferentes datos como los gestos faciales, la voz, la postura y las señales fisiológicas. Sobre esta base, diferentes autores han propuesto mapeos entre estados afectivos y patrones o variaciones observadas en estas modalidades. Nótese que el término modalidad se refiere a las fuentes de información. Así, la detección de emociones podría definirse como un problema de reconocimiento de patrones [33]. Sin embargo, como en cualquier problema de reconocimiento, el sistema necesita ser enseñado. El proceso de aprendizaje más común en el reconocimiento de emociones es mediante el uso de un clasificador supervisado, que se define como un proceso de aprendizaje supervisado a través de información etiquetada. Esta metodología re-

quiere tener diferentes muestras de información etiquetada o asignada a la emoción correcta. La información recogida de todas las modalidades se distribuirá en función de las etiquetas o clases asignadas a cada muestra adquirida. Por lo tanto, disponer de etiquetas o modelos emocionales precisos afectará en gran medida al rendimiento del sistema diseñado.

En los siguientes apartados se revisan las principales metodologías de clasificación de emociones humanas, así como sus ventajas e inconvenientes. Además, se resumen y analizan algunos factores clave a la hora de tratar la experiencia emocional y la personalidad y cómo pueden afectar al proceso de etiquetado de las emociones. Por último, se conectan las emociones relacionadas con el miedo utilizando los diferentes métodos de clasificación de emociones humanas, lo que proporciona un nuevo enfoque para tratar el reconocimiento del miedo en situaciones de violencia de género.

2.3.1 Clasificación discreta de las emociones

Ya en el siglo XIX, Darwin propuso que las emociones eran discretas o categóricas, es decir, que podían dividirse en módulos como el miedo, el asco, la ira, etc. [28]. Aunque no proporcionó ninguna especificación sobre el número exacto de esas emociones. Desde entonces, diferentes psicólogos y fisiólogos han utilizado el mismo o similar enfoque categórico para tratar las emociones. Este enfoque se basa en el concepto de emociones básicas, que son universalmente reconocibles y transculturales. Así, se afirma que cada una de estas emociones básicas tiene asociados diferentes patrones fisiológicos, así como diferentes efectos en la voz, los gestos faciales, la postura, etc. Sin embargo, a lo largo de los años, ha habido diferentes teóricos psicológicos y fisiológicos que aportan listas de emociones primarias, cada uno de ellos basándose en diferentes criterios para definir cuáles son emociones básicas y cuáles no. Por ejemplo, Ekman y Friesen [34, 35] aportaron datos para reafirmar la teoría expuesta por Darwin, y basaron la mayor parte de sus primeras investigaciones para las emociones básicas en expresiones faciales inequívocas entre culturas. Basándose en este criterio, afirmaron la existencia de seis emociones básicas: ira, asco, miedo, sorpresa, tristeza y alegría. En 1972, viajaron a Papúa Nueva Guinea y conocieron a la tribu Fori. Presentaron diferentes imágenes a la tribu, que fue capaz de identificar las seis emociones diferentes. Después, mostraron imágenes de expresiones faciales de la gente de la tribu Fori, con las mismas emociones, a personas de otras nacional-

Table 2.1: Main categorical models of emotions developed since 19th century.

Autores(s)	Afirmaciones	Basadas en
Ekman & Friesen [35]	Seis emociones básicas: ira, asco, miedo, alegría, tristeza y sorpresa	Gestos faciales universales
Carroll Izard [36]	Diez emociones básicas: ira, desprecio, asco, angustia, miedo, culpa, interés, alegría, vergüenza y sorpresa	Derivadas del comportamiento
Robert Plutchik [37]	Ocho emociones básicas: ira, aceptación, alegría, anticipación, miedo, disgusto, tristeza, y sorpresa	Evolución biológica y supervivencia
Nico Frijda [38]	Seis emociones básicas: el deseo, la felicidad, interés, sorpresa, asombro y tristeza	Cambio en acción
Oatley & Laird [39]	Cinco emociones básicas: ira, asco, ansiedad, felicidad y tristeza	Estados cognitivos

idades y culturas. Concluyeron que las emociones se interpretaban correctamente y afirmaron que las emociones son universalmente reconocibles por las expresiones faciales.

Cronológicamente después de Darwin, algunas de las principales contribuciones a este enfoque de clasificación de las emociones humanas se resumen en el cuadro 2.1. Entre estos autores, Robert Plutchik es conocido por haber creado la rueda de las emociones [37], que fue una de las primeras representaciones gráficas que intentó ilustrar cómo se relacionaban las emociones desde un punto de vista categórico. Algunos de los aspectos clave del modelo de Plutchik han influido en propuestas posteriores de clasificaciones discretas de las emociones humanas. Por ejemplo, introdujo el concepto de emociones opuestas, que no pueden experimentarse al mismo tiempo, e incluso propuso que las emociones pueden sentirse con diferente intensidad, lo que lleva a transformar la rueda en un modelo discreto multidimensional de las emociones. En realidad, el aspecto más discutido del enfoque de Plutchik es el primer concepto, ya que muchos científicos no están de acuerdo citando diferentes ejemplos en los que se pueden desencadenar emociones opuestas al mismo tiempo. Cabe destacar que los colaboradores citados en la tabla 2.1 no se ponen de acuerdo en el número y la naturaleza de las emociones básicas, incluso algunos de estos autores a lo largo de los años han modificado su número declarado de emociones básicas. En el caso de Ekman, considerado un pionero dentro del campo de la investigación emocional en el presente siglo, su investigación se basó en la evidencia existente de siete emociones básicas a las que posteriormente se incluyeron otras diez emo-

ciones adicionales agradables [40]. Otro psicólogo muy conocido e influyente es Nico Frijda, que identificó dieciocho emociones básicas en su primer modelo emocional propuesto [38], que posteriormente evolucionó a un total de seis [41]. Además del número de emociones básicas, estas propuestas difieren también en cuanto a la teoría emocional en la que se basan. Como puede observarse, la motivación partió de un punto de vista darwiniano (expresión facial universal), que evolucionó introduciendo factores conductuales y biológicos. Finalmente, Oatley y Laird introdujeron el concepto cognitivo al afirmar que las emociones son estados de base cognitiva encargados de coordinar procesos cuasi-autónomos dentro del sistema nervioso [39]. Nótese que este último hecho está en consonancia con algunas de las afirmaciones de la teoría de la valoración de las emociones explicada en la sección 2.2, que admite la posibilidad de una respuesta fisiológica no correlacionada con respecto a la emoción sentida.

Independientemente de la diversidad de modelos categóricos, la mayoría de ellos incluyen la ira, la felicidad, la tristeza y el miedo. Sin embargo, debido al origen cultural de cada uno de los autores, algunos postulan la ira y otros la rabia para referirse a la misma emoción, al igual que cuando utilizan el miedo y la ansiedad. A pesar de estos desacuerdos, la literatura suele referirse a estas emociones como respuestas primarias o básicas y universales a los estímulos. Hay que tener en cuenta que existen más modelos categóricos que los citados en la tabla 2.1. En relación con estos hechos, una encuesta realizada por Ekman en 2015 [42] supuso un avance clarificador hacia la estandarización de las emociones básicas. En concreto, se preguntó a 248 científicos del ramo, mediante la misma encuesta, con el objetivo de obtener alguna evidencia de universalidad en alguna faceta de las teorías emocionales y modelos categóricos. El mayor acuerdo se obtuvo con sólo cinco de las dieciocho emociones propuestas: ira (91%), miedo (90%), asco (86%), tristeza (80%) y felicidad (76%). La encuesta concluye afirmando que, aunque es necesario trabajar para reducir los desacuerdos, existe un acuerdo sobre las emociones básicas.

2.3.2 Clasificación dimensional de las emociones

Para paliar el problema derivado de la aplicación de diferentes términos categóricos a un mismo concepto emocional y el análisis de las emociones complejas mediante la combinación de diferentes emociones básicas, se hace necesaria la utilización de otras escalas y métodos de cuantificación más allá de los modelos categóricos. Así,

diferentes autores han introducido lo que se conoce como dimensiones del estado afectivo. Por ejemplo, Wundt fue el primero en introducir el uso de dos dimensiones para clasificar e identificar las emociones ya en 1896 [43]. Introdujo las dimensiones agradable-desagradable y baja-alta intensidad, que fueron utilizadas por muchos otros investigadores en los años siguientes. Otro de los primeros autores relevantes fue Osgood [44], que utilizó tres factores diferentes para evaluar los estados afectivos. Estos factores fueron definidos como evaluación, actividad y potencia. Además, como se ha comentado en el apartado anterior, Plutchik afirmaba que las emociones se sienten con diferente intensidad. Este hecho implica que incluso los autores que han contribuido al desarrollo de modelos categóricos de las emociones, necesitaban asumir la existencia de algún tipo de dimensión para distinguir las emociones complejas de las básicas. En este sentido, la inclusión de dimensiones cuantitativas permite la creación de un espacio multidimensional, en el que el sesgo categórico disminuye y las emociones básicas y complejas pueden ser igualmente identificadas. Además, la autocalificación o el autoinforme de estas dimensiones después de cada estímulo presentado a la persona, al igual que el autoinforme cuando se utilizan emociones discretas, tiene en cuenta tanto las diferencias culturales como las experiencias previas del mismo estímulo. Sin embargo, en este caso multidimensional, la perfecta comprensión de las distintas dimensiones y la autoevaluación emocional presentan una ardua tarea.

Uno de los modelos dimensionales más utilizados es el modelo circumplex, postulado por Russel [45]. Este modelo se basa en dos dimensiones diferentes, valencia o placer (P) y arousal (A), que pueden interpretarse como las dimensiones modernas de las propuestas por Wundt. En concreto, ambas están concebidas para medir diferentes aspectos clave del estado afectivo actual. Así, la dimensión de la valencia representa la naturaleza positiva o negativa del estado afectivo, mientras que el arousal indica la excitación o activación que proporciona ese estado afectivo. A pesar de que el modelo circumplex ha sido uno de los modelos dimensionales más utilizados, la adición de otros ejes ortogonales conduce a un espacio multidimensional más completo. Por ejemplo, Mehrabian [46] introdujo la dominancia como una nueva dimensión emocional y así propuso el modelo de placer, excitación y dominancia PAD. Posteriormente, este modelo ha demostrado ser útil para desentrañar las emociones

que se sitúan en el mismo cuadrante para un espacio emocional bidimensional (PA). En este sentido, Demaree et al. en [47] afirmaron que se requiere una clasificación tridimensional de las emociones (PAD) para identificar adecuadamente un estado afectivo. Compararon la distinción miedo-enojo utilizando el modelo PA y el modelo PAD. Como resultado, Demaree et al. aseguraron que sólo la dominancia puede describir emociones como el miedo y la ira, asociadas a la sumisión y la dominancia respectivamente.

Aunque en las últimas décadas, las clasificaciones dimensionales de las emociones humanas han ganado atención, todavía existe un profundo debate sobre la interpretación de estas dimensiones y cómo se explica y aplica dicha interpretación. Este último hecho es clave cuando se trata de comparar estudios de diferentes investigadores que han utilizado las mismas dimensiones pero que las han explicado a los voluntarios de forma diferente. Dentro de este contexto y tratando de aportar claridad en cuanto a la definición, comprensión y explicación de las diferentes dimensiones, Bakker et al. [4] relacionaron el modelo PAD con las respuestas afectivas, cognitivas y conativas o el modelo de afecto, cognición y comportamiento (modelo ABC de actitudes), Fig. 2-4. Llegaron a la conclusión de que el placer, la excitación y la dominancia pueden utilizarse junto con el modelo ABC y la distinción entre sentir, pensar y actuar, respectivamente, para comprender mejor las dimensiones originales y describir las experiencias ambientales posteriores. Esta última conclusión es clave para este trabajo de investigación, ya que pone de manifiesto la necesidad de incluir el proceso de valoración o evaluación de la emoción considerando los efectos del entorno externo. También indicaron la necesidad de realizar más investigaciones centradas en las experiencias el modelo PAD para concebirlo como un modelo de emoción dimensional sólido y probado.

A diferencia de la información emocional cualitativa proporcionada por las clasificaciones discretas, las clasificaciones dimensionales ofrecen métricas cuantitativas específicas sobre los estados afectivos. Esto puede considerarse una ventaja a la hora de diseñar cualquier sistema de reconocimiento automático de emociones, ya que las etiquetas dimensionales autodeclaradas son más específicas y tienen más probabilidades de ser utilizadas para el cálculo afectivo. Sin embargo, localizar en coordenadas dimensionales tanto las emociones básicas como las complejas no es

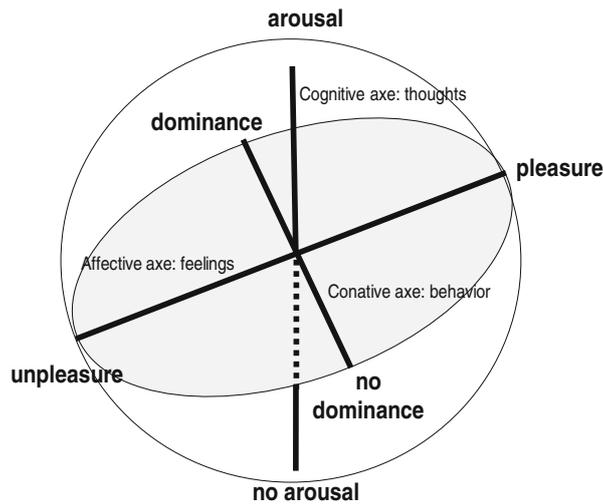


Figure 2-4: Relación entre los modelos PAD y ABC. Este último es el modelo de actitudes por Bakker et al.. Esta relación trata de dar una visión más clara que la propuesta inicial del modelo PAD [4].

una tarea fácil. Diferentes autores han realizado estudios con una población relativamente grande con el objetivo de definir las posiciones dimensionales exactas para las emociones dentro de los modelos dimensionales. Por ejemplo, Fontaine et al. [5] utilizaron cuatro dimensiones para localizar el espacio exacto de 24 emociones discretas utilizando más de 600 participantes, escalas Likert de 9 puntos para cada dimensión y considerando tres orígenes culturales diferentes. Estas emociones se tomaron del conocido instrumento GRID¹, que comprende 144 características de emoción representativas de los diferentes componentes de las emociones. Lograron mapear los 24 términos emocionales en su esquema propuesto de cuatro dimensiones, Figura 2-5, y señalaron que el número óptimo de dimensiones depende de lo que los investigadores estén preguntando o interesando. No obstante, concluyeron que su estudio no puede tomarse como una representación dimensional de la experiencia emocional, aseguraron que los modelos bidimensionales pasan por alto fuentes de variación emocional clave, como el dominio de la emoción, y animaron a la comunidad investigadora a aplicar tres o más dimensiones para desentrañar adecuadamente la complejidad de la emoción.

Sin tener en cuenta el profundo debate que supone tener en cuenta las perspectivas metateóricas discreta y dimensional, ya hay investigadores que proponen unirlos. De hecho, los autores de [48] afirmaron que ambos modelos existen, pero cada uno de ellos pretende explicar características diferentes de las emociones. Por ejemplo,

¹unige.ch/cisa/files/7214/9371/2318/Grid_questionnaire_Aug_2013.pdf

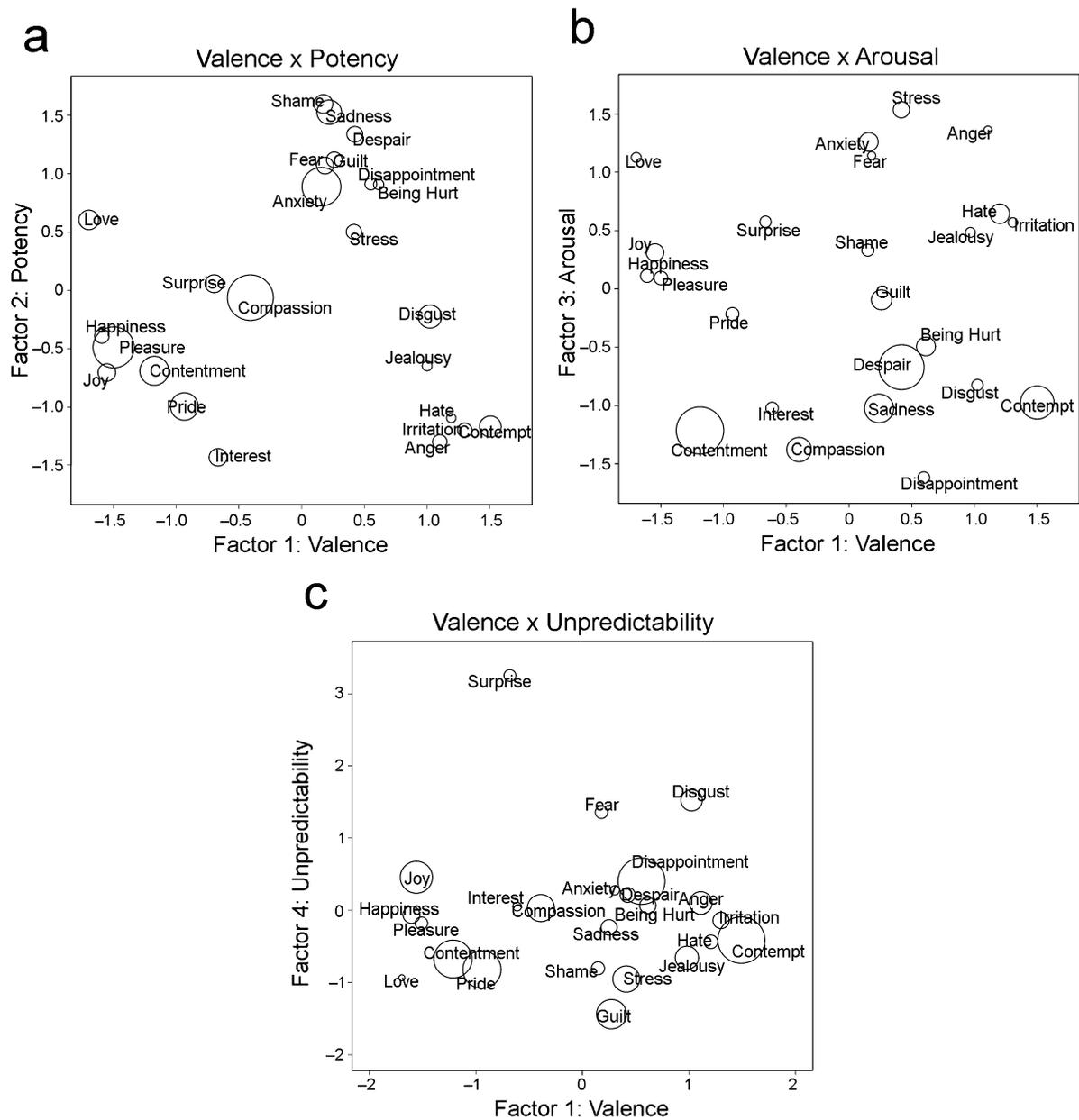


Figure 2-5: Los 24 términos emocionales se han mapeado en el esquema cuatridimensional propuesto por Fontaine et al. en [5].

afirmaron que la supremacía de un método de clasificación sobre otro nunca estará asegurada por la psicología de las emociones, lo que incluso se convierte en una ventaja a la hora de desentrañar el comportamiento de las emociones, ya que conduce a aumentar la comprensión de las mismas al no excluir ninguna perspectiva o información emocional. A lo largo de una revisión detallada, también afirmaron que las emociones discretas pueden ser descompuestas mediante la inclusión de dimensiones dentro de ellas. Aunque este hecho específico ya fue propuesto inicialmente por Plutchik [37] con su rueda multidimensional de las emociones y la variación de la intensidad dentro de las emociones básicas, la combinación de estos dos enfoques principales de clasificación podría ser explotada por la literatura hacia una mejor comprensión psicofisiológica de las emociones. En cuanto al efecto de esta cuestión a la hora de diseñar herramientas de reconocimiento de emociones para prevenir y evitar situaciones de violencia de género, la combinación de ambos enfoques podría impulsar una detección más consistente y una mejor comprensión de cualquier tipo de emoción relacionada con el miedo. La combinación específica seguida en esta investigación se detalla en la sección 2.3.4.

2.3.3 Rasgos personales, procesos cognitivos, atención y sesgo de género

Además de las teorías emocionales humanas y de las clasificaciones de las emociones mencionadas y revisadas en la sección anterior, los pensamientos y las creencias individuales son componentes que desempeñan un papel esencial dentro de las respuestas emocionales y fisiológicas que también podrían considerarse al abordar el diseño de cualquier sistema de reconocimiento de emociones. En realidad, estos elementos cobraron relevancia con la aceptación y el auge de la psicología cognitiva a partir de 1950. Sin embargo, a pesar de este hecho, la pregunta sin respuesta que todavía rodea a muchas de las teorías emocionales humanas es el efecto específico que la cognición tiene sobre las diferentes emociones [49]. Esto es aún más confuso cuando la cognición converge con la valoración. Esta última se basa en una asociación automática de un estado afectivo (asociación emocional) ya sea con baja o alta valencia, y es el núcleo de las diferentes teorías emocionales humanas como se ha comentado anteriormente. Desde una perspectiva psicológica, el camino entre un

proceso cognitivo o evaluativo y una reacción de appraisal puede derivar el uno del otro y viceversa, ya que una evaluación puede ser un raciocinio de una asociación emocional previa, y esta última puede ser también el producto de una evaluación emocional posterior [50]. En este contexto subjetivo, los rasgos personales y la atención e interpretación de los estímulos pueden afectar en gran medida a las respuestas afectivas. En los últimos años, algunos estudios han intentado relacionar los distintos grados principales de los cinco factores de la personalidad (extraversión, neuroticismo, apertura a la experiencia, amabilidad y conciencia) con los procesos y cambios emocionales de la vida cotidiana. Por ejemplo, Emma Komulainen et. al. en [51] realizaron un experimento con 104 estudiantes universitarios (18 hombres, 86 mujeres) siguiendo un método de muestreo de experiencias en el que los estudiantes informaron de diferentes métricas afectivas unas 10 veces al día a intervalos semi-aleatorios. Observaron y concluyeron que los rasgos de personalidad pueden influir en diferentes procesos emocionales. En concreto, destacaron que esos rasgos afectan en gran medida a los trastornos depresivos, de ansiedad y de estrés, que están relacionados con la respuesta afectiva negativa a los contextos de la vida cotidiana. Este hecho guarda una profunda relación con diferentes conceptos neurofisiológicos que están directamente relacionados con la cantidad de carga afectiva negativa que cada individuo puede manejar [52]. Dicha carga se conoce como carga alostática y es un factor crucial para empezar a entender las particularidades del perfil fisiológico y emocional de las víctimas de la violencia de género, ya que están sometidas a situaciones negativas crónicas (miedo, pánico, estrés) que conducen a la restricción afectiva en contextos traumáticos con el objetivo de recuperar la homeostasis fisiológica y el equilibrio conductual y proteger su integridad psicológica. Además, hay que considerar y contabilizar las diferencias de género al añadir el sesgo de género al problema del reconocimiento de emociones. Por ejemplo, está comprobado que las mujeres son más sensibles a las expresiones interpersonales durante las interacciones sociales que los hombres, lo que se acompaña de una diátesis del estado de ánimo e incluso Post-Traumatic Stress Disorder (PTSD) [53–55].

Todos estos componentes plantean diferentes incertidumbres que hacen que el diseño de un sistema inteligente de reconocimiento de emociones sea una tarea en la que se tengan en cuenta los diferentes factores subjetivos emocionales individuales

para conseguir un rendimiento óptimo. Por lo tanto, si se desarrolla un sistema de reconocimiento de emociones utilizando información fisiológica y física, el uso de los métodos de clasificación de emociones humanas revisados debería ir acompañado de diferentes pruebas o cuestionarios que faciliten la dilucidación de los efectos producidos por la cognición, la valoración, la atención, los rasgos de personalidad, el género y la edad. Aunque en este trabajo de investigación no se tratan, cuantifican o tienen en cuenta directamente estos factores subjetivos individuales, durante la realización de la base de datos UC3M4Safety se ha recogido un conjunto de cuestionarios. En el capítulo 6.

2.3.4 El mapa del miedo dentro de los métodos de clasificación de las emociones humanas

La emoción del miedo es una de las emociones básicas que son comunes a lo largo de la mayoría de las diferentes clasificaciones categóricas de las emociones e incluso representa una emoción distinguida clave para las clasificaciones dimensionales a la hora de explicar las ventajas de tales modelos para tratar el desentrañamiento de la emoción basado en la dominancia. Debido a la aplicación dirigida de este trabajo de investigación, la comprensión adecuada de la delimitación discreta y dimensional del miedo es esencial. En este sentido, la figura 2-6 ilustra una idea sobre este hecho. La

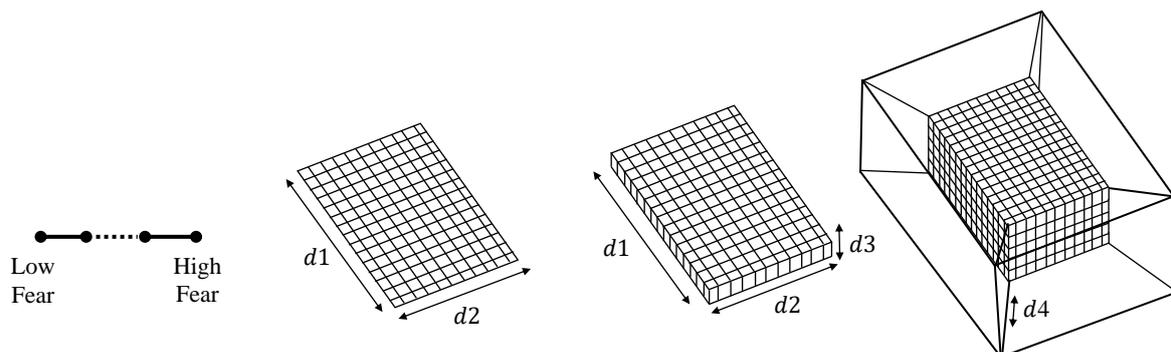


Figure 2-6: De izquierda a derecha: concepto de miedo unidimensional (niveles de intensidad discretos), miedo contenido en un espacio bidimensional (modelo PA), conceptos tridimensionales (modelo PAD) y cuatridimensionales (modelo PAD más cualquier dimensión intrínseca individual).

manera más fácil de considerar el miedo es adoptar una forma de factor unidimensional discreto. Este método viene determinado por el número de divisiones o niveles de intensidad del miedo que se desee. Al pasar a más dimensiones, disponemos de más información para determinar y caracterizar inequívocamente el sector exacto

del miedo. Por ejemplo, una percepción bidimensional, que puede estar relacionada con el arousal y la valencia, puede definir un cuadrante específico en el que se ubican las emociones negativas. En este caso, el número de niveles o divisiones en las diferentes dimensiones repercute directamente en la incertidumbre de localización exacta del miedo dentro de dicho cuadrante. Así, cuantas más divisiones, más limitada o acotada es el área en la que se puede encontrar el miedo. Este último hecho también supone un reto cuando los niveles de esas dimensiones se recogen directamente de los voluntarios mediante el uso de métricas autoinformadas, ya que no es factible ni pragmático ofrecer la posibilidad de elegir un número elevado de niveles. Finalmente, como se comentó en la sección 2.3.2, el uso de tres dimensiones puede beneficiar la separación de diferentes emociones que comparten dos de ellas. Hay que tener en cuenta que añadir más dimensiones conduce a un espacio de búsqueda multidimensional, lo que se traduce en un problema de optimización más complejo para encontrar el punto dulce del miedo.

Específicamente para este trabajo de investigación, se utiliza una nueva fusión entre modelos discretos y dimensionales para la detección del miedo. Así, tomando como referencia el modelo PAD, se define que el miedo se localiza en baja valencia, alta excitación y baja dominancia. A partir de esa localización multidimensional específica, se realiza un mapeo binario emocional discreto etiquetando esa localización como miedo y las otras como no miedo. Esta relación viene dada por la siguiente función de Heaviside,

$$\mathbf{H}(x_i) = \Theta(\varepsilon_i - x_i) \quad (2.1)$$

donde $\Theta : \mathbf{H} \rightarrow (0, 1)$ y ε_i es el umbral de miedo especificado de la dimensión i . Obsérvese que el resultado final del mapeo binario para el modelo propuesto se obtiene realizando la siguiente operación lógica,

$$\mathbf{H}(P, A, D) = \mathbf{H}(P) \wedge \mathbf{H}(A) \wedge \mathbf{H}(D). \quad (2.2)$$

Durante este trabajo de investigación, todos los experimentos realizados recogieron etiquetas emocionales autoinformadas que se utilizaron para el mapeo binario del miedo. En este caso, y siguiendo la literatura [49], se utilizó la misma escala Likert de 1 a 9 para calificar cada una de las tres dimensiones. Por lo tanto, la distinción

entre niveles bajos y altos se realizó en base a la mitad de la escala. Aunque se encuentra una limitación en cuanto a que este nuevo enfoque asume que todo el espacio cúbico formado por alta excitación, baja valencia y baja dominancia, está directamente relacionado con la emoción categórica del miedo, es la primera vez que se realiza una fusión entre los modelos discretos y dimensionales y se aplica utilizando datos reales y emociones no actuadas. Teniendo en cuenta esta investigación como base en este aspecto, se podrían desarrollar más investigaciones hacia una mejor delimitación multidimensional y categórica de la emoción objetivo. En este sentido, se podrían añadir más dimensiones para considerar factores individuales como los rasgos de personalidad, el efecto de la cognición, los procesos de atención y el sesgo de género. Dichas dimensiones producirían un cambio multidimensional del cubo del miedo, lo que puede conducir a un mejor desentrañamiento y detección del miedo.

2.4 Herramientas para el análisis científico de las respuestas emocionales humanas

Uno de los principales objetivos dentro de la comunidad de la computación afectiva es la generación de nuevas bases de datos para facilitar y potenciar la tarea de reconocimiento de emociones. Una base de datos de computación afectiva o de reconocimiento de emociones puede definirse como un experimento de elicitación de emociones con un conjunto de voluntarios, que autoinforman de las emociones sentidas por un conjunto específico de estímulos. Además, entre estos componentes esenciales, los distintos conjuntos de señales fisiológicas y físicas recogidos durante estos experimentos específicos controlados, basados en el laboratorio, son fundamentales para generar posteriormente modelos de reconocimiento de emociones utilizando esa información como datos. Nótese que el objetivo subyacente es desentrañar los patrones y variaciones de los datos fisiológicos y físicos observados en esos experimentos con la ayuda de las diferentes etiquetas recogidas. Sobre esta base, se han presentado y propuesto en la literatura diferentes herramientas, elementos y métodos para proporcionar una elicitación de emociones eficaz dentro de esos experimentos. En esta sección, se presentan y detallan estos factores para comprender el estado actual de la técnica en este sentido. Hay que tener en cuenta que los analizados aquí se aplican y utilizan en condiciones controladas o de laboratorio, lo que limita la apli-

cación directa de los sistemas inteligentes resultantes a la validación sobre el terreno. Este último hecho también conduce a la necesidad de generar bases de datos en la naturaleza. En el capítulo 3 se detallan estas últimas consideraciones y se ofrece una descripción y un análisis detallados de todos los componentes que intervienen en la generación de una base de datos de reconocimiento de emociones. Además, en esta sección, también se abordan y discuten los diferentes retos para proporcionar una verdad básica de etiquetado de estímulos fiable.

Como se ha analizado en los apartados anteriores, los factores individuales son clave en las emociones. Este hecho dificulta la obtención de la misma emoción para un grupo de personas que están bajo el mismo experimento de la base de datos. Aunque esto se suele abordar mediante protocolos experimentales bien definidos dentro de un entorno controlado de laboratorio, estas incertidumbres de personalidad siempre presentan un sesgo subjetivo introducido por los voluntarios cuando, por ejemplo, autoinforman de la emoción sentida. En general, podemos dividir el tipo de estímulos utilizados en dichos protocolos en dos grupos principales: actuados y no actuados. Los primeros son interpretados mayoritariamente por actores y actrices entrenados, que siguen un "guión de elicitación emocional" [56]. Independientemente de la capacidad de los actores y actrices para profundizar en el estado emocional solicitado, esto se traduce en una forma sintética de generar estados afectivos que conduce a una respuesta autónoma no plenamente emocional. Por ello, en la literatura se prefieren los estímulos de tipo no actuado. Estos y sus principales características se resumen en la tabla 2.2. Existen principalmente seis tipos diferentes de estímulos no actuados, que van desde la imaginería hasta el Virtual Reality (VR). Algunos de ellos proporcionan una sensación estática al no involucrar completamente a la persona en el entorno emocional deseado, mientras que otros proporcionan esa posibilidad. También se encuentran diferencias con respecto a los procesos cognitivos, conductuales y físicos desencadenados por dichos estímulos. De entre todos los estímulos, destaca el VR por ofrecer la sensación más cercana a los escenarios del mundo real, lo que se traduce en un alto grado de correlación entre las condiciones de la investigación y el fenómeno emocional estudiado (validez ecológica). Estos hechos llevaron al equipo de UC3M4Safety a desarrollar un entorno VR con estímulos 2D y 3D para ser utilizado durante la realización de los diferentes

Table 2.2: Revisión del tipo de estímulo utilizado en entornos de laboratorio controlados.

Tipo de estímulo	Características	Ejemplos
Imágenes	Estímulos estáticos, componentes cognitivos, la visualización de la duración es clave, mejores para la identificación emocional	[58], [59], [60] [61], [62], [63]
Videos	Estímulos estáticos, componentes comportamentales y cognitivos, la visualización de la duración es clave, pueden proporcionar mejor inmersión emocional que las imágenes	[64], [65], [66] [67], [68], [12]
Juegos	Estímulos dinámicos, componentes comportamentales y cognitivos, latencia entre las acciones de la usuaria y el juego son claves	[69], [70], [71]
Tests de Estrés	Estímulos estáticos y dinámicos, cognitivos, comportamentales, y físicos, alto acuerdo en la literatura sobre estos tests	[72], [73], [74] [75], [76]
VR	Estímulos estáticos y dinámicos, cognitivos y comportamentales cercanos a escenarios de la vida real, ofrecen la mejor validación ecológica	[77], [78], [79] [80]

conjuntos de datos [57]. En el capítulo 6 se ofrecen más detalles.

A pesar de la variedad de tipos de estímulos, uno de los mayores retos que implica la elicitación de emociones para el diseño de sistemas de reconocimiento de emociones es la obtención de una verdad de base fiable, es decir, determinar adecuadamente qué emociones (etiquetas) evocan qué estímulo. La evaluación de la verdad de base es en realidad una de las partes más críticas dentro del diseño de esos sistemas [81]. De hecho, ese proceso es de suma importancia y afectará en gran medida al entrenamiento y al rendimiento de la inferencia posterior del sistema consciente de los afectos. Por ejemplo, cuando se entrena un algoritmo de aprendizaje automático utilizando las etiquetas de la verdad sobre el terreno, la información que se suministra a dicho algoritmo hereda la distribución de dichas etiquetas, lo que puede sesgar en gran medida los patrones de datos originales subyacentes. Para hacer frente a este tipo de problemas, existen diferentes estrategias o metodologías para recoger la verdad sobre el terreno que incluso se utilizan conjuntamente dentro del experimento. Así, en la literatura se han propuesto diferentes métodos para obtener datos fiables de autoevaluación emocional. Una de las metodologías más utilizadas y fiables para recoger la verdad del terreno se basa en una conocida técnica pictórica no verbal, se trata del SAM [6]. La representación original puede verse en la figura 2-7. Se basa en el espacio PAD (valencia, arousal y dominancia, respectivamente de la primera a la tercera fila) y en una escala de Likert de 1 a 9, en la que el centro de la escala se relaciona con un estado afectivo neutro. Sin embargo, se puede observar que esta

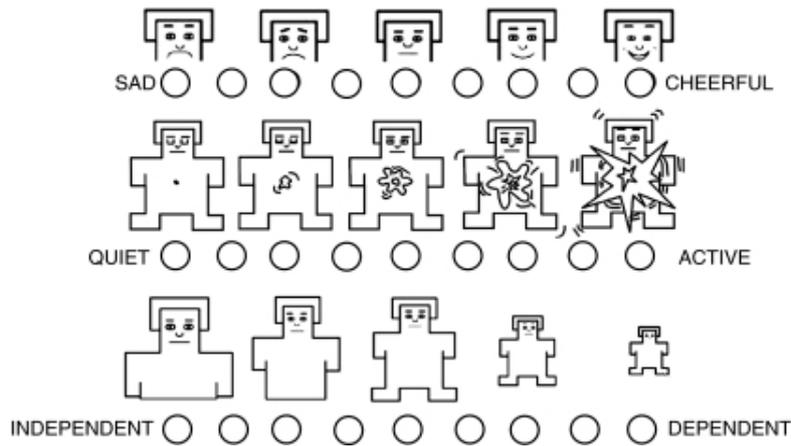


Figure 2-7: SAM original [6].

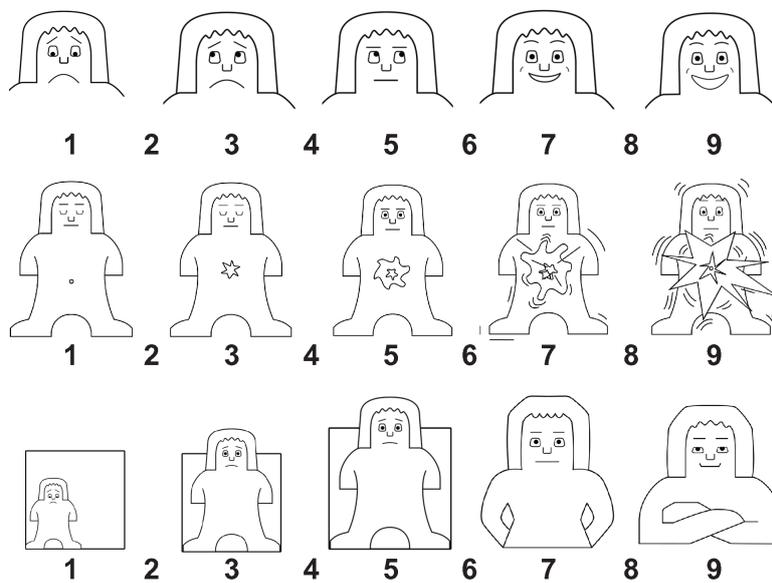


Figure 2-8: SAM modificado por el equipo UC3M4Safety.

representación original muestra mayoritariamente líneas rectas y una actitud muy masculina que puede afectar al etiquetado para las mujeres. Por lo tanto, el equipo de la UC3M4Safety modificó la SAM con el fin de proporcionar un menor sesgo de género. Hay que tener en cuenta que esta modificación se realizó en base a un panel de expertos en violencia de género [82]. El nuevo SAM resultante se muestra en la figura 2-8.

A modo de resumen de este apartado, podemos afirmar que las bases de datos de reconocimiento de emociones son necesarias para generar sistemas de computación afectiva, pero también son imprescindibles para estudiar las diferencias de respuestas emocionales en función de factores físicos, fisiológicos, de género, personales y de otro tipo de interés. Además, existe un amplio abanico de herramientas y métodos que facilitan dicho proceso de generación de sistemas de computación afectiva y, aunque

son mejorables para evitar posibles sesgos (por ejemplo, de género), el estado del arte en este sentido es sólido y fiable.

2.5 Indicadores fisiológicos de las respuestas de las emociones humanas

Como se detalla en la sección 1.2, uno de los principales objetivos de este trabajo de investigación es desentrañar las relaciones entre las señales fisiológicas y las emociones negativas, por ejemplo, el miedo, para proporcionar los primeros pasos hacia una detección automática de situaciones de riesgo en un contexto de violencia de género y/o acoso sexual. Para lograrlo, es necesario recopilar, comprender y aplicar a cualquier tecnología que se desarrolle un profundo conocimiento de la actividad fisiológica del cuerpo humano en el marco de las respuestas emocionales.

En primer lugar, las señales fisiológicas son manejadas por el Autonomous Nervous System (ANS) y, por tanto, no pueden ser manipuladas por la voluntad humana [83, 84]. Este hecho ha llevado a la literatura a proponer y proporcionar diferentes arquitecturas fiables de sistemas conscientes del afecto utilizando únicamente información fisiológica [85]. Sin embargo, en cuanto a los modelos y teorías emocionales, existen posturas contrarias en cuanto a la activación y comportamiento específico del ANS ante las diferentes emociones [86, 87]. Para este trabajo de investigación, seguimos una de las últimas afirmaciones sobre la activación del ANS ante las emociones [88], que se basa en la actividad diferenciada del ANS para la preparación de la conducta y la protección del cuerpo con respecto a las diferentes emociones, que es esencial para la adaptación humana. Esto está profundamente interrelacionado con el funcionamiento del cerebro al recibir estímulos externos. De hecho, la activación del ANS es consecuencia de los circuitos internos entre las diferentes partes del cerebro que se encargan de decodificar esos estímulos y desencadenar los mecanismos necesarios para adaptarse adecuadamente a ellos. Dos de las partes principales son la amígdala y el hipotálamo, Figura 2-9. La primera es la responsable del procesamiento emocional, mientras que el segundo funciona como un centro de mando. Así, en caso de un estímulo externo amenazante, la amígdala envía una señal de socorro al hipotálamo, que activa el Sympathetic Nervous System (SNS) a través de las glándulas suprarrenales. Obsérvese que el SNS es la rama

del ANS responsable de la conocida respuesta de lucha o huida. Finalmente, esas glándulas liberan diferentes catecolaminas (por ejemplo, epinefrina) que provocan una serie de cambios y reacciones fisiológicas. Una vez que la amenaza desaparece, el Parasympathetic Nervous System (PNS) toma el mando actuando como descanso de las reacciones fisiológicas anteriores (homeostasis). Nótese que el PNS es la rama del ANS responsable de la conocida respuesta de reposo y digestión. Aunque estos comportamientos y características biológicas están mayoritariamente consensuados en la literatura, sigue existiendo un gran interés investigador por aportar experimentos empíricos sobre alguno de los hechos comentados [89]. Como consecuencia de ello, el estudio del desentrañamiento emocional-fisiológico en la literatura se ha realizado de forma intensiva desde 1950, ya que diferentes investigadores intentaban abordar las teorías emocionales y comprender los cambios en las variables fisiológicas debidos a las respuestas emocionales [90].

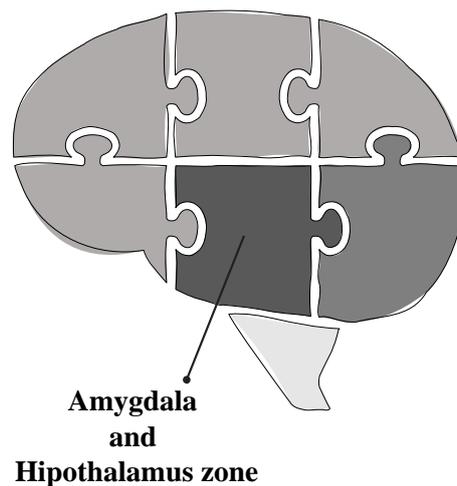


Figure 2-9: Location of the two main parts, amygdala and hypothalamus, involved in the emotional processing and autonomous nervous system regulation.

Esta sección proporciona una revisión detallada sobre el nexo entre las variaciones fisiológicas y las emociones negativas, destacando específicamente aquellos sistemas en la literatura dirigidos a la detección del miedo. Además, debido a la naturaleza portable del sistema presentado en este trabajo, sólo se cubren las tres variables fisiológicas que pueden ser hoy en día más aptas para ser vestibles o portables (aspecto discreto). A pesar de este estudio, todavía hay muchos factores, como los inherentes al individuo, que pueden afectar directamente a la morfología de la señal fisiológica y, por tanto, a los patrones subyacentes asociados a las emociones negativas. Estos componentes, como la edad, el género, las condiciones cardiovasculares,

la carga alostática y otros, no solían ser considerados en la literatura a la hora de diseñar sistemas de reconocimiento de emociones. Así, en esta sección y en los capítulos siguientes, analizamos y abordamos la influencia que esos elementos pueden tener en la morfología de las señales y damos algunas ideas para tratarlos desde la perspectiva de los sistemas digitales.

Antes de entrar en detalles sobre cada una de las señales fisiológicas que se van a analizar y estudiar, hay que destacar la naturaleza de la información fisiológica, ya que afecta en gran medida a las técnicas aplicadas para extraer los indicadores fisiológicos afectivos o métricas fisiológicas emocionales. Como cualquier otro sistema biológico complejo, las señales fisiológicas humanas poseen un comportamiento no lineal y no estacionario [91]. Sin embargo, al estar destinadas a ser procesadas digitalmente dentro de una plataforma embebida específica, se utilizan ventanas de procesamiento de longitud fija para extraer los diferentes indicadores o métricas, lo que lleva a considerar la fisiología como cuasi-estacionaria cuando se trata de ventanas de procesamiento cortas. Últimamente, aunque este último hecho puede restringir el uso de técnicas de procesamiento lineal, la aplicación de técnicas no lineales está teniendo mucho éxito en los actuales sistemas de reconocimiento de emociones basados en información fisiológica [26] y está impulsando la comprensión de los sistemas biológicos complejos tanto en salud como en enfermedad [92]. Por ello, en los siguientes apartados, así como en los siguientes capítulos, se considera esencial el comportamiento fisiológico no lineal.

2.5.1 Actividad cardiovascular

La actividad cardíaca es una de las informaciones fisiológicas más utilizadas para generar sistemas de reconocimiento de emociones [26, 49]. Las diferentes fases de los latidos, que se traducen en diferentes presiones sanguíneas dentro de las paredes musculares de los vasos sanguíneos, permiten monitorizar las variaciones o cambios tanto simpáticos como parasimpáticos [8]. Desde una perspectiva puramente fisiológica, por un lado, la presión arterial más alta se alcanza durante la fase sistólica, en la que el corazón se contrae para forzar la sangre a través de las arterias. Por otro lado, la presión más baja se alcanza durante la fase diastólica, en la que el corazón vuelve a llenarse de sangre. Esta información se ve muy afectada por la dieta del individuo, la edad y las posibles enfermedades cardíacas [93, 94]. Independiente-

mente del tipo de estos factores intrasujeto, todos ellos conducen a modificaciones morfológicas respecto a una forma de onda ideal esperada. Estas modificaciones se deben principalmente a los cambios de resistencia de los vasos sanguíneos periféricos, que van desde diferentes niveles de vasoconstricción a diferentes niveles de vasodilatación [95]. Para este trabajo de investigación, siendo el miedo la emoción objetivo a detectar, es necesaria la comprensión de estos principios fisiológicos, ya que está demostrado que los estímulos de miedo aumentan la resistencia periférica total conduciendo a un aumento de la vasoconstricción. Esto último es esencial para distinguir adecuadamente los patrones fisiológicos basados en el miedo de los basados en cualquier otra emoción [85, 96].

La adquisición de esta información fisiológica puede realizarse mediante diferentes sensores de forma no invasiva como el ECG, la fotopletismografía (PPG), y de forma invasiva como el catéter arterial. Debido a los requisitos de portabilidad, bajo consumo y discreción del sistema propuesto, nos centramos en los sensores PPG. Se basan en un método de medición óptica que emplea una fuente de luz (un único Light Emitting Diode (LED) o una matriz de LEDs) y un fotodetector que se sitúan en la superficie de la piel para medir el BVP. Existen dos tipos de sensores de PPG, de reflexión y de transmisión. La figura 2-10 muestra un ejemplo de estos métodos ilustrando cuál es la diferencia con respecto al recorrido que tienen a través de las diferentes capas de la piel. En el modo de reflexión, el fotodetector recibe la luz emitida que ha sido retrodispersada o reflejada por el efecto banana de las capas interiores [97, 98], mientras que en el modo de transmisión, el fotodetector está completamente opuesto al LED y recibe la luz transmitida que atraviesa todas las capas de la piel. La principal diferencia a la hora de obtener la señal de ambos métodos es el comportamiento invertido que presenta la reflexión PPG debido al retroceso de la luz reflejada recibida. Para este trabajo de investigación, nos centramos principalmente en el modo de reflexión debido al aspecto wearable y a que la mayoría de los sensores PPG disponibles son de este tipo.

Por lo tanto, centrándonos en el modo de reflexión, cabe destacar que la intensidad de la luz que recorre las diferentes capas decae exponencialmente. En concreto, esta afirmación se explica por la Ley de Lamberts-Beer [99], que se aplica para modelar adecuadamente la intensidad de luz recibida por el fotodetector de la siguiente

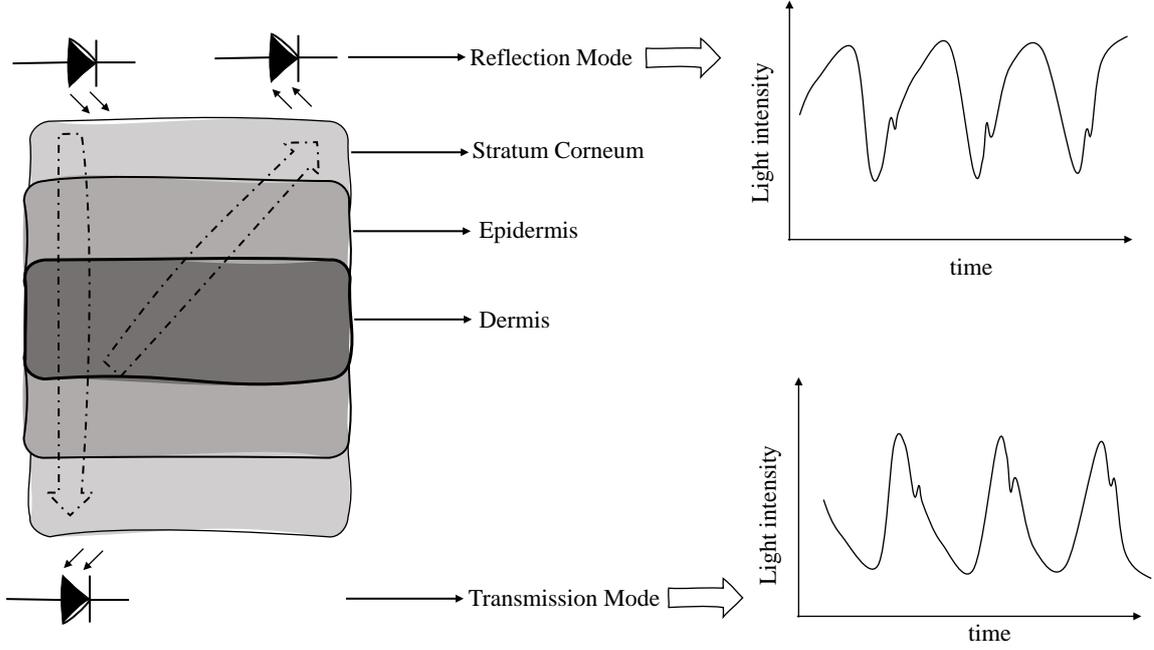


Figure 2-10: Ilustración de las dos técnicas de medición del PPG, reflexión y transmisión. Obsérvese que la señal obtenida está invertida en un método con respecto al otro.

manera:

$$I = I_{in}e^{-\lambda t}, \quad (2.3)$$

donde λ es la longitud de onda de una luz específica, I es la luz total detectada por el fotodetector y I_{in} es la luz transmitida o incidente. Sabiendo que λ puede expresarse como una relación directa entre el coeficiente de absorción del medio y la longitud del trayecto, y que el primero puede dividirse en contribución tisular no pulsátil (componente DC) y pulsátil (componente AC), la ecuación 2.3 también puede expresarse como

$$I = I_{in}e^{-(\mu_{AC}d(t)+\mu_{DC}m)}, \quad (2.4)$$

donde μ_{AC} y μ_{DC} son los coeficientes de absorción para los tejidos pulsátiles y no pulsátiles, respectivamente, y $d(t)$ y m son las longitudes del recorrido de la luz a través de dichos componentes. Además, la intensidad de la luz incidente también puede separarse en la intensidad reflejada estática, I_{rf} , y la intensidad del efecto banana, I_b , como sigue:

$$I = I_{rf} + I_b e^{-(\mu_{AC}d(t)+\mu_{DC}m)}. \quad (2.5)$$

Así, a partir de la ecuación 2.5, la relación entre la componente de CC y CA viene dada por:

$$\frac{AC}{DC} = \frac{I_b e^{-(\mu_{DC} m) \mu_{AC} d(t)}}{I_{rf} + I_b e^{-(\mu_{DC} m)}}. \quad (2.6)$$

En caso de suponer que la luz reflejada es despreciable, la amplitud de la componente AC normalizada sería directamente proporcional a la longitud del recorrido de la luz arterial dinámica. Este supuesto es el caso ideal, en el que la relación CA/CC se maximiza, sin embargo, en las aplicaciones reales, el hueco espacial entre el acLED y el fotodetector y entre el sensor y la piel (hueco de aire) afectará al componente de CC y minimizará la contribución de CA. Este problema se solía abordar aplicando estructuras de bloqueo de la luz en los sensores PPG y minimizando el entrehierro [100, 101]. Hay que tener en cuenta que la ubicación del sensor es igualmente importante en este asunto [102]. Estos conceptos revisados y los fundamentos para la medición de PPG son esenciales para diseñar adecuadamente sistemas vestibles eficientes sometidos a integrar dicha tecnología de sensores. En el capítulo 6 se ofrecen más detalles sobre cómo tratar el ruido de las señales de ácnúmeros.

Desde el punto de vista del procesamiento de señales, una señal PPG contiene diferentes características o métricas que pueden ser extraídas y analizadas para decodificar su relación con las emociones. En este trabajo, distinguimos entre características temporales, frecuenciales y no lineales. Independientemente del tipo específico de características a extraer, se requiere el análisis morfológico de la señal para obtener los puntos característicos necesarios PPG. La figura 2-11 muestra un ejemplo morfológico de dos periodos de frecuencia cardiaca en los que aparecen las dos fases de actividad cardiaca comentadas anteriormente: sistólica y diastólica. Además de los picos sistólico y diastólico, existen otros puntos característicos que afectarán al proceso de delineación de esta señal. Por ejemplo, la predicción o incisura, que es el producto de las reflexiones de la pared arterial, puede verse en la señal de PPG también justo antes de la muesca dicroica. Esta morfología sensible y variable hace que la monitorización del PPG sea un reto. De hecho, recientemente en [103], los autores presentaron un estudio comparativo con un grupo de 53 individuos que recogían datos de PPG de seis wearables diferentes de consumo y de investigación. Compararon la frecuencia cardíaca proporcionada por estos dispositi-

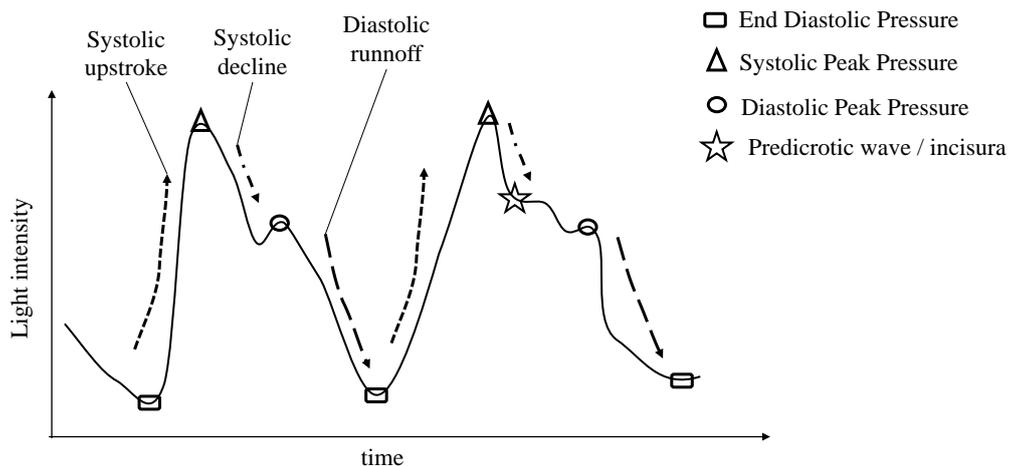


Figure 2-11: Ejemplificación de los diferentes puntos característicos a extraer dentro de la morfología de la señal PPG.

tivos con la frecuencia cardíaca real obtenida mediante un electrocardiograma. El experimento dio como resultado un error medio máximo de $15,9 \pm 8,1$ Beats Per Minute (BPM), siendo el tipo de dispositivo y la actividad de ejercicio los factores que más afectan a la frecuencia cardíaca estimada. Nótese que este error es relevante si se quiere aproximar a una norma análoga de equipos médicos como la UNE-EN 60601-2-27, que establece que el error máximo para equipos clínicos es de 5 BPM. Este problema observado se debe a diferentes aspectos:

- La mayoría de los wearables, ya sean de consumo o de investigación, no están pensados ni diseñados para adquirir datos PPG clínicos o de diagnóstico (donde la morfología está totalmente preservada), sino que obtienen datos PPG básicos de calidad. Este hecho se traduce en una morfología muy variable que incluso depende del dispositivo por consideraciones electromecánicas específicas.
- Cada wearable utiliza un algoritmo propio para extraer los puntos característicos y calcular la frecuencia cardíaca. Este hecho se traduce en una variabilidad entre las mediciones de los distintos dispositivos.
- Los artefactos de movimiento modifican fuertemente la morfología de la señal PPG. Algunos de estos dispositivos implementan técnicas para hacer frente a ello, mientras que otros no lo hacen.

Por lo tanto, a pesar de la proliferación de los sensores de PPG y su aceptación por parte del sector privado debido a la mejor integrabilidad y rentabilidad que el ECG,

sigue existiendo una necesidad metodológica de delineación y estándares de Motion Artifact Removal (MAR). Generalmente, cuando se quiere extraer la frecuencia cardíaca o el periodo de la señal PPG, los picos sistólicos o los valles diastólicos finales son una opción válida para obtener dicha periodicidad. Obsérvese que el ancho de banda máximo de la frecuencia cardíaca es de aproximadamente 0,6-3,5 Hz, lo que equivale a 36-210 BPMs. Por lo tanto, en el peor de los casos para la frecuencia de actividad cardíaca más lenta cuando se dispone de recursos digitales limitados, es decir, dispositivos wearables limitados, el uso de una ventana de procesamiento de dos segundos asegura encontrar al menos dos picos sistólicos o dos picos diastólicos finales. Como ya se ha comentado, la mayoría de las características se calculan a partir de estos puntos. Para este trabajo de investigación, se ofrecen más detalles sobre los algoritmos específicos de delineación, la extracción de características y las técnicas MAR utilizadas en los capítulos. 4 and 5.

En cuanto a la relación entre la actividad cardíaca y la emoción del miedo, existe una amplia gama de publicaciones en la literatura [104–110]. Algunas de las publicaciones intentaron diferenciar entre emociones positivas y negativas basándose únicamente en la información extraída de la frecuencia cardíaca, mientras que otras consideraron indicadores afectivos más fisiológicos a partir de diferentes variables fisiológicas, por ejemplo, indicadores electrodérmicos o de cardiorrespiración. Por un lado, la mayoría coincidió en que la emoción de miedo provoca un aumento de la aceleración cardíaca, vasoconstricción, disminución del flujo sanguíneo y aumento de la presión arterial tanto sistólica como diastólica. Por otro lado, los que incluyeron más variables fisiológicas reclamaron la necesidad de considerar más información que la actividad cardíaca debido a la relación directa observada entre métricas específicas de la frecuencia cardíaca, como la variabilidad de la misma, y el aumento de la frecuencia respiratoria o los diferentes niveles de actividad electrodérmica. Hay que tener en cuenta que, aunque existe un conocimiento bien establecido en la literatura con respecto a los efectos de la actividad cardíaca producidos por el miedo, los experimentos se realizan en laboratorio, donde las condiciones están bajo control.

2.5.2 Actividad Electrodérmica

La actividad electrodérmica (EDA) o respuesta galvánica de la piel (GSR) es, junto con la actividad cardíaca, una de las señales fisiológicas más estudiadas que, además,

ha recibido un importante avance en su comprensión y conexión con las respuestas emocionales [111, 112]. Aunque hay más de un tipo de glándulas implicadas en este proceso, las principales responsables de la EDA son las glándulas sudoríparas ecricinas o merocrinas, que están controladas por el SNS. Éstas se encuentran en la piel y están inervadas únicamente por axones de la rama simpática (fibras nerviosas sudomotoras largas). Obsérvese que cada axón inerva alrededor de $1,28\text{cm}^2$ de piel [113]. La figura 2-12 muestra un ejemplo ilustrativo de estas entidades distribuidas en las diferentes capas de la piel. El hecho de que estas glándulas estén inervadas únicamente por el SNS hace que el EDA sea el candidato perfecto para cuantificar la actividad del SNS (lucha y huida) y, aunque la sudoración también juega un papel importante en la termorregulación para conseguir una homeostasis adecuada, está demostrado que los diferentes cambios en la conductividad de la piel están fuerte y directamente correlacionados con la intensidad de la emoción evocada por los estímulos externos. Muchos autores aseguran que dichos cambios están relacionados con el nivel de arousal [85, 112, 114]. En cuanto a la evolución de la forma de esos cambios con respecto al tiempo, el EDA está formado por un componente tónico y otro fásico. El primero es un componente que varía lentamente, el Skin Conductance Level (SCL), mientras que el segundo es el Skin Conductance Response (SCR) rápido en el tiempo. Obsérvese que la teoría fisiológica que subyace a estos cambios o variaciones de la EDA se basa en la difusión y la apertura de los poros que plantea el modelo de válvula poral de Edelberg [115].

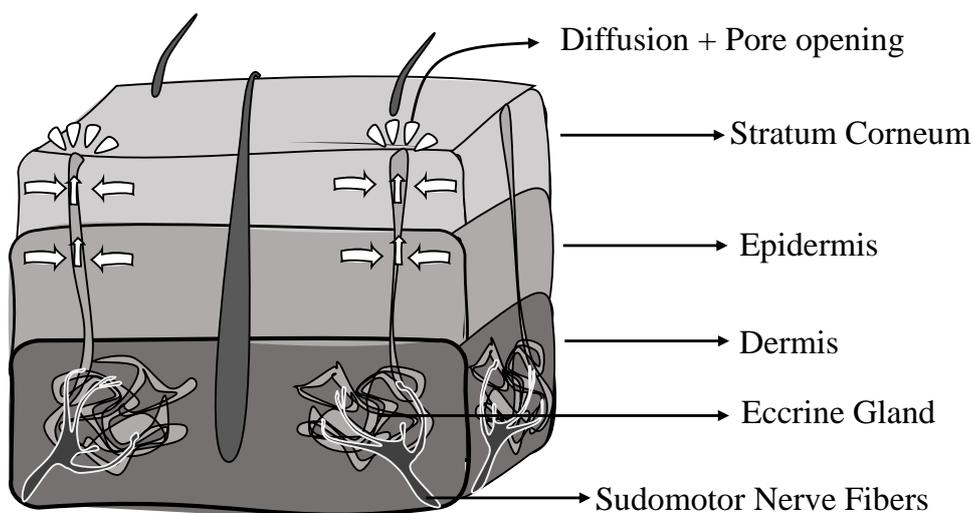


Figure 2-12: Ilustración del comportamiento de las glándulas merocrinas y del proceso de difusión a través de las diferentes capas de la piel.

La adquisición de esta información fisiológica puede realizarse principalmente me-

diante dos técnicas diferentes: la endosomática y la exosomática. Por un lado, la primera se basa en la medición del potencial electrodérmico mediante dos electrodos sin aplicar ni corriente ni tensión entre ellos. Por otro lado, la segunda se basa en la medición de la resistencia o conductancia electrodérmica mediante dos electrodos aplicando un pequeño voltaje o corriente entre ellos. A lo largo de los años, se han estudiado ambos métodos, aunque la complejidad de las ondas y la difícil interpretación de la metodología endosomática han llevado a una amplia aceptación y uso de las mediciones exosomáticas [116]. Así, las técnicas exosomáticas se caracterizan por utilizar una fuente de electricidad directa o alterna a través de circuitos activos o pasivos [117]. En primer lugar, debido a la fuerza electromotriz en la superficie de los electrodos, el uso de corriente continua puede provocar la polarización de los mismos. Este problema puede mitigarse con CA. Sin embargo, la aplicación de una tensión de fuente superior a 100 mV y el uso de electrodos de Ag/AgCl también minimizan los problemas de polarización cuando se utiliza la CC. En segundo lugar, las técnicas de CA conducen a una implementación de circuitos más compleja, lo que se debe principalmente al hecho de que hay que preservar tanto la información independiente como la dependiente de la frecuencia, así como a la aplicación de técnicas digitales posteriores para recuperar las partes real e imaginaria de dichas mediciones. Nótese que el término dependiente de la frecuencia se refiere en realidad al comportamiento de la susceptancia de la piel [118]. Por último, cabe señalar que la cantidad de resultados de investigación que consideran las técnicas exosomáticas de CC es excepcional en comparación con las de CA y, aunque la CA podría superar a la CC, es necesario realizar más investigaciones para demostrar este predominio. De hecho, hoy en día las técnicas de CC se han establecido como un estándar de facto para la adquisición de EDA [119]. La tabla 2.3 resume las principales diferencias analizadas entre ambas técnicas exosomáticas.

Table 2.3: Principales diferencias entre las mediciones exosomáticas de CC y CA.

Propiedades	CC	CA
Polarización de los electrodos	≈	✓
Circuitería sencilla	✓	✗
Información de la conductancia	✓	✓
Información de la susceptancia	✗	✓
Independiente de la frecuencia	✓	✗
Cantidad de investigación	✓	≈

Por lo tanto, centrándose en las mediciones exosomáticas de CC, se pueden utilizar diferentes circuitos electrónicos activos y pasivos. Una de las implementaciones más sencillas se realiza utilizando divisores de tensión simples compuestos por una resistencia fija y otra variable. Esta última es la piel humana. Sin embargo, esta técnica es propensa a errores de medición muy perceptibles debido a la diferencia entre la fuente de tensión y la tensión a medir, que hace que esta última no sea constante. De hecho, para la monitorización de la EDA se utiliza ampliamente la circuitería activa, ya que mitiga estos problemas y proporciona un mayor control sobre la medición. Convencionalmente, se emplean métodos de corriente y tensión cuasi-constantes aprovechando las configuraciones de amplificadores operacionales inversores y no inversores. Por ejemplo, la Figura 2-13 representa una configuración inversora como un posible ejemplo de tales disposiciones. En el caso de que los electrodos se coloquen dejando la resistencia de entrada R_i como la piel, se aplica una tensión casi constante que produce que el valor de la conductancia de la piel sea proporcional a la tensión de salida del circuito. Por el contrario, si la resistencia de retroalimentación R_f es la piel, se aplica una corriente cuasi-constante sobre ella y la tensión de salida resultante es proporcional al valor de la resistencia. Nótese que para ambas configuraciones, el límite de corriente debe ajustarse ya sea sintonizando las tensiones de entrada y referencia o la resistencia de entrada, respectivamente. Estos ajustes deben asegurar una corriente en todo el cuerpo no superior a $10\mu A/cm^2$, que es el nivel de densidad de corriente recomendado para las medidas de EDA [120]. Además, hay que destacar también la referencia común a la salida y a la entrada, que tiene por objeto evitar cualquier contaminación endosomática de la medición exosomática a realizar. En la literatura se pueden encontrar diferentes propuestas de circuitos EDA basados en configuraciones de op-amp inversor [121–123]. Obsérvese que estos circuitos solían ir seguidos de otros circuitos de acondicionamiento de op-amp para ajustar la señal y filtrarla antes de la adquisición.

Independientemente de la circuitería activa utilizada, y además del compromiso de sintonización para asegurar un límite de corriente de seguridad, la relación entre el rango y la sensibilidad es especialmente relevante cuando se mide esta señal fisiológica debido al rango relativamente amplio (de $0\mu S$ a $25\mu S$) y la sensibilidad de $0.01\mu S$ que hay que satisfacer para registrar adecuadamente todo el SCR dentro de los com-

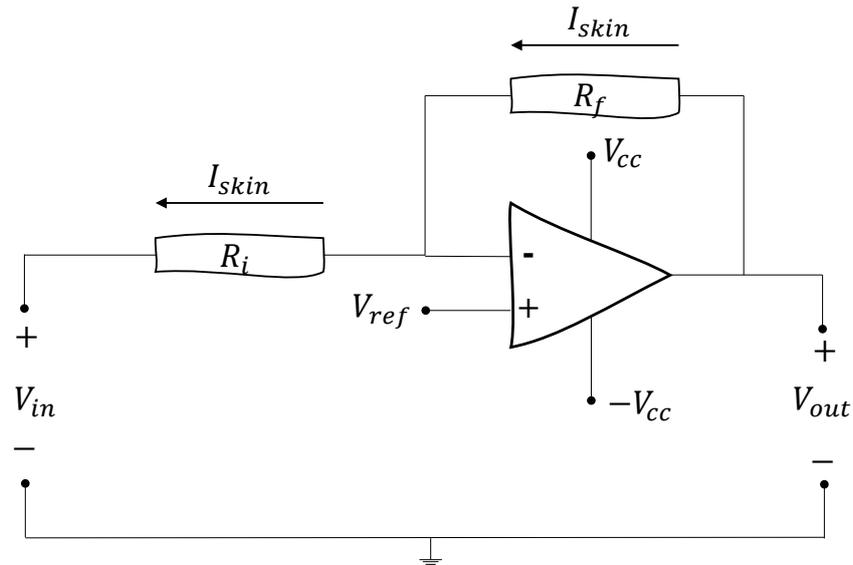


Figure 2-13: Ejemplo de configuración de un amplificador operacional inversor para la adquisición de DC EDA exosomática.

ponentes tónico y fásico de la señal EDA [112]. Para solucionar este problema, se puede aplicar un circuito de puente de Wheatstone. En ese caso, llegando a la calibración del puente entre las dos ramas, es decir, mediante el ajuste de un potenciómetro en la rama del puente opuesta a la resistencia de la piel humana, de forma que la diferencia de potencial sea cero o una determinada tensión deseada [124]. Una vez que el circuito se encuentra en ese estado, las perturbaciones de la diferencia de potencial son el SCR, y el SCL puede obtenerse a partir de la calibración del puente. Aunque este método puede, de hecho, proporcionar una medida fiable y asegurar un rango y una sensibilidad adecuados ajustando en tiempo de ejecución algunas de las resistencias (potenciómetro), no ha sido totalmente adoptado ni ampliamente utilizado en la literatura. Para este trabajo de investigación, se adopta un circuito activo exosomático de CC. Más detalles se dan en el capítulo 6. Hay que tener en cuenta que existen otras opciones para la adquisición exosomática de CC, como los amplificadores acoplados a la CA y los circuitos de retroalimentación, sin embargo, ofrecen una mayor complejidad de los circuitos.

Desde el punto de vista del procesamiento de la señal, una de las primeras tareas a realizar tras la adquisición es aplicar un filtrado básico de paso bajo y separar adecuadamente los componentes tónicos y fásicos (SCL y SCR). Ambos son igual de importantes en lo que respecta al desentrañamiento de la emoción, por lo que se desea su conservación a lo largo de la adquisición analógica y la manipulación digital.

Sin embargo, el componente fásico es el que contiene el ERSCR, que se traduce en diferentes ráfagas de EDA que se relacionan emocionalmente con los estímulos externos y se caracterizan por diferentes métricas basadas en el nivel de excitación real evocado. Así, la descomposición tónica y fásica está destinada principalmente a la correcta identificación y análisis de los ERSCRs. Nótese que el componente fásico también puede presentar un Nonspecific Skin Conductance Response (NSSCR), que se produce en ausencia de un estímulo elicitor identificable. Se pueden asumir diferentes umbrales para la métrica de cada pico SCR detectado para determinar la distinción entre ERSCRs y NSSCRs [125]. La figura 2-14 muestra un ejemplo de un ERSCR y algunas de las métricas que pueden extraerse de él. Para este trabajo de investigación, a lo largo de los siguientes capítulos se ofrecen más detalles sobre las características específicas extraídas.

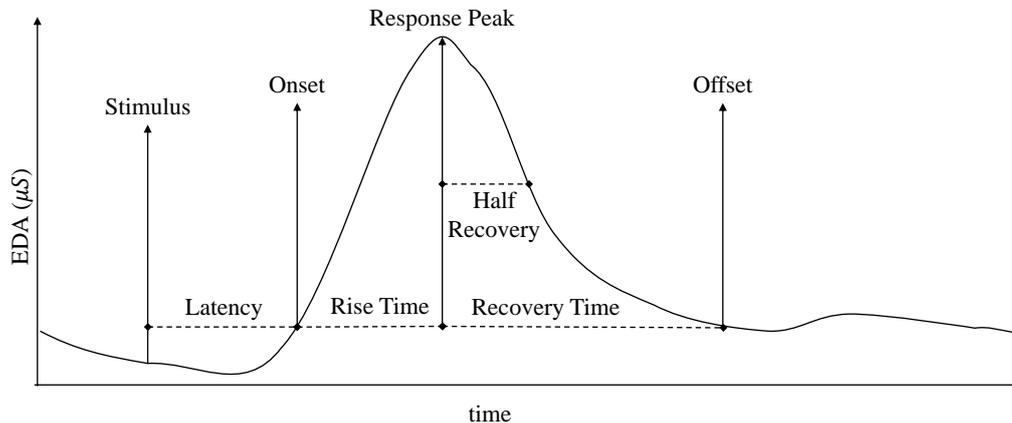


Figure 2-14: Un ejemplo ilustrativo de una ERSCR y algunas de las métricas que se pueden extraer de ella (*Stimulus - Estímulo, Latency - Latencia, Onset - Punto de inicio, Rise Time - Tiempo de subida, Response Peak - Pico de respuesta, Recovery time - Tiempo de recuperación, Half Recovery - Mitad de recuperación, Offset - Punto de parada*).

Uno de los métodos más sencillos para superar la descomposición tónica y fásica de la señal EDA es asumiendo una combinación lineal de estas dos, tal y como da la siguiente aproximación:

$$EDA_{total} \approx EDA_{tonic} + EDA_{phasic}, \quad (2.7)$$

donde EDA_{total} es la señal filtrada, EDA_{tonic} es el componente de baja frecuencia o la tendencia asociada al SCL, y EDA_{phasic} es la señal resultante que contiene los diferentes SCR. Así, restando la tendencia de la señal filtrada y aplicando una técnica de *trough to peak* (basada en los picos), se pueden extraer todos los picos

relevantes de la señal. Obsérvese que mediante dicha sustracción se obtiene una componente de la señal pseudofásica. En concreto, se puede obtener EDA_{total} aplicando un filtro FIR de paso bajo con una frecuencia de corte de 1,5Hz, que se selecciona en base a que la información de EDA queda por debajo de ella [112]. Posteriormente, la obtención de la EDA_{tonic} puede realizarse implementando un filtro de mediana móvil utilizando una ventana lo suficientemente amplia como para capturar la tendencia por debajo de 0,05Hz [125]. Aunque esta técnica puede implementarse de forma sencilla y no tiene un efecto negativo en el almacenamiento ni en la limitación de recursos, es sólo una aproximación y se enfrenta a diferentes problemas. Por un lado, el componente fásico resultante puede ser negativo, lo que nunca debería ocurrir desde una perspectiva fisiológica. Por otro lado, este método no tiene en cuenta ni trata el solapamiento de SCRs, lo que puede llevar a una subestimación de las diferentes amplitudes de pico de respuesta. Por lo tanto, este método se recomienda como punto de partida. A lo largo de los últimos años han aparecido en la literatura otros desarrollos acompañados de herramientas automatizadas, como Ledalab [126], que recogen los diferentes algoritmos más utilizados para potenciar su aplicabilidad en la investigación de EDA. Por encima de estos algoritmos encontramos cvxEDA [127] y SparsEDA [128]. El primero está motivado por el método de deconvolución introducido por Alexander et. al. [129], en el que afirmaron que el Sudomotor Nerve Activity (SMNA) posee una constante de tiempo más corta que la propia señal EDA y produce ráfagas (difusión de poros) que llegan como eventos separados y discretos. Aplicaron una técnica de deconvolución mediante una función biexponencial que abordaba el problema del solapamiento de SCR. Así, considerando esa base y manejando el problema de la racionalidad negativa de la componente fásica, cvxEDA utiliza una optimización convexa que está restringida por la sparsity y la no negatividad de la SMNA, que modifica la ecuación 2.7 como sigue:

$$EDA_{total} = IRF * (Driver_{tonic} + Driver_{phasic}), \quad (2.8)$$

donde IRF se identifica como la función biexponencial de respuesta al impulso de Bateman, y los *Drivers* son la información procedente del SMNA. Este algoritmo se ha aplicado con éxito a diferentes casos de investigación de EDA. Sin embargo, aunque la operación de convolución por sí misma necesita pocos recursos

computacionales, el procedimiento de optimización convexa necesita afinar diferentes hiperparámetros, lo que conlleva un tiempo computacional elevado. En cuanto al algoritmo SparsEDA, que es uno de los últimos métodos de descomposición EDA publicados recientemente en 2017, se basa en los trabajos de deconvolución anteriores, pero introdujo características diferentes, como la aplicación de la versión no negativa de la contracción mínima absoluta y el operador de selección mediante el uso del algoritmo de regresión de ángulo mínimo que hacen que la deconvolución sea más rápida, eficiente e interpretable que sus predecesores. A pesar de estas ventajas, su aplicabilidad y rendimiento para segmentos pequeños (inferiores a 70 segundos) sigue siendo objeto de debate. Así, aunque estos dos algoritmos aportan diferentes ventajas relacionadas principalmente con la interpretación fisiológica de la EDA, su aplicabilidad en dispositivos wearables con restricciones multimodales, como el brazalete de Bindi, es una tarea difícil debido a los altos recursos computacionales derivados de operaciones específicas, como la optimización convexa. Por lo tanto, se necesitan alternativas que se sitúen entre los métodos trough-to-peak y los convexos. Por ejemplo, algunos autores [130, 131] han utilizado un método Regularized Least-Squares Detrending (RLSD) [132] en el que el componente tónico se aproxima a un componente de tendencia aperiódica de baja frecuencia mediante

$$EDA_{tonic} = \frac{EDA_{total}}{(I + \lambda D_2^T D_2)}, \quad (2.9)$$

donde $\lambda D_2^T D_2$ es el término de regularización que sesga el SCL a una tendencia suave, I es la matriz de identidad, y D_2 es una aproximación discreta del operador de la 2ª derivada. Obsérvese que cuanto mayor sea la λ , más suave será la componente SCL. Una vez obtenida esta componente tónica aproximada, se aplica la misma sustracción a la señal EDA original para obtener la componente fásica. Para este trabajo de investigación se han utilizado todos los métodos revisados, aunque sólo se han incrustado los métodos trough-to-peak y RLSD. Más detalles sobre los resultados obtenidos se encuentran en el capítulo 5 y en el capítulo 6.

Como para cualquier señal fisiológica, durante su adquisición pueden observarse artefactos de ruido debidos al movimiento, a transitorios rápidos e incluso a electrodos sueltos. La figura 2-15 representa un ejemplo real de las diferentes fuentes de ruido que pueden encontrarse en las mediciones. Esta imagen muestra la difer-

encia entre los electrodos secos y los húmedos, ya que estos últimos se ven más afectados por el ruido debido a la inexistencia de Ag/AgCl que hace que la interfaz piel-electrodo sea menos robusta y sea efectiva sólo a través del sudor. Hay que tener en cuenta que este problema es especialmente relevante para el Bindi, ya que se basa en electrodos secos. Para combatir este tipo de fuentes de ruido y mitigar sus posibles efectos negativos durante el procesamiento de la EDA, se han aplicado diferentes pasos de preprocesamiento, como los filtros de media móvil y de mediana. En los capítulos 4.

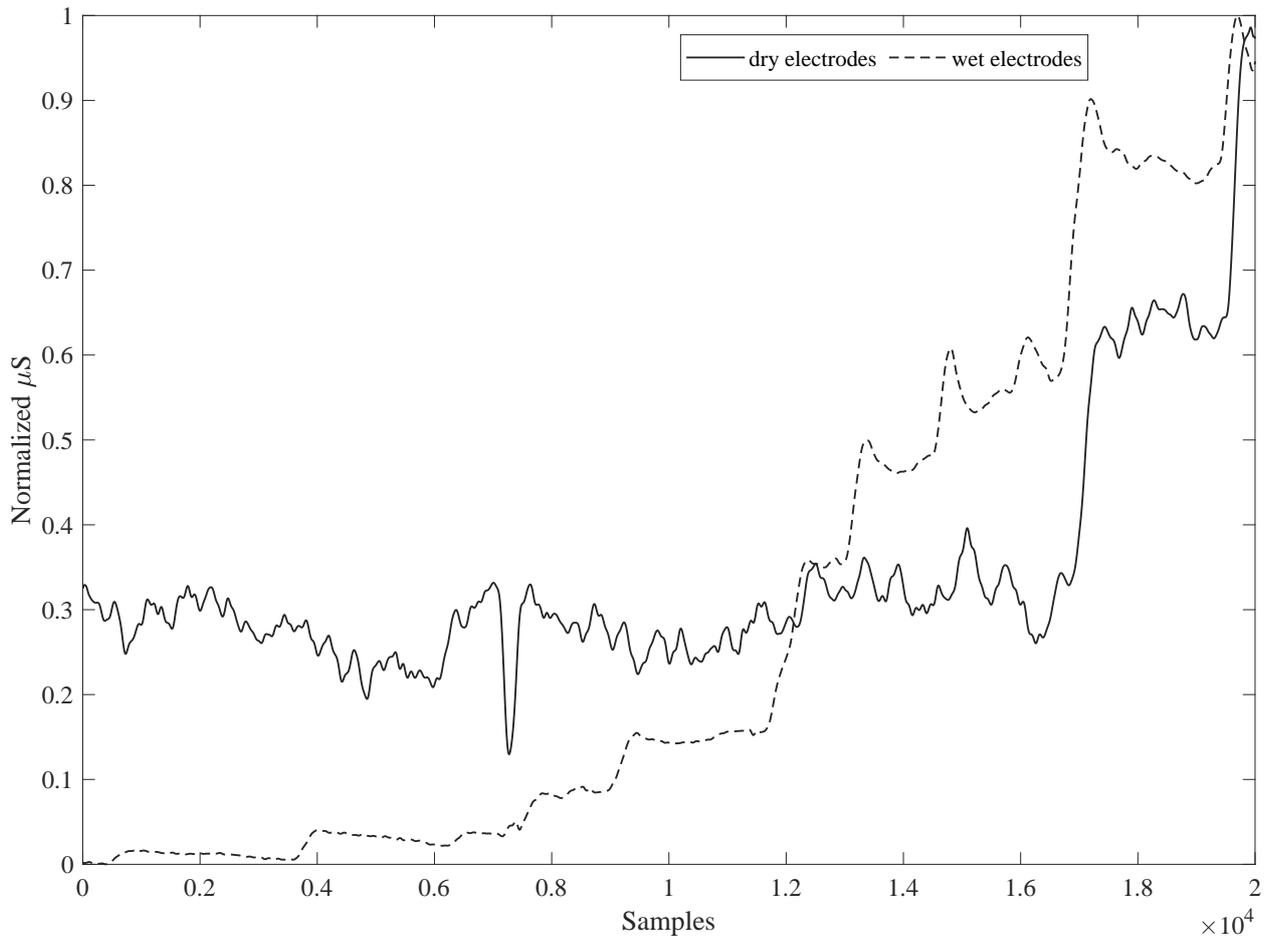


Figure 2-15: Diferencia entre los electrodos secos y húmedos que miden en la parte ventral sobre la parte derecha (húmeda) e izquierda (seca) de la muñeca. Obsérvese que las unidades están normalizadas μS y que los electrodos húmedos contienen 0,5% de sal de cloruro.

En cuanto a la relación que tiene la EDA con la emoción de miedo, diferentes grupos de investigación estudiaron este aspecto [85,133]. Como se ha explicado anteriormente, los cambios observados en la señal EDA se pueden relacionar directamente con la intensidad de la emoción, pero no con el tipo de la misma. De hecho, los estudios que consideran únicamente esta señal se dirigen a la detección del estrés, a

la cuantificación del arousal o incluso a la evaluación de la función simpática, pero no a la identificación de la emoción. En resumen, la emoción de miedo no puede detectarse utilizando únicamente la información de la EDA. En cualquier caso, la información extraída de esta señal puede proporcionar una excelente visión sobre los cambios de la activación simpática. Por ejemplo, ante situaciones muy estresantes, el SNS segrega diferentes hormonas catecolamínicas (adrenalina y noradrenalina) que hacen que la señal EDA se caracterice por un aumento del SCL y un incremento de las diferentes métricas que se extraen del SCR excepto la latencia que tiende a disminuir. Obsérvese que en estas situaciones aumenta la transpiración, que está directamente relacionada con la homeostasis y no con el proceso emocional. La interpretación de esta información en cuanto a su implicación emocional en el caso de las víctimas de violencia de género es más compleja debido a la posible sobrecarga alostática, que se refiere a los efectos acumulativos de las situaciones estresantes de la vida diaria que experimentan los individuos y que pueden llegar a inhibir la desconexión de la activación simpática [134]. Aunque el tratamiento de este último hecho específico está fuera del alcance de este trabajo de investigación, los detalles y la base se dan en los capítulos 5 and 6.

2.5.3 Temperatura de la piel

La temperatura de la piel del cuerpo no es tan popular en comparación con las dos señales fisiológicas ya detalladas. Sin embargo, existen varios investigadores que se ocupan de la identificación emocional utilizando también esta información [26, 135, 136]. Los fundamentos fisiológicos de esta señal están fuertemente interrelacionados con el flujo sanguíneo y las respuestas electrodérmicas del cuerpo. De hecho, la temperatura de la piel está fuertemente relacionada con los cambios de vasomoción por medio de las fibras simpáticas noradrenérgicas que regulan dicho proceso. Como se ha dicho anteriormente, el ANS no sólo proporciona mecanismos para hacer frente a los estímulos externos amenazantes, sino que también es el principal responsable de los diferentes procesos de homeostasis. Concretamente, en el caso de la temperatura corporal, el hipotálamo es el principal controlador termorregulador [137]. En todo el cuerpo tenemos diferentes receptores de temperatura que permiten al hipotálamo percibir y analizar continuamente la temperatura corporal. Una vez que esta parte de nuestro cerebro recoge la información necesaria, las acciones posteriores varían en

función de esa retroalimentación negativa, como para cualquier sistema de control fisiológico. Así, como en el caso de un termostato doméstico, a partir de un valor normal preestablecido, se activarán las diferentes defensas termorreguladoras para preservar, en el caso de nuestro cuerpo, 37 . Por ejemplo, debido a esas defensas, la temperatura corporal no se desvía más que unas décimas de grado del valor preestablecido. Es más, existe incluso un rango denominado interumbral, por encima del cual no se activa ninguna acción termorreguladora, que se sabe que está en torno a los 0,2°C. En realidad, esas defensas termorreguladoras son las siguientes:

- La sudoración y la vasodilatación son las defensas que se desencadenan cuando se produce una situación de calentamiento.
- Por el contrario, la vasoconstricción se desencadena para disminuir la pérdida de calor, principalmente por la disminución de la radiación superficial de la piel.

Por lo tanto, el proceso de termorregulación autonómica opera de forma sincronizada con otros sistemas fisiológicos de control de retroalimentación negativa, como el flujo sanguíneo y el manejo autónomo electrodérmico. Además, hay que destacar que existe una fuente interna de variabilidad de la temperatura marcada por cambios físicos, mentales y de comportamiento que siguen un ciclo diario conocido como ritmo circadiano. La comprensión y el entendimiento de estos factores fisiológicos son esenciales para evaluar y valorar adecuadamente la información que se desea recoger.

Desde la perspectiva del procesamiento de señales, esta señal no presenta la misma complejidad que las anteriores. Por el contrario, su información está contenida en frecuencias muy bajas, por debajo de 0,5Hz. Por lo tanto, el uso de un filtro FIR ordinario es suficiente para obtener una señal limpia. Después, la literatura tiende a extraer de ella características estándar, como la media, la mediana, la desviación estándar y otros estadísticos de alto orden. Además, su información de frecuencia extraída se divide comúnmente en diferentes bandas como cualquier otra señal fisiológica [12, 138]. Sin embargo, así como el procesamiento es uno de los más fáciles entre las señales fisiológicas, su integración no lo es. Para implementar un sensor de temperatura dentro de un dispositivo wearable restringido es necesario tener en cuenta algunas consideraciones. Por ejemplo, los autores de [139] elaboraron

una encuesta teniendo en cuenta 172 estudios desde 1960 hasta 2016, en la que revisaron todos los factores que afectan a la medición de la temperatura cuando se trata de termometría de contacto. Concluyeron con una serie de recomendaciones y compensaciones entre todos estos factores (interfaz piel-sensor, fijación, protección y sesgo ambiental, presión del sensor en la piel, etc.) que pueden afectar en gran medida a la temperatura corporal de la piel que se va a medir. Estos requisitos técnicos hacen que la integración de los sensores de temperatura de la piel sea una tarea difícil. De hecho, ningún dispositivo portátil disponible en el mercado (tipo smartwatch) integra un sensor de temperatura de la piel del cuerpo. Existen dispositivos wearables de grado de investigación que integran sensores infrarrojos de termopila, como el E4 de Empatica^{®2} [140]. Sin embargo, actualmente estos últimos tienen un coste elevado, lo que hace que su integración no sea tan sencilla como la termometría de contacto.

En cuanto a la relación entre la emoción del miedo y las variaciones de la temperatura corporal, diferentes estudios en la literatura se ocuparon de ello. Las primeras investigaciones sobre este tema se encuentran en [141–143], en las que, aunque los experimentos se realizaron con diferentes procedimientos experimentales, coincidieron en que la temperatura corporal disminuye bajo la elicitación del miedo. Nótese que en los tres estudios, los sensores de temperatura se colocaron en la palma de la mano. Recientemente, las investigaciones que se centran en las variaciones de la temperatura corporal con respecto a las emociones se centran más en el mapeo térmico facial utilizando imágenes térmicas infrarrojas funcionales. Por ejemplo, los autores de [135] utilizaron 60 imágenes de [58] y pidieron a veinticuatro estudiantes (19 mujeres) que calificaran las imágenes basándose en las escalas SAM mientras medían la temperatura de la piel facial y la EDA. Observaron que el mayor descenso de la temperatura se producía en las imágenes con mayor excitación. Así, afirmaron que la regulación autónoma del arousal se realiza en realidad mediante dos respuestas cutáneas simpáticas, la térmica y la electrodérmica. Sin embargo, una de las principales desventajas de la temperatura corporal de la piel en comparación con otras informaciones fisiológicas es la gran latencia de la señal. Esto supone una limitación a la hora de utilizar únicamente esta información para inferir el estado emocional.

²<https://www.empatica.com/en-eu/research/e4/>

Por ello, su integración solía ir acompañada, es decir, compensada, de otras señales fisiológicas, como EDA y BVP [133].

A pesar de estos patrones y características termo-emocionales observados, hasta donde yo sé, no existe ninguna investigación que trate sobre las variaciones de la temperatura de la piel del cuerpo en la muñeca ni sobre su comportamiento ante situaciones de violencia de género relacionadas con el miedo. Estos hechos son esenciales para este trabajo de investigación teniendo en cuenta la propuesta de pulsera dentro del sistema Bindi, ya que el sensor de temperatura está directamente unido a la muñeca debido al propio factor de forma. En los capítulos siguientes se ofrecen más detalles sobre los resultados obtenidos 5 y 6.

2.6 Conclusión

En este Capítulo, hemos proporcionado los fundamentos necesarios para plantear un sistema de reconocimiento de emociones. Hay que tener en cuenta que los aspectos técnicos relacionados con el diseño específico para el entrenamiento de dicho sistema se proporcionan en el siguiente Capítulo.

Así, se han revisado y detallado las diferentes teorías emocionales y técnicas de clasificación de emociones humanas. Específicamente y orientado al caso de uso particular de esta investigación, se propone un nuevo enfoque pragmático para fusionar las clasificaciones discretas y dimensionales de las emociones humanas hacia la identificación de la emoción miedo. Además, se ofrece un análisis exhaustivo de los factores intrapersonales que afectan a la modulación de la emoción, como los rasgos personales, la atención y el sesgo de género, para establecer futuras posibilidades de investigación que se desarrollarán como una extensión de este trabajo de investigación. Además, se presentan y comparan las diferentes herramientas de elicitación de emociones utilizadas dentro de la comunidad de la computación afectiva, destacando la reciente inclusión de VR, que está superando los experimentos de elicitación de emociones. Por otro lado, se han revisado y analizado las señales fisiológicas de interés para este trabajo de investigación, detallando su comportamiento y características y estudiando su relación con la emoción miedo. Nótese que la comprensión de dicha información fisiológica es esencial para cuantificar y distinguir adecuadamente los diferentes patrones fisiológicos que son producto de una reacción

emocional.

Sobre esta base, podemos concluir que, aunque las emociones sentidas están sesgadas por diferentes factores intrapersonales, la información fisiológica puede ser utilizada como una cuantificación o medición indirecta de dichos estados afectivos, ya que estas señales son controladas por el ANS, junto con sus evaluaciones subjetivas autoinformadas. En este contexto, la conjunción de diferentes señales fisiológicas, más que el uso de una sola de ellas, puede ser utilizada para dar lugar a un sistema informático afectivo inteligente capaz de distinguir diferentes estados afectivos. En la búsqueda de un sistema de reconocimiento de emociones de este tipo, que puede ser ampliado para su uso cotidiano, se destacan dos factores principales que se consideran esenciales para el desarrollo de Bindi en este caso. En primer lugar, la necesidad de tener en cuenta las dos clasificaciones de las emociones humanas, la discreta y la dimensional, puede ser una ventaja para explicar las diferentes características de las emociones. En segundo lugar, el análisis de múltiples fuentes fisiológicas de información en tiempo real es una tarea compleja ya que, desde la perspectiva de los wearables, están sometidas a diferentes fuentes de ruido que afectan directamente a la calidad de las señales y, por tanto, a la inferencia del reconocimiento de emociones. En los siguientes capítulos se ofrecen más detalles sobre la aplicación de todos los aspectos detallados relativos a las clasificaciones de las emociones humanas, las herramientas de elicitación de emociones y el desenrañamiento fisiológico y de las emociones.

Bases de datos y aprendizaje automático para el reconocimiento de emociones

En este capítulo, por un lado, ofrecemos un análisis completo sobre la estructura y los procedimientos experimentales utilizados para la generación de bases de datos diseñadas para el reconocimiento de emociones. Hay que tener en cuenta que, tal y como se especifica en el capítulo 2, estas bases de datos son esenciales para recoger las respuestas emocionales y entrenar los sistemas de reconocimiento de emociones. Además, también se explica cada parte de la cadena completa de procesamiento de datos para el sistema de computación afectiva que utiliza dichas bases de datos. Hay que tener en cuenta que la comprensión del estado actual de la generación de bases de datos ha sido esencial para diseñar adecuadamente la base de datos presentada en este trabajo. Además, se realiza una revisión crítica a lo largo de los diferentes apartados, aportando recomendaciones sobre lo que se debería considerar para la generación de una base de datos de reconocimiento de emociones y se explica lo que finalmente se ha aplicado para la generación de la nuestra, que se detalla en su totalidad en el Capítulo 6.

Antes de entrar en detalles de cada una de las partes implicadas tanto en la generación de la base de datos como en el diseño del sistema de computación afectiva, se muestra en la Figura 3-1 una representación general de dichos elementos y acciones. Como se indica en el capítulo 2, una base de datos para el entrenamiento de un sistema de computación afectiva se compone de los siguientes elementos principales 1) estímulos, 2) señales físicas y fisiológicas, 3) etiquetas, y 4) voluntarios. El segundo

y el tercer elemento se utilizarán para entrenar y validar el sistema de computación afectiva, mientras que el primero es necesario para provocar reacciones emocionales en los voluntarios. En este contexto, el proceso de construcción de una base de datos implica las siguientes tareas en relación con estos elementos:

- a) Antes de la generación de la base de datos:
 - Se recolecta un pool de estímulos.
 - En caso de enfrentarse a un experimento de tiempo limitado, se aplican diferentes métodos para reducir el número de estímulos del pool anterior.
 - El conjunto final de estímulos se organiza para ser utilizado durante la generación de la base de datos.
 - Los diferentes sensores se seleccionan validados para recoger toda la información durante la elicitación de la emoción.
- b) Durante la generación de la base de datos:
 - Se recogen y almacenan las diferentes variables a medir durante la recepción de los estímulos.
 - Se recogen y almacenan los datos autodeclarados (etiquetas de emoción) para identificar la información fisiológica y física con respecto a los estímulos específicos.
- c) Después de la generación de la base de datos:
 - El filtrado y acondicionamiento digital se utiliza para limpiar las diferentes señales.
 - Se realiza un análisis exploratorio de los datos para identificar anomalías e incluso problemas físicos o fisiológicos.
 - Extracción de diferentes métricas sintéticas y/o características de los datos. A partir de ellos, se aplica la reducción, selección y optimización.
 - En caso de estar bajo un caso de uso multimodal, se pueden abordar diferentes alternativas hacia la fusión de datos.
 - Aplicación de un proceso iterativo entre la propia arquitectura de fusión de datos, el algoritmo de clasificación y los procesos de ajuste fino de los hiperparámetros.
 - Liberación del modelo con el mejor rendimiento.

Este capítulo está estructurado de la siguiente manera. En la primera sección, se

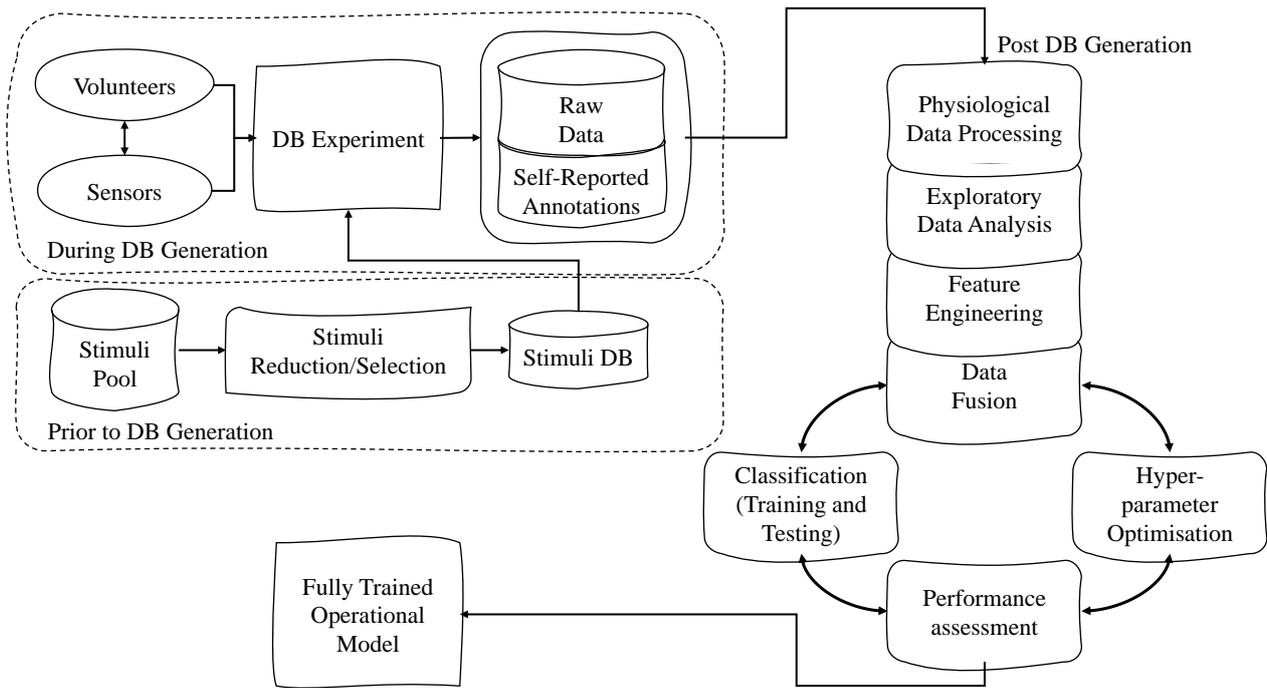


Figure 3-1: Elementos, procesos y acciones comunes necesarios para la generación de una base de datos de reconocimiento de emociones (*Volunteers - Voluntarios, Sensors - Sensores, DB Experiment - Experimento de la BBDD, Raw Data - Datos crudos, Self-Reported Annotations - Anotaciones reportadas, Stimuli Pool - Conjunto de Estímulos, Stimuli Reduction/Selection - Selección/Reducción de estímulos, Physiological Data Processing - Procesamiento de los datos fisiológicos, Exploratory Data Analysis - Análisis exploratorio de los datos, Feature Engineering - Ingeniería de los datos, Classification - Clasificación, Performance assessment - Evaluación del desempeño, Data Fusion - Fusion de Datos, Hyperparameter Optimisation - Optimización de hiperparámetros, Fully trained operational model - modelo completamente entrenado*).

explican y revisan uno a uno los elementos, procesos y acciones comunes necesarios para la generación de una base de datos de reconocimiento de emociones. La siguiente sección proporciona un resumen detallado de las diferentes bases de datos multimodales de reconocimiento de emociones que están disponibles en la literatura. Esta Sección también detalla cómo dichas bases de datos han abordado los diferentes puntos explicados en la Sección anterior, así como sus limitaciones y aplicabilidad a nuestro caso de uso. En la tercera sección, se destacan las diferencias entre las configuraciones de laboratorio y de campo para la generación de la base de datos. Hay que tener en cuenta que, para este trabajo de investigación, la única base de datos presentada en este documento se basa en configuraciones de laboratorio. Sin embargo, las conclusiones de esta última sección pueden utilizarse para la generación de bases de datos sobre el terreno en un futuro próximo.

3.1 Metodología general de las bases de datos

Los elementos y procesos que intervienen en la generación de una base de datos de reconocimiento de emociones se esquematizan en la Figura 3-1. En las siguientes subsecciones se explican sus particularidades, ventajas e inconvenientes. Hay que tener en cuenta que el resultado final es un modelo operacional completamente entrenado, basado en procesos no en línea (*offline*) que se realizan después de la generación de la base de datos. Por lo tanto, en este capítulo no se aborda ninguna optimización digital integrada para cada proceso. La integración y la optimización integradas para los dispositivos portátiles en tiempo real se detallan en el capítulo 5.

3.1.1 Selección y análisis de estímulos

La generación de un conjunto de estímulos adecuados es la primera etapa para cualquier base de datos centrada en el reconocimiento de emociones. Este paso es esencial, ya que cuanto mejor estén preetiquetados los estímulos, mejor será la verdad de la investigación del experimento. Esta última puede ser incluso una herramienta útil para ser comparada y analizada con las anotaciones de cada voluntario dentro de la base de datos, como se detalla en la sección 2.4. Así, la situación ideal es que la metodología de etiquetado fuera la misma durante la selección de los estímulos y durante el experimento de la base de datos. Obsérvese que en este tipo de experimentos, en realidad tenemos dos tipos de etiquetas (verdad absoluta o *ground truth*), las que provienen de la selección de estímulos y las que provienen de las valoraciones autoinformadas de los voluntarios.

En la literatura ya existe una amplia gama de opciones de bases de datos de estímulos. Por ejemplo, en 1997 el Instituto Nacional de Salud Mental lanzó el sistema internacional de imágenes afectivas como una base de datos que contenía cientos de imágenes etiquetadas [58]. Este sistema ha crecido desde entonces, hasta llegar a más de 1000 imágenes etiquetadas, e incluso se ha adaptado a otras culturas, como el español [144]. Las etiquetas contenidas en esta base de datos se caracterizan por la media y la desviación estándar de las tres dimensiones del espacio PAD. Aparte de los grupos de estímulos de imagen disponibles públicamente, también se encuentran bases de datos basadas en vídeo [12, 138, 145]. Independientemente

del tipo de estímulo, el análisis y la selección de estímulos son esenciales, ya que algunas de estas bases de datos no son prácticas para ser aplicadas en su totalidad (todos los estímulos) dentro de un experimento. Por lo tanto, se pueden observar diferentes enfoques en la literatura para filtrar y seleccionar conjuntos más pequeños de estímulos. Algunos de ellos [138] utilizaron la llamada puntuación de realce emocional basada en el espacio PA y dada por

$$|e| = \sqrt{a^2 + v^2}, \quad (3.1)$$

donde a y v son los valores de excitación y valencia obtenidos de las etiquetas recogidas respectivamente. Así, los estímulos con $|e|$ más altos serán los que proporcionen una mayor intensidad emocional en términos de las dimensiones a y v . Obsérvese que esta ecuación puede ampliarse para que sea válida y aplicable a otras dimensiones. Las puntuaciones obtenidas para todos los estímulos pueden clasificarse para seleccionar un grupo más pequeño con los valores más altos. Otros, como [12], realizaron un experimento de preetiquetado en el que partían de un gran conjunto de estímulos y pedían a la gente que etiquetara esos estímulos mediante un sistema de anotación afectiva en línea. Aseguraron un número mínimo de etiquetas por vídeo, y finalmente seleccionaron las que obtuvieron el mayor grado de acuerdo. La evolución de este último método incluye estadísticas de orden superior aplicadas a los experimentos preetiquetados para evaluar el acuerdo entre diferentes anotadores. Por ejemplo, los autores de [145] calcularon las distancias de Jaccard para cada par de anotadores y calcularon la desviación media absoluta de la distribución de distancias acumuladas para, finalmente, considerar como valores atípicos aquellos que se desviaban más de un umbral específico. No obstante, uno de los pasos seguidos junto con la selección de estímulos es la evaluación del equilibrio de los estímulos. Este proceso asegura que los estímulos seleccionados se distribuyen de forma equitativa a lo largo de todas las diferentes emociones que hay que detectar o clasificar. Por ejemplo, un enfoque común es tratar con el modelo PA, sólo en aras de la simplicidad en comparación con el modelo PAD, y normalizar las calificaciones utilizando la media y la desviación estándar (μ/σ), para luego trazar la excitación normalizada frente a la valencia normalizada y proceder a evaluar el estado de equilibrio de los estímulos.

En realidad, este paso está muy condicionado por el objetivo de la experimentación. La mayoría de las bases de datos de estímulos disponibles para el reconocimiento de emociones están pensadas y diseñadas desde una perspectiva emocional general, es decir, con el objetivo de identificar emociones en general sin dirigirse a modelos emocionales binarios especializados. Incluso los procedimientos preetiquetados suelen ser realizados por el público en general, sin tener en cuenta ninguna evaluación de expertos. Este enfoque es totalmente comprensible desde una perspectiva de uso general y masivo, sin embargo, para trabajos de investigación como el que se aborda en este documento puede no ser adecuado. Teniendo en cuenta que el objetivo principal de este trabajo de investigación se basa en la generación de sistemas de computación afectiva para la detección del miedo en situaciones de violencia de género, la selección de los estímulos debe hacerse con especial cuidado y tales bases de datos pueden no ser adecuadas para cumplir con los requisitos en términos de elicitación de emociones específicas. En el capítulo 6 se ofrecen detalles más específicos sobre cómo hemos abordado la selección y el análisis de los estímulos para la generación de nuestra base de datos.

3.1.2 Procesamiento y adquisición de los sensores

Durante el experimento de cualquier base de datos, diferentes sensores están adquiriendo señales fisiológicas y/o físicas del voluntario mientras se le aplican estímulos. Estos sistemas sensoriales deben diseñarse adecuadamente teniendo en cuenta los siguientes aspectos:

- Como la generación de una base de datos no es más que una enorme recolección de datos para luego crear sistemas inteligentes a partir de ella, se recomienda tener frecuencias de muestreo relativamente altas. Esto permite experimentar con cualquier frecuencia de muestreo más baja en la etapa de entrenamiento para observar cómo eso limita y afecta a los diferentes modelos de clasificación.
- Independientemente de la frecuencia de muestreo, hay que garantizar la sincronización entre los diferentes sensores durante el experimento. Sin embargo, para aliviar este proceso, otra opción es almacenar las marcas de tiempo globales de cada uno de los datos de los sensores recibidos para asegurar aún más que corresponden a la misma franja horaria del experimento.
- Se recomienda el uso de un conjunto de herramientas de detección homologado

o estándar, que se utilizará como sistema de medición dorado. Esto permitirá la comparación posterior con otras bases de datos, así como la detección de fallos en el sistema de sensores propuesto.

- Los sensores deben estar ubicados preferentemente en la mejor posición de medición, lo más cerca posible de la ubicación final del cuerpo y, en caso de que se prevea una posterior integración vestibular del sistema informático afectivo resultante, donde se haya previsto.

Estos factores son recomendaciones basadas en el estado del arte [26] y en los conocimientos recogidos a lo largo del desarrollo de este trabajo de investigación.

En cuanto al procesamiento de los datos, la primera tarea es aplicar un filtrado básico mediante el uso de filtros digitales de paso de banda baja y alta. Para las señales que poseen una alta sensibilidad al ruido, como PPG, se pueden aplicar procedimientos Signal Quality Assessment (SQA) específicos y algoritmos MAR [146, 147]. Lo mismo se aplica a las señales que necesitan algoritmos especiales de separación de componentes, como EDA, como se explica en la sección 2-14. En el capítulo 2-14 se ofrecen más detalles sobre las diferentes técnicas y algoritmos diseñados y aplicados en este trabajo. Nótese que hoy en día ya existen en la literatura herramientas de libre acceso diseñadas para el procesamiento fisiológico. Por ejemplo, Soleymani et. al. en [148] diseñó una caja de herramientas abierta para el procesamiento de un conjunto completo de señales fisiológicas y la extracción de características relacionadas con las emociones. Además, se han encontrado diferentes cajas de herramientas de procesamiento de señales fisiológicas, especializadas en una sola señal fisiológica, [149, 150]. Sin embargo, estas cajas de herramientas están pensadas y orientadas a un proceso de diseño de sistemas basado en PC o fuera de línea, dejando de lado las limitaciones de los wearables integrados.

Junto con el procesamiento de datos, también se aplica la segmentación de datos sobre las diferentes señales. De hecho, la mayoría de los sistemas de reconocimiento de emociones en la literatura utilizan ventanas de procesamiento segmentadas para tratar y analizar los datos fisiológicos adquiridos. A la hora de abordar la segmentación de los datos, hay que tener en cuenta aspectos relacionados con las ventanas, como su resolución temporal y de frecuencia y la latencia emocional. Por un lado, la resolución temporal tiene una relación directa con la resolución de frecuencia.

Esto se debe a que es necesario garantizar una resolución de frecuencia específica para extraer información emocional útil para algunos rasgos fisiológicos [8]. Por otro lado, la latencia emocional está relacionada con el hecho de que una persona no experimenta la misma respuesta fisiológica (emoción) durante toda la recepción de un estímulo [151, 152]. Este último aspecto puede afectar definitivamente al rendimiento del sistema, ya que está relacionado con el posible etiquetado incorrecto de las muestras.

3.1.3 Análisis exploratorio de los datos

Una vez que las señales han pasado por todo el procesamiento de datos necesario, se recomienda realizar un análisis exploratorio de los datos. Este proceso puede realizarse utilizando los datos filtrados y/o las características extraídas. Este proceso puede proporcionar una visión excelente de lo que realmente, a primera vista, está ocurriendo durante el experimento. Además, también puede dar una idea de los casos en los que el sensor está funcionando mal y las etapas de filtrado o procesamiento no pudieron solucionarlo. Este tipo de análisis exploratorio de datos nos permite determinar algunos de los comportamientos fisiológicos durante las diferentes etapas de los experimentos y llevar a cabo acciones específicas para tratar algunos problemas, como que la recuperación fisiológica no funcione como se esperaba o alargar los estímulos ya que la latencia emocional estaba afectando a algunas de las respuestas fisiológicas. En el capítulo 6.

Las diferentes bases de datos de reconocimiento de emociones disponibles públicamente no informan de este análisis exploratorio de datos, los trabajos publicados se centran en la generación de la base de datos (el proceso de recogida de datos). El análisis exploratorio de datos es una tarea que requiere mucho tiempo, pero el efecto fisiológico del experimento es muy útil desde el punto de vista de la detección de emociones. Por ello, otras investigaciones han realizado este análisis tras la publicación de las diferentes bases de datos. Por ejemplo, los autores de [153] utilizaron una de las bases de datos públicas abiertas en la literatura [138], siete años después de su lanzamiento, y concluyeron que las emociones inducidas eran más fuertes en la parte final de los estímulos, basándose en un análisis exploratorio de datos sobre los datos fisiológicos filtrados. Esa conclusión les llevó a entrenar su sistema propuesto utilizando sólo los últimos 20 segundos de cada estímulo. Así, estas y otras

consideraciones técnicas resultaron en un aumento significativo de la precisión de la tasa de reconocimiento de emociones en comparación con las técnicas de clasificación de emociones existentes en el estado del arte. Sin embargo, a pesar de las ventajas que este proceso puede aportar en relación con la eficacia de los sistemas de computación afectiva generados, implica una cantidad considerable de tiempo, así como la necesidad de un buen conocimiento de las señales fisiológicas. Esta última consideración es, en realidad, el factor más desafiante, ya que los diferentes patrones fisiológicos pueden variar en gran medida entre sujetos y sesiones experimentales. De hecho, las recientes revisiones de reconocimiento de emociones en la literatura ni siquiera abordan nada relacionado con este tema [26, 49, 133, 154, 155].

3.1.4 Ingeniería de características

La ingeniería de características implica la utilización de diferentes mecanismos para mejorar el rendimiento del modelo de reconocimiento de emociones. Obsérvese que sólo se aplica a las estrategias de aprendizaje automático convencionales y a las estrategias de aprendizaje profundo en las que las entradas son las características extraídas. Por lo tanto, hay que hacer una distinción esencial antes de entrar en detalles sobre la ingeniería de características. Por un lado, el aprendizaje automático convencional y el aprendizaje profundo que utiliza la extracción de características requieren técnicas de extracción ad hoc, así como la optimización, Figura 3-2. Por otro lado, existen métodos de aprendizaje profundo que no necesitan una etapa de extracción de características, ya que pueden aprender patrones y principios inherentes directamente de los datos procesados para extraer características ya optimizadas automáticamente. Estos últimos métodos se conocen como soluciones end-to-end, y parecen ser muy prometedores para el reconocimiento de emociones en problemas fisiológicos y multimodales [156–158]. Sin embargo, las metodologías de aprendizaje profundo, ya sea que dependan de características elaboradas a mano o aprendidas, siguen requiriendo una cantidad considerable de recursos. Por ejemplo, TensorFlow Lite, que es hoy en día uno de los marcos de aprendizaje automático de código abierto más utilizados para dispositivos de baja potencia y muy restringidos, puede desplegar modelos de aprendizaje profundo con un tamaño de 300 KB a 1 MB¹. Desafortunadamente, cuando se considera el diseño de dispositivos wearables de ul-

¹<https://www.tensorflow.org/lite/guide> (consultado: 01/03/2022)

tra bajo consumo utilizando la tecnología actual de System on Chip (SoC), estos tamaños de memoria pueden perjudicar otras tareas críticas que se deben realizar dentro de dichos dispositivos. Sin embargo, cabe mencionar que se está aplicando un enorme esfuerzo para potenciar el aprendizaje profundo en los sistemas de computación de borde, como la fundación TinyML ² o la tecnología optimizada de potencia subumbral ³ de Ambiq Micro Inc. Así, para este trabajo de investigación, nos centramos en la arquitectura de aprendizaje automático convencional, dejando el aprendizaje profundo embebido y/o los sistemas profundos de extremo a extremo para futuras investigaciones. En las siguientes subsecciones se discuten los diferentes procesos que se pueden realizar para la estrategia de ingeniería de características. Nótese que estos se llevan a cabo una sola vez durante el entrenamiento del sistema, posterior a la generación de la base de datos, pero antes del despliegue del sistema.

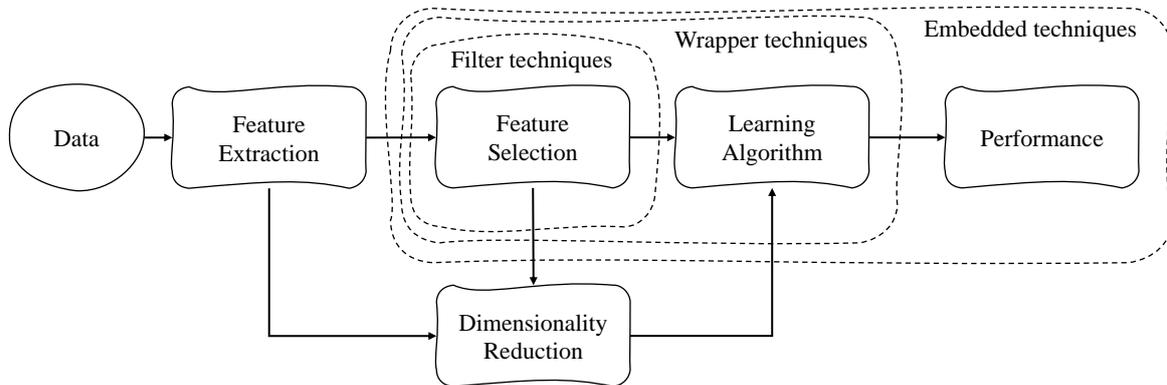


Figure 3-2: Procesos convencionales de ingeniería de rasgos para la selección supervisada de rasgos.

3.1.4.1 Extracción de características

La primera tarea dentro del proceso de ingeniería de características es la extracción de métricas sintéticas a partir de los datos filtrados previamente. Independientemente de la gran cantidad de técnicas de extracción de características que se pueden encontrar en la literatura para el reconocimiento de emociones [26, 49, 154], pueden ser divididas en tres categorías:

- Dominio temporal. Estas características poseen la menor complejidad computacional entre los diferentes tipos de técnicas de extracción de características. Obsérvese que la mayoría de las características del dominio temporal pueden implementarse de forma lineal ($\mathcal{O}(n)$). En concreto, proporcionan

²<https://www.tinyml.org/> (Consultado: 01/03/2022)

³<https://ambiq.com/> (Consultado: 01/03/2022)

información sobre los aspectos estacionarios y lineales de la serie temporal analizada. La mayoría de ellos se extraen mediante cálculos estadísticos de orden superior. Una de las mayores desventajas de estas características es la incapacidad de captar el comportamiento fisiológico no estacionario.

- Dominio de la frecuencia. El objetivo de estas características es obtener la PSD en bandas de frecuencia específicas para las diferentes señales fisiológicas. El método habitual para obtener la PSD se basa en la Discrete Fourier Transform (DFT) utilizando el algoritmo Fast Fourier Transform (FFT). En este caso concreto, la complejidad temporal suele ser de hasta $\mathcal{O}(n \log n)$ en comparación con las características temporales. Además, tratar el dominio de la frecuencia es sinónimo del problema de la resolución de la frecuencia temporal. Este último hecho es de especial relevancia en el caso de la información fisiológica, ya que algunas de ellas son señales que cambian lentamente, como la EDA, que se sabe que tiene respuestas que varían en el tiempo de 1 a 30 s en función del tipo de estímulo [119]. Además de este problema, existe información fisiológica que es una señal muestreada de forma desigual o no uniforme, lo que hace imposible la aplicación del algoritmo FFT. Este problema se aborda en la literatura empleando diferentes técnicas como la interpolación previa o el periodograma de Lomb-Scargle [159]. Por lo tanto, aunque el contenido frecuencial ha demostrado ser una medida fiable para rastrear las emociones, hay que tener en cuenta diferentes compensaciones en cuanto a la optimización de estas técnicas, así como las necesidades de resolución de la frecuencia (tamaño de almacenamiento de la ventana temporal y capacidad de procesamiento).
- Métodos no lineales. Para desentrañar las propiedades dinámicas y no estacionarias de las señales fisiológicas, se utilizan diferentes métodos. Hay que tener en cuenta que este tipo de características también se denominan características caóticas. De hecho, los trabajos que han contemplado, utilizado e incluso comparado las características no lineales frente a las temporales o frecuenciales han obtenido una mejora considerable del rendimiento en su objetivo específico de reconocimiento de emociones [160, 161]. Además, en los últimos años, la aplicabilidad del aprendizaje profundo a los problemas de reconocimiento de emociones ha aumentado debido a los prometedores resultados

obtenidos [153, 162, 163], lo que es un indicativo del componente emocional no lineal fisiológico también. Dentro de este contexto y en consonancia con la importancia no lineal, diferentes estudios y revisiones recogieron y analizaron el comportamiento fisiológico no lineal [91, 92], como se señala en la sección 2.5. La principal y mayor desventaja de estas técnicas es la complejidad temporal que tienen, ya que puede ser de hasta $\mathcal{O}(n^2)$.

La mayoría de los sistemas de reconocimiento de emociones basados en señales fisiológicas y que utilizan etiquetas emocionales bien conocidas se basan en la extracción convencional de características temporales y de frecuencia. Por lo tanto, la combinación de los tres dominios (temporal, de frecuencia y no lineal) debería ampliarse en la literatura. Este enfoque podría aprovecharse para comprender mejor las variaciones fisiológicas y los cambios relativos a las métricas autodeclaradas que se utilizan como etiquetas en este tipo de sistemas. Hay que señalar que, la categorización proporcionada por este trabajo de investigación se basa en las diferentes revisiones comentadas en la literatura. Sin embargo, puede haber técnicas de extracción de características más específicas o incluso nombres diferentes para las categorías propuestas. Por ejemplo, las características morfológicas [164], siendo aquellas que se refieren a propiedades específicas de la señal fisiológica (amplitudes, tiempos, número de picos, etc.), solían emplearse también indistintamente a las técnicas de dominio temporal. En el caso de este trabajo de investigación, hemos elaborado un compendio de las características más relevantes y exitosas considerando las tres categorías revisadas. En el capítulo 4 se ofrecen más detalles sobre su aplicación específica.

Una vez que se han extraído con éxito las características, es el momento de optimizarlas. Esta optimización puede realizarse mediante la selección y/o la reducción de características [165]. La primera se basa en identificar las características más relevantes y crear nuevos subconjuntos de características con ellas, mientras que la segunda se ocupa de la reducción de la dimensionalidad del problema mediante distintos tipos de transformación de bases. Obsérvese que el proceso de transformación de bases se refiere a la conversión de las características extraídas de alta dimensión, es decir, de alto número de características, en un espacio de baja dimensión con una pérdida mínima de información. Ambos métodos de optimización de características son esenciales para simplificar el modelo (menos almacenamiento, mejor

visualización, reducción de datos, la navaja de Occam), para evitar la maldición de la dimensionalidad y para reducir el tiempo de entrenamiento.

3.1.4.2 Selección de características

Para el procedimiento de selección de características, diferenciamos tres técnicas comunes, que se esquematizan en la Figura 3-2 y se aplican para el reconocimiento de emociones según la relación con los métodos de aprendizaje [166]. En primer lugar, podemos encontrar las técnicas más sencillas conocidas como métodos de filtro. Estos se basan en métricas estadísticas generales, como la correlación con la variable dependiente, mediante las cuales se clasifican las diferentes características para seleccionar posteriormente el nuevo subconjunto. Aunque poseen la menor complejidad computacional, son más propensos a fallar en la selección de las mejores características, ya que no se considera ni la interacción entre ellas ni el efecto del nuevo subconjunto en el rendimiento del clasificador. En segundo lugar, para evitar los problemas de los métodos de filtrado, encontramos los métodos de envoltura. Estos utilizan el clasificador para verificar el efecto del rendimiento de los nuevos subconjuntos generados de forma iterativa. Dos de los métodos envolventes más conocidos y utilizados son Sequential Forward Selection (SFS) y Sequential Backward Elimination (SBE). En concreto, el primero se inicializa con un subconjunto vacío de características y comienza a combinarlas hasta que no se observa ninguna mejora, mientras que el segundo realiza la misma operación al revés, comenzando con todas las características y eliminándolas una a una. Se sabe que los métodos de envoltura proporcionan un mejor rendimiento a costa de: (1) altas necesidades de cálculo cuando el número de características es relativamente alto, y (2) riesgo de sobreajuste cuando el número de muestras de entrada es relativamente bajo. Además, están fuertemente condicionados al tipo de clasificador utilizado durante las diferentes iteraciones de envoltura. Por último, el tercer tipo de estos métodos se conoce como métodos integrados. Se crearon para hacer frente a los diferentes inconvenientes de las dos técnicas anteriores y mantener sus ventajas. En este caso, el mecanismo de selección de características está integrado en el núcleo del algoritmo de clasificación y aprovecha su selección de características y su clasificación al mismo tiempo. Esto proporciona una complejidad computacional y una velocidad incluso comparada con las técnicas de filtro y siendo mucho menos propensa al sobreajuste.

Nótese que las técnicas comentadas utilizan las etiquetas o la variable objetivo, lo que se conoce como selección de características supervisada. Sin embargo, también hay métodos que no necesitan la variable objetivo, como las técnicas basadas en la correlación. Estas últimas técnicas pueden proporcionar información sobre la relación entre las diferentes características para descartar aún más la información redundante.

3.1.4.3 Reducción de dimensionalidad

Como ya se ha comentado, otra posibilidad para optimizar el espacio de características es aplicar la reducción de características. Este método se basa en una transformación no supervisada de las características extraídas en un espacio de características completamente nuevo. Por ejemplo, una de las técnicas más comunes es Principal Component Analysis (PCA), en la que cada nueva característica se obtiene mediante una combinación lineal de las características originales. En concreto, PCA calcula las matrices de covarianza de las características originales para luego extraer sus vectores propios y cada valor propio correspondiente. A continuación, los vectores propios se clasifican por los valores propios en orden descendente (de más a menos información transportada) y sólo se guardan los de interés. Dichos eigenvectores almacenados se juntan dando lugar a la matriz de proyección, que se utilizará para la proyección de los datos originales. Una de las principales desventajas es que PCA puede producir que las variables independientes sean menos interpretables, ya que las características originales se convierten en componentes principales. Este método ha sido ampliamente utilizado para la reducción de características en el reconocimiento de la emoción y otros problemas de aprendizaje automático [155, 167, 168]. Nótese que, además de PCA, existe una gran variedad de métodos en la literatura en cuanto a la reducción de características, como el t-Distributed stochastic neighbour embedding, el análisis discriminante generalizado, o el análisis de componentes independientes [169].

Teniendo en cuenta que la búsqueda del subconjunto ideal de características, ya sea por selección o por reducción, es un problema NP-hard, la única forma de obtener una solución óptima es realizando una búsqueda exhaustiva dentro del espacio de la solución o dentro de la aplicación de diferentes técnicas de reducción de características. Sin embargo, incluso considerando que este proceso puede realizarse durante

el entrenamiento del sistema sin implicar ninguna restricción digital incorporada, se trata de una tarea difícil. Además, la gran variedad de técnicas y la investigación activa en este campo introducen aún más complejidad al problema. Por lo tanto, la propuesta, el desarrollo y/o la implementación de nuevas técnicas de reducción de la dimensionalidad están fuera del alcance de este documento. En su lugar, a lo largo del desarrollo de este trabajo de investigación, se han aplicado diferentes métodos de selección de características comúnmente utilizados para nuestro caso de uso específico. En los capítulos 4 y 6 se ofrecen más detalles sobre su implementación.

3.1.5 Optimización de Hyperparámetros

El término hiperparámetro se refiere a los valores que intervienen en el proceso de aprendizaje de los distintos algoritmos de aprendizaje automático y que no pueden estimarse a partir de los datos. Cuando se trata de aprendizaje automático convencional, el proceso de ajuste de hiperparámetros puede mejorar mucho el modelo de clasificación durante el entrenamiento. Sin embargo, en cuanto a la selección de características, este proceso también es un problema de dificultad NP, ya que los hiperparámetros perfectos se obtienen después de haber verificado todas las combinaciones diferentes y posibles. Para simplificar, imaginemos un problema de aproximación por mínimos cuadrados (un problema de regresión lineal), en el que el ajuste del modelo se evalúa mediante el residuo de cada punto dado por

$$r = y - f(x), \quad (3.2)$$

donde r es el residuo obtenido para la muestra observada y al considerar el modelo definido por $f(x)$. Suponiendo que la aproximación del modelo es una línea recta, la ecuación anterior resulta en

$$r = y - (b + mx), \quad (3.3)$$

donde b es la intercepción con la variable dependiente y m es la pendiente del modelo de línea recta. De hecho, estos son los parámetros del modelo que afectan directamente al ajuste de los puntos de datos observados. Sin embargo, para encontrar el ajuste óptimo, hay que evaluar una función de pérdida para todas las combina-

ciones posibles. Por ejemplo, las técnicas de mínimos cuadrados utilizan funciones de pérdida cuadráticas para minimizar los residuos. Dado que la ejecución de todas las combinaciones posibles es una tarea que consume mucho tiempo, en la literatura se utilizan diferentes técnicas para optimizar esta búsqueda y proporcionar una máquina bien optimizada, es decir, el algoritmo [170]. Uno de los métodos más sencillos para hacerlo es establecer un número máximo de iteraciones para verificar dicha función de pérdida en función de un tamaño de paso específico o una tasa de aprendizaje mientras se avanza hacia el mínimo de dicha función de pérdida. Estos últimos valores se fijan antes de ejecutar el modelo y son externos a él, identificándose como hiperparámetros. Aunque hay muchas técnicas de optimización de hiperparámetros, hemos revisado tres de ellas: la búsqueda

- Búsqueda en cuadrícula. Esta técnica se basa en una cuadrícula predefinida de combinaciones de hiperparámetros, es decir, un espacio preestablecido de posibles combinaciones, que se ejecutan y prueban secuencialmente. Este método suele ser muy exhaustivo, pero al mismo tiempo consume mucho tiempo. Por ejemplo, si tomamos tres hiperparámetros y comprobamos 50 valores para cada uno, eso da como resultado un total de 125.000 combinaciones a probar. Por lo tanto, la búsqueda en cuadrícula se puede utilizar para una primera aproximación al problema, sabiendo que no va a ser ni la mejor ni la más barata en términos de consumo de recursos y tiempo [171].
- Búsqueda aleatoria. Esta técnica sigue el mismo concepto que la búsqueda en cuadrícula, es decir, se realiza una búsqueda sobre un espacio preestablecido de combinaciones posibles. Sin embargo, en lugar de evaluar esas combinaciones secuencialmente, la técnica utiliza combinaciones aleatorias dentro de dicho espacio. La cantidad de iteraciones está limitada explícitamente por el diseñador. En general, se ha demostrado que este método proporciona mejores modelos en la mayoría de los casos y requiere menos tiempo de cálculo [172].
- Optimización bayesiana. Uno de los puntos débiles de las dos técnicas anteriores es que la evaluación de nuevos puntos o combinaciones de hiperparámetros dentro de la malla no considera ninguna información sobre la evolución de la puntuación a lo largo del proceso de optimización. Así, el ajuste bayesiano de hiperparámetros se conoce como una técnica Sequential Model-Based Optimi-

sation (SMBO) que utiliza el conocimiento de las iteraciones anteriores para concentrarse en las mejores puntuaciones de pérdida de la función, es decir, se basa en un modelo probabilístico gaussiano continuamente actualizado que permite elegir la siguiente combinación de hiperparámetros de forma informada para potenciar la evaluación de valores más prometedores [173].

Sobre el uso específico de estas técnicas para este trabajo de investigación se ofrecen más detalles sobre el uso de las mismas en los capítulos 4 y 6.

3.1.6 Fusión de datos

El carácter interdisciplinario de los problemas de la informática afectiva, cuyo objetivo es el reconocimiento de las emociones, junto con los avances tecnológicos abren un sinfín de posibilidades en cuanto a la observación de las modalidades. Obsérvese que el término modalidad se refiere a la adquisición de datos multisensorial, en la que cada sensor está destinado a captar datos de fuentes de información totalmente diferentes (por ejemplo, auditiva, fisiológica, textual, visual). Por ejemplo, nuestro cerebro ya trabaja sobre la base de información multisensorial y toma decisiones basadas en la fusión de datos. De hecho, los autores de [174] realizaron una revisión detallada y exhaustiva de los experimentos multimodales en la literatura en comparación con los uni-modales. A través de esa revisión, confirmaron que los sistemas multimodales superan a los unimodales. Además, también señalan que las técnicas de aprendizaje profundo están ganando terreno al aprendizaje automático convencional mediante el uso de modelos de aprendizaje profundo de extremo a extremo, que no necesitan los pasos de extracción de características, ya que pueden alimentarse directamente utilizando los datos en bruto [158].

Dentro de este contexto, podemos categorizar a Bindi como un sistema multimodal, en el que tenemos dos modalidades diferentes: fisiológica y de audio. Estas pueden fusionarse empleando diferentes metodologías de fusión de datos, que se describen a continuación:

- Fusión temprana. Este método se basa en realizar o aplicar la tarea de fusión en la fase inicial del problema, es decir, utilizando los datos o incluso las características. Lo primero puede hacerse eliminando la información correlacionada entre modalidades, mientras que lo segundo fusiona las diferentes características (de diferentes modalidades) en un solo vector de características. Por

ejemplo, la figura 3-3 representa un posible ejemplo de fusión temprana, en el que la extracción de características se aplica de forma independiente sobre ambas modalidades y el vector de características resultante es sólo la concatenación de las mismas. Este último proceso es uno de los métodos más rápidos para fusionar características, sin embargo, puede realizarse aplicando otras técnicas como la adición por puntos. Una de las principales ventajas de este método de fusión es que sólo es necesario entrenar un modelo de clasificación.

- Fusión tardía de elementos. En este caso, las fuentes de información siguen caminos totalmente independientes, que incluso pueden no tener los mismos componentes o procesos, hasta dar una salida de clasificación por modelos de clasificación diferentes e independientes según la modalidad. La figura 3-3 representa un posible ejemplo de fusión tardía, en el que ambas modalidades tienen modelos de clasificación independientes y la salida de los mismos se fusiona. Tanto si el modelo proporciona una etiqueta blanda (cualquier métrica de salida que proporcione información sobre la probabilidad predicha de pertenencia a una clase, por ejemplo, un 50% de probabilidad de pertenencia a la clase positiva) como una etiqueta dura (clase predicha sin ninguna información de probabilidad, por ejemplo, etiqueta "1" y "0" para la clase positiva y negativa), pueden utilizarse diferentes técnicas para realizar dicha fusión de datos. Por ejemplo, una de las técnicas más comunes es la realización de un esquema de ponderación [155] dado por

$$c = \arg \max \left\{ \prod_{m=1}^M P_i(X|C_m)^{\alpha_m} \right\}, \quad (3.4)$$

donde M es la cantidad total de modalidades, X es la entrada de datos, $P_i(X|C_m)$ es la probabilidad de que X pertenezca a la clase i y la proporcione el clasificador de una modalidad específica C_m . Los diferentes pesos para cada modalidad α_m se determinan durante la etapa de entrenamiento y sólo necesitan satisfacer $\sum_{m=1}^M \alpha_m = 1$. La principal ventaja de la fusión tardía es el diseño ad-hoc que se puede realizar para las diferentes modalidades de forma independiente, sin embargo, ese hecho también conduce a la necesidad de más de un clasificador.

- Fusión intermedia. Esta fusión de datos implica la transformación de las características extraídas en una nueva representación de los datos originales, es decir, el cambio de base. Se suele aplicar sobre todo cuando se trata de modelos de aprendizaje profundo, en los que la fusión de datos puede producirse en cualquier parte de las capas internas de la red neuronal. Esta fusión es más flexible en comparación con las otras dos, en las que la información se fusiona tanto al principio como al final. Sin embargo, hay muy pocos ejemplos de esta técnica en la literatura, al contrario que la anterior.

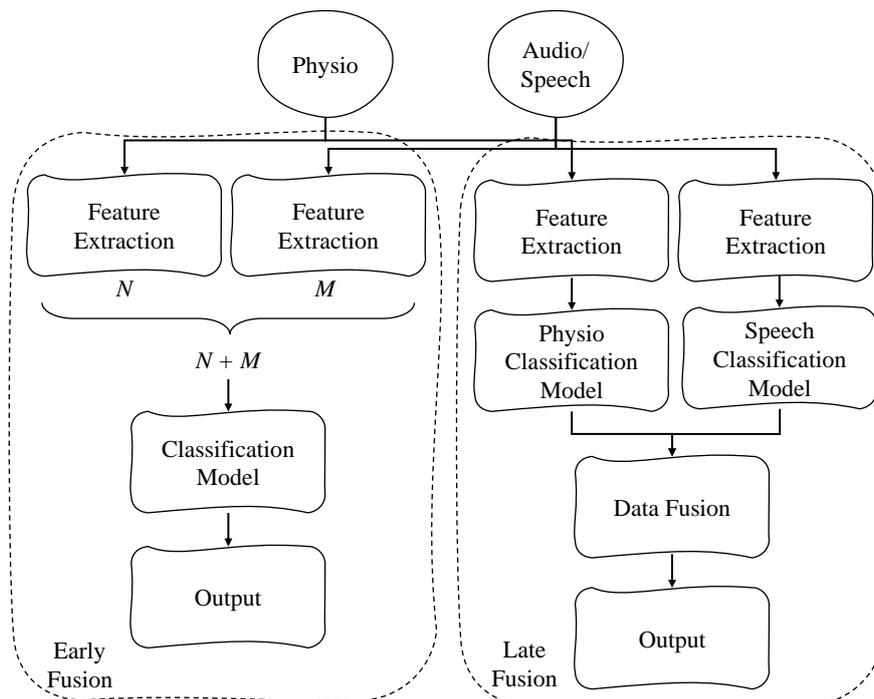


Figure 3-3: Técnicas de fusión temprana y tardía de datos para características fisiológicas y de audio/voz, con dimensiones N y M respectivamente.

En el capítulo 6 se ofrecen más detalles sobre el uso específico de estas técnicas para este trabajo de investigación.

3.1.7 Clasificación emocional

Esta etapa, junto con la fusión de datos, es una de las últimas que se llevan a cabo para conseguir un modelo de computación afectiva totalmente entrenado y probado, véase la Figura 3-1.

3.1.7.1 Compensación del sesgo (*bias*) y varianza (*variance*)

Antes de describir los diferentes modelos que son de interés para este trabajo, se podría explicar y abordar adecuadamente el trade-off Bias-Varianza para entender

todos los conceptos asociados dentro de esta etapa.

El rendimiento de los algoritmos de aprendizaje automático se define principalmente por su sesgo y su varianza. La relación entre estas métricas está directamente relacionada con los problemas de infra y sobreajuste. Por ejemplo, consideremos la ecuación 3.2, pero suponiendo que $f(x)$ define la verdadera relación entre y y x . En ese caso, y suponiendo que creamos o diseñamos una función $f'(x)$ que corresponde a nuestro modelo de aprendizaje automático, la calidad de dicho algoritmo en puntos de prueba no vistos puede medirse por el error cuadrático medio (MSE) como

$$MSE_{f'} = E[(y - f'(x))^2], \quad (3.5)$$

lo cual se puede descomponer en

$$\begin{aligned} MSE_{f'} &= E[(f(x) + r - f'(x))^2] \\ &= E[(f(x) + r - f'(x) + E[f'(x)] - E[f'(x)])^2] \\ &= E[(f(x) - E[f'(x)])^2] + E[(E[f'(x)] - f'(x))^2] + E[r^2] + \\ &\quad + 2E[(E[f'(x)] - f'(x))(f(x) - E[f'(x)])] \end{aligned} \quad (3.6)$$

$$Note : E[E[f'(x)]] = f'(x)$$

: el último término se cancela, es cero.

$$\begin{aligned} &= E[(f(x) - E[f'(x)])^2] + E[(E[f'(x)] - f'(x))^2] + E[r^2] \\ &= bias[f'(x)]^2 + variance[f'(x)] + \sigma_r^2, \end{aligned}$$

donde r es el ruido residual o aleatorio con media cero y varianza σ_r^2 ($E[r^2]$), *bias* es la diferencia entre el valor medio esperado de la predicción y el valor real, y *varianza* cuantifica la consistencia del valor de predicción de salida en función de la variación de los puntos de datos de entrenamiento. Un ejemplo ilustrativo de estos conceptos se muestra en la Figura 3-4, de la que se pueden obtener diferentes conclusiones:

- Un modelo con un sesgo alto y una varianza baja se encuentra en la zona de infraajuste, siendo incapaz de ajustarse a los datos de entrenamiento. Este hecho conduce a elevados errores de entrenamiento y prueba.
- Un modelo con un sesgo bajo y una varianza alta está dentro de la zona de sobreajuste, ajustándose demasiado a los datos de entrenamiento y siendo

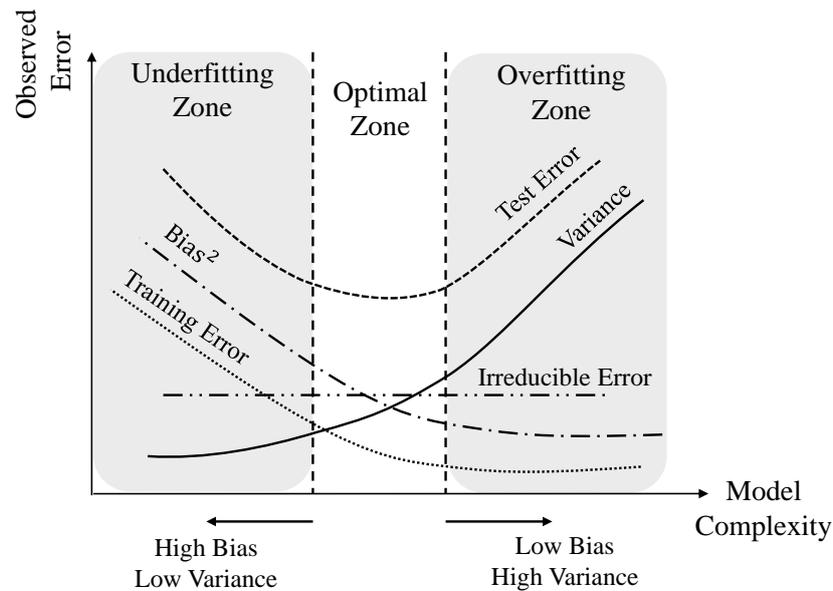


Figure 3-4: Compensación entre el sesgo y la varianza con el infraajuste (*Underfitting Zone*), el sobreajuste (*Overfitting Zone*) y las zonas óptimas (*Optimal Zone*).

incapaz de generalizar o ajustarse a nuevos datos de prueba no vistos. Este hecho conduce al menor error de entrenamiento a costa de un elevado error de prueba.

- El mejor modelo es el que minimiza los errores de los valores predichos erróneos (bajo sesgo) y presenta una consistencia robusta a las variaciones de los datos de entrenamiento (baja varianza). Esto se identifica como la zona óptima, en la que el modelo logra el equilibrio perfecto entre el error de entrenamiento y el de prueba.
- Incluso cuando se consigue el menor sesgo y varianza dentro de la zona óptima, la calidad del modelo vendrá determinada por el error irreducible, que es irrelevante para el modelo y está relacionado con el ruido inherente a los datos.

Hay que tener en cuenta que es esencial considerar el comportamiento habitual del sesgo y la varianza de los distintos algoritmos de aprendizaje automático que se van a evaluar. De hecho, los algoritmos tradicionales de aprendizaje automático sufren este problema de compensación a medida que aumenta su complejidad.

3.1.7.2 Algoritmos de aprendizaje máquina

Para este trabajo de investigación, se utilizan diferentes modelos de aprendizaje automático bien conocidos, basados en revisiones actuales centradas en el reconocimiento de emociones [26]. Se describen como sigue:

- Máquinas de vectores de apoyo (SVM) |citesvmreview. Este algoritmo de clasificación supervisada es uno de los algoritmos de aprendizaje automático más populares. Aunque originalmente fue propuesto únicamente para problemas de clasificación binaria, a lo largo de los años se ha extendido y aplicado también para problemas multiclase. La idea principal de este clasificador se basa en encontrar un hiperplano que separe mejor los datos en las diferentes clases. Hay que tener en cuenta que los datos son las diferentes características extraídas que se introducen en el clasificador. En este contexto, es necesario definir dos elementos principales para entender el concepto de hiperplano: los vectores de soporte y los márgenes. Como se muestra en la Figura refch3:svmhyperplane, suponiendo una clasificación binaria con dos características, los vectores de soporte o puntos del vector de soporte son los más cercanos al hiperplano (parte media del margen). Desde una perspectiva 2D, el hiperplano puede conceptualizarse como la línea que separa ambas clases dada por la ecuación 3.3.

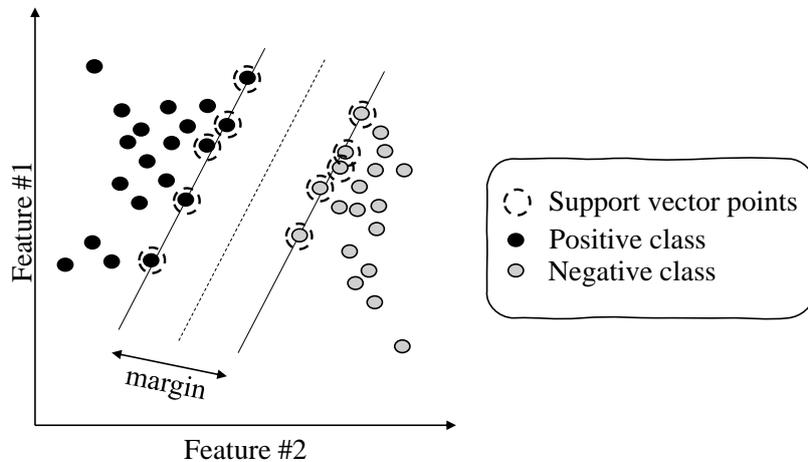


Figure 3-5: Ilustración del hiperplano para el clasificador SVM para la clasificación binaria (los puntos negros son la clase positiva, y los puntos grises son la clase negativa).

Sin embargo, para definir el hiperplano completo, dicha ecuación se expande o se generaliza a las M dimensiones del problema como sigue,

$$\begin{aligned}
 y &= w_0 + w_1x_1 + w_1x_1 + \dots + w_Mx_M \\
 &= w_0 + \sum_{n=1}^M w_nx_n \\
 &= b + w^T X,
 \end{aligned}
 \tag{3.7}$$

donde w^T son los puntos del vector soporte, X son los puntos de entrenamiento proporcionados, b es el término sesgado o el desplazamiento de dicho hiperplano, y y es la etiqueta de clase (positiva o negativa para un problema binario). Así, podemos definir cualquier hiperplano como el conjunto de puntos que satisfacen

$$w^T X + b = 0. \quad (3.8)$$

Obsérvese que, teniendo en cuenta estas ecuaciones, el problema de optimización para obtener el hiperplano óptimo se basa en maximizar el margen para separar mejor los datos en las diferentes clases, como ya se ha comentado anteriormente. Por tanto, dicho problema de optimización consiste en realidad en seleccionar dos hiperplanos iniciales que cumplan las siguientes restricciones:

$$w^T X + b \geq 1, \text{ for the positive class} \longrightarrow y = 1, \quad (3.9)$$

$$w^T X + b \leq -1, \text{ for the negative class} \longrightarrow y = -1. \quad (3.10)$$

These constraints can be rearranged and expressed by the following,

$$y * (w^T X + b) \geq 1. \quad (3.11)$$

No obstante, las ecuaciones y supuestos anteriores sólo son válidos si los datos son linealmente separables, lo que no ocurre cuando se trata de información fisiológica debido a la naturaleza no lineal de la misma. En estos casos, la ecuación anterior se modifica añadiendo un parámetro extra, ζ , que permite o contabiliza el error de clasificación durante el entrenamiento. Esto nos lleva a los márgenes suaves, en lugar de los márgenes duros, con la siguiente formulación:

$$y_i * (w^T X + b) \geq 1 - \zeta. \quad (3.12)$$

Además, se utiliza un hiperparámetro C para manejar ese coste de clasificación errónea y mantener el control de los márgenes suaves. Sin embargo, en la mayoría de los casos en que los datos no son linealmente separables, la aplicación de márgenes suaves no es suficiente, y es necesario aplicar diferentes kernels.

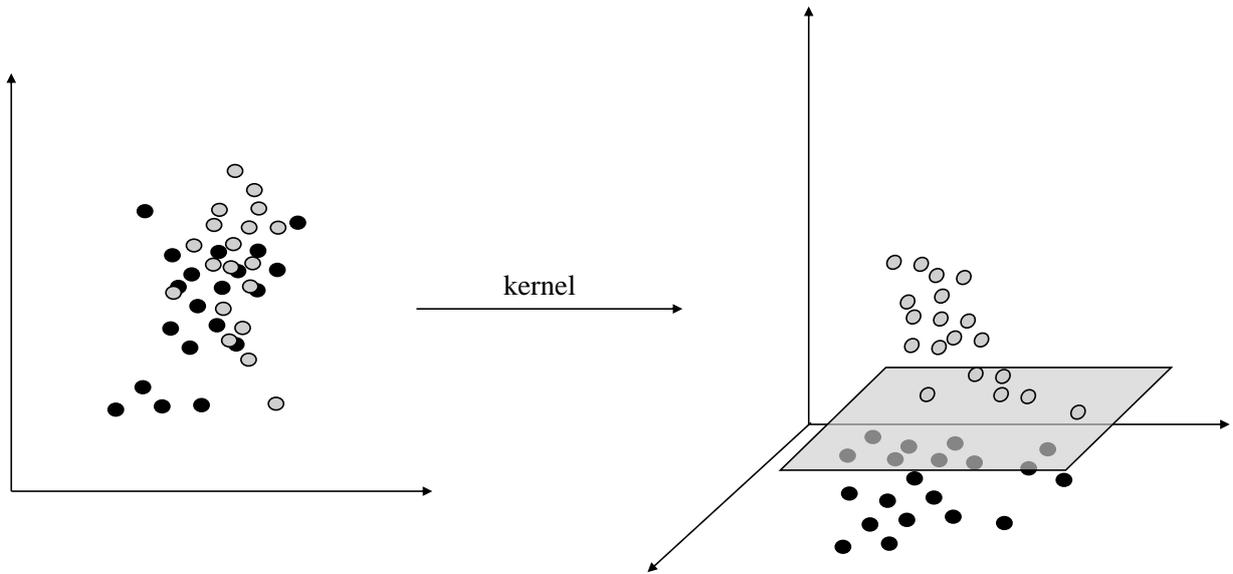


Figure 3-6: Ilustración del truco del núcleo para un problema binario.

La aplicación de un kernel puede considerarse como un mapeo de los datos en dimensiones más altas, de modo que puedan ser linealmente separables en un nuevo espacio de características de mayor dimensión, Figura 3-6.

Además, estos kernels utilizan el llamado truco del kernel mediante el cual no es necesario conocer ni preocuparse por estas transformaciones de mayor dimensión, ya que las funciones del kernel admiten entradas en el espacio dimensional inferior original y devuelven el producto punto de los vectores transformados en el espacio dimensional superior. Por ejemplo, uno de los núcleos comúnmente empleados en los sistemas de reconocimiento de emociones y también utilizado en este trabajo de investigación es el núcleo gaussiano o de función de base radial (RBF), que viene dado por

$$K(X_1, X_2) = \exp\left(-\frac{\|X_1 - X_2\|}{2\sigma^2}\right), \quad (3.13)$$

donde $\|X_1 - X_2\|$ es la distancia euclidiana (norma L2) entre los puntos de datos (puntos de datos de características) X_1 y X_2 , y σ es el hiperparámetro que se debe ajustar para considerar que dos puntos son similares (pertenecen a la misma clase). Nótese que este kernel está acotado superiormente por 1, ya que la distancia entre dos puntos que son extremadamente similares es cero. En función del valor de σ , la región de similitud (zona donde $K(X_1, X_2)$ es superior a cero) entre los puntos cambiará, Figura 3-7. Este algoritmo

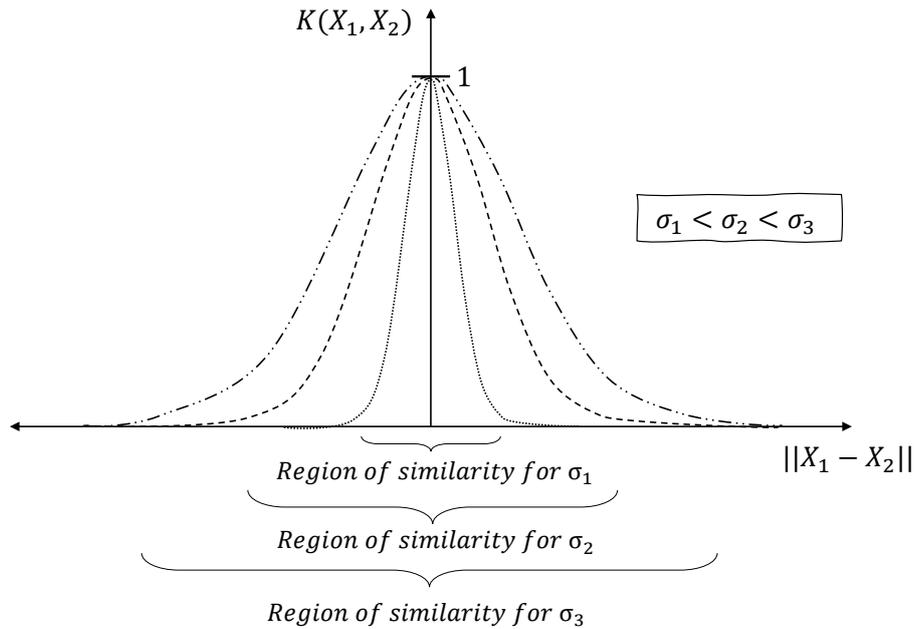


Figure 3-7: RBF valores del núcleo en función de la distancia entre los dos puntos que se evalúan para diferentes σ .

es un clasificador discriminativo cuyo sesgo y varianza están determinados por los hiperparámetros C y σ para los márgenes suaves y el kernel RBF respectivamente. La principal ventaja es que presenta una mayor eficiencia de memoria en comparación con otros clasificadores (sólo necesita almacenar los vectores de soporte, no todos los puntos de datos de entrenamiento), pero no funciona bien cuando se lleva con demasiado solapamiento entre las diferentes clases.

- K-Nearest Neighbours (KNN) [175]. Este es también un algoritmo de clasificación supervisado, sin embargo, se llama un clasificador *lazy*. Desde un punto de vista matemático, no hay un proceso de aprendizaje real dentro del algoritmo. En su lugar, busca la mejor distancia d y el número de vecinos k que maximiza la separación de las clases. Por lo tanto, la realización de predicciones ante la llegada de nuevos datos requiere el cálculo de dicha distancia con cada uno de los puntos de datos de entrenamiento y la posterior comparación con los k vecinos circundantes para determinar la clase de pertenencia. Este algoritmo asume que existen cosas similares en estrecha proximidad. Hay que tener en cuenta que se pueden utilizar diferentes tipos de distancias (euclidiana, Minkowski, bloque de ciudades, Mahalanobis, etc.), así como diferentes algoritmos de ordenación para encontrar los k vecinos más cercanos tras el

cálculo de las distancias.

Desde un punto de vista práctico, el KNN es uno de los algoritmos más sencillos de aplicar. Por lo tanto, es una elección correcta para un primer enfoque de prueba de concepto. Sin embargo, se vuelve significativamente más lento a medida que aumenta el número de muestras de entrenamiento, además de afectar a la eficiencia de la memoria.

- Métodos de conjunto (ENS) [176]. Estos métodos son en realidad un conjunto de técnicas de aprendizaje automático, más que un clasificador. Se basan en la combinación de diferentes modelos base o clasificadores débiles para producir un clasificador óptimo o fuerte. Dicha combinación se suele realizar en forma de "bagging" o "boosting". En el bagging, cada modelo se entrena de forma independiente con el mismo conjunto de entrenamiento, mientras que en el boosting, cada clasificador débil se entrena teniendo en cuenta el rendimiento del clasificador anterior aplicando un mecanismo de ponderación de datos (se asignan pesos más altos a las instancias clasificadas incorrectamente).

Para este trabajo de investigación se utilizan métodos de boosting ensemble y, en concreto, se aplica el clasificador Adaptive Boosting o AdaBoost. Este clasificador es muy popular para la clasificación binaria y los clasificadores débiles que se emplean para aplicarlo suelen ser árboles de decisión (árboles con un solo nodo o árboles de decisión de un solo nivel) o árboles poco profundos (árboles con una profundidad muy limitada). Obsérvese que este tipo específico de árboles mejora la comprensibilidad. Así, para cada aprendiz débil (m) y para todas las instancias del conjunto de entrenamiento (N), este clasificador calcula el error de clasificación ponderado como

$$\epsilon_m = \frac{\sum_{i=1}^N w_i^{(m)} I(f_m(x_i) \neq y_i)}{\sum_{i=1}^N w_i^{(m)}}, \quad (3.14)$$

donde $w_i^{(m)}$ es el peso de la instancia i para el aprendiz m , y I es la función de pérdida definida por

$$I(f_m(x), y) = \begin{cases} 0, & \text{if } f_m(x_i) = y_i \\ 1, & \text{if } f_m(x_i) \neq y_i \end{cases} \quad (3.15)$$

Tras el entrenamiento, este clasificador predice la etiqueta de la nueva información no vista siguiendo una combinación lineal ponderada de todos los clasificadores débiles considerados (M), que viene dada por lo siguiente

$$g(x) = \text{sign} \left(\sum_{m=1}^M \alpha_m f_m(x) \right), \quad (3.16)$$

donde α_m es el peso total asignado a cada aprendiz débil dado por

$$\alpha_m = \frac{1}{2} \log \frac{(1 - \epsilon_m)}{\epsilon_m}. \quad (3.17)$$

Desde una perspectiva práctica, la etapa de inferencia de este algoritmo de aprendizaje automático requiere menos almacenamiento y posee una menor complejidad computacional y temporal en comparación con los dos algoritmos revisados anteriormente. Sin embargo, es más sensible a los datos ruidosos y a los valores atípicos, por lo que requiere que los datos estén debidamente filtrados y libres de ruido antes de alimentar a la máquina.

3.1.7.3 Técnicas de validación cruzada

Además del modelo específico que se va a aplicar la separación de los datos en conjuntos de entrenamiento, validación y prueba debe realizarse antes del proceso de entrenamiento. Aunque la separación de estos conjuntos de datos suele estar integrada en las tareas del procedimiento de clasificación, también puede conceptualizarse como una operación adicional, véase "Evaluación del rendimiento" en la figura 3-1. En este contexto, el conjunto de entrenamiento puede definirse como el conjunto del que el modelo va a aprender los patrones subyacentes y ajustar sus hiperparámetros. El conjunto de validación es el afectado por las técnicas de validación cruzada, a través del cual se puede obtener una estimación del rendimiento del modelo. Hay que tener en cuenta que este conjunto es en realidad parte del conjunto de entrenamiento (el modelo ve estos conjuntos en el entrenamiento). Por último, la parte de prueba es la que no es vista por el modelo durante el entrenamiento y proporciona la evaluación final e imparcial de un modelo completamente entrenado. Sin embargo, existen diferentes métodos para separar estos conjuntos, y la selección de una técnica u otra depende principalmente de la cantidad de datos y de la necesidad de afinar los hiperparámetros. Estas técnicas de separación se agrupan bajo el

término CV. Para este trabajo de investigación se han utilizado diferentes técnicas de CV adaptadas a las bases de datos de reconocimiento de emociones. Se describen a continuación:

- Retención de datos. Este es el método más sencillo de CV en el que los datos se dividen en dos conjuntos (de entrenamiento y de validación). Durante este CV, el modelo se ajusta con el primero y se evalúa utilizando los datos dentro del segundo. Obsérvese que el modelo final entrenado se obtiene utilizando todo el conjunto de datos (entrenamiento y validación). Aunque también se puede obtener un tercer conjunto de datos para considerarlo como el conjunto de datos de prueba, datos no vistos que no se utilizan en absoluto durante la etapa de entrenamiento. Una proporción típica de división es del 80% para el entrenamiento y del 20% para la prueba, aunque esta proporción depende del conjunto de datos. La principal desventaja de este método es el riesgo de sobreajuste (alta varianza) ya que los diferentes conjuntos (diferentes distribuciones de la división) pueden incluso afectar a los resultados obtenidos. Hay que tener en cuenta que, al reducirse el conjunto de datos de entrenamiento cuando se utiliza esta técnica, se puede incluso correr el riesgo de perder patrones inherentes a las señales o datos. No se puede asegurar un control estricto de los datos específicos dentro del conjunto de entrenamiento después de la división, lo que incluso afecta a los resultados para diferentes conjuntos (diferentes distribuciones de la división).
- La técnica anterior se basa en la disminución del tamaño del conjunto de entrenamiento original, lo que puede dar lugar a conjuntos de datos más pequeños y aumenta el riesgo de perder los patrones inherentes a las señales o los datos. Para superar las limitaciones de la técnica anterior y por lo tanto, para hacer frente a esos problemas y disminuir la varianza del entrenamiento, este método se basa en dividir el conjunto de entrenamiento en k particiones, que pueden proporcionar hasta k posibilidades diferentes para entrenar y validar el sistema. En comparación con la técnica anterior, se suele preferir este método, ya que puede ofrecer una medida más realista (menos optimista) del rendimiento del modelo. La principal desventaja de este método es el tiempo de cálculo necesario para ejecutar k veces el entrenamiento del modelo.

- Dejar una muestra para probarla (LOTO). En este método, se deja una muestra fuera del proceso de entrenamiento para posteriormente probar el modelo con ella. Sin embargo, para el caso de uso de reconocimiento de emociones, esta técnica puede modificarse para identificar una muestra como prueba del experimento. Por ejemplo, en un experimento basado en el registro fisiológico mientras se visualizan diferentes imágenes, un ensayo se identificaría como los datos fisiológicos capturados durante la visualización de una de las imágenes. Además, el hecho de que el número de posibles combinaciones de entrenamiento esté definido por el número de ensayos, hace que esta técnica tenga las mismas ventajas y desventajas que para el $k - fold CV$ con k igual al número de ensayos.
- Leave-One-Subject-Out (LOSO). Esta técnica sigue el mismo concepto que LOTO, pero en este caso la muestra que se deja fuera del entrenamiento es un sujeto o voluntario completo. Como se ha dicho anteriormente, considerando el mismo ejemplo de reconocimiento de emociones por imágenes, todos los datos recogidos de un sujeto se utilizan para las pruebas mientras que el entrenamiento se realiza con el resto de sujetos o voluntarios. La principal diferencia de esta técnica con respecto a LOTO es la variabilidad de datos que se observa en el conjunto de prueba. De hecho, el conjunto de pruebas en LOTO se basa en un solo ensayo, que se identifica con una etiqueta, mientras que el conjunto de pruebas en LOSO se basa en diferentes ensayos de los mismos voluntarios. Así, mientras que LOSO puede asegurar, al menos para un sujeto, una distribución de prueba representativa, LOTO está siempre sujeto a la incertidumbre de tener un conjunto de pruebas representado por una sola etiqueta.

Para este trabajo de investigación, se implementan algunas de estas técnicas para manejar la generación de los diferentes conjuntos (entrenamiento, validación y prueba) para los modelos de aprendizaje automático. En concreto, el problema del reconocimiento de emociones requiere la aplicación de estas estrategias para generar dos tipos de modelos: los dependientes del sujeto y los independientes. Los primeros se entrenan, validan y prueban utilizando los datos de un solo voluntario, mientras que los segundos utilizan los datos de todos los voluntarios para crear un modelo

global. La principal diferencia entre estos modelos es la personalización. De hecho, la mayor parte de la variabilidad entre los sujetos reside en la naturaleza dinámica de sus estados afectivos y su experiencia previa. Este hecho puede demostrarse por la superioridad de los modelos dependientes del sujeto sobre los modelos independientes del sujeto en la literatura [154]. Por lo tanto, en consonancia con el capítulo 2, la interpretación de los estímulos y los cambios fisiológicos dependen en gran medida de los voluntarios. Por lo tanto, surge la personalización, como se hizo en [26], en la que los autores concluyeron que se podría desplegar un modelo de reconocimiento de emociones independiente del sujeto pero, en algún momento, será necesaria la personalización del usuario para mejorar el sistema. Por estos motivos, existe la necesidad en la literatura, a la hora de afrontar el reconocimiento de emociones mediante machine learning convencional, de plantear nuevas técnicas de CV que aporten algún tipo de personalización. De hecho, observando otros campos que utilizan también información humana, se puede observar la aplicación de técnicas CV híbridas que básicamente combinan modelos [7] independientes y dependientes del sujeto. Dentro de este contexto y, hasta donde yo sé, no hay ningún trabajo de investigación de reconocimiento de emociones que aplique técnicas CV híbridas. Por ello, además de aplicar algunas de las técnicas revisadas, este trabajo de investigación propone la utilización de la técnica denominada Leave-half-Subject-Out (LASO) CV. Obsérvese que la representación gráfica de las técnicas LOSO, LOTO, y LASO CV se muestra en la Figura 3-8.

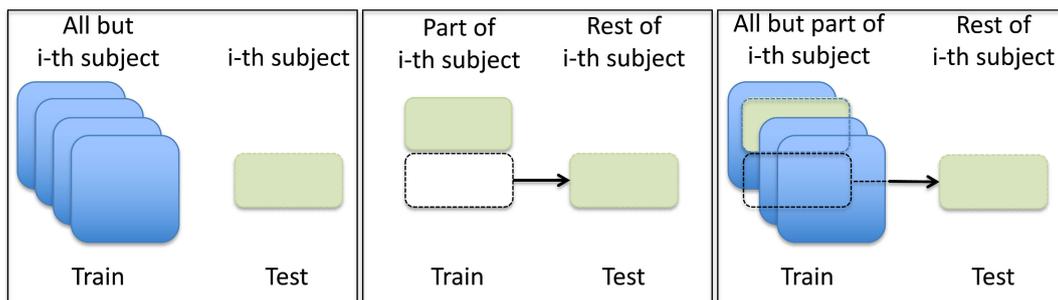


Figure 3-8: Representación gráfica de las técnicas LOSO, LOTO y LASO [7].

Independientemente del tipo de modelo y como se indica en la sección 3.1.2, los datos fisiológicos recogidos durante los experimentos se segmentan en ventanas de procesamiento. Éstas son sometidas a un solapamiento para aumentar el rendimiento de la delineación fisiológica, lo que puede afectar en gran medida a la interpretación de los resultados obtenidos cuando se utilizan las técnicas de CV detalladas. Por

ejemplo, en el caso de aplicar un CV de k pliegues sobre un vector de características extraídas de señales fisiológicas filtradas, con ventanas y superpuestas, podrían existir pliegues, es decir, ventanas de procesamiento, que contengan parte de la información del pliegue anterior. Este hecho puede llevar a una interpretación demasiado optimista de los resultados y, aunque depende de la longitud del solapamiento, debe evitarse. Así, centrándonos en el reconocimiento de emociones mediante el uso de señales fisiológicas, podrían preferirse estrategias que no indaguen en este problema, como LOTO, LOSO y LASO. Hay que tener en cuenta que esta última recomendación puede verse afectada por la cantidad de datos disponibles para el entrenamiento, lo que puede imposibilitar la aplicación de algunas técnicas como LOTO para modelos dependientes del sujeto.

3.2 Bases de datos abiertas

Dentro de la comunidad de la computación afectiva, diferentes conjuntos de datos se ocupan del reconocimiento de emociones utilizando señales fisiológicas. Las más comunes son MIT [177], DEAP [138], MAHNOB [12], DECAF [145], ASCERTAIN [178], y WESAD [179]. La tabla 3.1 resume los principales detalles de dichas bases de datos. Estas bases de datos abiertas son consideradas como un sólido punto de referencia por la comunidad científica. En esta sección se revisan las bases de datos de reconocimiento de emociones de interés para este trabajo de investigación. Nótese que anteriormente, en [180] y [181], realizamos un análisis detallado de algunas de estas bases de datos abiertas disponibles y aportamos conclusiones sobre sus metodologías y enfoques de reconocimiento emocional. Debido a la similitud con respecto a algunos de los mecanismos de elicitación emocional, las metodologías experimentales y, sobre todo, la información fisiológica de interés para este trabajo, se eligen sólo dos de ellas para seguir realizando pruebas de concepto de detección de miedo basadas en sus señales y estímulos, lo que se detalla en el capítulo 4.

Por un lado, las primeras pruebas de concepto de este trabajo de investigación se desarrollaron utilizando DEAP [138]. Esta base de datos contiene información fisiológica de 32 voluntarios (16 mujeres). El experimento consiste en un total de 40 video-clips de un minuto de duración cada uno. Los estímulos se seleccionaron a partir de un conjunto más amplio o de una etapa de preetiquetado basada en

Table 3.1: Las bases de datos de reconocimiento de emociones más comunes con una configuración de laboratorio utilizadas dentro de la comunidad científica de la computación afectiva.

Base de datos	Sujetos/as (M/F)	Etiquetas	Caso de Uso	Precisión	Año
MIT [177]	1 (0/1)	Discrete	General	81.00%	2005
DEAP [138]	32 (16/16)	Arousal/Valence	General	57.00/62.70%	2012
MAHNOB-HCI [12]	30 (13/17)	Arousal/Valence	General	46.20/45.50%	2012
DECAF [145]	30 (16/14)	PAD	General	55.00/60.00/50.00%	2015
ASCERTAIN [178]	58 (37/21)	Arousal/Valence	General	66.00/68.00%	2017
WESAD [179]	15 (12/3)	Arousal/Valence	Estres	86.46%	2018

las valencias, excitaciones y clasificaciones de dominancia recogidas por SAM. Los sensores periféricos (fisiológicos) incluidos son el electroencefalograma, el electro-miograma, la amplitud de la respiración, el GSR, el electrooculograma, el PPG y el SKT. En cuanto a algunas de las limitaciones de este conjunto de datos, debido a la configuración de laboratorio, los voluntarios estaban muy limitados en términos de movimiento, por lo que los modelos entrenados no son válidos para las condiciones de la vida real. Hay que tener en cuenta que se realizó un registro de línea de base de cinco segundos entre los estímulos utilizando una cruz de fijación en la pantalla. Por último, los autores de la base de datos crearon tres sistemas binarios, cada uno de los cuales infiere un nivel bajo o alto de arousal, valencia y agrado, en los que utilizaron las calificaciones autoinformadas como verdad de base (etiquetas). Presentaron estos resultados como evaluación comparativa y obtuvieron las siguientes métricas de precisión media (ACC) y puntuación F1: 57,00% (ACC) y 53,30% (puntuación F1) para la excitación, 62,70% (ACC) y 60,80% (puntuación F1) para la valencia, y 59,10% (ACC) y 53,80% (puntuación F1) para el gusto. Esta base de datos es de especial interés principalmente por dos factores. En primer lugar, contiene la misma información fisiológica que la pulsera de Bindi. En segundo lugar, las etiquetas autodeclaradas recogidas durante los experimentos contienen espacio PAD. En el capítulo 4 se ofrecen más detalles técnicos sobre el sistema de computación afectiva DEAP y los propuestos tras su publicación.

Por otro lado, el conjunto de datos MAHNOB-HCI incluye datos fisiológicos de 30 participantes en el estudio (17 mujeres) [12]. Este conjunto de datos de reconocimiento de emociones basado en el laboratorio contiene datos de un total de 20 clips de vídeo por voluntario, que se seleccionaron a partir de un conjunto mayor o de una etapa de preetiquetado como DEAP y que tenían una duración media de

aproximadamente 81 segundos, entre 34,9 y 117 segundos ($M = 81,4s$; $SD = 22,5s$). Las respuestas fisiológicas registradas se adquirieron utilizando el sistema Biosemi active II, e incluían ECG, GSR, amplitud de la respiración, SKT, electroencefalograma, mirada de los ojos y vídeos de la cara y el cuerpo. En cuanto al conjunto de datos DEAP, la configuración de laboratorio hace que los modelos entrenados no sean válidos para las condiciones de la vida real. Sin embargo, una de las principales diferencias con DEAP es que en MAHNOB los autores tuvieron en cuenta las recuperaciones emocionales de los voluntarios entre los estímulos, en lugar de limitarse a esperar cinco segundos entre ellos. De hecho, antes de ver cualquier vídeo emocional, se mostraron a los participantes diferentes clips neutros. Este proceso se utilizó para recuperar un nivel fisiológico basal, disminuir el sesgo emocional después de experimentar una emoción y, en definitiva, manejar las diferencias fisiológicas intrasujeto. Así, todas las respuestas registradas para cada estímulo contenían 30 segundos de datos al principio y al final del intervalo correspondiente a este proceso de recuperación. Por último, los autores de la base de datos crearon dos sistemas no binarios de reconocimiento de emociones, cada uno de los cuales infería un nivel bajo, medio y alto de arousal y valencia, respectivamente. Para obtener la verdad básica, utilizaron un mapeo entre las calificaciones de emoción discreta autoinformadas y las dimensiones emocionales basadas en [5]. Consiguieron un promedio de métricas de ACC y de puntuación F1 de hasta el 46,20

A pesar de los beneficios que estas bases de datos aportan a este trabajo de investigación, no están pensadas para elicitación específicamente el miedo y así poder detectar mejor la situación de riesgo en contextos de violencia de género. Así, pueden ser utilizadas para generar y estudiar pruebas de concepto para el motor de aprendizaje automático del miedo en el que se centra este trabajo de investigación e incluso aportar conclusiones preliminares para la amplia casuística dentro de esta compleja tarea. Sin embargo, tal y como se ha comentado en los capítulos anteriores, al ser el desentrañamiento entre las reacciones fisiológicas y el miedo en situaciones de violencia de género uno de los principales objetivos de este trabajo, se podría crear una nueva base de datos dirigida realmente a nuestro caso de uso específico. Además, dicha base de datos podría utilizar VR para proporcionar experimentos inmersivos de elicitación de emociones más fuertes. Más detalles sobre la base de datos creada

durante este trabajo de investigación y sus particularidades se encuentran en el capítulo 6.

3.3 Conclusión

En este capítulo, hemos proporcionado una revisión y un análisis completos para la generación y el procesamiento de bases de datos de reconocimiento de emociones, desde un punto de vista experimental hasta los procedimientos de procesamiento de datos que pueden aplicarse una vez terminada la base de datos.

En primer lugar, concluimos que no existe un protocolo estándar para el análisis y la selección de los estímulos. Todos los conjuntos de datos disponibles públicamente están pensados desde una perspectiva emocional general, es decir, con el objetivo de identificar las emociones en general sin dirigirse a modelos emocionales binarios especializados. Este hecho hace que la evaluación de los estímulos por parte de expertos no sea tan crítica. Pero, para trabajos de investigación como el que se aborda en este documento, esta estrategia no puede aplicarse y puede no ser adecuada. La situación condicionada por los estímulos de las víctimas de la violencia de género, así como sus posibles episodios de TEPT, hacen necesaria la ayuda de expertos para ajustar y seleccionar los estímulos que se presentarán durante nuestros experimentos. En segundo lugar, se recomienda encarecidamente realizar un análisis exploratorio de los datos para determinar algunos de los comportamientos fisiológicos y llevar a cabo acciones específicas para tratar algunos problemas, como la recuperación fisiológica de la elicitación de la emoción. En tercer lugar, hemos detallado diferentes recomendaciones sobre las técnicas de CV a aplicar cuando se trata de problemas de reconocimiento de emociones. Este hecho es de especial relevancia debido a la variabilidad inter e intra que puede existir entre los diferentes voluntarios en dichos experimentos. Esto hace que la validación, y otros procesos clave como la normalización, sean parámetros a explorar bajo la amplia casuística que presenta el problema de reconocimiento de emociones mediante el uso de información fisiológica y física. Así, se seleccionan preferentemente nuevas técnicas de CV que consideren la variabilidad intra e inter, como LASO, para ser utilizadas y aplicadas sobre las técnicas comunes.

El trabajo de investigación que se presenta en este documento versa sobre la

propuesta, estudio, diseño e implementación de una nueva base de datos de reconocimiento de emociones, diseño de aprendizaje automático del miedo y desarrollo de un dispositivo wearable edge. Esto hace que el conocimiento de este Capítulo sea esencial para entender los temas de los Capítulos siguientes.

Parte II

Clasificación del miedo usando el estado del arte

Prueba de concepto para clasificar miedo

Una vez revisado todo el estado del arte respecto a las emociones, la información fisiológica, las bases de datos para el reconocimiento de emociones, y los diferentes procedimientos de post-procesamiento para diseñar un modelo de aprendizaje automático totalmente probado, aplicaremos dichos conocimientos para diseñar diferentes sistemas de reconocimiento de emociones binarias de miedo utilizando las dos bases de datos detalladas en la Sección 3.2. En concreto, en este capítulo, las arquitecturas propuestas se basan únicamente en la parte fisiológica uni-modal de Bindi, teniendo en cuenta la descripción de Bindi en la Sección 5.2. Así, estas propuestas pretenden impulsar las primeras implementaciones embebidas de toda la cadena de procesamiento de datos, incluyendo el motor de aprendizaje automático, dentro de la pulsera inteligente de Bindi. Hay que tener en cuenta que la casuística y las posibilidades multimodales se tratan y detallan en el Capítulo 6.

En los siguientes apartados comenzaremos abordando tres sistemas iniciales desarrollados sobre la base de datos DEAP. Posteriormente, debido a algunas limitaciones observadas en DEAP, se utilizará la base de datos MAHNOB para diseñar otros dos sistemas de reconocimiento binario de emociones de miedo. Por último, se compararán todas las métricas de rendimiento generadas con respecto al estado del arte actual en materia de reconocimiento de emociones y, más concretamente, de detección de miedo. Además, para contextualizar el alcance de los resultados obtenidos, se tratan y discuten aspectos clave como el balance de clases, la selección de características y otros procesos. Cabe destacar que los diferentes sistemas presentados en este capítulo han sido diseñados y validados en un ordenador personal. En concreto,

se utilizó Matlab® como plataforma de software y todo el código desarrollado tomó Toolbox for Emotional feature extraction from Physiological signals (TEAP) [148] como referencia, que es una caja de herramientas de código abierto disponible actualmente para el procesamiento de datos fisiológicos y la extracción de características. De hecho, hemos estado en contacto con los desarrolladores de TEAP y hemos contribuido con correcciones a su repositorio. Finalmente, se ha desarrollado una nueva caja de herramientas totalmente automatizada a partir de esa base, que da cuenta desde el preprocesamiento de la señal hasta el entrenamiento y las pruebas de aprendizaje automático. Esta herramienta se ha aplicado para diseñar los sistemas presentados en este capítulo, así como para diferentes experimentos con otros conjuntos de datos y proyectos dentro del departamento bajo el que se ha realizado este trabajo de investigación.

Los sistemas de reconocimiento de emociones propuestos se ordenan de menor a mayor complejidad dentro de este Capítulo. De esta forma, las estrategias de investigación y desarrollo seguidas a lo largo de este trabajo han alimentado de forma incremental las diferentes implementaciones realizadas con Bindi. Así, independientemente de la complejidad de la arquitectura, la mayoría de ellas se basan en los componentes mostrados en la Figura 4-1, que representa una descripción global y general del entrenamiento del sistema de reconocimiento de miedos propuesto. Incluye los pasos típicos de la cadena de procesamiento analizada en el capítulo 3, desde el análisis del conjunto de datos de señales fisiológicas hasta el preprocesamiento de los datos brutos, la extracción de características y la clasificación de las emociones. De hecho, la mayoría de los sistemas de reconocimiento de emociones en la literatura siguen esta arquitectura pero centrándose en la clasificación de las emociones desde un punto de vista general, detectando un conjunto de emociones sin tener en cuenta si el usuario es hombre o mujer [133]. Sin embargo, si se dirige a la identificación de una sola emoción que podría estar relacionada con una situación específica y se tienen en cuenta las particularidades relacionadas con el género, se podría aprovechar para conseguir un sistema más preciso. Esta última afirmación se basa en la idea de que las mujeres reconocen la comunicación no verbal o la prosodia emocional con mayor precisión [54], como se revisa en la sección 2.3.3. Estos conceptos no se tienen en cuenta en ningún sistema actual de reconocimiento de

emociones mediante señales fisiológicas presentado en la literatura. Actualmente, hasta donde yo sé, no existe ningún sistema de detección de emociones desarrollado para identificar diferentes situaciones sociales críticas, como los episodios de violencia de género. En este contexto, se podría diseñar un sistema de detección de miedo especializado que activara un protocolo de protección que podría incluir una conexión con un círculo de confianza o incluso con las fuerzas del orden, para proporcionar inmediatamente la ayuda necesaria. Este último es uno de los principales objetivos de Bindi, tal y como se indica en el capítulo 1. Por lo tanto, el valor añadido de la arquitectura propuesta en este capítulo es doble: (1) la generación de una primera prueba de concepto para un sistema de reconocimiento binario especializado en el miedo utilizando únicamente información fisiológica (hasta ahora los enfoques del estado de la técnica tratan con varias emociones), y (2) la consideración de las limitaciones de procesamiento digital para seguir adaptando adecuadamente dicho sistema para integrarlo en una plataforma de dispositivo de borde vestible que permita la protección de las personas vulnerables.

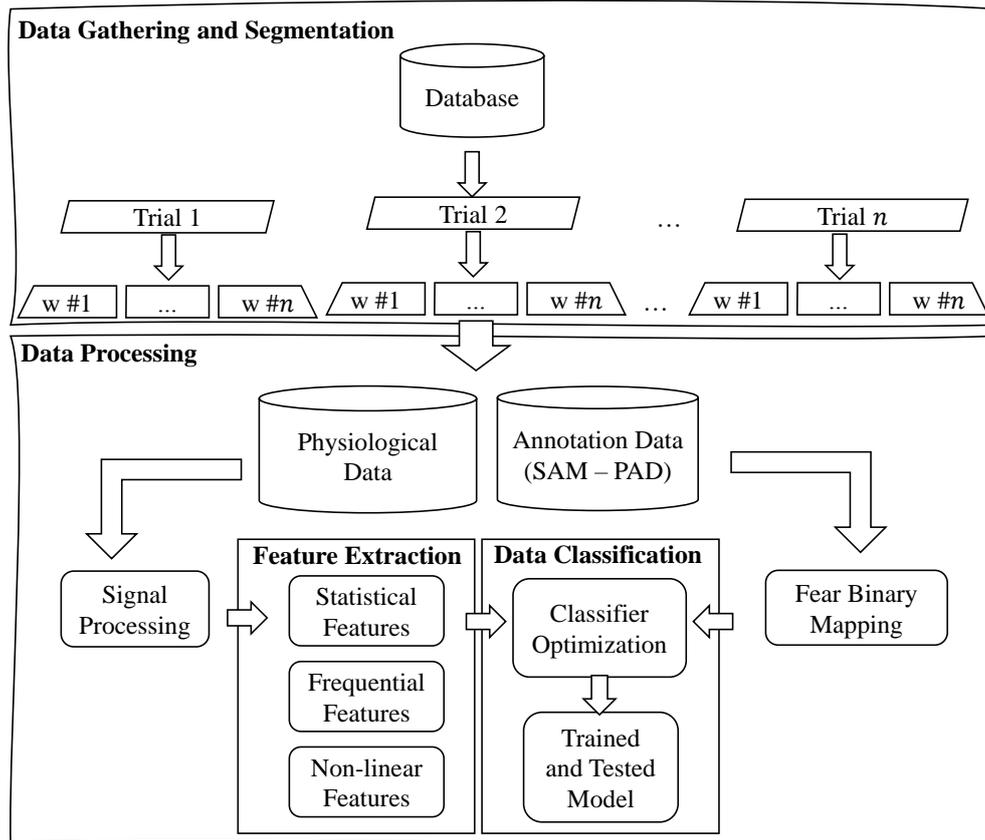


Figure 4-1: Visión general del proceso de entrenamiento para el sistema de reconocimiento de miedo propuesto que emplea datos de sensores fisiológicos y etiquetado de emociones de enfoque dimensional. Este último se introduce en el procedimiento de mapeo binario del miedo. Obsérvese que $w\#n$ denota las diferentes ventanas obtenidas tras la segmentación de los datos, si procede.

Cabe destacar dos consideraciones específicas de diseño relativas al tipo de modelos generados y a las restricciones específicas de incrustación de Bindi. Por un lado, a lo largo del diseño de estas pruebas de concepto iniciales, se ha abordado y perseguido la necesidad de un modelo robusto y fiable independiente del sujeto. El diseño de un modelo totalmente independiente del sujeto permitiría potenciar el proceso de despliegue inicial de cualquier herramienta tecnológica capaz de detectar cualquier emoción mediante el aprendizaje automático. Este hecho puede concretarse en Bindi para la detección de miedo en situaciones de Violencia de Género. Esta consideración es esencial para entender algunas de las decisiones tomadas y los parámetros estudiados en este capítulo y en los siguientes. Hay que tener en cuenta que este sistema independiente del sujeto, que se despliega en un proceso de configuración inicial, está sujeto a ser adaptado y personalizado posteriormente para el sujeto concreto para mejorar su rendimiento, tal y como se indica en la Sección 3.1.7.2. Por otra parte, el

proceso de diseño de todos los diferentes sistemas de reconocimiento de emociones binarias de miedo presentados en esta investigación también estuvo sesgado por restricciones específicas de recursos y capacidades integradas. Éstas se fijaron en 64 kB de RAM y 512 kB de Flash. Hay que tener en cuenta que estos recursos fueron impuestos por el equipo de investigación para reducir el diseño a una implementación ligera, sin embargo, se pueden establecer diferentes limitaciones teniendo en cuenta la respectiva mejora o empeoramiento del rendimiento. En el capítulo 6 se ofrecen detalles más específicos de la implementación integrada.

4.1 Clasificación del miedo usando DEAP

Como se detalla en la sección 3.2, la base de datos DEAP es una de las más utilizadas en la literatura en cuanto a reconocimiento de emociones con señales periféricas o fisiológicas. Aunque no es un conjunto de datos especializado en el miedo, es decir, los diferentes estímulos fueron seleccionados desde una perspectiva emocional general sin centrarse específicamente en ninguna emoción en particular, contiene los elementos necesarios para que diseñemos la primera prueba de concepto del sistema de detección de miedo basado exactamente en las mismas señales fisiológicas de nuestro interés, es decir, PPG, GSR, y SKT. El DEAP contiene datos de 32 participantes para un total de 40 videoclips, que se seleccionaron a partir de una etapa de preetiquetado tras las valoraciones de arousal, valencia y dominancia. Sin embargo, cabe destacar que el equipo de medición de esta base de datos fue el sistema Biosemi ActiveTwo¹, que es un equipo de medición profesional pensado para ser empleado en condiciones de laboratorio. Este hecho hace que las señales adquiridas se alejen de las mediciones reales obtenidas con dispositivos wearables. Así, los sistemas propuestos aquí sirven como prueba inicial de concepto y nos han permitido identificar diferentes aspectos clave a tener en cuenta tanto a la hora de diseñar una base de datos como de entrenar un modelo de aprendizaje automático a partir de dichos datos.

En cuanto a la metodología específica seguida durante los experimentos DEAP, la Figura 4-2 muestra un diagrama simplificado de la experimentación aplicada para cada voluntario y cada estímulo. Obsérvese que la línea de base de 2 minutos se

¹<http://www.biosemi.com>

aplicó justo al principio del experimento. De esta figura se aprecia una transición muy corta entre estímulos consecutivos y, por tanto, entre dos emociones elicítadas. Este hecho puede afectar en gran medida al estado emocional de un voluntario, y por tanto a la recuperación fisiológica, antes del siguiente videoclip. Además, se realizó un descanso obligatorio a la mitad del experimento (estímulo número 20), durante el cual se ofrecieron al voluntario galletas y bebidas sin cafeína ni alcohol. Esta metodología experimental puede introducir un sesgo muy perjudicial en función del orden de los estímulos y de la emoción a la que van dirigidos. En los siguientes subapartados se analizan, desde el punto de vista fisiológico, los posibles efectos detectados en los datos y etiquetas recogidas. Independientemente de estos hechos, hasta donde yo sé, el DEAP fue la primera base de datos que propuso estímulos de selección y metodología experimental de laboratorio bien documentados, junto con un número relativamente alto de voluntarios, y lo hizo todo de acceso totalmente abierto.

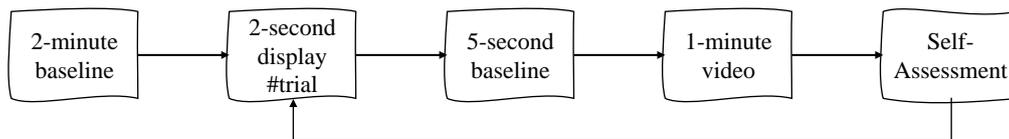


Figure 4-2: Diagrama simplificado de la experimentación aplicada para cada voluntario y cada estímulo para la base de datos DEAP.

Antes de entrar en detalles sobre el análisis realizado y los sistemas diseñados utilizando los datos del DEAP en este trabajo de investigación, se podría hacer una revisión de las técnicas de procesamiento de datos y aprendizaje automático aplicadas por el trabajo original de la base de datos, y por las investigaciones posteriores que la utilizan. Los autores del trabajo original de la base de datos DEAP aplicaron procedimientos básicos de preprocesamiento para eliminar las derivas temporales de baja frecuencia de algunas señales y suavizarlas utilizando filtros de media móvil. Extrajeron 106 características fisiológicas y emplearon un método de selección de características por filtro para utilizar sólo las de mayor rango. En concreto, aplicaron la puntuación discriminante lineal de Fisher dada por la ecuación 4.1,

$$J_f = \frac{|\mu_1 - \mu_2|}{\sigma_1^2 + \sigma_2^2}, \quad (4.1)$$

donde μ_k y σ_k representan la media y la varianza de la clase k para cada caracterís-

tica f . Obsérvese que esta ecuación es válida para $k = 2$, es decir, un problema de clasificación binaria. Cuanto mayor sea esta puntuación, más importante será esa característica específica. Así, el objetivo es maximizar la puntuación para obtener una gran varianza entre clases (numerador) y una pequeña varianza dentro de la clase (denominador). Sin embargo, esta metodología no tiene en cuenta la combinación de rasgos y tampoco maneja los redundantes, lo que conduce a un espacio de rasgos seleccionados subóptimo con un umbral de discriminación empírico. Para la clasificación, utilizaron un clasificador gaussiano de Bayes ingenuo para un problema de dos clases y tres casos de uso diferentes, niveles bajos y altos de excitación, valencia y agrado. Este clasificador específico se caracteriza por ser un modelo generativo, es decir, tiene un sesgo alto y una varianza baja derivada de las distribuciones gaussianas asumidas aprendidas de las características, lo que puede producir problemas de infraajuste. El resultado de este clasificador considerando N clases viene dado por la ecuación 4.2,

$$y = \operatorname{argmax}_y P(y) \prod_{i=1}^N P(x_i|y), \quad (4.2)$$

donde podemos obtener la clase inferida y para un conjunto dado de características o vector de características x_i . Nótese que este clasificador hace dos supuestos clave al considerar que las características son independientes y se distribuyen normalmente. Siendo esta última dada por

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right), \quad (4.3)$$

donde μ_y y σ_y^2 son la media y la varianza de los valores en x asociados a la clase i . Por último, el CV aplicado fue LOTO considerando los 40 estímulos audiovisuales utilizados durante los experimentos. Al emplear todas las señales periféricas, proporcionaron métricas promedio de Accuracy (ACC) y F1-score y obtuvieron un 57,00% (ACC) y un 53,30% (F1-score) para el arousal, un 62,70% (ACC) y un 60,80% (F1-score) para la valencia, y un 59,10% (ACC) y un 53,80% (F1-score) para el gusto. Nótese que no proporcionaron las desviaciones estándar asociadas a dichos valores medios.

Desde la publicación de la base de datos DEAP, se han propuesto en la literatura diferentes sistemas de aprendizaje automático utilizando sus datos. Algunas

publicaciones estudiaron la importancia de las características aplicando diferentes métodos y mejorando dicho proceso. Por ejemplo, los autores de [182] consideraron el mismo problema de clasificación que los autores de DEAP, pero utilizaron la eliminación recursiva de características para abordar la información mutua y redundante. Aplicaron un clasificador SVM y obtuvieron un 66,36% (ACC) y un 63,99% (F1-score) para la excitación, y un 68,71% (ACC) y un 63,25% (F1-score) para la valencia, lo que superó el trabajo original de DEAP. Utilizaron todas las características de todas las modalidades, incluidas las extraídas de los sensores que no se pueden llevar puestos. De hecho, concluyeron que las señales del electroencefalograma desempeñaban un papel clave en la separación de clases (distinción entre clases). Aunque su caso de uso no puede extrapolarse directamente al nuestro, ya que sólo nos basamos en tres señales periféricas y hacia un sistema de detección binaria del miedo. La mejora de los resultados en comparación con el trabajo original debido a la aplicación de técnicas de selección de características menos restrictivas y de un clasificador discriminativo es valiosa y puede ayudar en nuestra investigación.

También hay otras publicaciones que no consideraron todo el conjunto de señales y, en cambio, redujeron su número hacia un concepto más apto para el uso. Por ejemplo, [153] es uno de los últimos sistemas de reconocimiento de emociones basado en DEAP. Los autores diseñaron un sistema de reconocimiento de emociones de cinco clases (Feliz, Relajado, Asco, Triste, Neutral) utilizando el modelo PA. Aplicaron una técnica de nivel de fusión de características aprovechando una arquitectura de red de creencia profunda junto con la extracción de características estadísticas convencionales sobre sólo tres señales fisiológicas (PPG, EDA y EMG). Por último, entrenaron un clasificador SVM y obtuvieron hasta un 89,53% de precisión media para un modelo independiente del sujeto siguiendo una configuración LOSO, que superó el estado de la técnica. En su trabajo, no consideraron ninguna restricción de implementación real relacionada con la segmentación de los datos, la resolución de la frecuencia, el almacenamiento y la complejidad aplicada o necesaria. Además, sólo tomaron los datos fisiológicos registrados durante los últimos 20 segundos de cada estímulo basándose en su hipótesis de que la inmersión emocional era mayor al final del videoclip. Esta hipótesis no se ha demostrado con un método estadístico, objetivo y/o cuantificable, sino que sólo se ha evaluado mediante la exploración

visual fisiológica.

Entre el resto de investigaciones realizadas sobre la base de datos DEAP y en lo que respecta específicamente al caso de uso de reconocimiento de miedo, se encuentran cuatro sistemas en la literatura. Por un lado, los dos primeros [181, 183] son nuestras publicaciones y se detallan en los siguientes subapartados. En adelante se denominarán DEAP-b1 y DEAP-b2 para este capítulo y los siguientes. Por otro lado, los autores en [184] y en [185] emplearon el mismo paradigma de miedo binario que se describe en la sección 2.3.4. En [184], utilizaron todos los voluntarios DEAP y todas las señales fisiológicas disponibles, incluyendo las que proporcionaban datos de electrooculograma y electroencefalograma. Realizando un Design Space Exploration (DSE) para diferentes técnicas de selección de características, así como para nueve máquinas de clasificación diferentes, incluyendo las redes neuronales profundas, lograron hasta un 90,07% de precisión media para un modelo independiente del sujeto sin selección de características, sólo utilizando los datos filtrados, y siguiendo una estrategia de Hold-Out con una proporción de división de entrenamiento-prueba de 70/30. Nótese que [184] se publicó después de nuestro trabajo de reconocimiento binario de emociones de miedo, DEAP-b1 [183], que, hasta donde yo sé, fue la primera investigación que aplicó dicho paradigma de etiquetado de miedo al reconocimiento de emociones a través de señales fisiológicas. En [185], tomaron nuestra investigación de [181, 186] como referencia principal y elaboraron un análisis exhaustivo que comprendía un estudio detallado de los efectos para el reconocimiento de la emoción binaria del miedo cuando se utilizan diferentes métodos y técnicas de elementos de aprendizaje automático. A diferencia de la investigación en [184], en ésta aplicaron etapas de filtrado, segmentación de datos con y sin solapamiento, selección de características, reducción de la dimensionalidad y ajuste de desequilibrios con Synthetic Minority Over-sampling TEchnique (SMOTE). También se basaron en los datos de los voluntarios del DEAP, pero descartaron la mayoría de las señales y sólo emplearon GSR y PPG. Por último, utilizaron ventanas de procesamiento de datos de 20 segundos y lograron una tasa máxima de precisión en el reconocimiento del miedo de hasta el 93,50% para un clasificador SVM junto con PCA considerando una estrategia de no solapamiento y 5 k -fold CV. Aunque esta última investigación es un valioso trabajo hacia la exploración de la amplia

DSE respecto al reconocimiento del miedo, su principal limitación es la técnica CV aplicada ya que podrían existir pliegues, es decir, ventanas de procesamiento, que contengan parte de la información del pliegue anterior, ver Sección 3.1.7.3. Por lo tanto, este hecho puede conducir a resultados demasiado optimistas.

4.1.1 Consideraciones sobre el equilibrio de los estímulos y las etiquetas

Como se indica en la sección 3.1.1, uno de los enfoques comunes que se siguen durante la generación de una base de datos está relacionado con la evaluación del equilibrio de los estímulos. Esto se refiere a la representación estadística de las diferentes clases. Por ejemplo, en cualquier problema de clasificación es deseable tener la misma cantidad de instancias para todas las clases. De lo contrario, el algoritmo de clasificación podría derivar en favorecer el aprendizaje de la clase con mayor representación frente al resto de las clases. Por ello, el análisis de las etiquetas durante toda la generación y el procesamiento de la base de datos es fundamental para contextualizar y comprender tanto la elicitación de emociones como los resultados obtenidos.

En DEAP, los 40 videoclips utilizados durante el experimento se eligieron de un conjunto de estímulos más amplio. Se inició una etapa de preetiquetado con 120 videoclips y se reunieron unas 14 valoraciones por vídeo. Tras este proceso y mediante la ecuación 3.1, los autores seleccionaron los vídeos que se encontraban en las esquinas extremas de los cuadrantes normalizados dentro del espacio PA, lo que dio como resultado un conjunto de 40 videoclips con etiquetado extremo y utilizados para provocar emociones en los voluntarios mientras se medía su señal fisiológica. Tras visualizar estos videoclips, los voluntarios etiquetaron la emoción sentida por ellos. Así, esta metodología condujo a la generación de dos conjuntos diferentes de etiquetas, las de la etapa de preetiquetado y las autoinformadas y recordadas durante el experimento. Por lo general, estas últimas son las preferidas para entrenar modelos de aprendizaje automático basados en los datos fisiológicos y/o físicos recopilados. Sin embargo, la distribución de las etiquetas autodeclaradas puede ser muy diferente con respecto a las etiquetas de la fase de preetiquetado. Por ejemplo, en la figura 4-3 se muestran las diferencias entre las etiquetas previas al

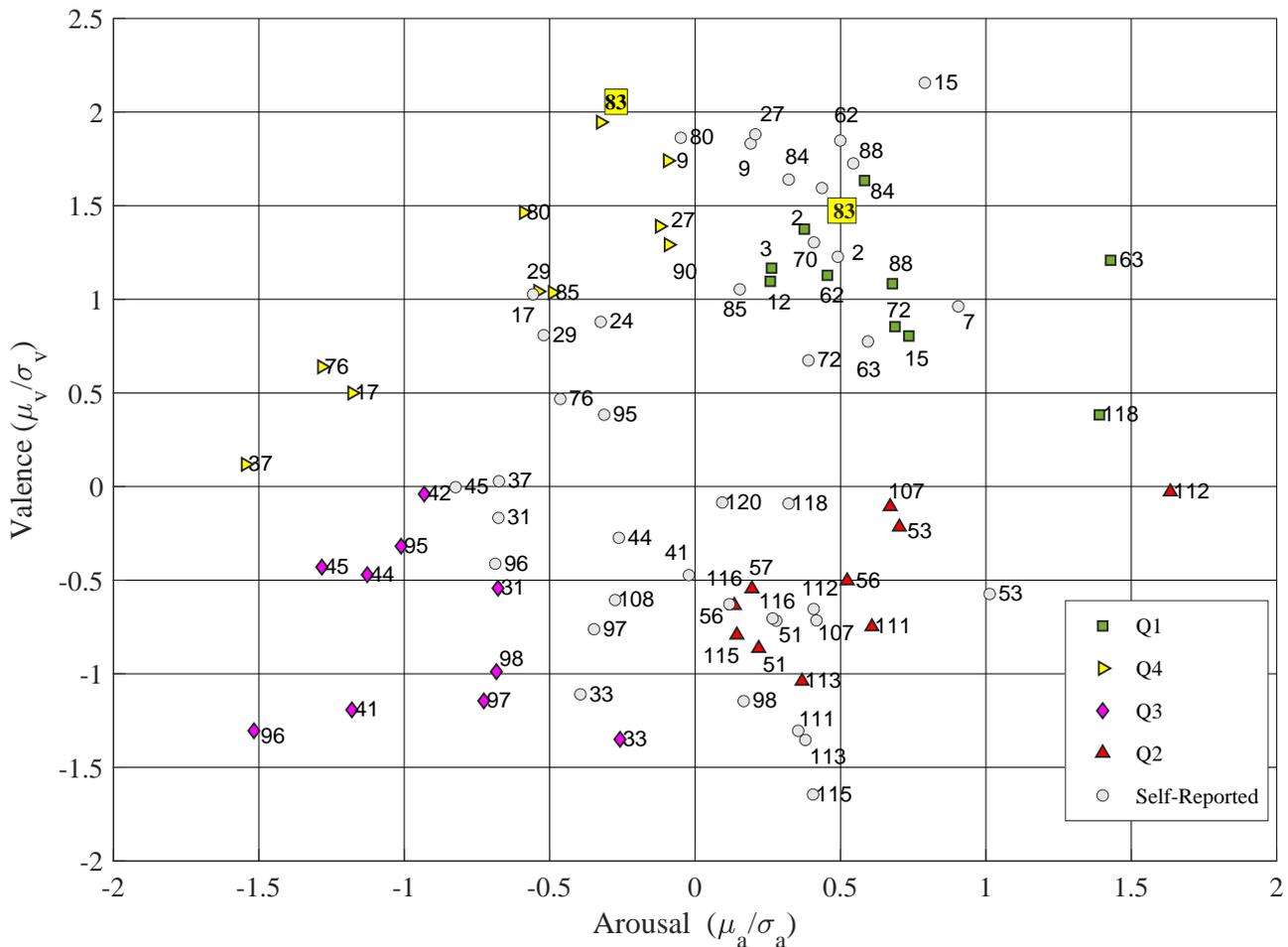


Figure 4-3: Diferencias de etiquetado para la base de datos DEAP y la numeración original de los videoclips seleccionados.

etiquetado y las autodeclaradas para la base de datos DEAP y los videoclips seleccionados. En el caso de los estímulos preetiquetados, las etiquetas obtenidas se clasifican con un símbolo diferente en función de la ubicación del cuadrante normalizado (Q1 -excitación positiva, valencia positiva-, Q2 -excitación positiva, valencia negativa-, Q3 -excitación negativa, valencia negativa-, Q4 -excitación negativa, valencia positiva-). Por el contrario, las etiquetas autoinformadas se representan con el mismo símbolo y color. Obsérvese que los autores del DEAP seleccionaron los estímulos que alcanzaron las medias más altas y las variaciones más pequeñas entre las diferentes valoraciones. Como se puede observar, las valoraciones previas al etiquetado no siguen la misma distribución que las valoraciones autodeclaradas, lo que hace que incluso los mismos estímulos se sitúen en cuadrantes emocionales diferentes, por ejemplo el estímulo 83.

La tabla 4.1 presenta los vídeos que se encuentran en un cuadrante diferente. Sin tener en cuenta los que se encuentran en distintas ubicaciones dentro del mismo

Table 4.1: vídeos que se encuentran en un cuadrante diferente con respecto al pre-etiquetado frente a las etiquetas autodeclaradas.

Estímulo ID	Cuadrante Pre-etiquetado	Cuadrante Reportado
9	4	1
27	4	1
45	3	4
83	4	1
85	4	1
95	3	4
98	3	2
118	1	2

cuadrante y los que están justo en las líneas fronterizas, el 20% de los estímulos no evocan la emoción objetivo o preetiquetada durante la realización de los experimentos. Esto se traduce en una distribución de etiquetas distinta y, por tanto, puede dar lugar a diferencias de rendimiento del sistema cuando se entrena con etiquetas preetiquetadas o autoetiquetadas. Sin embargo, los estímulos relacionados con el segundo cuadrante (Q2), basados en calificaciones de preetiquetado, son los únicos que presentan una concordancia completa en comparación con las etiquetas autoinformadas. Así, se puede concluir que los estímulos que pretenden evocar emociones negativas (alta excitación, valencia negativa), como las emociones relacionadas con el miedo, son menos propensos a ser confundidos con otros cuadrantes. Aunque esta conclusión apoya firmemente el enfoque de una estrategia de etiquetado basada en el 2D a la hora de enfrentarse al desarrollo de un sistema de reconocimiento binario del miedo, también podría considerarse y explorarse la consideración de más dimensiones.

Desde una perspectiva 3D y teniendo en cuenta el modelo PAD, las valoraciones autodeclaradas y el mapeo de Miedo propuesto en la sección 2.3.4, sólo cinco estímulos se sitúan dentro del cubo emocional relacionado con el miedo (51, 98, 111, 115), Figura 4-4. Este hecho es especialmente relevante debido al desentrañamiento de los estímulos contenidos dentro de Q2 en la Figura 4-3. Así, desde el punto de vista de la valoración subjetiva, la ampliación a una dimensión más (dominancia) puede beneficiar la localización emocional de los estímulos y, por tanto, su posterior asignación de etiquetas. Sin embargo, dicho beneficio se ve ensombrecido en este caso concreto al saber que esos cinco estímulos representan sólo el 12,5% de la cantidad total de videoclips. Esta última consideración representa una situación

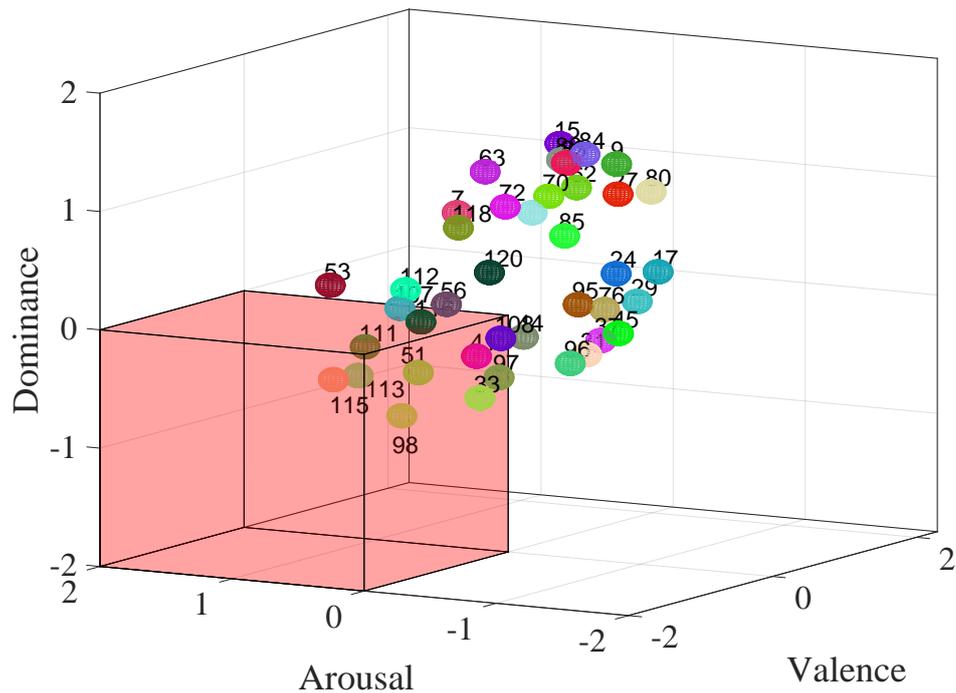


Figure 4-4: PAD para las etiquetas autodeclaradas por los voluntarios. El mapeo del miedo propuesto en la sección 2.3.4 está marcado con un cubo de color.

muy desequilibrada en el caso de un sistema de reconocimiento binario de emociones de miedo, pero puede mitigarse utilizando técnicas de sobremuestreo sobre la clase minoritaria, como se explica en las siguientes subsecciones.

Para proporcionar una cuantificación específica de la distribución desequilibrada autodeclarada, la tabla 4.2 muestra los ratios de desequilibrio autodeclarados para cada tipo de etiqueta recogida. Por un lado, estos ratios de desequilibrio se calculan sobre la base de un problema de dos clases, dividiendo cada dimensión en dos niveles (alto y bajo), lo que equivale a lo que se hizo en la publicación original de DEAP. Por otro lado, como la selección de estímulos en el DEAP se hizo exclusivamente en base a la excitación y la valencia, el desequilibrio resultante para las otras dos valoraciones reunidas (dominancia y agrado) es mayor. Por ejemplo, como esta base de datos no se centró en la elicitación de emociones negativas, la proporción de desequilibrio observada en la dominancia indica la presencia de más estímulos positivos en los que los voluntarios valoraron un alto grado de control sobre la emoción evocada.

No obstante, en el caso de abordar el diseño de un sistema de reconocimiento binario de emociones de miedo, en el que se realiza una transformación binaria a partir

Table 4.2: Ratios de desequilibrio reportados por las voluntarias y los voluntarios para la base de datos DEAP.

Clase Baja:Alta	Arousal	Valencia	Dominancia	Liking
Ratio de Balanceo	1.4:1	1.2:1	1:1.6	2:1

de los espacios PA o PAD en lugar de dirigirse a la clasificación de múltiples niveles mediante modelos unidimensionales, los ratios de desequilibrio se detallan en la Tabla 4.3. Estos ratios de desequilibrio deben contextualizarse en función del equilibrio individual obtenido a través de las clasificaciones específicas autodeclaradas. Por ejemplo, la Figura 4-5 muestra el balance de clases para cada voluntario después de haber aplicado el mapeo binario de miedo de un espacio PA, que dio como resultado "1" o la clase positiva para el Q2 (alta excitación y baja valencia) y "0" o la clase negativa para el resto de los cuadrantes. Obsérvese que el umbral del 25% está resaltado como marca de referencia que indica que, basándose en la verdad básica original esperada por los investigadores, esta cantidad de estímulos debería evocar una emoción situada en el Q2. Como puede observarse, 17 del total de voluntarios (32) presentan menos del umbral esperado dentro de la clase positiva, lo que afecta fuertemente al ratio de desequilibrio de este mapeo binario. Además, hay incluso un voluntario (23) que no ha informado de ninguna valoración dentro de la clase positiva. Estos hechos, además de explicar la elevada tasa de desequilibrio obtenida, permiten comprender la complejidad de las diferencias interindividuales. No obstante, los porcentajes medios de las clases considerando los 32 voluntarios son de 76,50% y 23,50% para las clases negativas y positivas respectivamente. Obsérvese que el porcentaje medio de la clase positiva se acerca al 25% esperado. Este último hecho apoya las conclusiones obtenidas con la Figura 4-3, por la que afirmábamos que se puede abordar una estrategia de etiquetado basada en 2D para diseñar un sistema de reconocimiento binario de emociones de miedo utilizando DEAP. Para el mapeo binario aplicado al considerar un espacio PAD, la clase positiva está determinada por la baja dominancia, el alto arousal y la baja valencia, mientras que la negativa viene dada por las otras combinaciones posibles. La figura 4-6 muestra el balance de clases por sujeto en tal caso. Nótese que en este gráfico no se espera que se alcance un umbral, ya que la etapa de selección de estímulos previa al etiquetado se basó únicamente en el arousal y la valencia. En este caso, hay tres voluntarios

Table 4.3: Ratios de balanceo PA y PAD para la base de datos DEAP.

Clase NoMiedo:Miedo	PA	PAD
Ratio de balanceo	3.2:1	6.3:1

(23, 27, 28) que no mostraron ninguna valoración de clase positiva y los porcentajes medios de clase son del 86,33% y del 13,67% para las clases negativas y positivas respectivamente. Estos hechos explican claramente el mayor ratio de desequilibrio con respecto a la binarización PA e indican que para esta base de datos el equilibrio de la dimensión de dominancia no era crucial. Esto último es esencial para nuestro caso de uso debido a la necesidad de distinguir entre emociones específicas que sólo difieren en la dimensión de dominancia, como el miedo y la ira, tal y como se comentó en el capítulo 2. No obstante, a pesar de este problema, se podría explorar un sistema de reconocimiento de emociones binarias de miedo utilizando esta base de datos como prueba de concepto. Además, como ya se ha señalado anteriormente, se pueden aplicar diferentes técnicas de sobremuestreo para tratar estas condiciones de desequilibrio extremo.

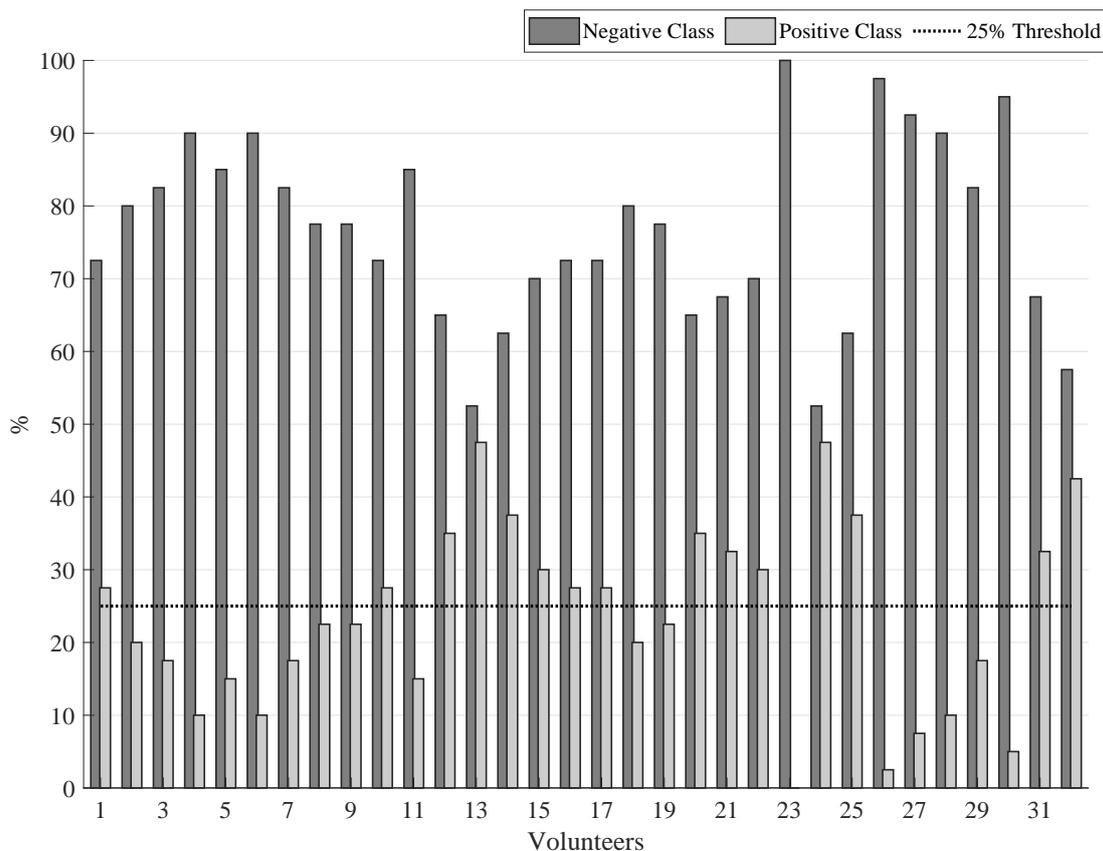


Figure 4-5: Balance de clases por voluntario después de haber aplicado el mapeo binario de miedo de un espacio PA.

Otro proceso esencial a la hora de evaluar la consistencia del etiquetado en los

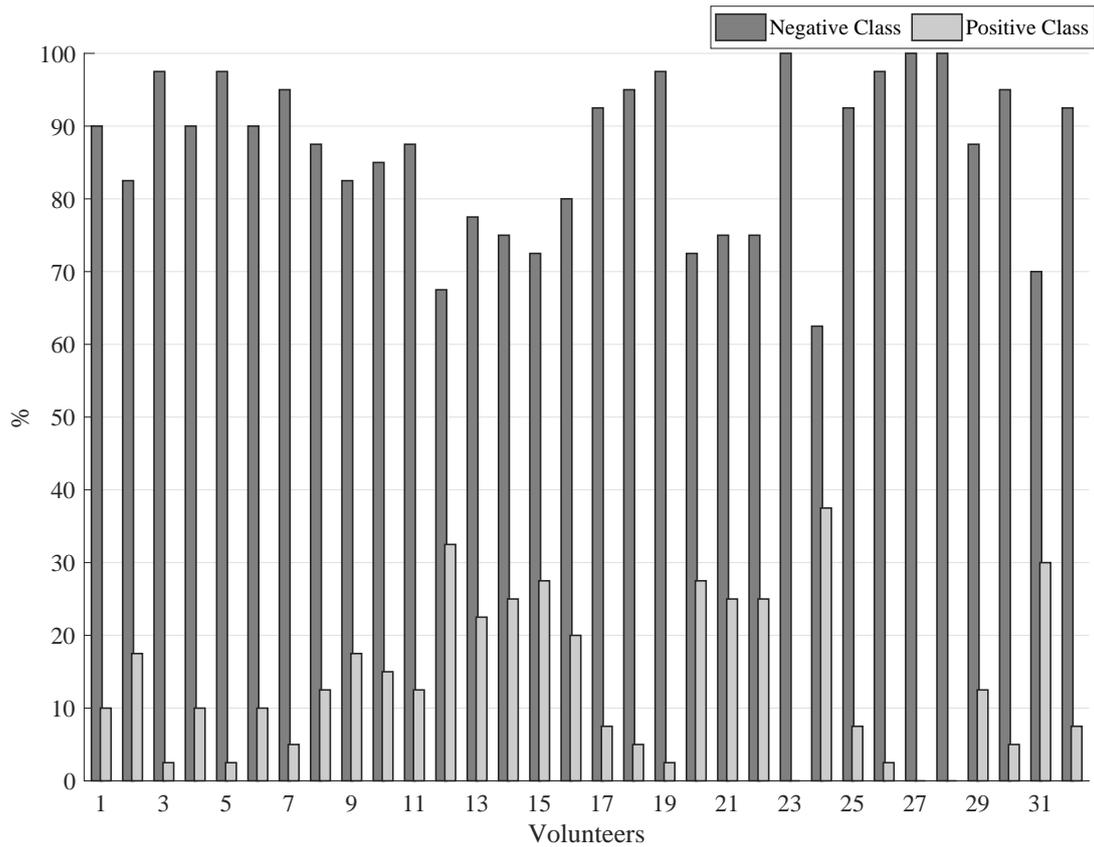


Figure 4-6: Balance de clases por voluntario después de haber aplicado el mapeo binario de miedo de un espacio PAD.

diferentes voluntarios es observar las correlaciones interindividuales de las etiquetas. Esta tarea proporciona información que puede relacionarse directamente con los resultados obtenidos de los diferentes modelos de aprendizaje automático. Por ejemplo, considerando los dos mapeos binarios de miedo realizados, es decir, de PA y de PAD, los resultados obtenidos tras una prueba de Levene y una prueba de Kruskal-Wallis rechazaron la hipótesis nula de que las varianzas son iguales en todos los voluntarios ($p < 0,001$). Obsérvese que ambos conjuntos de etiquetas binarizadas presentan una distribución no normal y que el nivel de significación se fijó en $p < 0,05$. Estos hechos conducen a la evaluación y aplicación de pruebas de correlación e independencia para estudiar el comportamiento de etiquetado de los distintos voluntarios. Así, la Figura 4-7a y la Figura 4-7b muestran los valores p promediados para la correlación de Spearman y la prueba de Chi-cuadrado de independencia para el mapeo basado en el miedo PA, respectivamente. Nótese que ambos son métodos no paramétricos para evaluar las diferentes asociaciones entre variables. Sin embargo, el primero responde a asociaciones monótonas, mientras que

el segundo proporciona información relacionada con la independencia de las variables considerando cualquier tipo de asociación. Los resultados arrojados por ambos procesos son cercanos, de hecho, ambos no rechazan la hipótesis nula. Esto indica que no existe ninguna diferencia estadística entre los distintos grupos, es decir, la correlación se considera no significativa y las distintas variables son independientes. Por lo tanto, podemos concluir que no hay pruebas suficientes para sugerir que exista una asociación entre la etiqueta binaria de los voluntarios. Además, algunos voluntarios (4, 8, 16, 21, 26) muestran valores p elevados en comparación con los demás, lo que puede interpretarse como una mayor decorrelación e independencia de sus etiquetas. Las figuras 4-8a y 4-8b presentan los valores p promediados para el mapeo basado en el miedo binario PAD y las mismas pruebas de correlación e independencia. En este caso, podemos observar una mayor decorrelación e independencia para los distintos voluntarios en comparación con las pruebas anteriores. Esto permite conocer la distribución del conjunto de datos de las etiquetas e incluso puede orientar el proceso de diseño. Así, de la comparación de estas cifras y del estudio de consistencia de las etiquetas se pueden extraer dos conclusiones principales (1) estos procesos permiten identificar a los voluntarios que proporcionaron un etiquetado muy distinto durante los experimentos, y (2) aunque el espacio PAD proporciona más información en términos de modelado emocional, cuantas más dimensiones se añaden, menor es la concordancia que se puede obtener de las calificaciones autodeclaradas por los voluntarios.

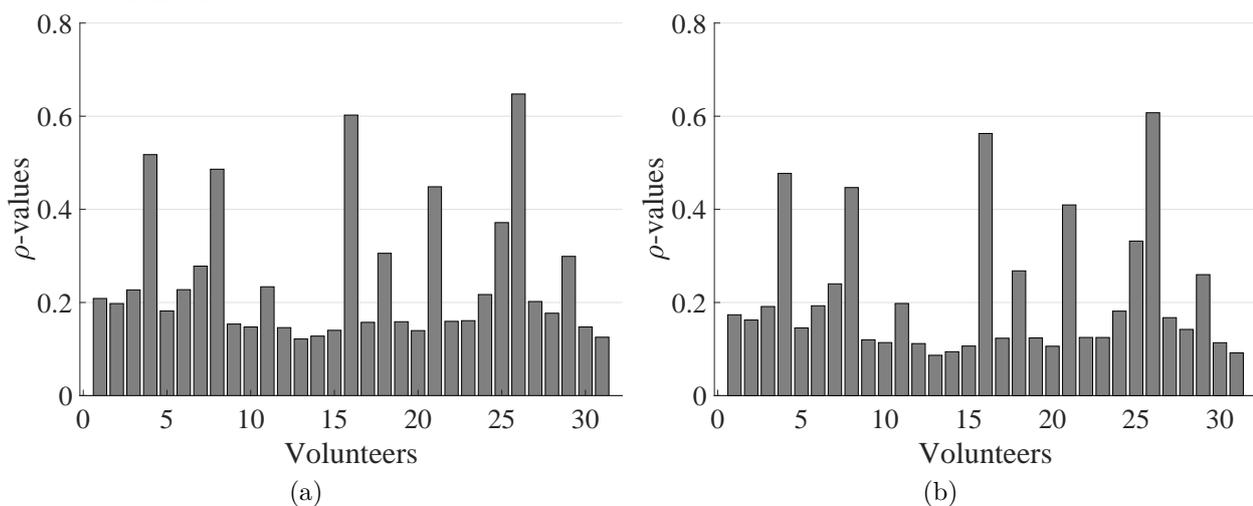


Figure 4-7: Valores p promediados para todos los voluntarios considerados y sus etiquetas aplicando: a) la correlación de Spearman, y b) para la prueba Chi-cuadrado de independencia. En este caso, las etiquetas se binarizan utilizando el mapeo basado en el miedo binario.

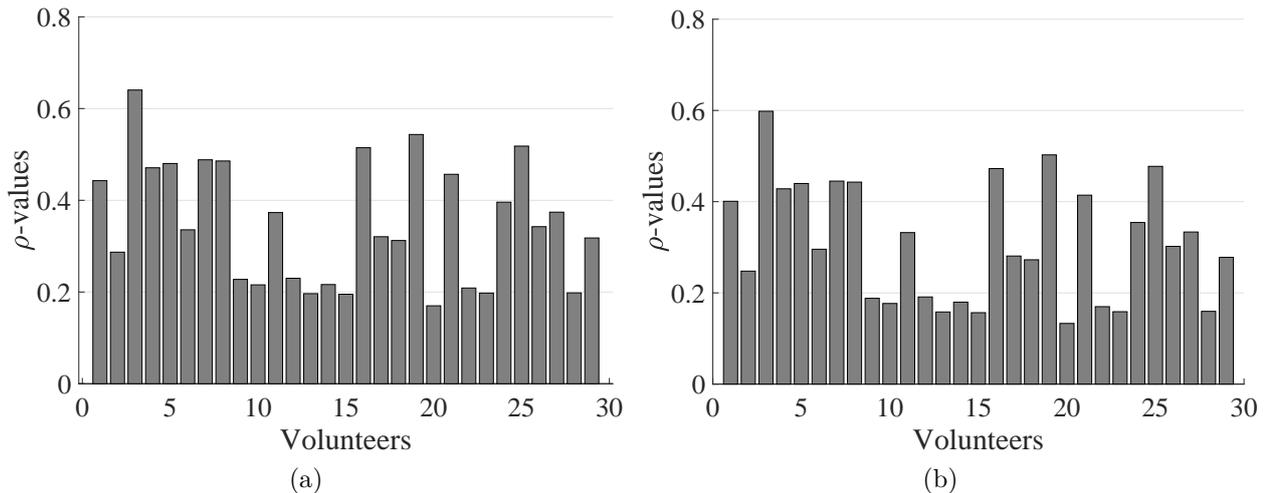


Figure 4-8: Valores p promediados para todos los voluntarios considerados y sus etiquetas aplicando: a) la correlación de Spearman, y b) para la prueba Chi-cuadrado de independencia. En este caso, las etiquetas se binarizan utilizando el mapeo basado en el miedo binario.

A pesar de las diferencias de equilibrio y concordancia observadas al aplicar la transformación binaria de miedo de ambos modelos emocionales, el uso de la dimensión de dominancia para distinguir adecuadamente la emoción de miedo, nos llevó a diseñar el DEAP-b1 utilizando la transformación binaria de miedo del espacio PAD. Los resultados obtenidos con ese sistema, al considerar una perspectiva independiente del sujeto, no fueron prometedores [183]. Por ello, decidimos simplificar el problema diseñando un sistema que utilizara un mapeo binario de miedo del espacio PA, DEAP-b2. Este último mejoró los primeros resultados y demostró que un sistema de reconocimiento binario de emociones relacionadas con el miedo era factible utilizando únicamente información fisiológica. Los detalles específicos de cada uno de estos sistemas se presentan en la sección 4.1.4. Además, hay que tener en cuenta que los diferentes resultados recogidos en el estudio de equilibrio de estímulos y consideración de etiquetas que se ofrece en esta sección estuvieron siempre presentes durante el diseño de dichos sistemas.

4.1.2 Análisis exploratorio de datos y procesamiento de filtros

Durante la generación de la base de datos del DEAP, los investigadores tuvieron algunos problemas que afectaron a la adquisición de los sensores de algunos voluntarios y a la recuperación fisiológica diseñada. En esta sección, generamos diferentes

gráficos sincronizados con la metodología experimental para realizar un análisis exploratorio de los datos y evaluar el comportamiento tanto desde el punto de vista fisiológico como del funcionamiento del sensor. Por ejemplo, la Figura 4-9 muestra un ejemplo de una de las representaciones gráficas para la evaluación visual fisiológica realizada durante este paso. En concreto, los gráficos corresponden al experimento completo del voluntario número 22. Las señales representadas son GSR, BVP, y SKT, de alto a bajo orden respectivamente. La señal de sincronización indica los diferentes estados del experimento: 20 estímulos representados por cada diente de sierra y 20 tareas de asignación de etiquetado en cada decadencia de los mismos, a los que sigue una pausa y los 20 estímulos finales con su respectivo etiquetado. Cabe destacar que los datos mostrados se obtuvieron directamente de los archivos ".bdf" (formato de datos de BioSemi generado por el software de grabación Actiview) proporcionados por la base de datos. Nótese que también subieron una versión preprocesada de los datos, sin embargo, sólo aplicando un down-sampling sin ninguna otra etapa de filtrado adicional. Así, del análisis visual exploratorio de los datos brutos de todos los participantes, obtuvimos tres conclusiones principales:

- Mientras que GSR y SKT mostraron una calidad aceptable, BVP necesitó ser filtrado para eliminar no sólo los ruidos de alta frecuencia, sino también el wander de la línea de base, es decir, la tendencia de muy baja frecuencia que se produce por el efecto de la respiración en la adquisición de PPG.
- La expectativa de una pausa fisiológica no controlada (tras el estímulo 20), expuesta durante el análisis de la metodología de la base de datos en la sección 4.1, se confirma a primera vista al menos por el GSR. Obsérvese el incremento del nivel tónico durante la pausa y que este comportamiento se repite en todos los voluntarios. Desde el punto de vista de una base de datos de laboratorio en la que las condiciones deben estar debidamente controladas, este tipo de recuperación puede dar lugar a efectos desconocidos para los sistemas de reconocimiento de emociones que se van a entrenar. Por lo tanto, en caso de realizar una recuperación o pausa fisiológica, podrían aplicarse otras estrategias que dieran cuenta de una estabilización o desviación real de las señales fisiológicas.
- Se observaron incoherencias en la temperatura fisiológica de la piel en difer-

entes voluntarios. Éstas se referían a valores de temperatura cutánea muy bajos. Por ejemplo, la señal SKT de la Figura 4-9 presenta una variación de 29 °C a 25 °C, que no está dentro de los rangos SKT normales y/o válidos en condiciones controladas de laboratorio. Este problema puede deberse a diferentes factores, como un mal funcionamiento de la adquisición del sensor o una fijación incorrecta del mismo al cuerpo.

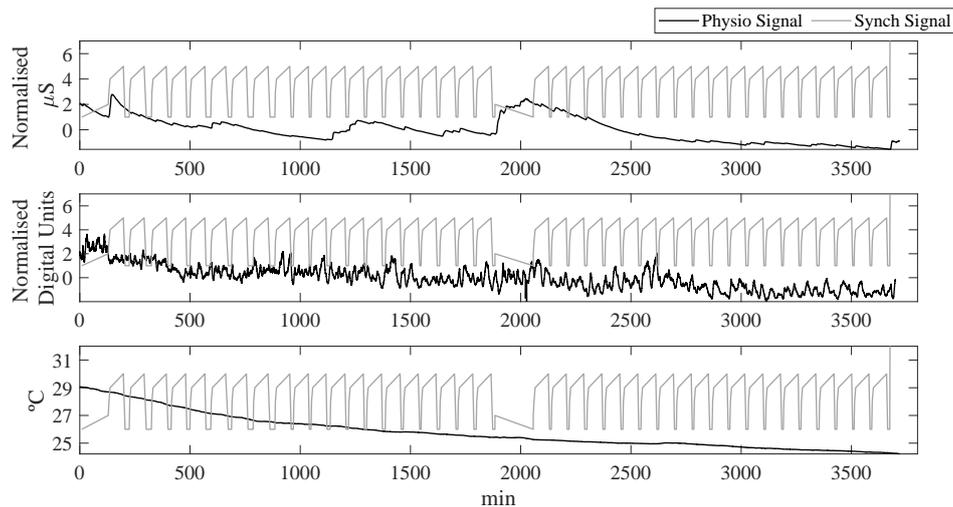


Figure 4-9: Ejemplo de una de las representaciones gráficas de la evaluación visual fisiológica realizada.

Para hacer frente a los diferentes problemas de ruido observados en el BVP, se pueden diseñar diferentes filtros. Por un lado, el ruido de alta frecuencia puede filtrarse mediante un filtro Finite Impulse Response (FIR) de paso bajo de forma directa. Por otro lado, el efecto residual de la línea de base o de la deriva de baja frecuencia que se presenta en la señal puede eliminarse mediante una etapa de filtrado Butterworth Infinite Impulse Response (IIR) de paso bajo hacia delante y hacia atrás. En concreto, la técnica forward-backwards maneja la fase no lineal de dichos filtros. Por ejemplo, un ejemplo de la aplicación de estos diferentes procesos de filtrado para un fragmento específico de la señal anterior se muestra en la Figura 4-10. Cabe señalar que estas dos etapas de filtrado detallado pueden ser independientes, es decir, no tienen que aplicarse estrictamente una tras otra. De hecho, sólo hemos empleado el filtro FIR para DEAP-b1 y DEAP-b2. Esta consideración se basó en el hecho de observar derivas de baja frecuencia dentro de la señal durante la extracción de características. El filtro FIR diseñado dio como resultado una frecuencia de corte de 3,5 Hz con una atenuación de -6 dB. Obsérvese que durante el proceso de diseño se utilizó una ventana Hamming para minimizar adecuadamente el primer lóbulo

lateral.

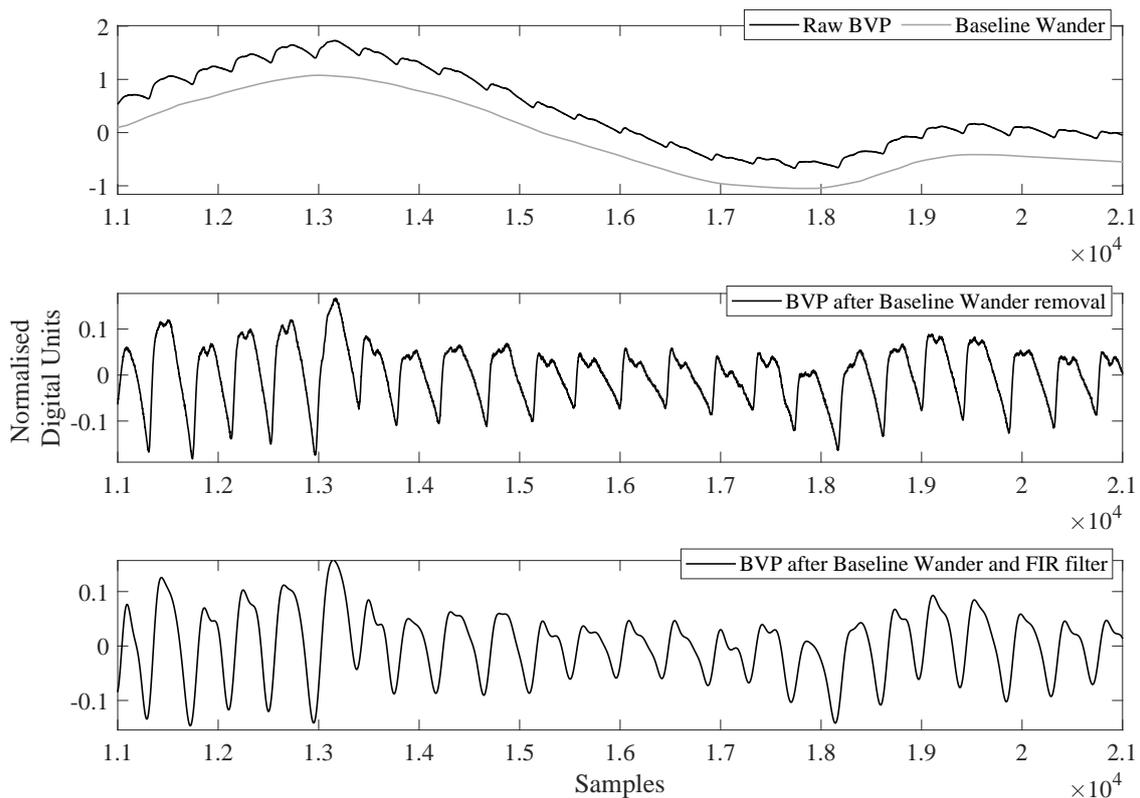


Figure 4-10: Ejemplo de filtrado para la extracción y eliminación de la línea de base mediante el filtrado IIR, y la eliminación del ruido alto.

En cuanto al problema de SKT con algunos de los voluntarios, 11 de 32 voluntarios estaban afectados. Por lo tanto, sólo se consideraron 21 voluntarios válidos para el DEAP-b1. Sin embargo, después, la necesidad de aumentar el conjunto de datos llevó a considerar el conjunto completo de voluntarios para el DEAP-b2 a expensas de omitir SKT y utilizar sólo GSR y BVP.

Como ya se destacó en el capítulo 3, la mayoría de las bases de datos de reconocimiento de emociones disponibles al público no realizan un análisis exploratorio exhaustivo de los datos durante o después de la generación del conjunto de datos. Este hecho puede dar lugar a comportamientos inesperados a la hora de diseñar sistemas de reconocimiento de emociones. Por lo tanto, este paso es necesario para garantizar la calidad de los datos proporcionados.

4.1.3 Extracción de características

El diseño de DEAP-b1 fue nuestro primer sistema de reconocimiento binario de emociones de miedo [183], que siguió un enfoque de cálculo aproximado al no aplicar ningún procedimiento de extracción de características, es decir, reducir la comple-

alidad del sistema a cambio de disminuir la precisión [187]. En su lugar, solo se consideró el valor filtrado de cada variable fisiológica. Este enfoque de fuerza bruta es opuesto a otros enfoques, que emplearon más de cien características extraídas de las señales fisiológicas, como el trabajo original de DEAP [138]. La decisión de un sistema de extracción de características bare-metal de este tipo fue motivada por una exploración inicial de las posibilidades de diseño al tratar directamente con datos filtrados y/o crudos, y por los limitados recursos integrados que presentaba la primera versión de Bindi (64KB de RAM). Nótese que otras publicaciones recientes, como [184], también utilizaron directamente los datos en bruto para generar motores de aprendizaje automático de miedo.

Al contrario que el DEAP-b1, nuestro segundo sistema de reconocimiento binario de miedo propuesto, el DEAP-b2 [181], implementó técnicas convencionales de extracción de características usualmente empleadas en la literatura. Se extrajeron utilizando toda la duración del clip de vídeo, es decir, ventanas de procesamiento de 60 segundos. La tabla 4.4 presenta la lista completa de características para las dos señales fisiológicas consideradas en este sistema. En las siguientes subsecciones se ofrecen detalles específicos y la justificación de las características.

Es necesario aplicar diferentes procesos de delineación para obtener puntos fisiológicos específicos para cada señal antes de extraer las características de las señales, como se explica en la sección 2.5. Por ejemplo, la señal BVP requiere la identificación de picos y valles. Para el sistema DEAP-b2, esto se hace implementando el mismo algoritmo de delineación BVP que se propone en [148]. Para la señal GSR hay que extraer también los componentes tónico y fásico, SCL y SCR. En este caso, asumimos una combinación lineal de estos dos componentes representada en la ecuación 2.7. La tendencia de la señal GSR se obtiene mediante un filtro de mediana móvil con una ventana deslizante de \pm cuatro segundos, que se basa en sustituir cada entrada por la mediana de las entradas vecinas para dicha ventana. Después, la tendencia se resta directamente a la señal GSR, lo que da el componente SCR.

4.1.3.1 Dominio Temporal

Las características del dominio del tiempo pueden dividirse en dos grupos principales: estadísticas de orden superior y características morfológicas.

Dentro del primer grupo, el bloque principal es el cálculo del promedio de la señal

Table 4.4: Características extraídas del sistema DEAP-b2.

Sensor	Domain	Features
PPG/BVP (13)	Time-domain: (3)	Average of filtered signal Mean of Inter-Beat-Interval Heart Rate Variability
	Frequency-domain: (5)	Power spectral density of four bands (0–0.1 Hz, 0.1–0.2 Hz, 0.2–0.3 Hz and 0.3–0.4 Hz) Inter-Beat-Interval spectral density ratio between 0–0.08 Hz and 0.15–0.5 Hz bands
	Non-linear domain: (5)	Inter-Beat-Interval Multi Scale Entropy (five levels)
GSR (7)	Time-domain: (7)	Average of filtered signal
		Number of ERSCR peaks per second
		Average relative amplitude of ERSCR peaks per second
		Average rise time of ERSCR peaks per second
		Standard deviation of filtered signal
		25th percentile value 75th percentile value

dentro de una ventana de procesamiento, en la que se adquieren un total de N muestras a una frecuencia de muestreo específica f_s . El promedio sigue

$$\mu_X = \frac{1}{N} \sum_{n=1}^N X_n, \quad (4.4)$$

donde X representa las señales BVP o GSR. Para el caso de la BVP, el valor medio está relacionado con la resistencia periférica, que es responsable del tono vascular, como se indica en el capítulo 2. Además, cuando no se aplica el método de eliminación de la línea de base, esta información se mezcla con los efectos de la amplitud de la respiración, que pueden afectar a las partes de CC y de muy baja frecuencia de la señal (ecuación 2.6). Para el caso GSR, el promedio de la señal contiene información de la parte estacionaria del nivel tónico de la señal o SCL. Por lo tanto, está fuertemente relacionado con la cuantificación del arousal. Para tener en cuenta la variabilidad de dicha información, también se considera la desviación estándar del GSR dada por la raíz cuadrada de la varianza, ecuación 4.5

$$s_X = \sqrt{\frac{1}{N-1} \sum_{n=1}^N |X_n - \mu_X|^2}. \quad (4.5)$$

Por último, los restantes procedimientos estadísticos de orden superior son los cuartiles 25 y 75 de los elementos de procesamiento de la ventana actual. Estos también

se aplican únicamente a la señal GSR y se obtienen mediante un algoritmo basado en la ordenación.

En cuanto a los rasgos morfológicos, se caracterizan por la identificación de los puntos fisiológicos de delineación dentro de la ventana de procesamiento actual. La señal BVP se somete a la extracción de dos rasgos morfológicos: la media y la variabilidad del IBI. Esta métrica es la diferencia temporal entre los diferentes picos sistólicos identificados. Su media y variabilidad están relacionadas con la respuesta del ANS. En concreto, indican los cambios de la variabilidad cardíaca en respuesta a factores de estrés agudos (vídeos). Obsérvese que esta información permite rastrear la respuesta del sistema cardiovascular, y esta variabilidad se espera de una persona sana. Por el contrario, podría ocurrir que para personas con una condición de estrés crónico, como el PTSD, presente una variabilidad mínima o nula. La serie temporal del IBI viene dada por la ecuación 4.6

$$IBI_n = t_{sys_{n+1}} - t_{sys_n}, \quad (4.6)$$

donde $t_{sys_{n+1}}$ y t_{sys_n} son las posiciones temporales de los picos sistólicos $n + 1$ y n , respectivamente. Suele expresarse en milisegundos y oscila entre 1000 ms y 600 ms (60-100 BPM) cuando está en condiciones de reposo. La variabilidad del IBI o Heart Rate Variability (HRV) se calcula como la desviación estándar de los IBI recogidos a lo largo de la ventana de procesamiento, que también se conoce como SDNN. Nótese que ésta es una de las diferentes opciones posibles para calcular el HRV. Cuando se trata de la señal GSR, el número, la amplitud y el tiempo de subida de los diferentes picos ERSCR se extraen utilizando el componente SCR obtenido tras la sustracción de la tendencia. Por lo tanto, para extraer estas tres características, se ejecuta un método de trough-to-peak sobre el SCR. Para esta implementación, asumimos que el criterio de amplitud normalmente aceptado para discernir ERSCRs sobre un estímulo externo es 0.01 microS [112]. Nótese que estas tres características se expresan en $\mu S/seg$, es decir, se calculan y normalizan por el tiempo de procesamiento de la ventana, en este caso la duración del vídeo.

Aunque las características basadas en el dominio del tiempo no pueden tratar la información fisiológica no estacionaria, proporcionan un punto de partida fuertemente respaldado y validado para cualquier sistema de reconocimiento de emociones.

4.1.3.2 Dominio Frecuencial

Antes de abordar cualquier extracción de información en frecuencia, la resolución en frecuencia debe ajustarse en consecuencia para poder obtener todas las bandas de densidad espectral de potencia (PSD) establecidas. De hecho, dicha resolución sólo depende de la longitud temporal de la ventana de procesamiento. Por ejemplo, en este caso, un tamaño de ventana de 60 segundos da como resultado una resolución de frecuencia de 0,016 Hz/bin dada por la ecuación 4.7

$$f_{res} = \frac{f_s}{f_s * T_{len}} = \frac{1}{T_{len}}, \quad (4.7)$$

donde T_{len} es el tamaño de la ventana en segundos, f_s es la frecuencia de muestreo de la señal discreta, y f_{res} es la resolución de frecuencia en Hz/bin . Obsérvese que esta última se refiere a la diferencia de frecuencia entre cada bin, es decir, los resultados o bins de un algoritmo FFT indican la respuesta de la magnitud de la frecuencia para frecuencias centradas específicas separadas por f_{res} , Figura 4-11. En nuestro caso, el primer bin de frecuencias está centrado en 0,016 Hz, el segundo está centrado en 0,033 Hz, y así se hace para los siguientes bins consecutivos. Así, teniendo en cuenta que la banda de PSD más baja a extraer está limitada de 0 a 0,1 Hz para la señal de BVP, el uso de esta longitud de ventana temporal es suficiente para tratar todas las bandas de PSD necesarias y obtener una separación adecuada entre ellas.

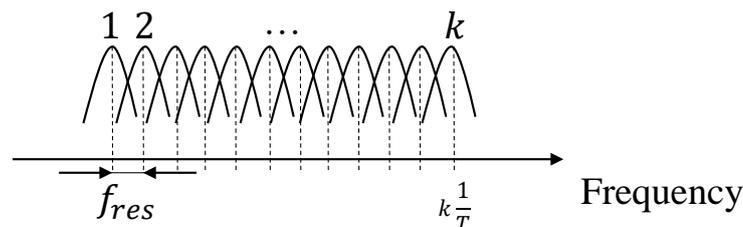


Figure 4-11: Ilustración de la resolución de la frecuencia y ubicación de los bins de frecuencia basados en una ventana de procesamiento de T segundos.

Las características en el dominio de la frecuencia para el DEAP-b2 sólo se contemplaron utilizando la señal BVP. Se extrajeron cuatro bandas de baja frecuencia PSD de la señal filtrada junto con la relación PSD de la contribución de baja frecuencia y la de alta frecuencia de los IBIs extraídos. Por un lado, las cuatro bandas de baja frecuencia van de 0 a 0,4 Hz en pasos de 0,1 Hz. Esto permite recoger información sobre los componentes de baja frecuencia dentro del BVP, es decir, principalmente los efectos respiratorios. Por otro lado, la relación PSD de las bandas de baja y

alta frecuencia para los IBI extraídos se basa en la activación simpática y parasimpática. Por ejemplo, en el caso de que la varianza de IBI observada fuera muy baja, la actividad cardíaca sería estable o constante, lo que desde el punto de vista frecuencial implica que las bandas de muy baja frecuencia tienen más potencia que las de alta frecuencia. Hay que tener en cuenta que este estado fisiológico podría ser desencadenado por factores de estrés agudos, es decir, en nuestro caso las emociones negativas que conducen a la activación simpática. Sin embargo, cuando se está en condiciones de reposo, la varianza del IBI será alta, lo que conduce a la activación de las bandas de alta frecuencia. Esto se representa en la Figura 4-12, que muestra una representación y relación ideal entre las partes de baja frecuencia (LF) y de alta frecuencia (HF) del IBI PSD, dada por la Task Force de la Sociedad Europea de Cardiología [8]. Como se puede observar, aunque la parte de HF ocupa un mayor rango espectral en comparación con la parte de LF, se obtiene un incremento evidente en LF al no estar en condiciones de reposo y normalizar ambos factores. Hay que tener en cuenta que se pueden dividir en más bandas internas, proporcionando información relativa a las frecuencias ultrabaja, muy baja, muy alta y ultraalta. En este caso, utilizamos una distinción gruesa agrupándolas en dos bandas principales: de 0 a 0,08 Hz para LF y de 0,15 a 0,5 Hz para HF.

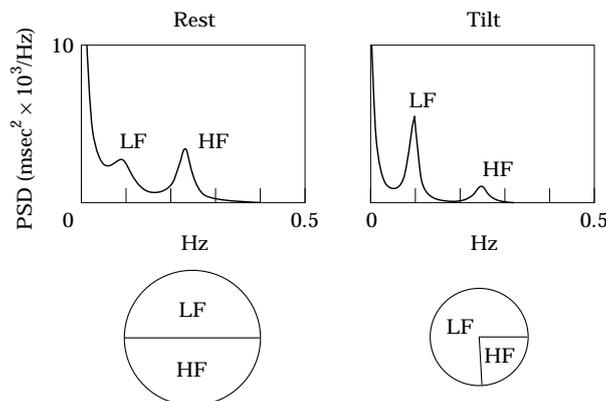


Figure 4-12: Representación ideal y relación entre las partes de baja frecuencia (LF) y de alta frecuencia (HF) de la IBI PSD [8].

La PSD se calculó utilizando el estimador de promedio de segmentos superpuestos de Welch. Hay que tener en cuenta que la resolución de frecuencia para la IBI no es la misma que para la señal BVP filtrada, ya que la IBI es una señal muestreada o adquirida de forma irregular. Esto implica que para una ventana temporal fija, el número de IBIs recogidos puede no ser el mismo para diferentes ventanas de

procesamiento. Por ejemplo, considerando una ventana de procesamiento fija de 60 segundos, un 40 BPM estable llevaría a una resolución de frecuencia de 0,025 Hz/bin, mientras que un 100 BPM estable llegaría hasta 0,01 Hz/bin. Obsérvese que el peor escenario será la frecuencia cardíaca más baja. Para mantener una resolución de frecuencia fija independientemente de la variabilidad de las BPMs, los puntos de IBI obtenidos se interpolan y remuestran a 8 Hz, con lo que se consigue 0,031 Hz/bin. Esto último es suficiente para tratar todas las bandas de PSD necesarias y obtener una separación adecuada entre ellas.

4.1.3.3 Dominio No-lineal

El último conjunto de características a extraer se basa en información no lineal. En el caso del sistema DEAP-b2, éstas se aplican únicamente a la señal IBI extraída. Así como la información proporcionada por la señal GSR está más directamente relacionada con la activación SNS debida a las glándulas sudoríparas ecrinas, la información obtenida de la señal BVP puede presentar una amplia gama de comportamientos que se producen por diferentes combinaciones o no linealidades fisiológicas (SNS y PNS) y físicas. Éstas pueden identificarse como factores vasculares o hemodinámicos que se modifican por factores estresantes externos o incluso por la homeostasis hacia la termorregulación en diferentes condiciones físicas. Así pues, las características no lineales pueden aportar información que los métodos lineales pierden. De hecho, la superioridad de los métodos no lineales cuando se aplican al reconocimiento de emociones utilizando información fisiológica es un tema candente en la actualidad [188].

Se sabe que el comportamiento fisiológico no lineal puede verse en diferentes escalas de tiempo, como los ritmos circadianos. Por lo tanto, para este sistema utilizamos la Multi-Scale Entropy (MSE) introducida en [189] para considerar el aspecto no lineal de la serie temporal IBI y una dependencia de la escala temporal. Esta métrica amplía la entropía de la muestra a diferentes escalas temporales para proporcionar una perspectiva adicional cuando la escala temporal de relevancia es desconocida, como es nuestro caso. Todos los cálculos se basan en la entropía estadística dada por la ecuación 4.8

$$H(X) = - \sum_i p(X_i) \log(p(X_i)), \quad (4.8)$$

donde $p(X_i)$ es la función de masa de probabilidad para el bloque de datos i . Ob-

sérvese que se trata de una medida de la incertidumbre media de la señal analizada, es decir, una serie temporal con fluctuaciones no periódicas generará valores más altos que una señal sinusoidal pura (no caótica). Así, el cálculo de la entropía de la muestra comienza segmentando la serie temporal de N puntos por una dimensión de incrustación m , donde $m < N$. Esto lleva a $N - m + 1$ tales segmentos. A continuación, se calcula la distancia d entre los distintos puntos m -dimensionales y se compara con un umbral predefinido r . En caso de que $d < r$, los dos segmentos se consideran similares y se almacena una clasificación positiva de "1" m , de lo contrario se anota una clasificación nula de "0". Esto se hace igualmente para $m + 1$. Finalmente, los resultados se expresan mediante las matrices: expresadas en las ecuaciones 4.9 y 4.10

$$A(m, r) = \frac{1}{N - m} \sum_{i=1}^{N-m} \Theta(r - d_i), \quad (4.9)$$

$$B(m + 1, r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m} \Theta(r - d_{i+1}), \quad (4.10)$$

que se utilizan para proporcionar la entropía final de la muestra expresada en la ecuación 4.11

$$SampEn(X) = -\log \left(\frac{A(m, r)}{B(m + 1, r)} \right). \quad (4.11)$$

La extensión que MSE introduce dentro de la entropía de la muestra es el grano grueso o el muestreo descendente de la serie temporal en diferentes escalas de tiempo. Así, en cada nivel (escala temporal), la serie temporal de grano grueso se obtiene promediando los respectivos puntos de la serie temporal. Esto se ilustra en la figura 4-13 y se expresa matemáticamente mediante la ecuación

$$y_j^\tau = \frac{1}{\tau} \sum_{i=(j-1)\tau}^{j\tau} x_i, 1 \leq j \leq N\tau, \quad (4.12)$$

donde τ es la escala de tiempo o nivel. Por último, se calcula la entropía de la muestra sobre los y_j^τ obtenidos. En nuestro análisis, empleamos un MSE de cinco niveles ($\tau = 5$), fijamos $m = 2$, y $r = 0,2\sigma$, donde σ es la desviación estándar de la serie temporal IBI. Obsérvese que estos parámetros se eligieron basándose en trabajos anteriores en la literatura que tratan sobre el reconocimiento de emociones

y este tipo de características no lineales [190].

$$\begin{array}{c}
 \text{Time scale 1: } \underbrace{x_1 \ x_2}_{y_1} \ \underbrace{x_3 \ x_4}_{y_2} \ \underbrace{x_5 \ x_6}_{y_3} \ \underbrace{x_7 \ x_8}_{y_4} \ \dots \ \underbrace{x_{N-1} \ x_N}_{y_N} \\
 \\
 \text{Time scale 2: } \underbrace{x_1 \ x_2 \ x_3}_{y_1} \ \underbrace{x_4 \ x_5 \ x_6}_{y_2} \ \underbrace{x_7 \ x_8 \ x_9}_{y_3} \ \dots \ \underbrace{x_{N-2} \ x_{N-1} \ x_N}_{y_N} \\
 \\
 \dots \\
 \\
 \text{Time scale 5: } \underbrace{x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6}_{y_1} \ \underbrace{x_7 \ x_8 \ x_9 \ x_{10} \ x_{11} \ x_{12}}_{y_2} \ \dots \ \underbrace{x_{N-2} \ x_{N-1} \ x_N}_{y_N}
 \end{array}$$

Figure 4-13: Cada serie temporal de grano grueso obtenida para cada nivel de la técnica o algoritmo de extracción de características MSE.

4.1.4 Sistemas de clasificación del miedo

En los siguientes apartados se detallan y explican los resultados obtenidos con los sistemas DEAP-b1 y DEAP-b2. Obsérvese que todas las especificaciones relativas a la transformación del mapa de etiquetado, el análisis exploratorio de datos, el procesamiento de datos y la extracción de características ya se han detallado en las secciones anteriores de este capítulo.

4.1.4.1 Sistema DEAP-b1

Este sistema [183] fue, hasta donde yo sé, el primero en la literatura en proponer y validar el mapeo binario específico del miedo utilizando el espacio PAD y sólo tres variables fisiológicas. Como ya se ha dicho en apartados anteriores, sólo se emplearon 21 de los 32 voluntarios del DEAP, y no se aplicó ninguna extracción de características como tal, sino que se consideró el valor filtrado de cada variable fisiológica. Cada voluntario fue sometido a un escalado de 0 – 1 para el conjunto completo de valores fisiológicos recogidos durante el experimento. Para la clasificación, de cara a una primera prueba de concepto de implementación embebida del sistema, consideramos un algoritmo perezoso, concretamente, un KNN. Tenga en cuenta que el valor de k se fijó a la raíz cuadrada del tamaño del conjunto de entrenamiento, que es una práctica comúnmente aplicada. Para simplificar el cálculo, se considera la distancia euclidiana para comparar dos muestras. Para ser justos en

esta comparación y evitar los problemas relacionados con los valores en diferentes unidades y escalas, cada uno de los tres valores de una muestra se normaliza como se ha indicado anteriormente (escala de 0 – 1). Por último, para la validación, implementamos una estrategia de Hold-Out CV y realizamos un barrido experimental de parámetros para la relación de Hold-Out. Además, para hacer frente a la situación de fuerte desequilibrio en el etiquetado, decidimos aplicar el aprendizaje sensible a los costes mediante el ajuste de un parámetro de coste de clasificación errónea. Esta práctica se utiliza habitualmente en los problemas de clasificación binaria de desequilibrio. En este caso, dicho parámetro define una penalización que da más importancia (peso) a los falsos negativos producidos. Por lo tanto, el uso de esta penalización es útil para reducir la tasa de falsos negativos en nuestro sistema, que es fundamental para la aplicación de la violencia de género. Por ejemplo, al considerar a Bindi, el brazalete se encuentra en realidad en la parte inferior de una cascada de dispositivos más potentes, por lo que este dispositivo wearable restringido podría actuar como disparador para ejecutar algoritmos más complejos en capas superiores si fuera necesario. De ahí que sea esencial reducir el número de falsos negativos en este primer paso, aunque penalice la precisión.

Como primer paso en la evaluación de DEAP-b1, nos centramos en los datos de un voluntario arbitrario, el número 18 (*p18*). Para este voluntario concreto, hay un total de 256.358 muestras o instancias de cada variable fisiológica recogidas durante todo el experimento (40 clips). El algoritmo KNN se entrena utilizando diferentes valores de Hold-Out y costes de clasificación errónea. Para cada combinación de Hold-Out y miss-classification, se generan aleatoriamente 30 sistemas de inteligencia independientes para tener validez estadística en los resultados obtenidos. Se analizan y comparan diferentes métricas como la precisión, la especificidad (o tasa de verdaderos negativos), la sensibilidad (o tasa de verdaderos positivos) y la media geométrica entre la sensibilidad y la especificidad. Cabe mencionar que el uso de la memoria se ve fuertemente afectado por la proporción de Hold-Out, ya que es necesario almacenar el espacio de entrenamiento completo para procesar más muestras y proporcionar futuras inferencias. Así, el espacio de diseño para el ratio de Hold-Out va de 0,7 a 0,9999, lo que lleva a un tamaño del conjunto de entrenamiento del 30% al 0,01%.

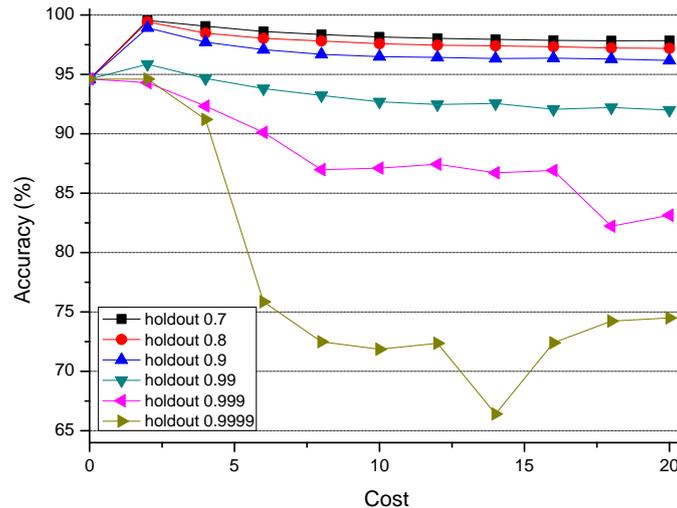


Figure 4-14: Exactitud frente al coste de la clasificación errónea para p_{18} .

La figura 4-14 muestra la precisión frente al coste de la clasificación errónea para los diferentes valores de Hold-Out en p_{18} . Analizando esta figura, comprobamos que i) la precisión es mejor para valores más bajos de Hold-Out (el conjunto de entrenamiento es mayor y entonces, el sistema está mejor caracterizado) y ii) la precisión suele disminuir a medida que aumenta la penalización por clasificación errónea (el número de falsos negativos se reduce, pero también aumenta el número de falsos positivos). La figura 4-15 muestra la sensibilidad frente al coste de la clasificación errónea para los diferentes valores de Hold-Out en p_{18} . Analizando esta figura, comprobamos que la sensibilidad aumenta con el coste de la clasificación errónea en función del mecanismo de esta penalización. La figura 4-16 muestra la especificidad frente al coste de la clasificación errónea para los diferentes valores de Hold-Out en p_{18} . En esta figura, la especificidad disminuye con el coste de clasificación errónea basado en esta penalización. Nótese que la leyenda de las dos últimas figuras es la misma que la primera.

Aplicando este análisis al resto de los voluntarios y observando comportamientos similares, determinamos que un coste de clasificación errónea de 8 unidades puede ser adecuado para el conjunto de datos actual. Sin embargo, en el caso de la relación Hold-Out, esta decisión no es inmediata ni se basa puramente en el rendimiento, ya que es necesario estudiar el impacto en la memoria y el cálculo de este parámetro. Así, la Tabla 4.5 muestra, para cada valor de Hold-Out para el voluntario p_{18} , el tamaño del conjunto de entrenamiento, la memoria necesaria considerando dicha asignación del conjunto de entrenamiento y la memoria consumida por el algoritmo

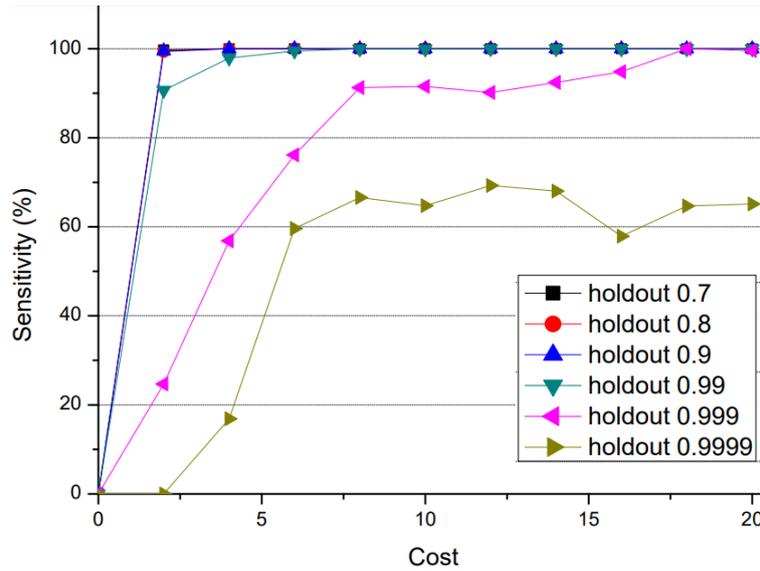


Figure 4-15: Sensibilidad frente al coste de la clasificación errónea para $p18$.

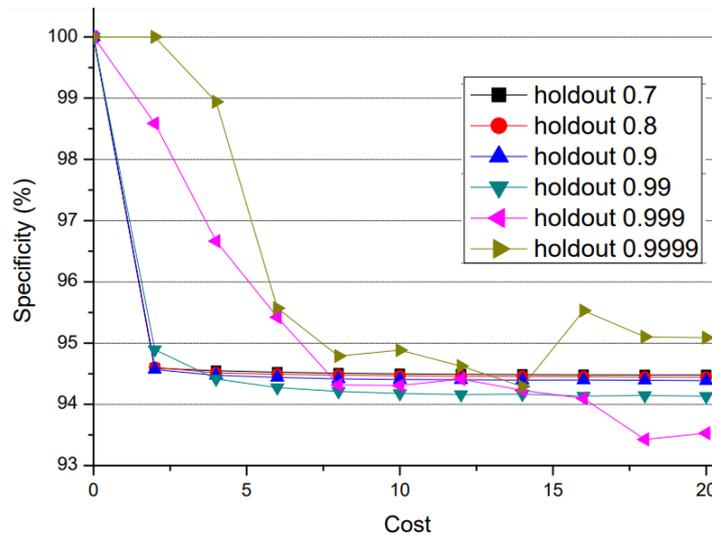


Figure 4-16: Especificidad frente al coste de la clasificación errónea para $p18$.

KNN, así como una estimación del número de operaciones. Nótese que la memoria utilizada en KB se basa en un tipo de datos enteros de 32 bits, y el número de operaciones se basa en la complejidad computacional media del método de ordenación rápida que suele encontrarse en las implementaciones de KNN. Esta complejidad es $\mathcal{O}(n \log n)$, donde n es el número de elementos a ordenar, es decir, el tamaño del conjunto de datos de entrenamiento. Analizando esta tabla, comprobamos que el valor de Hold-Out tiene un impacto importante en la memoria utilizada y en el número de operaciones a computar. De hecho, pasar de 0,999 a 0,99 conlleva más de 13 veces más operaciones. Estos aspectos son críticos para un sistema de edge-computing como Bindi. En consecuencia, después de analizar las tendencias en las

Hold-Out	tamaño entrenamiento	memoria usada (kB)	operaciones
0.7	76907	976.36	375767
0.8	51272	650.91	241483
0.9	25636	325.45	113024
0.99	2564	32.54	8739
0.999	256	3.25	618
0.9999	26	0.32	36

Table 4.5: Impacto del tamaño del conjunto de entrenamiento en la memoria y el cálculo para $p18$. Enfoque dependiente del sujeto.

figuras 4-14, 4-15 y 4-16, el valor de Hold-Out se fija en 0,99.

Sobre la base de los valores definidos anteriormente para el coste de retención y de clasificación errónea, la tabla 4.6 muestra las métricas de precisión, sensibilidad, especificidad y media geométrica para todos los voluntarios considerados en el conjunto de datos. Obsérvese que todos los voluntarios tienen un tamaño comparable a $p18$, por lo que fijamos el mismo valor de Hold-Out. En este caso, también se han entrenado 30 sistemas independientes para que cada modelo tenga validez estadística. Analizando esta tabla, comprobamos que, por término medio, la precisión, la sensibilidad, la especificidad y la media geométrica fueron de 0,85, 0,99, 0,81 y 0,81 respectivamente. Obsérvese que la media más alta y la varianza más baja de la sensibilidad en comparación con las demás métricas se debe al coste de la clasificación errónea, es decir, el sistema está más sesgado hacia la clase positiva.

En caso de considerar un enfoque independiente del sujeto y aplicar el mismo coste de clasificación errónea, el Hold-Out podría reevaluarse en términos de consumo de memoria. Así, la tabla 4.7 muestra, para cada valor de Hold-Out, el tamaño del conjunto de entrenamiento al mezclar los datos de todos los voluntarios considerados en el conjunto de datos, la memoria utilizada en kB se basa en un tipo de datos enteros de 32 bits, y el número de operaciones sigue el mismo enfoque de cálculo que para el caso dependiente del sujeto. Analizando esta tabla, comprobamos que, para este sistema, un valor de Hold-Out igual a 0,999 es comparable a un valor de 0,99 para el caso dependiente del sujeto en la Tabla 4.5, es decir, necesitamos reducir el ratio de Hold-Out para conseguir un número similar de puntos en el conjunto de datos de entrenamiento. La tabla 4.8 muestra las métricas de precisión, sensibilidad, especificidad y media geométrica para el caso independiente del sujeto asumiendo diferentes valores de Hold-Out y considerando el coste de clasificación errónea de 8 de antes. En concreto, las métricas obtenidas para el Hold-Out de 0,999 son signi-

Precisión	Sensibilidad	Especificidad	Media Geométrica	Voluntarios/as
0.71	0.97	0.67	0.81	p_1
0.93	1.00	0.85	0.92	p_2
0.80	0.97	0.77	0.86	p_3
0.79	0.97	0.85	0.90	p_4
0.96	0.97	0.97	0.97	p_5
0.91	0.99	0.94	0.96	p_6
0.99	1.00	0.97	0.98	p_7
0.86	0.97	0.89	0.92	p_8
0.86	1.00	0.80	0.89	p_9
0.83	0.98	0.88	0.92	p_{10}
0.94	1.00	0.87	0.93	p_{11}
0.69	1.00	0.52	0.72	p_{12}
0.84	1.00	0.73	0.85	p_{13}
0.76	0.99	0.70	0.83	p_{14}
0.79	1.00	0.66	0.81	p_{15}
0.71	0.99	0.74	0.85	p_{16}
0.87	1.00	0.86	0.92	p_{17}
0.93	1.00	0.94	0.96	p_{18}
0.82	0.99	0.79	0.88	p_{19}
0.84	1.00	0.81	0.90	p_{20}
0.77	0.99	0.74	0.85	p_{21}
0.84 (0.08)	0.99 (0.01)	0.81 (0.11)	0.88 (0.06)	$\mu(\sigma)$

Table 4.6: Métricas de precisión, sensibilidad, especificidad y media geométrica para cada voluntario asumiendo un coste de retención y de clasificación errónea de 0,99 y 8, respectivamente. Enfoque dependiente del sujeto.

ficativamente inferiores a las del caso dependiente del sujeto. Además, comprobamos que las tres métricas mejoran al considerar valores más bajos de holdout, como para el caso dependiente del sujeto. Sin embargo, su alto impacto en la complejidad espacial y temporal hace que su implementación en dispositivos de borde restringido no sea factible.

Por lo tanto, basándonos en estos experimentos, podemos concluir que una implementación dependiente del sujeto puede mejorar significativamente el rendimiento de la inferencia del estado emocional en un dispositivo wearable de tamaño reducido. En concreto, el enfoque dependiente del sujeto proporciona hasta 0,84, 0,99, 0,81 y 0,88 de precisión, sensibilidad, especificidad y media geométrica de media, mientras que el enfoque independiente del sujeto proporciona hasta 0,54, 0,88, 0,47 y 0,62, para las configuraciones elegidas. Para este último enfoque, en el caso de Bindi, que pretende proporcionar un motor de aprendizaje automático del miedo para ser

desplegado en la vida real, la sensibilidad debería ser cercana a 1,00 para maximizar la detección de verdaderos positivos, como ocurre con el enfoque dependiente del sujeto.

A pesar de la evidencia de los resultados, la estrategia Hold-Out utilizada para ambos modelos, dependiente e independiente del sujeto, puede conducir a resultados demasiado optimistas. Esto se debe a que las entradas del sistema son los valores fisiológicos filtrados y el Hold-Out aplicado no tuvo en cuenta si pertenecen al mismo vídeo. Así, los procesos de entrenamiento y de prueba podrían estar utilizando información del mismo conjunto de datos fisiológicos recogidos durante la visualización de un determinado videoclip.

Hold-Out	tamaño entrenamiento	memoria usada (kB)	operaciones
0.98	76555	971.89	373892
0.99	38277	485.94	175423
0.999	3828	48.59	13715
0.9999	383	4.86	989

Table 4.7: Impacto del tamaño del conjunto de entrenamiento en la memoria y el cálculo. Enfoque independiente de la materia.

Precisión	Sensibilidad	Especificidad	Media geométrica	Hold-Out
0.66	0.96	0.52	0.71	0.980
0.64	0.95	0.50	0.69	0.990
0.54	0.88	0.47	0.62	0.999
0.50	0.81	0.45	0.60	0.9999

Table 4.8: Exactitud, sensibilidad, especificidad y media geométrica para cada retención probada asumiendo un coste de clasificación errónea de 8. Enfoque independiente del sujeto.

4.1.4.2 Sistema DEAP-b2

Durante la investigación y el desarrollo del sistema DEAP-b1, identificamos cinco inconvenientes diferentes:

1. El bajo número de voluntarios podría afectar a la variabilidad de los datos.
2. La complejidad y el desequilibrio del mapeo binario de miedo de PAD fue particularmente alto para esta base de datos.
3. El hecho de no haber tenido en cuenta la extracción de características podría

llevar a la pérdida de información fisiológica de interés.

4. La aplicación de una estrategia de Hold-Out sobre los valores fisiológicos filtrados podría dar lugar a métricas demasiado optimistas.
5. La complejidad espacial de este algoritmo perezoso de KNN era considerablemente alta cuando se consideraban valores de Hold-Out bajos.

El primer inconveniente puede solucionarse considerando el conjunto completo de voluntarios a expensas de omitir la señal SKT de todos ellos. Obsérvese que durante el análisis exploratorio de los datos se encontraron imprecisiones en SKT para un total de 11 voluntarios. La segunda deficiencia puede paliarse aplicando el mapeo binario del miedo a partir de PA, que mostró una relación de desequilibrio menor. La tercera razón es una de las más sensibles desde el punto de vista fisiológico, ya que al considerar sólo los valores filtrados en bruto, estamos perdiendo toda la información temporal, morfológica, basada en la frecuencia y no lineal. El cuarto inconveniente debe solucionarse para evaluar adecuadamente el rendimiento del sistema y garantizar que no se proporcione a la fase de entrenamiento ninguna información relacionada con las pruebas o incluso con las mismas. Por último, el quinto inconveniente afecta al tamaño del conjunto de entrenamiento y motiva la evaluación de diferentes algoritmos de clasificación que proporcionen menos requisitos de espacio. En base a estos problemas identificados, el sistema DEAP-b2 [181] intentó superarlos considerando los 32 voluntarios del DEAP, sólo los datos de los sensores PPG y GSR, un mapeo binario del miedo utilizando el espacio PA, y un conjunto completo de 20 características que incluyen el dominio temporal, frecuencial y no lineal, (enumeradas en la Tabla 4.4 y detalladas en la Sección 4.1.3). Además, el proceso de extracción de características se aplicó considerando una ventana de procesamiento de 60 segundos, que se correspondía con la duración del estímulo y proporcionaba un conjunto de 20 características por vídeo. De este modo, nos aseguramos de que no se da información dentro del mismo vídeo tanto en el entrenamiento como en la prueba al realizar la estrategia de Hold-Out.

Además de simplificar el problema de etiquetado y reducir el ratio de desequilibrio mediante la elección de un mapeo binario de miedo basado en el espacio PA, se aplicó una técnica de sobremuestreo sobre los datos de la clase minoritaria (miedo). En concreto, se aplicó SMOTE para tratar los problemas de equilibrio observados [191].

Esta técnica se basa en un enfoque de sobremuestreo de la clase minoritaria que genera nuevas muestras considerando los vecinos k más cercanos, en lugar de un sobremuestreo con reemplazo. Así, en lugar de tener 1280 instancias (32 voluntarios x 40 vídeos) con un ratio de equilibrio de clase de aproximadamente 76/24% (negativo/positivo), conseguimos un ratio de equilibrio de clase de hasta 50/50% con un total de 1800 instancias. Obsérvese que el valor k para el SMOTE se fijó en 5.

En cuanto a los clasificadores específicos utilizados en este sistema, se utilizaron Gaussian naive Bayes (ecuación 4.2) y SVM con RBF kernel. Esta decisión se basó en la consideración de dos hechos principales:

- Decidimos utilizar el mismo clasificador utilizado por DEAP (Gaussian naïve Bayes) para proporcionar una comparación justa con respecto al conjunto de datos original.
- Para superar la complejidad espacial de KNN, se aplica un clasificador SVM con RBF kernel, ya que conserva todas las ventajas del algoritmo KNN, almacenando sólo los vectores de soporte durante el entrenamiento en lugar de todo el espacio de entrenamiento.

Además del escalado de 0–1 realizado en el sistema anterior, se aplicó el z-score por voluntario para normalizar los datos en este caso. Para la metodología de prueba del sistema DEAP-b2, se ejecutó la estrategia Hold-Out desde 0,01 (1%) hasta 0,9 (90%) para 100 iteraciones cada paso de 0,01. Además, para la validación del clasificador SVM durante el entrenamiento, se implementa un $k - fold$ con $k = 5$.

La topología del sistema, en este caso, se basa en un enfoque independiente del sujeto, ya que uno dependiente del sujeto no era factible debido a la pequeña cantidad de datos (40 conjuntos de 20 características por voluntario). Además, aparte de las métricas de clasificación utilizadas para el DEAP-b1, en este caso también se da el Area Under the Curve (AUC), que proporciona una medida del rendimiento a través de todos los posibles umbrales de clasificación y presenta la probabilidad de que el modelo clasifique un positivo aleatorio más alto que un negativo aleatorio.

Todas estas consideraciones de DEAP-b2 se estructuran y combinan en seis configuraciones diferentes para proporcionar un DSE acotado para el caso de uso independiente del sujeto. Éstas son las siguientes:

- Caso 1. El sistema se implementa sin utilizar ninguna selección de caracterís-

ticas y aplicando el clasificador gaussiano de Bayes ingenuo.

- Caso 2. El sistema utiliza el mismo método de selección de características de filtro que DEAP (puntuación discriminante lineal de Fisher, ecuación 4.1), pero implementa un clasificador SVM con kernel RBF. Este último se toma de [148] con $\gamma = 0,15$ y $C = 1$.
- Caso 3. Esta configuración del sistema sigue la misma estructura que el caso 2, pero sin emplear la selección de características.
- Caso 4. Implementa SMOTE para tratar el problema de equilibrio, utiliza la puntuación discriminante lineal de Fisher para seleccionar las características relevantes y ejecuta el clasificador gaussiano naïve Bayes.
- Caso 5. Presenta la misma configuración que el Caso 4, pero utilizando el clasificador SVM del Caso 2.
- Caso 6. Emplea SMOTE, puntuación discriminante lineal de Fisher para seleccionar las características relevantes y un clasificador SVM con kernel RBF. Además, se aplica una búsqueda en cuadrícula para encontrar los hiperparámetros óptimos para dicho clasificador.

En la tabla 4.9 se ofrece un análisis comparativo de estas seis implementaciones diferentes, en el que se pueden apreciar los resultados de tipo aleatorio por las AUC obtenidas para todos los casos, excepto el caso número seis. Nótese que algunos de los valores de esta tabla han sido modificados en comparación con los obtenidos en [181], ya que se ha realizado un mayor ajuste de los modelos después de dicha publicación. En primer lugar, el clasificador gaussiano naïve Bayes del caso 1 obtiene un rendimiento pobre. Esto puede verse afectado por la independencia de las características extraídas, ya que se sabe que este tipo de clasificador proporciona un buen rendimiento cuando las características son independientes entre sí. Entonces, llama la atención que una alta puntuación de precisión no signifique que el modelo tenga un rendimiento adecuado. Por ejemplo, los casos 2 y 3 tienen la tasa de clasificación más alta, pero no hay sensibilidad, sin embargo la especificidad es cercana al 100%. Por lo tanto, se da la paradoja de la exactitud, por lo que la exactitud sólo refleja la distribución de clases subyacente. También para estos dos casos, la técnica de selección de características específicas por sí misma, es decir, la puntuación discriminante lineal de Fisher, no proporciona ninguna ventaja. Esto

Case	Precisión ($\mu(\sigma)$)	Sensibilidad ($\mu(\sigma)$)	Especificidad ($\mu(\sigma)$)	Media Geométrica ($\mu(\sigma)$)	AUC ($\mu(\sigma)$)
1	52.48 (0.34)	50.84 (2.14)	53.36 (0.90)	52.08 (1.38)	52.55 (1.13)
2	76.47 (0.34)	0.12 (0.12)	99.87 (0.11)	3.46 (0.11)	50.00 (0.10)
3	76.54 (0.37)	0.03 (0.07)	99.96 (0.07)	1.73 (0.07)	50.00 (0.10)
4	51.80 (0.54)	52.73 (0.94)	50.86 (0.90)	51.78 (0.91)	52.24 (0.80)
5	53.27 (0.73)	58.80 (2.83)	47.72 (1.77)	52.97 (2.23)	53.50 (2.20)
6	62.80 (4.75)	62.27 (4.14)	66.99 (5.79)	62.62 (4.73)	62.79 (4.72)

Table 4.9: Métricas de precisión, sensibilidad, especificidad y AUC para cada caso asumiendo las condiciones especificadas, respectivamente. Enfoque independiente del sujeto.

puede deberse al conjunto de características no óptimo que genera esta técnica, como se ha indicado anteriormente en la sección 4.1.4. Los casos 4 y 5, en comparación con el 6, demuestran que es esencial encontrar los hiperparámetros correctos para lograr el mejor equilibrio entre sesgo y varianza. Así, aplicando una búsqueda en cuadrícula, el rendimiento del esquema de clasificación mejora. Por último, el caso 6 combina un método de sobremuestreo sintético, una selección de rasgos clasificados, un clasificador no lineal y un proceso de ajuste de hiperparámetros de búsqueda en cuadrícula, logrando hasta un 62,79% de AUC, que supera al resto de los casos.

Aunque la media geométrica de este último caso es inferior a la obtenida para el modelo dependiente del sujeto del DEAP-b1, hay que destacar de nuevo el inconveniente de validación que se observa al aplicar la estrategia Hold-Out directamente sobre los valores fisiológicos filtrados. Además, al comparar los resultados del DEAP-b2 con el DEAP-b1 para el sujeto-independiente, podemos observar que ambos sistemas alcanzan una media geométrica similar, pero el DEAP-b2 supera al DEAP-b1 en especificidad en más de un 15% y en precisión en más de un 8%. Esto indica que el DEAP-b2 presenta un mejor equilibrio entre los falsos positivos y los falsos negativos, lo que da lugar a un sistema de mejor rendimiento. Finalmente, desde un balance de complejidad temporal y espacial, el SVM supera claramente al anterior KNN implementado. Este último presenta $\mathcal{O}(n \log n)$ y $\mathcal{O}(n)$ para la complejidad temporal y espacial respectivamente, siendo n el tamaño del conjunto de entrenamiento. Por el contrario, el SVM con el kernel RBF alcanza $\mathcal{O}(n_{sv}d)$ y $\mathcal{O}(n_{sv})$ para la complejidad temporal y espacial respectivamente, donde n_{sv} es el tamaño o número de los vectores de soporte y d es el número de atributos o características a emplear. En el peor de los casos, considerando las 20 características y el

millar de vectores de soporte elegidos sobre el conjunto completo de 1.800 instancias, habrá un total de 20.000 operaciones por predicción, lo que es considerablemente superior a las dependientes e independientes del sujeto de DEAP-b1. Sin embargo, debido a la selección de características que realiza DEAP-b2, la mayoría de las iteraciones terminaron con la mitad de las características, lo que se traduce en 10.000 operaciones. Esta medida se encuentra entre los modelos DEAP-b1 dependiente del sujeto y el independiente del sujeto que utiliza KNN. En cuanto a la complejidad espacial, el SVM en este caso requiere hasta 39,06 kB (1000 *sv* por cada característica), que también se encuentra entre los dos modelos DEAP-b1. Obsérvese que la memoria utilizada en kB se basa en un tipo de datos enteros de 32 bits para todos los vectores de soporte a almacenar.

Cabe destacar que pueden implementarse y aplicarse otras optimizaciones algorítmicas SVM, así como otras alternativas de selección de características, para conseguir una menor cantidad de vectores de soporte, lo que reduciría tanto los requerimientos de tiempo como de espacio de este sistema conduciendo incluso en algunos casos a un mejor rendimiento de reconocimiento. De hecho, la tabla 4.10 muestra los resultados para la misma configuración de entrenamiento, validación y prueba que el caso número seis pero cambiando el método de selección de características a mrMR. Esta técnica [192] se basa en la suposición de que dentro de todo el conjunto de características dadas hay un conjunto mínimo-óptimo en el que dichas características son mutuamente tan disímiles entre sí como sea posible, pero también marginalmente tan similares a la variable de clasificación como sea posible. Con esta técnica, queremos seleccionar las características que tengan la máxima relevancia para la variable de clasificación (objetivo) y que presenten una redundancia mínima en comparación con el resto de las demás características evaluadas. Así, para medir dichas propiedades entre dos variables (X y Y), se emplea la información mutua, que viene dada por la ecuación 4.13

$$I(X, Y) = \sum_{x \in X} \sum_{y \in Y} p_{(X,Y)}(x, y) \log \frac{p_{(X,Y)}(x, y)}{p_X(x)p_Y(y)}, \quad (4.13)$$

donde $p(x, y)$ es la distribución probabilística conjunta, y $p(x)$ y $p(y)$ son las funciones de densidad de probabilidad marginal para cada variable respectivamente. A

partir de esta información, el nivel de similitud (o disimilitud) entre dos características (i y j) se codifica mediante la condición de mínimo; enunciada en la ecuación 4.14

$$\min W_I, W_I = \frac{1}{|S|^2} \sum_{i,j \in S} I(i, j), \quad (4.14)$$

donde S es el subconjunto de características mínimas-óptimas. Del mismo modo, el poder discriminante de los rasgos o la relevancia para la variable de clasificación h (objetivo) viene dado por la condición máxima establecida en la ecuación 4.15

$$\max V_I, V_I = \frac{1}{|S|} \sum_{i \in S} I(h, i). \quad (4.15)$$

A partir de las ecuaciones 4.14 y 4.15, se puede obtener el conjunto de características mínimo-óptimo optimizándolas de forma simultánea y aplicando diferentes funciones de criterio. En concreto, la implementación particular de mrMR para este sistema aplicó el criterio Mutual Information Difference (MDI) para realizar el proceso de clasificación. Ecuación 4.16

$$\max(V_I - W_I). \quad (4.16)$$

Dado que esta técnica de selección de características es un método de selección de características por filtro que clasifica las características existentes basándose en dichas premisas, debe indicarse el número de características finales, K , que se considerarán después de dicha clasificación. En este caso, la selección se realiza en base a un barrido experimental de parámetros teniendo en cuenta los requisitos de rendimiento y almacenamiento. Finalmente, K se fija en 10. Por lo tanto, los resultados obtenidos utilizando esta técnica específica de selección de características superan en más de un 18% el AUC del modelo independiente del sujeto de DEAP-b2, manteniendo los mismos requisitos de almacenamiento. Este último experimento demuestra que las posibilidades de optimización son elevadas en este complejo problema.

Caso	Precisión	Sensibilidad	Especificidad	Media Geométrica	AUC
	$(\mu(\sigma))$	$(\mu(\sigma))$	$(\mu(\sigma))$	$(\mu(\sigma))$	$(\mu(\sigma))$
6+	81.54 (8.69)	70.93 (14.92)	94.59 (3.89)	81.55 (10.21)	81.60 (8.70)

Table 4.10: Exactitud, sensibilidad, especificidad y métrica AUC para la selección de características mrMR y SVM con kernel RBF. Enfoque independiente del sujeto.

4.2 Clasificación del miedo mediante MAHNOB

Después de haber presentado los resultados de tres sistemas diferentes de detección de miedo (uno dependiente del sujeto y dos independientes del sujeto) utilizando el conjunto de datos DEAP y siendo conscientes de las limitaciones encontradas, se requirió con fuerza la necesidad de un nuevo conjunto de datos en el que se resolvieran o aliviaran estos problemas. Algunos de estos problemas identificados se referían a la recuperación fisiológica entre estímulos, a las imprecisiones de los datos de temperatura de la piel y al desequilibrio de clases del mapeo del miedo y la binarización.

Como ya se ha revisado en la sección 3.2, la base de datos MAHNOB supera las limitaciones de recuperación fisiológica de DEAP, mantiene la misma información fisiológica recogida, y presenta incluso más etiquetas autoinformadas por los voluntarios. Además, no se observan problemas de medición del SKT, ni de ninguna otra variable fisiológica, con ninguno de los voluntarios válidos. Estas afirmaciones se ven incluso reforzadas por la literatura; por ejemplo, los autores en [193] realizaron un DSE para los vectores de características de DEAP y MAHNOB para investigar la relevancia de las características fisiológicas dentro de ambos conjuntos de datos. Uno de sus experimentos concluyó que los estímulos de MAHNOB eran más inmersivos emocionalmente que los de DEAP. De hecho, dicho trabajo ha motivado la realización de posteriores y recientes investigaciones como [194]. Sobre esta base, en este apartado se detallan los resultados obtenidos para un sistema de reconocimiento binario de miedo dependiente e independiente del sujeto, basado en el conjunto de datos MAHNOB. En concreto, y tratando de diseñar un sistema más especializado hacia el objetivo a largo plazo de este trabajo de investigación, se fijan dos aspectos clave de diseño:

- Sólo se emplean mujeres voluntarias. Esta restricción de diseño permite el desarrollo de sistemas de reconocimiento de emociones muy especializados debido a las particularidades emocionales entre hombres y mujeres, tal y como se revisa en el Capítulo 2.
- Se contempla la binarización del miedo a partir del espacio PAD. El factor de dominancia, como se revisa en el capítulo 2, es esencial para distinguir entre algunas de las principales emociones negativas (miedo y rabia).

Además, la segmentación de datos y otros procesos como la extracción de características se modifican o amplían para mejorar los sistemas presentados en la sección anterior.

Para la metodología específica seguida durante los experimentos de la base de datos MAHNOB, la figura 4-17 muestra un diagrama simplificado de la experimentación aplicada para cada voluntario y cada estímulo. A diferencia de DEAP, MAHNOB tuvo en cuenta la reducción del sesgo emocional después de cada visualización de estímulo y, por tanto, de las respuestas de emoción. De hecho, los clips neutros utilizados se seleccionaron al azar de un conjunto más amplio proporcionado por el laboratorio de psicofisiología de Stanford [195]. Esta consideración, junto con la grabación de 30 segundos antes y después del ensayo, proporcionó una recuperación fisiológica destinada a aislar la activación emocional entre los estímulos.

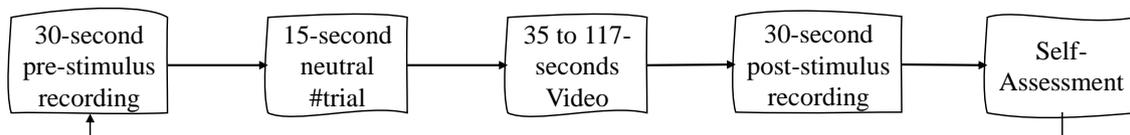


Figure 4-17: Metodología seguida durante los experimentos de la base de datos MAHNOB.

En cuanto a las diferencias técnicas concretas entre los sistemas presentados en este capítulo y el que podría integrarse realmente en Bindi, cabe destacar dos de ellas. Por un lado, el conjunto de datos de MAHNOB incluye la información de la actividad cardíaca medida con un sensor de ECG. Por tanto, los diferentes algoritmos de delineación o detección de picos, que se utilizan para extraer la información morfológica para calcular las diferentes características o métricas, deben ser diseñados específicamente para la morfología ECG y no para la PPG revisada en la sección 2.5.1. Este hecho afecta directamente a cualquier posible opción de integración del preprocesado para aprovechar esa parte algorítmica en la plataforma embebida de la pulsera inteligente de Bindi. Sin embargo, está demostrado en la literatura que PPG es un sustituto válido de ECG para diferentes métricas o características como HRV [196–198]. Por lo tanto, la extracción de características y los procesos posteriores pueden aplicarse de la misma manera independientemente de si el sensor es ECG o PPG. Por otra parte, el equipo utilizado durante el experimento de MAHNOB fue el mismo que en DEAP. Además, incluso en el caso de disponer de datos del sensor PPG y poder obtener características morfológicas al utilizar dicho equipo,

debido a la tarea extremadamente difícil de obtener señales PPG de calidad clínica (la morfología está totalmente preservada) con dispositivos wearables, eso no tendría ningún valor debido al altísimo ruido en la vida real. Así pues, al igual que en el apartado anterior, los sistemas aquí propuestos sirven como prueba de concepto y facilitan los diferentes procesos DSE que hay que realizar para el diseño e integración de un sistema óptimo de reconocimiento binario de emociones de miedo en el borde.

Precediendo a la presentación de los métodos empleados y los resultados obtenidos, se detalla una revisión del estado del arte, en cuanto a la utilización de MAHNOB para la generación de sistemas de reconocimiento de emociones. En primer lugar, como ya se ha descrito en la sección 3.2, el trabajo original del conjunto de datos MAHNOB comprende la adquisición de diferentes señales fisiológicas a una tasa de frecuencia de muestreo de 256 Hz durante la visualización de diferentes estímulos audiovisuales (20 clips emocionales intercalados con 20 clips neutros). En primer lugar, utilizaron procedimientos básicos de preprocesamiento para eliminar las derivas temporales de baja frecuencia de algunas señales y suavizarlas mediante filtros de media móvil. Extrajeron un total de 102 características de todas las señales recogidas y aplicaron un método de selección de características por filtro para utilizar sólo las de mayor rango. En concreto, utilizaron el método Analysis of Variance (ANOVA) de una vía y rechazaron cualquier característica no significativa ($p > 0,05$). Para la tarea de clasificación, proporcionaron dos sistemas de reconocimiento de emociones basados en niveles bajos, medios y altos de detección de la excitación y la valencia, para utilizarlos como punto de referencia en futuras investigaciones con estos datos. Los niveles resultaron del mapeo entre las palabras clave emocionales y las clases siguiendo [5]. En cuanto al clasificador, emplearon un SVM con RBF kernel y ajustaron γ utilizando un CV de 20 veces. Por último, la estrategia de prueba aplicada fue LOSO. Al emplear todas las señales periféricas, proporcionaron métricas promedio de ACC y F1-score y obtuvieron 46,20% y 38,00% para arousal y 45,50% y 39,00% para valencia, respectivamente. Cabe destacar que también realizaron una fusión de datos multimodal utilizando datos de EEG y de recolección de la mirada y obtuvieron mejores resultados, 67,70% y 62,00% para arousal y 76,10% y 74,00% para valencia.

Desde su publicación, se han propuesto en la literatura diferentes sistemas de apren-

dizaje automático utilizando sus datos. Por ejemplo, se destaca el trabajo de [194] por su similitud con nuestra investigación. Utilizaron el time warping dinámico multidimensional como técnica no lineal para tratar la dinámica fisiológica seguido de un clasificador de apilamiento. Sus resultados alcanzaron hasta un 94,00% y un 93,60% de precisión para un modelo de tres clases emocionales independiente del sujeto, utilizando todas las señales fisiológicas de la base de datos MAHNOB y una estrategia de *CV k - fold*. Aunque intentaron disminuir el posible efecto de sesgo combinando ambas metodologías de etiquetado, mapeando el espacio dimensional de la excitación y la valencia en una emoción discreta específica, su modelo no fue capaz de capturar la diferencia entre el miedo y la ira. Este hecho es esencial para nuestro caso de uso.

Entre el resto del estado del arte basado en la base de datos MAHNOB y en lo que respecta específicamente al caso de uso de reconocimiento del miedo, el único sistema propuesto en la literatura es nuestra publicación [186]. Este es el que se detalla en los siguientes subapartados.

4.2.1 Consideraciones sobre el equilibrio de los estímulos y las etiquetas

Al igual que la base de datos DEAP, los estímulos de MAHNOB se basaron en un pool de estímulos anterior más amplio. En concreto, el estudio preliminar contiene 155 videoclips de diferentes películas [199]. Cada videoclip recibió una media de 10 anotaciones utilizando una escala Likert de 9 puntos para las dimensiones de excitación y valencia mediante el SAM y etiquetas emocionales discretas. Basándose en el acuerdo acumulativo de estas últimas, los investigadores seleccionaron hasta 14 estímulos de este estudio previo. Por ejemplo, se seleccionó el vídeo con mayor número de etiquetas de miedo para elicitarlo. Los seis vídeos restantes hasta llegar a los 20 vídeos del experimento se eligieron basándose en contenidos audiovisuales populares en línea. Así, la mayoría de los estímulos seleccionados para esta base de datos se eligieron siguiendo un criterio emocional de tipo discreto. Nótese que, como ya se ha señalado, los 20 vídeos neutros utilizados para la recuperación fisiológica fueron validados por el laboratorio de psicofisiología de Stanford.

En este contexto de etiquetado, los investigadores de MAHNOB no tuvieron en

Table 4.11: Mapeo de dimensiones discretas para el arousal y la valencia basado en [5] y adoptado por MAHNOB [12].

Arousal classes	Emotional keywords
Calm	sadness, disgust, neutral
Medium arousal	joy and happiness, amusement
Excited/Activated	surprise, fear, anger, anxiety
Valence classes	Emotional keywords
Unpleasant	fear, anger, disgust, sadness, anxiety
Neutral valence	surprise, neutral
Pleasant	joy and happiness, amusement

cuenta los aspectos de la dimensión emocional (arousal, valencia y dominancia), por lo que el conjunto de etiquetas de verdad básica generadas se basó en emociones discretas. Sin embargo, proporcionaron y utilizaron un mapeo discreto-dimensional para la excitación y la valencia basado en [5] como se muestra en la Tabla 4.11. Desgraciadamente, como el estudio preliminar de [199] no es de acceso público, no podemos realizar el mismo análisis de etiquetado exploratorio (informes de la verdad en el terreno frente a los autoinformes de los voluntarios durante el experimento) que con el DEAP. Por lo tanto, los autoinformes de los voluntarios pueden compararse con la verdad de campo sólo mediante la binarización del miedo de estos últimos basada en las etiquetas emocionales discretas. Tras realizar la binarización de las etiquetas de miedo en MAHNOB utilizando los autoinformes proporcionados para arousal, valencia y dominancia y siguiendo el mapeo de miedo propuesto en la sección 2.3.4, se analizó la distribución obtenida para todas las voluntarias consideradas, resultando asimétrica. Esto significa que la aparición de las etiquetas de miedo no fue uniforme para todas las participantes. Así, la figura 4-18 muestra que el 60% de las voluntarias informaron de más del 30% de etiquetas de miedo binarias, mientras que el resto de las voluntarias estaban por debajo de esa cantidad. Obsérvese que, en esta figura, la notación Vx se refiere al voluntario x , con $x \in 1 \dots 12$, y la notación G se refiere a la distribución original de miedo binario del experimento (el número real de estímulos destinados a provocar miedo; es decir, sólo el 20% de la cantidad total de vídeos). Esta situación desequilibrada es especialmente relevante para V11, con sólo un 5% de datos de miedo. Este análisis respalda los supuestos ya destacados en capítulos anteriores, como que la interpretación de los estímulos depende en gran medida de los voluntarios.

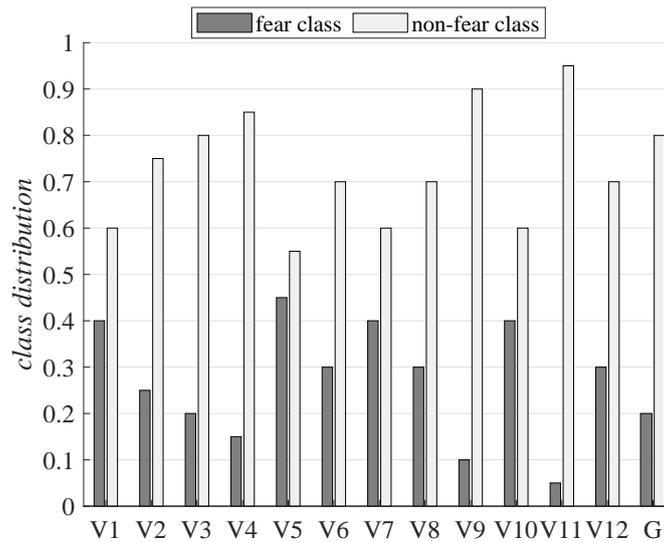


Figure 4-18: Distribución de clases para el mapeo binario del miedo sobre los autoinformes subjetivos en MANHOB para todas las diferentes voluntarias consideradas, y la distribución de clases original prevista en el experimento.

Sin embargo, en el caso de la evaluación del balance medio o de los porcentajes medios de clase de las 12 voluntarias, la relación de desequilibrio es de 1:2,6 (Clase Sin Miedo:Clase con Miedo) y los porcentajes de clase consecuentes ascienden al 72,50% y al 27,50% para las clases negativas y positivas respectivamente. Sobre esta base, se pueden obtener dos conclusiones. Por un lado, la media de clases positivas en este caso es incluso superior a la que se espera conseguir siguiendo la verdad de campo. Aunque la diferencia es inferior al 10%, este hecho debe contextualizarse con respecto a los voluntarios masculinos. Por ejemplo, los porcentajes medios de las clases de los nueve voluntarios masculinos válidos son del 81,11% y del 18,89% para las clases negativas y positivas respectivamente. Sin considerar la realización de ninguna prueba estadística para valorar si la diferencia entre hombres y mujeres es significativa, desde el punto de vista emocional y teniendo en cuenta las diferencias de procesamiento emocional de las mujeres expuestas en el apartado 2.3.3, éste podría ser uno de los factores que están influyendo. Por otro lado, el equilibrio obtenido para esta base de datos utilizando el mapeo binario de miedo desde el espacio PAD es menor que el ratio de desequilibrio observado al realizar el mapeo binario de miedo desde el espacio PA con DEAP. Esta conclusión no puede interpretarse directamente como que MAHNOB es mejor que DEAP, pero proporciona una visión de las diferencias en cuanto a la percepción o eficacia de los estímulos de ambas bases de datos, lo que está en línea con trabajos de investigación anteriores [193].

Siguiendo el mismo análisis de esquemas para esta base de datos que el aplicado a la DEAP, se evalúan las correlaciones interindividuales de las etiquetas. En este caso, los resultados obtenidos tras una prueba de Levene y una prueba de Kruskal-Wallis rechazaron la hipótesis nula de que las varianzas son iguales en todos los voluntarios ($p < 0,001$). Nótese que el conjunto de etiquetas binarizadas presenta una distribución no normal y que el nivel de significación se fijó en $p < 0,05$. Tras estos procesos, se aplican la correlación de Spearman y la prueba de independencia de Chi-cuadrado, Figuras 4-19a y 4-19b. Los resultados obtenidos se aproximan y no rechazan la hipótesis nula en promedio para cada uno de los 12 voluntarios, lo que indica que la correlación media se considera no significativa y las diferentes variables son independientes. Por lo tanto, se puede concluir que no hay pruebas suficientes para sugerir que exista una asociación entre las etiquetas binarias de miedo de los voluntarios. Además, al comparar estas gráficas con las obtenidas para el estudio de correlación e independencia de los sistemas anteriores utilizando DEAP y el binario de miedo también de PAD, se observa una mayor concordancia en este caso.

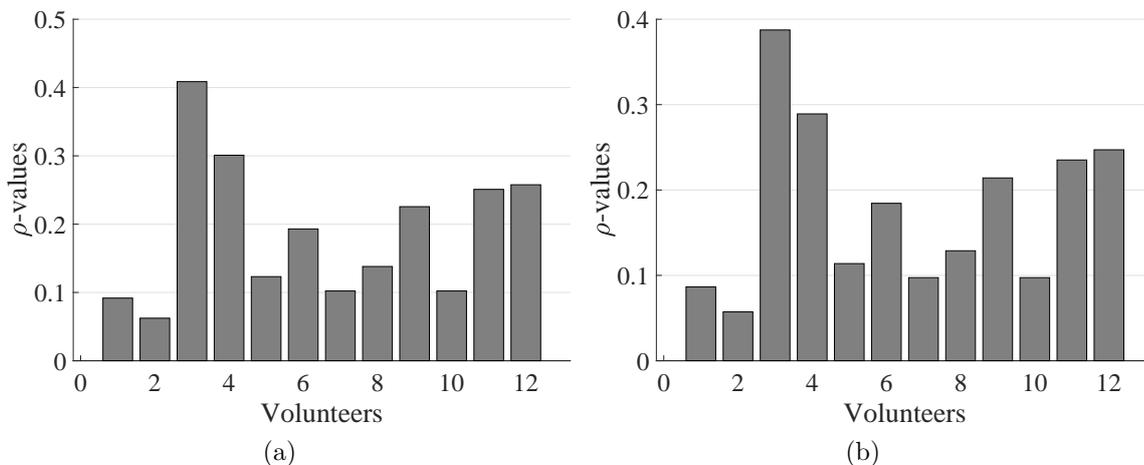


Figure 4-19: Valores p promediados para todos los voluntarios MAHNOB considerados y sus etiquetas aplicando: a) la correlación de Spearman, y b) para la prueba Chi-cuadrado de independencia. En este caso, las etiquetas se binarizan utilizando el mapeo basado en el miedo binario.

A partir del análisis aportado en este apartado, se demuestra que el mapeo binario de miedo a partir de PAD con MAHNOB es equivalente al realizado a partir de PA utilizando DEAP, lo que beneficia a los objetivos de este trabajo de investigación ya que se pueden aplicar para el actual las mismas o similares técnicas utilizadas para el paradigma de reconocimiento de emociones de este último. Hay que tener en cuenta que los diferentes resultados recogidos en el estudio de balance de estímulos y

consideración de etiquetas aportados en este apartado estuvieron siempre presentes durante el diseño de dichos sistemas.

4.2.2 Análisis exploratorio de datos, segmentación y filtrado de datos

El análisis exploratorio de datos realizado con MAHNOB siguió el mismo procedimiento que con el DEAP. Se generaron diferentes gráficos sincronizados con la metodología experimental para comprobar las recuperaciones fisiológicas o el clip neutral previo al estímulo, así como los rangos fisiológicos normales para todos los voluntarios considerados. Tras este análisis, se llegó a la conclusión de que, por término medio, los 30 segundos de datos al principio y al final del intervalo de videoclips junto con los clips neutros se comportaban realmente como se esperaba, lo que conducía a la estabilización de las señales fisiológicas y apuntaba al aislamiento emocional entre estímulos. Por lo tanto, se eliminaron los períodos de 60 segundos correspondientes a los 30 segundos anteriores y posteriores al estímulo en esta experimentación específica.

Como se indica en la sección 3.1.2, se utilizan métodos de segmentación de datos o basados en ventanas para extraer la información relacionada con la emoción en relación con los instantes de tiempo. A diferencia del sistema presentado anteriormente en este capítulo, este sistema opera sobre una base de segmentación de datos siguiendo los procedimientos típicos de segmentación de datos en la literatura [85]. En cuanto a la DSE que se enfrenta en esta etapa, se debe elegir una longitud de ventana adecuada para asegurar: que (1) la resolución de la frecuencia es suficiente para tratar todas las características basadas en la frecuencia, y (2) la longitud de cada ventana es la mínima posible para facilitar las tareas de procesamiento del host.

Para nuestro caso de uso específico y en función de las características que se van a extraer, que se describen y detallan más adelante, la distinción de frecuencias mínima requerida entre bandas es de 0,05 Hz, lo que se puede garantizar utilizando un tamaño de ventana de 20 segundos. Con esta duración de la ventana, se cumplen estas dos condiciones. Además, se emplea un solapamiento del 50%. Para seleccionar la duración óptima de la ventana y el solapamiento, hay que evaluar diferentes

consideraciones:

- Tanto el tiempo (cuanto más grande es la ventana, más largo es el procesamiento) como la complejidad computacional (cuanto más grande es el solapamiento, más operaciones se necesitan en el mismo tiempo).
- Hechos fisiológicos. Están relacionados con la naturaleza no estacionaria de estas señales, que pueden ser borrosas para ventanas muy grandes.
- Tamaño de entrenamiento del aprendizaje automático. Se refiere al número final de muestras o instancias proporcionadas después de la creación de ventanas, ya que el vector de características se extrae de cada ventana y, por tanto, el número de puntos de entrenamiento y de prueba varía en función del número de ventanas obtenidas de los datos.

En nuestro caso, se asumen algunas limitaciones fisiológicas cuando se trata de ventanas de 20 segundos. Por ejemplo, una duración de ERSCR superior a 20 segundos no puede ser capturada en una sola ventana. Obsérvese que, como se indica en la sección 2.5, los ERSCRs pueden variar entre 1 y 30 segundos, aunque la configuración inicial de un solapamiento del 50% permite una compensación equilibrada entre la cantidad de información ERSCR que se pierde y los requisitos de memoria. Basándonos en la duración de nuestras ventanas y en el solapamiento, la segmentación media por vídeo dio lugar a cinco ventanas o instancias, que tenían la misma clase o etiqueta.

En cuanto al almacenamiento de las señales adquiridas en una plataforma integrada, suponiendo, por ejemplo, una anchura máxima de 32 bits para cada punto de datos, los parámetros establecidos conducirían a una necesidad de memoria de 60 KB ($256 \text{ muestras por segundo} \times 20 \text{ s} \times 3 \text{ sensores para muestras de 32 bits}$). Este espacio de almacenamiento podría ser proporcionado por los actuales sistemas en chip que se utilizan para muchos dispositivos vestibles. No obstante, estos requisitos dependen de la aplicación y pueden modificarse y ajustarse en función de las capacidades de la plataforma integrada.

Independientemente de la longitud de la ventana, los datos se encapsulan en ranuras de tiempo fijas para ser procesados cuando se llenan, Figura 4-20. Estos datos segmentados (ventanas) obtenidos se preprocesan para eliminar el ruido y otros componentes no útiles para los siguientes pasos. Así, la calidad global de la señal se

mejora mediante filtros de denoising, centrándose en sus características fisiológicas específicas. En concreto, la señal ECG en bruto se somete a un filtro FIR de paso de banda a través de una cascada de filtrado de paso bajo y alto para facilitar la complejidad. Además, la línea de base residual se elimina utilizando una etapa de filtrado Butterworth IIR, que dio lugar a un filtro IIR de tercer orden con -6 dB a 0,5 Hz. Hay que tener en cuenta que utilizamos una transformación bilineal con prewarping de frecuencia para generar los coeficientes digitales. Después, se aplica Automatic Gain Control (AGC) para limitar la señal y mejorar la detección de picos. Para las señales GSR y SKT, se emplean filtros FIR de paso bajo para eliminar los ruidos de alta frecuencia.

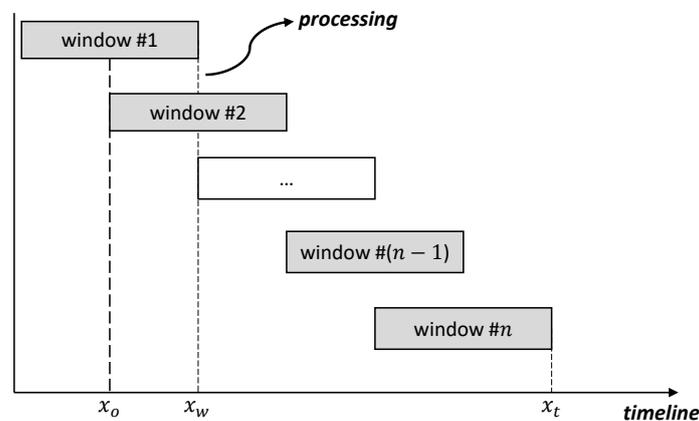


Figure 4-20: Proceso típico de segmentación de datos en los sistemas de reconocimiento de emociones basados en el aprendizaje automático.

4.2.3 Extracción de características

Para mejorar los resultados obtenidos con el DEAP-b1 y el DEAP-b2, la propuesta presentada en este capítulo considera y amplía las características de los tres grupos principales: características en el dominio del tiempo, en el dominio de la frecuencia y no lineales. Este conjunto de características comprende un total de 48 características que se detallan en las tablas 4.12, 4.13, y 4.14, para los tres sensores fisiológicos respectivamente. En concreto, se incluyen 25 características para el ECG (dos en el dominio del tiempo, nueve en el dominio de la frecuencia y 14 características no lineales), 17 características para el GSR (seis en el dominio del tiempo, tres en el dominio de la frecuencia y ocho características no lineales) y seis características para el SKT (cuatro en el dominio del tiempo y dos en el dominio de la frecuencia). Nótese que todas las características consideradas se basan en la

literatura fisiológica aceptada y conocida que trata de características relacionadas con las emociones [8, 200, 201], así como las características implementadas anteriormente (DEAP-b1 y DEAP-b2). Además, para este sistema, hemos aumentado considerablemente las características no lineales consideradas en nuestro modelo, que se basan en publicaciones recientes que incluían también estas métricas sintéticas en los sistemas de reconocimiento de emociones [160, 161]. En las siguientes subsecciones se detallan las diferentes características específicas extraídas para los tres dominios y señales diferentes.

Table 4.12: Características extraídas para la señal ECG y la propuesta de reconocimiento binario de emociones de miedo utilizando el conjunto de datos MAH-NOB.

Sensor	Domain	Features
ECG (25)	Time-domain:	Mean of Inter-Beat-Interval
	(2)	Heart rate variability
	Frequency-domain:	Power spectral density of four bands
	(9)	(0–0.1 Hz, 0.1–0.2 Hz, 0.2–0.3 Hz and 0.3–0.4 Hz)
		Inter-Beat-Interval Power spectral density for
		Low frequency (LF) (<0.08 Hz)
		Medium frequency (MF) (0.08–0.15 Hz)
		High frequency (HF) (0.15–0.5 Hz)
		Total energy ratio for MF
		Spectral density ratio between
		LF and HF band
	Non-linear:	Multiscale entropy at five levels
	(14)	Detrended fluctuation for filtered data
		Detrended fluctuation for Inter-Beat-Interval
		Recurrence rate
		Determinism
		Laminarity
		Longest RP diagonal line
		Diagonal lines entropy
		Trapping time
		Correlation dimension

Table 4.13: Características extraídas para la señal GSR y la propuesta de reconocimiento binario de emociones de miedo utilizando el conjunto de datos MAHNOB.

Sensor	Domain	Features	
GSR (17)	Time-domain: (6)	Filtered data mean value	
		ERSCR including number of peaks	
		ERSCR Amplitude and rise time	
		Standard deviation	
		First quartile	
		Third quartile	
	Frequency-domain: (3)	Power spectral density of two bands for SCL and SCR components (0–0.05 Hz, 0.05–1.5 Hz)	
		Spectral density ratio for 0–0.05 Hz	
		Non-linear: (8)	Detrended fluctuation for filtered data
			Recurrence rate
			Determinism
			Laminarity
			Longest RP diagonal line
			Diagonal lines entropy
			Trapping time
			Correlation dimension

Table 4.14: Características extraídas para la señal SKT y la propuesta de reconocimiento binario de emociones de miedo utilizando el conjunto de datos MAHNOB.

Sensor	Domain	Features
SKT (6)	Time-domain: (4)	Filtered data mean value
		Standard deviation
		Skewness
		Kurtosis
	Frequency-domain: (2)	Power spectral density of two bands (0–0.1 Hz, 0.1–0.2 Hz)

Antes del proceso de extracción de características, tienen lugar las tareas de delimitación fisiológica. Para este sistema, la señal de ECG sin procesar se somete a la identificación de picos para determinar el IBI y extraer una estimación válida de la

frecuencia cardíaca y los parámetros relacionados con la variabilidad de la misma. En concreto, se aplicó un detector de picos ECG basado en el algoritmo desarrollado por Pan y Tompkins en [202]. La figura 4-21 muestra la arquitectura de dicho algoritmo, que se alimenta de la señal ECG filtrada. Las diferentes etapas se describen como sigue:

- Diferenciador. Suele concebirse como un filtro derivador que se encarga de proporcionar información sobre la pendiente del patrón de onda morfológica del ECG. También atenúa los componentes de baja frecuencia, que se refieren a la despolarización auricular y a la repolarización ventricular. Para nuestro caso, este proceso se realiza a partir de la primera diferencia de la señal de ECG filtrada de entrada.
- Elevación al cuadrado. Es una operación no lineal que enfatiza los picos del ECG amplificando el resultado de la derivada anterior.
- Elemento Integrador. Como la salida de la derivada cuadrada puede presentar múltiples picos dentro de la duración de un solo periodo de ECG, se utiliza un filtro de integración de ventana móvil para suavizar dicha señal. La anchura de este filtro suele ser de 150 ms. La señal de salida de este proceso se conoce como señal integrada.
- Comprobación del umbral y búsqueda de la señal. Estos últimos procedimientos tienen por objeto identificar y corroborar la correcta localización de los picos locales dentro de la señal integrada. Se aplican diferentes restricciones fisiológicas para asegurar la detección fisiológica de los picos del ECG, como el tiempo de bloqueo de 200 ms entre los picos identificados. Una vez identificados todos los picos de la señal integrada, se aplica un proceso de búsqueda de vuelta para descartar y corregir aquellos picos o intervalos RR que causen problemas potenciales. Por ejemplo, en nuestro caso, realizamos una doble iteración buscando cinco picos por delante y evaluando la evolución mediana del vector pico a pico.

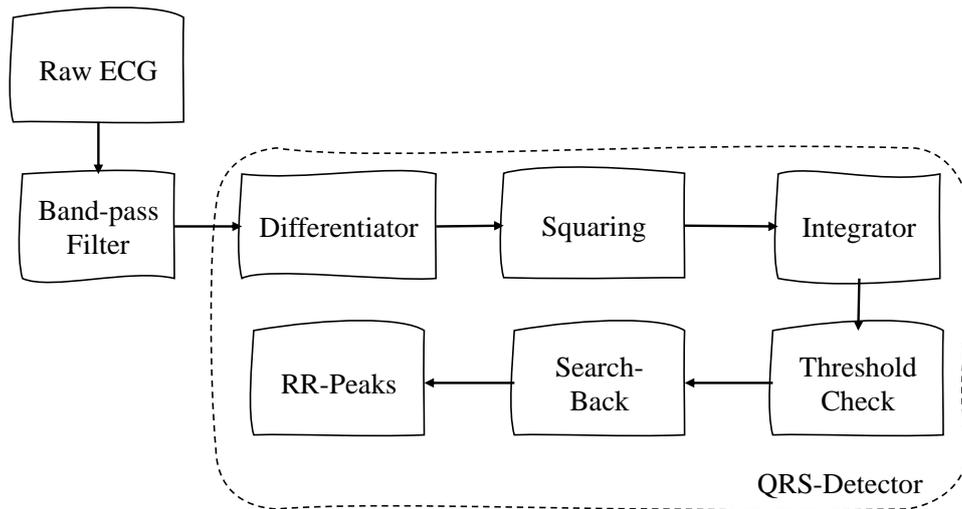


Figure 4-21: Esquema de la arquitectura del algoritmo de identificación de picos ECG aplicado en este trabajo.

Para la señal GSR, el FIR aplicado está diseñado para preservar la información por debajo de 1,5 Hz, que es la frecuencia máxima de la actividad SCR. Este filtro también se utiliza con la señal SKT para aprovechar el almacenamiento de un solo conjunto de coeficientes de filtro. En cuanto a la delineación de GSR, aplicamos los mismos procesos que se realizan con la base de datos DEAP. Así, se utiliza una combinación lineal seguida de la ecuación 2.7, a través de la cual se obtiene la tendencia de la señal GSR (SCL) mediante un filtro de mediana móvil con una ventana deslizante de cuatro segundos. Esa salida se resta a la señal GSR filtrada, obteniendo la componente SCR. Ambos componentes, así como la señal GSR filtrada, se utilizan para extraer las métricas o características sintéticas que se detallan en los siguientes apartados.

4.2.3.1 Dominio del tiempo y de la frecuencia

Para las características del dominio del tiempo que se van a extraer en este sistema, siguen la misma distinción o agrupación que las presentadas y detalladas en la sección 4.1.3, ya que se pueden dividir entre características estadísticas de orden superior y morfológicas. Sin embargo, el sistema propuesto en esta sección amplía las características estadísticas de orden superior en dos métricas adicionales: *skewness* y *kurtosis*. En concreto, se aplican a la señal SKT. Por un lado, la primera se refiere como un indicativo de la asimetría, positiva o negativa, que se desvía de una

distribución normal y viene dada por

$$s = \frac{1}{N} \sum_{i=1}^N \frac{(x_i - \bar{x})^3}{\sigma^3}, \quad (4.17)$$

donde x es el SKT filtrado con N muestras en este caso, y \bar{x} y σ son la media y la desviación estándar para la ventana de procesamiento actual. Por otra parte, *kurtosis* es la métrica estadística relacionada con la forma de una distribución de probabilidad mediante la medición del grado de concentración que se presenta alrededor de la media de la distribución de frecuencias para una variable aleatoria de valor real, también descrita como la medida de la cola. Esta medida estadística de orden superior viene dada por,

$$k = \frac{1}{N} \sum_{i=1}^N \frac{(x_i - \bar{x})^4}{\sigma^4}. \quad (4.18)$$

Estos momentos estadísticos permiten caracterizar la distribución temporal del SKT a lo largo del estímulo visualizado.

4.2.3.2 Dominio No-lineal

Para este sistema, el conjunto de características no lineales se amplía añadiendo hasta ocho nuevas características. La mayoría de ellas se basan en la teoría del caos y en técnicas de análisis de series temporales. Se describen y detallan a continuación.

- Detrended Fluctuation Analysis (DFA). Se trata de una potente técnica que puede aplicarse si se sospecha o se sabe que existen señales no estacionarias. Permite estimar el escalamiento de la ley de potencia (fractal) o el exponente de Hurst de una señal procedente de un sistema expuesto a dichas no estacionarias [203]. De hecho, en este caso, esta métrica proporciona una medida relativa a la autosimilitud fisiológica a diferentes resoluciones (tamaños de ventana), que puede traducirse en una evaluación de la complejidad fisiológica. Así, la serie temporal de longitud N se integra primero, y , y se encapsula en cajas o ventanas de longitud n . Estos segmentos no superpuestos se ajustan a un polinomio del que se obtiene la tendencia local y_n . Por último, la serie temporal integrada se detrae restando dicha tendencia local. La fluctuación

media cuadrática $F(n)$ viene dada por la siguiente ecuación:

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^N [y(k) - y_n(k)]^2}, \quad (4.19)$$

que se repite para todos los tamaños de ventana a evaluar. Obsérvese que en nuestro caso el ajuste polinómico es lineal (primer orden) y el número de tamaños de ventana en los que evaluar las fluctuaciones se establece empíricamente que $n \in t_w/10 \dots t_w$ con 10 pasos de muestra, donde t_w es el tamaño de muestra de la ventana de procesamiento.

- Tasa de recurrencia. Ésta y las siguientes características se basan en la interpretación matemática de los Recurrence Plot (RP)s. Estos se conceptualizan como gráficos bidimensionales en los que se pueden representar y cuantificar los estados de la trayectoria del espacio de fase de un sistema dinámico. Dichos estados se denominan las recurrencias que el sistema o señal presenta a lo largo de una determinada ventana de procesamiento temporal. Dicha representación bidimensional se obtiene mediante el cálculo de las distancias entre dos estados y la comparación con respecto a un umbral predefinido, siguiendo

$$R_{i,j} = \Theta(\epsilon_i - \|\vec{x}_i - \vec{x}_j\|), \quad \vec{x}_i \in R^m, \quad i, j = 1, \dots, N, \quad (4.20)$$

donde i y j son dos estados arbitrarios, ϵ_i es el umbral utilizado para la evaluación de la recurrencia, \vec{x}_i y \vec{x}_j son los módulos respectivos para cada estado, y m es la dimensión incrustada a considerar. Obsérvese que la separación t entre los espacios i y j puede ajustarse como se desee y sea necesario también. Para nuestro sistema, estimamos t y m utilizando la información mutua [204] y el falso vecino más cercano [205], respectivamente, y definimos ϵ como el 10% del diámetro promedio del espacio de fase de las observaciones [206]. Una vez obtenido el RP, la relación de la tasa de recurrencia puede derivarse de

$$RR = \frac{1}{N^2} \sum_{i,j=1}^N R_{i,j}, \quad (4.21)$$

que corresponde a la suma de correlación y cuantifica la cantidad de estados de recurrencia detectados.

- Dimensión de correlación. Esta técnica se utiliza habitualmente en el análisis de series temporales para caracterizar el atractor de un sistema dinámico, es decir, en este caso para medir la complejidad de un sistema fisiológico. Una dimensión de correlación aproximada o $D2$ puede calcularse como

$$D2 \approx \frac{\log(RR)}{\log(\epsilon)}. \quad (4.22)$$

- Determinismo. A partir del gráfico RP, las líneas diagonales proporcionan información sobre los patrones fisiológicos repetitivos de la serie temporal analizada. Esto se cuantifica mediante la siguiente ecuación:

$$DET = \frac{\sum_{l=l_{min}}^N LD(l)}{\sum_{i,j=1}^N R_{i,j}} = \frac{\sum_{l=l_{min}}^N LD(l)}{\sum_{l=1}^N LD(l)}, \quad (4.23)$$

donde $D(l)$ es el histograma de las diferentes longitudes de las líneas diagonales. Obsérvese que debe proporcionarse un parámetro mínimo de línea diagonal l_{min} , que en nuestro caso se fija empíricamente en 2.

- Laminaridad. Esta característica cuenta el porcentaje de puntos de recurrencia que forman líneas verticales dentro de la RP. Estos se denominan estados caóticos del sistema o no periódicos. Viene dada por

$$LAM = \frac{\sum_{l=l_{min}}^N lV(l)}{\sum_{i,j=1}^N R_{i,j}} = \frac{\sum_{l=l_{min}}^N lV(l)}{\sum_{l=1}^N lVl}, \quad (4.24)$$

donde $V(l)$ es el histograma de las diferentes longitudes de líneas verticales.

- Línea diagonal RP más larga. La cuantificación de la línea diagonal más larga dentro de la gráfica RP permite caracterizar la cantidad máxima de tiempo periódico dentro del sistema. Esto viene dado por

$$L_{max} = \max(l_i; i = 1, \dots, N_l), \quad (4.25)$$

donde N_l se refiere al número total de líneas diagonales dentro de la parcela RP. En nuestro caso, la implementación se realiza empleando un algoritmo de clasificación rápida utilizando las líneas diagonales previamente identificadas.

- Tiempo de captura. Como la característica anterior pretende caracterizar la

periodicidad de la señal, el tiempo de atrapamiento proporciona información sobre la cantidad de estados no estacionarios que se repiten dentro de la parcela RP. Se calcula como sigue:

$$TT = \frac{\sum_{l=l_{min}}^N lV(l)}{\sum_{l=l_{min}}^N l}. \quad (4.26)$$

- Entropía de las líneas diagonales. Finalmente, para considerar la incertidumbre de la periodicidad de la señal, se aplica la entropía de Shannon a la distribución de probabilidad de las longitudes de las líneas diagonales $p(l)$. Esto se calcula como:

$$ENT_R = - \sum_{l=l_{min}}^N p(l) \log(p(l)). \quad (4.27)$$

4.2.4 Sistemas de clasificación del miedo

En los siguientes apartados se detallan y explican los resultados obtenidos con el sistema propuesto utilizando MAHNOB. En concreto, se presentan dos sistemas: dependiente e independiente del sujeto. Ambos utilizan el mapeo binario de miedo del espacio PAD analizado en la sección 4.2.1. Además, a diferencia de los sistemas DEAP anteriores, cabe destacar algunas particularidades. En primer lugar, se aplican tres clasificadores diferentes. Dos de ellos son los mismos clasificadores empleados para los sistemas DEAP, SVM y KNN. El tercer clasificador es, en realidad, un conjunto de clasificadores que siguen un enfoque de aprendizaje ensemble. Para ello, se utiliza un algoritmo AdaBoost. Nótese que este último también ha sido revisado en la sección 3.1.7.

En este caso, y de cara a una mejor sintonización, la optimización de los hiperparámetros se realiza mediante una optimización bayesiana. Esta técnica pretende minimizar la tasa de error de clasificación a lo largo de las iteraciones, apoyándose en una estrategia CV. En concreto, se incluye una técnica SMBO. Así, la generación de nuevos hiperparámetros a evaluar se somete a procesos gaussianos, que aproximan la distribución de la función de coste $f(x) \sim GP$ (Gaussian Process). Esta distribución se actualiza a medida que se itera con los nuevos valores conocidos para los nuevos hiperparámetros. De este modo, se construye la función de distribución final $p(f(x)|f(x^*))$ donde x^* se refiere a los valores históricos. Con esta estimación, en el siguiente paso se calcula el punto que podría ser un candidato potencial. Para

ello, se utiliza una función $\alpha(\cdot)$ llamada de adquisición. Para la definición de esta función de adquisición, existen diferentes opciones. En este caso, se ha utilizado la estrategia de probabilidad de mejora, que trata de estimar la probabilidad de una mejora con la siguiente muestra.

En cuanto al procedimiento de validación, se han validado los modelos sujeto-dependiente y sujeto-independiente en base a un esquema estratificado de $k - fold$ CV ($k = 5$). Por un lado, para los modelos dependientes del sujeto, se calculó la media de todas las métricas para todos los voluntarios y la desviación media absoluta (DAM) a partir de los valores de CV obtenidos. Por otro lado, los modelos independientes del sujeto se dividieron en conjuntos de entrenamiento, validación y prueba, empleando una estrategia LOSO. Esta última nos permitió estudiar el rendimiento de varios sistemas independientes del sujeto entrenados con diferentes combinaciones de sujetos y probados con un único voluntario sobre el que el sistema no tenía información.

Los resultados presentados y el sistema fueron publicados en [186]. Nótese que en este caso no se implementó un proceso de selección o reducción de características ni se aplicó ninguna metodología de coste de clasificación errónea (aprendizaje sensible al coste). La justificación de esta última decisión se basó en la obtención de resultados de referencia para compararlos con futuras mejoras del sistema al añadir y aplicar diferentes técnicas.

4.2.4.1 Modelos dependientes de la voluntaria

La tabla 4.15 muestra las métricas de rendimiento de validación y la dispersión para los diferentes algoritmos de clasificación ligeros seleccionados para la generación de cada modelo dependiente del sujeto para todos los voluntarios. Tras analizar los resultados, se observa que no existe una relación de dependencia estricta entre la distribución de clases y el rendimiento. No obstante, el rendimiento de los modelos se vio directamente afectado por el tipo de clasificador utilizado. Además, otro factor clave que podría haber influido en el rendimiento estaba relacionado con la alineación de los patrones fisiológicos dependientes del sujeto y las etiquetas binarias de mapas de miedo obtenidas. Además, el uso de las puntuaciones Gmean y F1 nos permitió distinguir con mayor solidez los modelos de bajo rendimiento de los de mayor rendimiento.

Table 4.15: Métricas de rendimiento para cada modelo generado en función del sujeto y métricas de rendimiento medio y dispersión para cada algoritmo de clasificación.

Training Type	Trained Volunteers	SVM				KNN				ENS			
		ACC (MAD)	AUC (MAD)	Gmean (MAD)	F1 (MAD)	ACC (MAD)	AUC (MAD)	Gmean (MAD)	F1 (MAD)	ACC (MAD)	AUC (MAD)	Gmean (MAD)	F1 (MAD)
Subject dependent	V1	89.00%	90.30%	87.73%	85.71%	88.00%	88.67%	87.90%	85.37%	88.00%	79.32%	85.12%	83.33%
	V2	88.00%	92.43%	76.41%	71.43%	99.00%	99.89%	99.23%	87.72%	91.00%	97.47%	85.41%	80.85%
	V3	91.00%	94.44%	71.20%	74.29%	94.00%	96.19%	90.47%	85.00%	97.00%	95.31%	98.13%	93.02%
	V4	93.00%	95.29%	84.06%	75.86%	99.00%	96.67%	96.59%	96.55%	96.00%	99.69%	85.62%	84.62%
	V5	76.00%	84.97%	75.01%	72.09%	81.00%	91.47%	85.62%	84.62%	98.00%	99.88%	97.95%	97.78%
	V6	90.00%	93.67%	87.92%	83.33%	98.00%	98.60%	84.08%	82.86%	99.00%	99.90%	99.23%	98.36%
	V7	93.00%	98.54%	92.47%	91.14%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
	V8	85.00%	90.57%	81.22%	74.58%	94.00%	92.24%	90.58%	89.29%	93.00%	86.05%	92.12%	88.52%
	V9	96.00%	98.44%	83.16%	77.78%	99.00%	99.44%	99.40%	95.24%	100.00%	100.00%	100.00%	100.00%
	V10	89.00%	91.31%	87.73%	85.71%	94.00%	94.15%	93.24%	92.31%	100.00%	100.00%	100.00%	100.00%
	V11	95.00%	50.00%	00.00%	00.00%	99.00%	90.00%	89.44%	88.89%	100.00%	100.00%	100.00%	100.00%
	V12	77.00%	83.48%	62.91%	53.06%	91.00%	85.95%	84.97%	83.02%	94.00%	93.33%	90.58%	89.29%
		88.50% (4.66%)	88.62% (7.90%)	74.15% (14.72%)	70.42% 14.62%	94.66% (4.33%)	94.44% (4.02%)	91.80% (4.92%)	89.24% (4.53%)	96.33% (3.28%)	95.91% (4.94%)	95.51% (5.62%)	92.98% (6.38%)

Sin embargo, los resultados presentados podrían estar sesgados debido a la reducida cantidad de datos disponibles (100 muestras por voluntario, cinco ventanas de media por vídeo), así como por la asimetría detectada (datos desequilibrados). Centrándonos en la asimetría, este problema es especialmente relevante en V11. El efecto en el rendimiento debido a la asimetría para este voluntario se muestra en la Figura 4-22, que proporciona las matrices de confusión para V11 después de aplicar los tres algoritmos. Por el contrario, las matrices de confusión del voluntario V7 también se muestran en la Figura 4-23. Este voluntario mostró el mejor rendimiento en general, es decir, teniendo en cuenta las diferentes métricas para los tres clasificadores aplicados. En estas figuras, la clase positiva (miedo) está representada por el número dos, y la clase negativa (sin miedo) está representada por el número uno. Las filas corresponden a la clase predicha y las columnas a la clase verdadera o verdad fundamental. De izquierda a derecha y de arriba a abajo, cada matriz de confusión muestra los índices de verdaderos negativos, falsos positivos y falsas omisiones. La siguiente fila muestra los falsos negativos, los verdaderos positivos y la tasa de precisión. La última fila muestra la tasa de falsos negativos, la especificidad y la precisión global. El resto de las matrices de confusión para cada modelo dependiente del sujeto generado se muestran en [186].

Tras analizar estas matrices de confusión, también se comprobó que el rendimiento de los algoritmos para V11 era asimétrico. Así, por ejemplo, SVM proporcionó una alta precisión, de hasta el 95,00%, pero esta métrica estaba sesgada por el reducido número de muestras de este voluntario dentro de la clase positiva (sólo cinco muestras). En este caso, los resultados calculados de las métricas Gmean y F1 fueron del 0,00% debido a la tasa de predicción positiva nula, y el AUC fue del 50,00%, lo que demuestra que este modelo de clasificación no se comportó mejor que la adivinación aleatoria. El comportamiento mostrado por el SVM en este caso coincidió con el rendimiento poco fiable habitual de este algoritmo para distribuciones extremadamente desequilibradas; es decir, el SVM se orienta hacia la clase mayoritaria para optimizar la tasa de error durante la etapa de entrenamiento. Por el contrario, los algoritmos de boosting suelen ofrecer un mejor comportamiento para las distribuciones desequilibradas, tal y como muestra el ENS para este caso. No obstante, esta situación de desequilibrio debería evitarse durante la generación de la base de datos,

		SVM V11			KNN V11		
Predicted Class	1	95 95.0%	5 5.0%	95.0% 1.0%	95 95.0%	1 1.0%	99.0% 1.0%
	2	0 0.0%	0 0.0%	0% 0.0%	0 0.0%	4 4.0%	100% 0.0%
		100% 0.0%	0.0% 100.0%	95.0% 5.0%	100% 0.0%	80.0% 20.0%	99.0% 1.0%
		Ground Truth			Ground Truth		

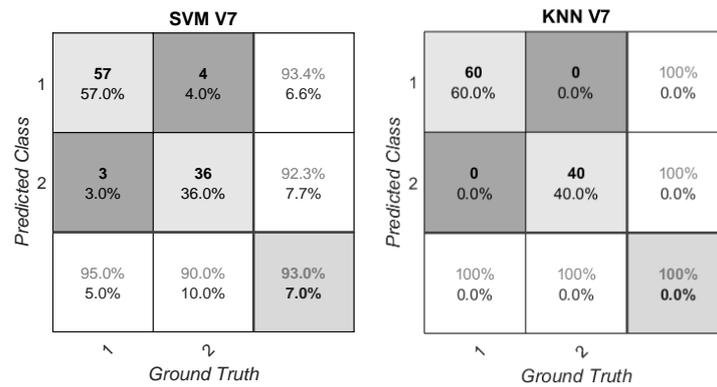
(a) SVM classifier

(b) KNN classifier

		ENS V11		
Predicted Class	1	95 95.0%	0 0.0%	100% 0.0%
	2	0 0.0%	5 5.0%	100% 0.0%
		100% 0.0%	100% 0.0%	100% 0.0%
		Ground Truth		

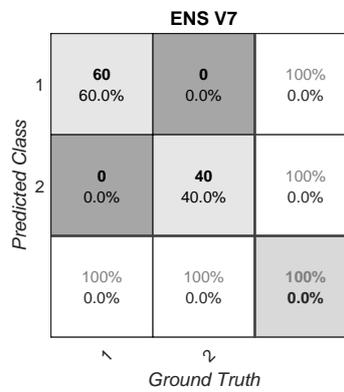
(c) ENS classifier

Figure 4-22: Matrices de confusión para un modelo dependiente del sujeto en V11, detectado como un problema de asimetría.



(a) SVM classifier

(b) KNN classifier



(c) ENS classifier

Figure 4-23: Matrices de confusión para un modelo dependiente del sujeto en V7.

y debería mejorarse la calidad y la diversidad de los estímulos considerados. En el caso de que esta situación no se abordara durante la generación de la base de datos, el sesgo generado en el rendimiento podría resolverse parcialmente seleccionando una técnica de clasificación adecuada, como se ha comentado anteriormente. Sin embargo, la falta de información de una de las dos clases no puede resolverse, lo que resulta en una posible clasificación incorrecta para las futuras muestras [207]. Otro posible enfoque para tratar este problema se basa en la aplicación de técnicas de aumento de datos o de clases ponderadas, como se aplicó anteriormente para el DEAP-b1 y el DEAP-b2. Por el contrario, en el caso de V7, el sistema mostró un 40,00% de información de clase positiva, lo que se traduce en un mejor rendimiento del SVM. KNN y ENS siguieron superando a SVM debido a las razones expuestas anteriormente para la optimización de la tasa de error de este clasificador.

4.2.4.2 Modelos independientes de la voluntaria

Centrándonos en el caso de uso independiente del sujeto, la combinación de todas las muestras individuales dio como resultado un conjunto de datos más grande con 1200 muestras (100 muestras por voluntario \times 12 voluntarios). Los rangos de las señales fisiológicas difieren para los distintos individuos debido a la naturaleza de cada uno y a las diferencias en la configuración de la medición (por ejemplo, la temperatura ambiente). Por lo tanto, los datos (características) de cada voluntario deben ser normalizados. Para ello, consideramos el método de la puntuación Z. Una vez normalizada la base de datos, se generó el sistema de reconocimiento de miedo binario utilizando un esquema $k - fold$ CV para la partición de validación y una metodología de prueba LOSO.

La tabla 4.16 muestra las métricas de rendimiento para cada algoritmo de clasificación en la generación del modelo independiente del sujeto. Obsérvese que el entrenamiento de estos modelos se realizó utilizando todos los voluntarios excepto el utilizado para la prueba en cada iteración (datos de prueba no vistos); es decir, se generaron y probaron un total de 12 modelos independientes del sujeto.

Table 4.16: Métricas de rendimiento para cada modelo generado independiente del sujeto y métricas de rendimiento medio y dispersión para cada algoritmo de clasificación. La etapa de entrenamiento se realiza utilizando todos los voluntarios excepto el voluntario probado en cada modelo generado (datos de prueba no vistos).

Training Type	Tested Volunteers	SVM				KNN				ENS			
		ACC (MAD)	AUC (MAD)	Gmean (MAD)	F1 (MAD)	ACC (MAD)	AUC (MAD)	Gmean (MAD)	F1 (MAD)	ACC (MAD)	AUC (MAD)	Gmean (MAD)	F1 (MAD)
Subject independent	V1	65.00%	60.83%	57.15%	47.76%	75.00%	71.25%	68.74%	62.69%	71.00%	65.83%	60.55%	52.46%
	V2	70.00%	61.33%	58.83%	42.31%	81.00%	74.00%	72.66%	61.22%	82.00%	86.72%	71.26%	60.87%
	V3	64.00%	66.00%	62.44%	40.00%	72.00%	61.88%	59.53%	39.13%	62.00%	61.19%	45.82%	24.00%
	V4	82.00%	71.01%	83.88%	59.09%	84.00%	87.84%	87.67%	63.64%	85.00%	91.61%	90.75%	66.67%
	V5	64.00%	70.55%	61.10%	55.00%	70.00%	71.74%	65.32%	59.46%	73.00%	75.58%	68.16%	63.01%
	V6	84.00%	88.57%	85.61%	77.14%	71.00%	68.81%	68.59%	56.72%	79.00%	87.86%	76.16%	66.67%
	V7	75.00%	90.54%	65.38%	59.02%	76.00%	91.83%	69.37%	63.63%	87.00%	99.46%	82.16%	80.60%
	V8	76.00%	81.90%	70.51%	60.00%	78.00%	72.86%	71.71%	62.07%	80.00%	85.00%	75.59%	66.67%
	V9	67.00%	69.67%	63.77%	21.82%	67.00%	59.44%	58.69%	18.87%	78.00%	84.78%	78.88%	42.11%
	V10	76.00%	79.63%	65.95%	60.00%	78.00%	72.92%	68.34%	63.33%	77.00%	82.30%	72.80%	67.61%
	V11	74.00%	90.53%	76.78%	23.53%	80.00%	89.47%	88.85%	40.00%	74.00%	86.32%	85.22%	27.78%
	V12	70.00%	72.05%	64.14%	51.61%	71.00%	66.90%	66.12%	53.97%	72.00%	67.72%	66.73%	54.84%
		72.25% (5.58)	75.22% (9.18)	67.96% (7.48)	49.77% (12.24)	75.25% (4.25)	74.07% (7.82)	70.47% (6.50)	53.73% (10.53)	76.67% (5.22)	81.20% (9.07)	72.84% (8.62)	56.11% (13.22)

Después de analizar esta tabla, los mejores resultados también los proporcionó ENS, con las métricas de rendimiento promedio más altas (81,20%, 72,84%, 56,11%) para el AUC, Gmean y la puntuación F1. Por el contrario, la SVM también proporcionó el peor rendimiento en general. Cabe destacar las diferencias entre todos los modelos independientes del sujeto generados. Por ejemplo, el mejor modelo alcanzó una Gmean de hasta el 90,75% cuando se probó con V4 y se entrenó con el resto de los voluntarios, y el peor modelo proporcionó una Gmean de hasta el 45,82% cuando se probó con V3 y se entrenó con el resto de los voluntarios. Este hecho pone de manifiesto la necesidad de contar con un conjunto de datos más amplio y equilibrado para hacer frente a estos problemas. En cuanto a la puntuación F1, se observa una gran variabilidad entre los distintos modelos. Por definición, esta puntuación es una media armónica ponderada entre la precisión y la recuperación, que deja fuera de la ecuación a los verdaderos negativos. Este hecho es clave cuando se presenta una incidencia positiva muy baja, pero una puntuación F1 alta no implica necesariamente un mejor rendimiento del sistema. Por ejemplo, las matrices de confusión de dos modelos probados independientes del sujeto para los clasificadores ENS se muestran en la Figura 4-24 para V4 y V7 con puntuaciones F1 de hasta 66,67% y 80,60% respectivamente. Basado en la aplicación de reconocimiento de miedo perseguido, podría ser más conveniente tener una clasificación errónea para el falso-positivo que sobre el falso-negativo. Por lo tanto, la comparación de la puntuación F1 para diferentes modelos independientes del sujeto debe ir acompañada de los requisitos y necesidades de la aplicación. Nótese que los dos ejemplos explicados no mostraron un rendimiento de clasificación perfecto. El resto de las matrices de confusión para cada modelo independiente del sujeto generado se proporcionan en [186].

		ENS / SI Tested with V4			ENS / SI Tested with V7		
Predicted Class	1	70 70.0%	0 0.0%	100.0% 0.0%	60 60.0%	13 13.0%	82.2% 17.8%
	2	15 15.0%	15 15.0%	50.0% 50.0%	0 0.0%	27 27.0%	100.0% 0.0%
		82.4% 17.6%	100.0% 0.0%	85.0% 15.0%	100.0% 0.0%	67.5% 32.5%	87.0% 13.0%
		Ground Truth			Ground Truth		
		(a)		(b)			

Figure 4-24: Matrices de confusión para los clasificadores ENS y los voluntarios probados (datos no vistos) sobre sus respectivos modelos independientes del sujeto: (a) probado V4, (b) probado V7.

En cuanto a la complejidad temporal y espacial de los modelos empleados, SVM y KNN ya fueron discutidos en las secciones anteriores. El modelo restante, ENS, es en realidad un clasificador AdaBoost, que se basa en un único aprendiz fuerte compuesto. Este último está formado por diferentes aprendices débiles que, en este caso, son árboles poco profundos. Así, dos parámetros son esenciales para estimar la complejidad temporal y espacial: el número de árboles y el número máximo de divisiones por árbol. Por un lado, la complejidad temporal solía definirse por $\mathcal{O}(feats * n_{trees})$ para este tipo de clasificador, donde $feats$ es el número de características y n_{trees} es el número total de árboles. Nótese que la complejidad temporal de los árboles no se incluye dentro de la complejidad temporal total de AdaBoost, ya que es insignificante en comparación con el tiempo total. Por otro lado, la complejidad espacial viene determinada por la cantidad de árboles superficiales entrenados y el número máximo de divisiones permitidas dentro de cada uno de ellos. Además, también hay que almacenar los pesos entrenados para los aprendices débiles. En el peor de los casos, cuando se trata de modelos independientes del sujeto, que son más complejos que los dependientes del sujeto, el número de árboles utilizados es una cuarta parte del conjunto de datos de entrenamiento, es decir, aproximadamente 300 árboles de media (1.200 instancias de entrenamiento), y el número máximo de divisiones permitidas por aprendiz débil es de media diez. Considerando estos valores y el conjunto total de 48 características, la complejidad temporal estimada alcanza hasta 14400

operaciones mientras que la complejidad espacial alcanza aproximadamente 13 kB (300 árboles \times 10 splits máximo + 300 pesos entrenados). Nótese que la memoria utilizada en kB se basa en un tipo de datos enteros de 32 bits para todos los parámetros a almacenar.

Cabe destacar que pueden implementarse y aplicarse otras optimizaciones algorítmicas de AdaBoost y del árbol, así como otras alternativas de selección de características, para lograr una menor complejidad temporal, que incluso puede conducir en algunos casos a un mejor rendimiento de reconocimiento. Aunque la información redundante no afecta tan negativamente a AdaBoost como a otros clasificadores, como SVM, la eliminación de la información irrelevante sí afecta y puede suponer un menor tiempo computacional. De hecho, en la Tabla 4.17 se muestran los resultados para la misma configuración de entrenamiento, validación y prueba que el caso independiente del sujeto empleando el clasificador AdaBoost, pero cambiando el método de selección de características a mrMR con $K = 10$. Nótese que esta técnica también se aplicó al sistema DEAP-b2. Se puede observar que las métricas obtenidas son similares a las presentadas sin selección de características. Sin embargo, la reducción de 48 a 10 características afecta directamente a la complejidad del tiempo de inferencia. Así, con esta configuración y considerando el mismo número de árboles en promedio, sólo es necesario realizar 3000 operaciones.

4.3 Discusión y Conclusión

En este capítulo se ha presentado el trabajo realizado para crear un sistema de detección de miedo utilizando conjuntos de datos disponibles públicamente. Parte del trabajo presentado también está contenido en artículos publicados [181, 183, 186]. A lo largo del diseño de los diferentes sistemas, se han identificado y abordado inicialmente los procesos esenciales que deben considerarse críticos para una implementación embebida. Por ejemplo, el principal foco de discusión es la complejidad temporal y espacial de los modelos resultantes en comparación con sus métricas de rendimiento. Como se indicó al principio del capítulo, el diseño de un modelo totalmente independiente del sujeto permitiría la primera generación de una herramienta tecnológica capaz de detectar cualquier emoción basada en el aprendizaje automático. Esta herramienta puede personalizarse durante la operación con los

Table 4.17: Métricas de rendimiento para cada modelo generado independiente del sujeto y métricas de rendimiento medio y dispersión para ENS tras la selección de características mrMR. La etapa de entrenamiento se realiza utilizando todos los voluntarios excepto el voluntario probado en cada modelo generado (datos de prueba no vistos).

Training Type	Tested Volunteers	ENS			
		ACC (MAD)	AUC (MAD)	Gmean (MAD)	F1 (MAD)
Subject independent (mrMR)	V1	90.00%	68.42%	59.37%	50.70%
	V2	87.27%	87.25%	76.18%	64.15%
	V3	87.73%	65.63%	45.00%	23.08%
	V4	88.64%	84.55%	84.48%	57.14%
	V5	90.45%	86.83%	61.50%	54.55%
	V6	92.27%	90.57%	85.22%	78.13%
	V7	87.27%	95.71%	78.58%	87.02%
	V8	90.91%	94.90%	92.38%	87.50%
	V9	88.18%	91.56%	81.10%	47.06%
	V10	85.00%	97.17%	79.06%	88.89%
	V11	85.00%	99.79%	88.85%	33.33%
	V12	93.64%	69.62%	53.67%	39.29%
		88.86% (2.16)	86.00% (9.30)	73.78% (12.60)	59.24% (18.25)

datos recogidos del sujeto. Por ejemplo, en el equipo UC3M4Safety, la detección del miedo en situaciones de Violencia de Género ha sido el germen de este trabajo de investigación. Los modelos dependientes del sujeto requieren tener suficientes datos para que los diferentes conjuntos de entrenamiento, validación y prueba sean estadísticamente significativos. En caso de tener suficiente información de un sujeto en particular, entonces se puede generar un modelo dependiente del sujeto e incluso perseguirlo, ya que archiva un mejor rendimiento que un modelo independiente del sujeto. Sin embargo, en la mayoría de los casos, cuando se trata de aplicaciones de la vida real, en las que durante el primer momento de despliegue no hay o hay poca cantidad de datos de ese sujeto en particular, entonces es necesario implementar un modelo independiente del sujeto.

La tabla 4.19 resume los mejores resultados obtenidos a lo largo de esta parte del trabajo para el reconocimiento binario de emociones de miedo cuando se trata de un modelo independiente del sujeto. Como se puede observar, se exploraron diferentes técnicas de optimización de hiperparámetros, clasificadores y configuración del sistema (con y sin selección de características). En primer lugar, el sistema DEAP-b1 utilizó 21 sujetos, empleó un KNN mediante el uso de una estrategia específica de

Hold-Out extremo, y logró una Gmean de hasta el 62,00%. La particularidad de dicho sistema fue que las señales filtradas se consideraron como entradas, ya que no se aplicó ninguna extracción de características. Por tanto, como ya se ha comentado en apartados anteriores, este hecho podría llevar a resultados demasiado optimistas. Nótese que en este sistema no se aplicó ninguna optimización de hiperparámetros como tal, ya que se utilizaron los valores obtenidos durante el barrido de parámetros para los modelos dependientes del sujeto. Debido a las limitaciones observadas para el DEAP-b1 en términos de complejidad espacial, se desarrolló el DEAP-b2 con el objetivo de obtener un clasificador más ligero. Este último consideró todo el conjunto de voluntarios del DEAP a costa de omitir una de las señales fisiológicas (SKT). En este caso, la singularidad se basó en el origen del mapeo binario del miedo, que se obtuvo directamente del espacio PA en lugar de utilizar el espacio PAD. Para aumentar el rendimiento del sistema DEAP-b2, la eliminación de los rasgos redundantes y la maximización de los relevantes a lo largo de mrMR condujo al sistema DEAP-b2+, que proporcionó un mejor rendimiento que DEAP-b2 y DEAP-b1. Así, en este caso, la aplicación de técnicas de selección de características resultó un paso vital para la mejora del sistema. Sin embargo, la complejidad espacial siguió siendo la misma. Por último, las limitaciones a las que se enfrentaba el DEAP se solucionaron utilizando la base de datos MAHNOB. Centrándonos en el clasificador basado en árboles, desarrollamos dos sistemas, con y sin selección de características. En este caso, la técnica de prueba CV aplicada fue LOSO, que ofrece una visión no demasiado optimista del rendimiento del sistema. Así, debido a las características específicas del clasificador, al aplicar la selección de características logramos métricas similares para Gmean y AUC, y obtuvimos el menor almacenamiento para el modelo.

Table 4.18: Los mejores resultados obtenidos a lo largo del capítulo 4 para el reconocimiento de la emoción binaria del miedo cuando se trata de un modelo independiente del sujeto.

System	DEAP-b1	DEAP-b2	DEAP-b2+	MAHNOB-fear	MAHNOB-fear+
Subjects	21	32	32	12	12
Signals	PPG, GSR, SKT	PPG, GSR	PPG, GSR	ECG, GSR, SKT	ECG, GSR, SKT
Hyp.Opt.	-	Grid Search	Grid Search	SMBO	SMBO
Classifier	KNN	SVM-RBF	SVM-RBF	ENS-AdaBoost	ENS-AdaBoost
CV Technique	<i>Hold – Out</i>	<i>k – fold</i>	<i>k – fold</i>	LOSO	LOSO
Space (kB)	48.59	39.06	39.06	13	13
AUC (MAD)	-	62.79 (4.72)%	81.60 (8.70)%	81.20 (9.07)%	86.00 (9.30)%
Gmean (MAD)	62.00%	62.62 (4.73)%	81.55 (10.21)%	72.84 (8.62)%	73.78 (12.60)%

Centrándonos en los últimos sistemas propuestos, MAHNOB-fear y MAHNOB-fear+, hay que tener en cuenta ciertas limitaciones. Por un lado, el enfoque de segmentación de datos utilizado presenta algunas desventajas cuando se trata de señales fisiológicas que cambian lentamente. Deberían aplicarse diferentes técnicas para tener en cuenta las distintas particularidades fisiológicas sin desperdiciar información. Por ejemplo, en el caso concreto de la RSG, el uso de la segmentación dinámica de datos y la superposición podría ser una solución válida. Sin embargo, cuando se trata de dispositivos con recursos limitados, una solución mejor podría ser llevar un registro de los arranques de las ERSCR y, al detectar los arranques para las sucesivas ventanas de procesamiento, calcular todas las métricas de las ERSCR. La principal ventaja de este último método es la independencia de la longitud de la ventana de procesamiento a costa de almacenar la información de seguimiento de los ERSCR hasta la finalización de los mismos (offset). Por otro lado, a pesar de utilizar una técnica de normalización específica (Z-score), se podrían aprovechar otros enfoques. Por ejemplo, ya estamos trabajando en la aplicación de diferentes técnicas de normalización, como el uso de intervalos de tiempo de recuperación para normalizar los datos del estímulo relacionado con la emoción y estudiar el efecto para el caso de uso del miedo analizado. Por último, cabe señalar que los resultados mostrados están limitados por el tamaño del conjunto de datos considerado (12 sujetos), que es el punto más débil de este tipo de modelos. Dado que no existe ningún otro conjunto de datos que se ajuste a nuestro caso de uso, se requiere un conjunto de datos más grande y mejor para crear un sistema más fiable. Por lo tanto, las limitaciones iden-

tificadas durante el desarrollo de estos sistemas confirman la pertinencia de crear un nuevo conjunto de datos centrado en la detección del miedo. Este conjunto de datos debería incluir algunos hechos clave, como el uso de tecnología de inmersión emocional, la modificación de la metodología de etiquetado para tener en cuenta la perspectiva de género, una distribución de estímulos adecuadamente equilibrada con respecto a las emociones objetivo y un mayor número de participantes. Más detalles sobre este último hecho y la nueva base de datos UC3M4Safety se recogen en el capítulo 6.

En cuanto a la comparación con otros trabajos de investigación, la amplia casuística del problema de reconocimiento de emociones es una tarea desafiante. Esto se debe a la gran cantidad de técnicas diferentes que se pueden aplicar dentro de la cadena de procesamiento de datos y la generación del modelo de aprendizaje automático. Sin embargo, podemos hacer una clara distinción utilizando cinco factores: a) CV utilizado para la validación y/o prueba, b) el número de sujetos contabilizados, c) el paradigma de clasificación de la emoción (binario, discreto y / o la detección de la emoción multidimensional), d) la cantidad y el tipo de señales utilizadas, y e) el uso de conjuntos de datos disponibles públicamente. Esto último es de gran importancia, ya que los trabajos basados en bases de datos abiertas pueden compararse directamente sin tener que profundizar en las discusiones sobre las diferencias metodológicas experimentales. En la tabla 4.19 se enumeran los factores anteriores con respecto a algunos de los principales trabajos del estado de la técnica que están directamente relacionados con éste y que han influido en él. A primera vista, observamos una gran variedad de técnicas, lo que dificulta la comparación. En primer lugar, sólo se han seleccionado los trabajos de investigación que están directamente relacionados con la detección del miedo o con la clasificación de las emociones. De hecho, dos trabajos se basan en la clasificación de emociones discretas, siete de ellos se centran en la clasificación de arousal y valencia (diferentes niveles) utilizando el modelo PA, y tres trabajos de investigación clasifican las emociones mediante el modelo PAD. De estos últimos, dos de ellos [184, 185] son los ya revisados en la sección 4.1 que utilizan nuestro paradigma de binarización del miedo propuesto. En segundo lugar, sólo seis de los trabajos emplearon una técnica de CV de dejar de lado (sujeto o ensayo). Los demás aplicaron $k - fold$ y Hold-Out, lo que, según la disposición de los datos,

puede llevar a resultados demasiado optimistas. Además, independientemente del paradigma de clasificación de emociones y de la técnica de CV aplicada, la mayoría de los trabajos no informaron de muchas métricas de rendimiento de aprendizaje automático, aparte de la precisión. Finalmente, teniendo en cuenta estos aspectos de contextualización, podemos concluir que las métricas obtenidas están en línea con el estado del arte.

Table 4.19: Los trabajos más recientes y principales del estado de la técnica que están directamente relacionados con esta investigación y que han influido en ella en lo que respecta a la computación afectiva utilizando información fisiológica.

	Subjects	Signals	Classifier	CV	Emotion	Dataset	Metrics
Lisetti and Nasoz [208]	14	ECG,GSR,SKT	KNN	LOO	Sadness, anger, fear, surprise, frustration, amusement	own	ACC(fear): 85.6%
Chanel et al. [209]	10	BP,EEG,GSR, PPG,RESP	SVM	LOSO	PA space calm-neutral vs. positive-excited	own	ACC: 66.00%
Valenza et al. [167]	35	ECG,GSR,RESP	QDA	40-fold CV	Five arousal and valence levels	own	ACC > 90%
Valenza et al. [161]	30	ECG	SVM-RBF	LOO	Two levels arousal and valence	own	ACC(V):79.00% ACC(A): 84%
Abadi et al. [145]	30	ECG,EOG,EMG	SVM	LOTO	Two levels arousal, valence, dominance	DECAF	ACC(A,V,D):50-60%
Rubin et al. [160]	10	ECG	SVM	<i>k - fold</i>	Binary Panic detection	own	ACC:73-97%
Rathod et al. [210]	6	GSR,PPG	SVM	Hold-Out	Normal, happy, sad, fear, anger	own	ACC < 87.00%
Zhao et al. [211]	15	PPG,GSR,SKT	NB,RF,SVM	LOSO	Four PA quadrant	own	ACC:76.00%
Marín Morales et al. [79]	60	EEG,ECG	SVM	LOSO	Two levels arousal and valence	own	ACC:75-82%
Santa Maria Granados et al. [163]	40	ECG,GSR	CNN	Hold-Out	Two levels arousal and valence	AMIGOS	ACC:71-75%
Miranda et al. [183]	15	PPG,GSR,SKT	RF	Hold-Out	Fear (PAD binarized)	DEAP	ACC:54.00%
Amani Albraikan et al. [194]	25	GSR,ECG,EEG, RESP,SKT	ENS	<i>k - fold</i>	Three levels arousal and valence	MAHNOB	ACC:94.00%
Miranda et al. [181]	32	PPG,GSR	SVM	<i>k - fold</i>	Fear (PA binarized)	DEAP	ACC:62.80%
Oana Balan et al. [184]	32	EEG and peripheral	RF	<i>k - fold</i>	Fear (our paradigm)	DEAP	ACC:89.96%
Miranda et al. [186]	12	ECG,GSR,SKT	ENS	LOSO	Fear (PAD binarized)	MAHNOB	ACC:76.67%
Oana Balan et al. [185]	32	PPG,GSR	Boosting	<i>k - fold</i>	Fear (our paradigm)	DEAP	ACC:91.70%

Parte III

Hacia un nuevo paradigma de detección del miedo en situaciones de violencia de género

Capítulo 5

Un nuevo sistema autónomo de reconocimiento de emociones: Bindi

Como se indica en el capítulo 1, uno de los principales objetivos de esta investigación se centra en proporcionar una solución tecnológica inteligente para prevenir y combatir la violencia de género. En base a ello, se propone el sistema 5-1. Se trata de un sistema multimodal autónomo que considera las tecnologías IoT para la detección de situaciones de riesgo en contextos de violencia de género. En concreto, la parte de edge-computing del sistema se concibe como una red ciberfísica inteligente capaz de detectar emociones relacionadas con el miedo. Esto se consigue mediante sensores inteligentes fisiológicos y físicos (audio y/o voz) que monitorizan continuamente al usuario. Esta tarea se completa con una fusión de datos multimodal basada en la niebla dentro de una aplicación ad-hoc para smartphones. Finalmente, en caso de confirmar una situación de riesgo, se activa una alarma a una red de protección predefinida. Además, la información se envía a servidores informáticos específicos en la nube, que se encargan de almacenar los datos recogidos para posteriores acciones legales. El diseño de este sistema impulsa la generación de nuevos mecanismos de prevención y lucha contra la violencia de género.

En este capítulo, en primer lugar, se realiza un estudio detallado sobre los sistemas y herramientas actuales para prevenir las agresiones violentas de género. Esto se hace considerando diferentes perspectivas como los dispositivos disponibles en el mercado, los sistemas de grado de investigación y las herramientas institucionales. Nótese que este último se centra en las instituciones españolas, debido al liderazgo



Figure 5-1: Arquitectura simplificada del sistema Bindi basado en las diferentes tecnologías del IoT.

mundial de España en este aspecto, tal y como se detalla en el capítulo 1. Además, se comparan y destacan las diferentes ventajas competitivas tecnológicas de Bindi. A este análisis le sigue una descripción exhaustiva del sistema Bindi. Así, se abordan los diferentes diseños de hardware y software dentro del brazalete de Bindi. En primer lugar, se detalla la arquitectura del sistema, tanto en su diseño como en su integración. Esto se acompaña de diferentes recomendaciones de integración de wearables fisiológicos a tener en cuenta para las siguientes versiones del sistema. En segundo lugar, se informa y explica la actual implementación integrada. Hay que tener en cuenta que los resultados ofrecidos en este capítulo se han presentado en diferentes publicaciones [11, 159, 183, 212].

5.1 Tecnología actual para luchar contra la violencia de género

El desarrollo de la tecnología a lo largo de los años ha hecho realidad la generación y aplicación de nuevas herramientas para prevenir la Violencia de Género [9, 22, 23]. Las ventajas de utilizar herramientas tecnológicas para ayudar a combatir este problema son múltiples:

- Accesibilidad de la protección. La tecnología puede facilitar y acercar el acceso a la protección de las víctimas.
- Centralización de la información. Diferentes instituciones y/o fuerzas pueden cooperar hacia un seguimiento conjunto de las circunstancias que rodean a las víctimas de la violencia de género.
- Recogida de información multimodal. La recopilación de diversas fuentes de in-

formación puede utilizarse para el análisis de predicción y prevención. Además, esto permite una mejor comprensión de la situación específica de la víctima.

- Tiempos de respuesta a la acción. Los puntos anteriores afectan directamente a los tiempos de decisión en la activación de los respectivos mecanismos institucionales.
- Refuerzo de la seguridad. Desde la perspectiva del usuario, la inclusión de una tecnología fiable y robusta puede proporcionar una mayor sensación de seguridad en las víctimas de la violencia de género.

Sin embargo, estas ventajas también van acompañadas de diferentes requisitos, consideraciones y cuestiones abiertas, que pueden resumirse en:

- La pseudoanonimización de los datos almacenados es crucial. Cualquier tecnología debe garantizar la protección y seguridad de todos los datos sensibles o identificables. La gestión y la propiedad de dicha información deben ser cuidadosamente consideradas. Así, debe garantizarse el estricto cumplimiento de las leyes de protección de datos. Además, cualquier solución tecnológica debe garantizar la cadena de custodia de la información recogida para que pueda ser utilizada posteriormente en cualquier proceso judicial.
- Las herramientas tecnológicas candidatas deben conectar directamente a las víctimas con profesionales especializados. Esto aboga por: 1) la necesidad de contar con más profesionales formados para tratar adecuadamente a las víctimas de la violencia de género, y 2) la elaboración de nuevos protocolos que tengan como objetivo evitar la revictimización.
- Es de suma importancia la alineación entre las soluciones tecnológicas propuestas, el gobierno y los actores privados. Hay que tener en cuenta que estos últimos desempeñan un papel fundamental en el desarrollo y la integración de las soluciones tecnológicas.
- La personalización tecnológica debe considerarse un aspecto esencial, ya que es muy necesario que la solución tecnológica se adapte y personalice a cada persona. Esto se debe a la adaptación a diferentes contextos y entornos heterogéneos. Sin embargo, esto podría enfrentarse a las limitaciones actuales de la tecnología para lograr dicha adaptación.
- Accesibilidad a las soluciones de base tecnológica propuestas. Se sabe que

existe una brecha de género en la posesión de móviles de aproximadamente un 7% en los países de ingresos bajos y medios [213]. Este hecho, acompañado de los menores ingresos percibidos por las mujeres, hace que el precio objetivo y la plataforma tecnológica sean factores críticos. El primero está relacionado con la asequibilidad de la solución, mientras que el segundo se refiere al hecho de que las soluciones sin necesidad de tecnología de telefonía móvil ayudarían a que la solución fuera más inclusiva.

Todos estos puntos justifican y fomentan el enfoque multidisciplinar que se reclama en el capítulo 1 y que es necesario para el diseño, el desarrollo y la integración de la tecnología que se ocupa de los contextos de la violencia de género. Aunque se trata de una tarea difícil, uno de los principales objetivos de esta investigación es proporcionar la base tecnológica necesaria para empezar a resolver estos problemas y preguntas abiertas. Cabe destacar que cualquier solución de base tecnológica diseñada y orientada a la casuística de la Violencia de Género puede ayudar a prevenir y combatir, pero nunca resolverá todo el problema. Eso, es una cuestión educativa.

Una de las tecnologías más empleadas son las aplicaciones basadas en el teléfono móvil. Hoy en día, esta tecnología es una de las más aceptadas a pesar de las consideraciones descritas anteriormente. Recientemente, los autores de [214] realizaron una revisión sistemática de hasta 171 aplicaciones cuyo objetivo era abordar la Violencia de Género a través de diferentes mecanismos. Independientemente de la aplicación específica, los autores concluyeron que la mayoría de ellas estaban principalmente enfocadas y diseñadas para soluciones de emergencia a corto plazo o puntuales. Este hecho deja de lado la perspectiva de la prevención y ofrece la posibilidad de identificar únicamente hechos aislados de Violencia de Género en lugar de ofrecer un seguimiento continuo y el autoempoderamiento de las Víctimas de Violencia de Género, que debería ser uno de los principales objetivos. Aunque los autores afirman que cada vez se incluyen más funciones educativas en las aplicaciones recientes, es necesario llevar a cabo más investigaciones relacionadas con la seguridad de los datos, la seguridad personal y la eficacia de dichas soluciones. Un ejemplo de una de estas aplicaciones es AlertCops [215]. Esta aplicación está promovida específicamente por el Ministerio del Interior español y permite notificar al instante cualquier

tipo de incidencia con las fuerzas del orden. Como característica diferenciadora respecto a otras aplicaciones existentes, en el último año se ha incorporado a esta aplicación el "Botón SOS", que permite reforzar la protección de los colectivos vulnerables. Este botón envía una alerta urgente al centro policial más cercano junto con su ubicación Global Positioning System (GPS) y una grabación de audio de 10 segundos de lo que está ocurriendo. Además, esta aplicación también incluye la función "Guardián", que últimamente han incluido muchas otras aplicaciones. En concreto, permite compartir la ubicación en tiempo real con los contactos seleccionados por el usuario. Aunque estas aplicaciones pueden explotar con éxito las diversas capacidades de la tecnología móvil, la toma de decisiones se basa en cualquier caso únicamente en medidas ambientales o relativas, pero nunca en medidas del propio usuario.

En este contexto, las partes interesadas del sector privado también han desarrollado herramientas tecnológicas que podrían utilizarse para abordar el caso de uso comentado. Sin embargo, la mayoría de estas soluciones se incluyen en la categoría de botones de pánico. Incluso en algunos países, como la India, se emitió una directiva relacionada con la inclusión obligatoria de un botón de pánico en cada teléfono móvil vendido a partir de 2017. Una de las soluciones de botón de pánico más destacadas que está pensada específicamente para hacer frente a situaciones de Violencia de Género es SaferPro de LeafWearables, una empresa india. Se trata de un dispositivo de pulsera que viene con una tarjeta de módulo de identidad de abonado de bajo consumo, lo que lo hace independiente del teléfono móvil. En concreto, una vez que el usuario pulsa el botón, se envía una alarma a un círculo seleccionado de respondedores y se inicia una grabación de audio. Sin embargo, los botones de pánico presentan importantes limitaciones en cuanto a la seguridad de las mujeres: 1) la exigencia de un papel activo en su autoprotección, lo que ciertamente no es posible bajo algunos tipos de agresión y/o bloqueo de las reacciones emocionales, 2) su falta de diseño discreto que puede provocar estigmas en las usuarias, y 3) la falta de apoyo de la infraestructura [216]. A pesar de los esfuerzos tecnológicos, este tipo de enfoque es cuestionado por varios expertos en Violencia de Género [21], que reclaman, entre otras cosas, una investigación y tecnología más avanzada en estas soluciones que se consideran anticuadas y un mayor grado de atención al pa-

pel de las víctimas. Además de los botones de pánico, también existen dispositivos comerciales que, aunque no están exactamente orientados al caso de uso de la Violencia de Género, permiten generar alarmas de forma automática sobre eventos anormales internos y externos detectados al usuario. Por ejemplo, el Apple Watch Series 4 y posteriores proporcionan detección de caídas y envían un SOS a contactos de emergencia predefinidos en caso de que el usuario no realice ninguna acción. El brazalete Embrace2 de Empatica es el único wearable de muñeca autorizado por la FDA para la epilepsia, que activa una alarma en caso de detección de convulsiones. Esto se hace mediante la monitorización GSR. Además, está equipado con otros tres sensores (SKT, acelerómetro y giroscopio) que también pueden ser adquiridos y almacenados con fines médicos. Este último sistema también abre la posibilidad de considerar el uso de una tecnología de detección similar para abordar la violencia de género. Entre los dispositivos de detección fisiológica avanzada lanzados recientemente, destacan Fitbit con FitbitSense y Oura con OuraRing. El primero es la única pulsera inteligente comercial que ofrece más de dos sensores fisiológicos integrados: GSR, ECG, PPG y SKT. Sin embargo, la actual integración electromecánica de algunos de estos sensores dificulta la aplicación de este dispositivo a otros casos de uso. Esto se debe principalmente al hecho de que la adquisición de una medición de GSR y/o ECG requiere que la mano libre esté encima del brazalete, ya que esto proporciona un circuito de bucle cerrado. Este último sistema se basa en un anillo inteligente y proporciona la adquisición de PPG y SKT con una precisión relativamente alta. Sin embargo, hay que tener en cuenta que el nicho de mercado de estos dispositivos se centra en el bienestar genérico, más que en cualquier otro caso de uso específico. Hasta donde yo sé, el único dispositivo comercialmente disponible orientado a proporcionar una herramienta para prevenir una condición específica relacionada con la fisiología es el mencionado Embrace2. Sin embargo, la proliferación de dispositivos comerciales vestibles con capacidades de detección fisiológica ha estado en auge en los últimos años y podría beneficiar el diseño y desarrollo de herramientas orientadas a la aplicación objetivo de esta investigación.

El sector público no ha sido ajeno a los avances tecnológicos. Cuando se trata de la vigilancia electrónica para ayudar a prevenir la Violencia de Género, España resulta ser uno de los países pioneros en el mundo en la promoción de este tipo de tecnología.

De hecho, como ya se repasó en el Capítulo 1, en 2013 se firmó un convenio entre el Ministerio del Interior, Justicia, Sanidad y Servicios Sociales e Igualdad del Consejo General del Poder Judicial y la Fiscalía General del Estado por el que se aprobó el "Protocolo de actuación del sistema de seguimiento por medios telemáticos de las medidas y sentencias de alejamiento en materia de violencia de género". Estas medidas obligan al agresor y a la víctima a llevar diferentes dispositivos, Figura 5-2. Además, todas las diferentes alarmas generadas por el sistema son monitorizadas y centralizadas por un centro especializado llamado Cometa, gestionado por una empresa privada (Securitas Direct) subcontratada por el gobierno español. Las partes interesadas consideradas, así como la centralización de la información que proporciona este sistema, están en consonancia con las ventajas y los requisitos anteriores. Sin embargo, en el caso concreto de este sistema, la tecnología empleada es obsoleta y se basa únicamente en la monitorización GPS, lo que en algunos casos provoca que el agresor acose aún más a la víctima. Aunque este protocolo y solución tecnológica han ayudado en la lucha contra la Violencia de Género durante los últimos años, su tecnología de detección basada en GPS, sumada a su baja batería y a los fallos intermitentes que se reportan, hacen que esta solución sea muy limitada.



Figure 5-2: Dispositivos considerados para el sistema de vigilancia electrónica dentro del "Protocolo de actuación del sistema de vigilancia por medios telemáticos de las medidas y penas de alejamiento en materia de violencia de género". DLI: Dispositivo que lleva el agresor; DLV: Dispositivo que lleva la víctima [9].

En el caso de buscar propuestas en el ámbito académico sobre el diseño de sistemas y herramientas para evitar la Violencia de Género, existe literatura [217]. Sin embargo, en cuanto a la mayoría de los dispositivos comerciales, éstos se cen-

tran en el diseño y optimización de sistemas basados en botones de pánico. Cabe mencionar que dejando de lado el resultado final físico del wearable, también existe literatura que aborda directamente la problemática de la Violencia de Género mediante el aprendizaje automático aplicado a la información centralizada. Por ejemplo, los autores de [218] utilizaron el aprendizaje automático para diseñar modelos que predijeran con precisión el riesgo de reincidencia de un agresor de violencia de género. Emplearon 40.000 informes de violencia de género extraídos de VioGen y superaron el algoritmo de evaluación de riesgo preexistente basado en técnicas estadísticas clásicas. Aparte de eso, en la literatura faltan propuestas de sistemas dirigidos a la prevención y lucha contra la Violencia de Género.

De este análisis podemos concluir que ninguna de las soluciones tecnológicas públicas, de investigación o privadas para combatir la Violencia de Género se beneficia de los avances clave del estado del arte y de la electrónica de consumo actuales, como la analítica fisiológica y física y la computación afectiva. Estos avances pueden ser aprovechados para conseguir una herramienta tecnológica de prevención de la Violencia de Género mejor, autónoma y más discreta, que es el objetivo del equipo UC3M4Safety mediante el sistema Bindi. Además, el diseño de una herramienta de este tipo para la seguridad de las mujeres requiere que éstas sean co-creadoras de la solución, algo que este equipo está considerando fuertemente mediante una estrecha colaboración con diferentes asociaciones de mujeres y grupos de discusión de mujeres profesionales expertas en la materia. Hasta donde yo sé, Bindi es el único sistema que propone una herramienta tecnológica para ayudar a prevenir y combatir la Violencia de Género mediante la informática afectiva [11, 183, 219, 220].

5.2 Bindi

Cronológicamente, Bindi ha pasado por varias fases de diseño y desarrollo, Figura 5-3. La primera prueba de concepto fue el *iGlove*, que fue una tesis de máster co-supervisada [221]. La idea de este sistema era diseñar e implementar un primer sistema de monitorización fisiológica continua wearable. En concreto, estaba equipado con tres sensores fisiológicos (BVP, GSR y SKT) y permitía la transmisión continua de datos mediante Bluetooth Low Energy (BLE) a un teléfono móvil. Este dispositivo estaba basado en [222]. Además, el SoC integrado en el *iGlove* era un

ARM[®]Cortex-M0 de 32 bits con 32KB de RAM y 256KB de Flash. Este dispositivo cumplió con éxito su objetivo de crear la primera herramienta para impulsar la investigación en computación afectiva dentro del equipo UC3M4Safety. Posteriormente, se diseñó la primera versión formal de Bindi, Bindi 1.0, utilizando la mayor parte del hardware de *iGlove* como sólido punto de partida. Como ya se introdujo al principio de este capítulo, Bindi 1.0 es un sistema de red de área personal formado por tres dispositivos: una pulsera, un colgante y una aplicación para smartphone. En concreto, el SoC integrado dentro de Bindi 1.0 era un ARM[®]Cortex-M4 de 32 bits con 64KB de RAM y 256KB de Flash. Específicamente para este sistema, fui responsable de algunas de las principales tareas relacionadas con el brazalete, tales como 1) la supervisión de los esquemas, Printed Circuit Board (PCB), y el diseño de la maqueta, 2) el diseño del firmware y la integración del sistema, y 3) la coordinación de las diferentes validaciones y aplicación de bancos de pruebas para asegurar la funcionalidad. Cabe destacar que, independientemente de estas tareas específicas, la mayor parte del diseño, integración, implementación y validación se ha realizado de forma conjunta, ordenada y organizada por un grupo de personas pertenecientes al equipo de UC3M4Safety. Esta primera versión de Bindi es la que se aborda en este capítulo. Además, Bindi 1.0 es uno de los sistemas sensoriales empleados durante la grabación del conjunto de datos WEMAC que se explica en el Capítulo 6. Tras las diferentes limitaciones identificadas durante el desarrollo y uso de Bindi 1.0, la UC3M4Safety diseñó Bindi 2.0 durante los dos últimos años. Este nuevo sistema sufrió un drástico proceso de miniaturización que aprovechó la integración del hardware de Bindi 1.0. Además, en esta nueva versión se han incluido nuevos sensores y diferentes mejoras de hardware, así como nuevas funcionalidades de firmware. En este caso, el SoC integrado dentro de Bindi 2.0 era un ARM[®]Cortex-M4 de 32 bits con 256KB de RAM y 1MB de Flash. Hay que tener en cuenta que, a medida que se ha ido mejorando la tecnología de Bindi, también han aumentado sus necesidades computacionales, lo que ha hecho que se necesite más capacidad de almacenamiento en particular. Sin embargo, este hecho no ha supuesto un aumento considerable en el consumo de energía si comparamos específicamente Bindi 1.0 y Bindi 2.0 [11, 223].

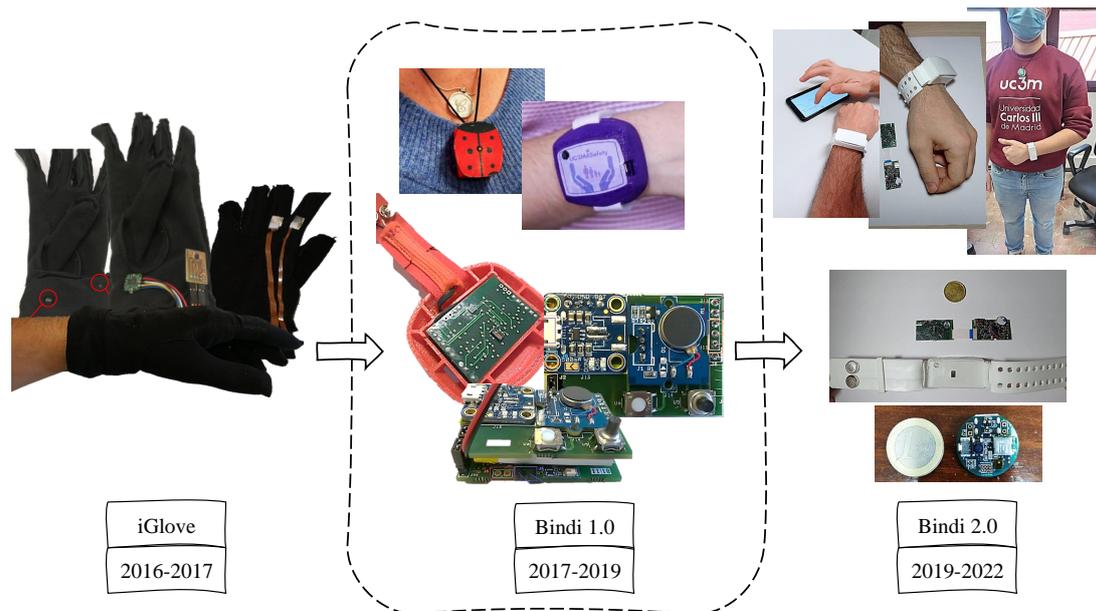


Figure 5-3: Evolución de la tecnología Bindi desde 2016 hasta 2022.

Una vez abordado el estado del arte de la tecnología aplicada a la lucha contra la Violencia de Género y el contexto tecnológico de Bindi, los siguientes apartados se centran en ofrecer una perspectiva integrada sobre los diferentes procesos, técnicas y métodos de señalización digital diseñados e implementados en la pulsera de Bindi durante la evolución de esta investigación. En primer lugar, se presenta un análisis detallado de la arquitectura de la pulsera, tanto desde el punto de vista del hardware como del software. En segundo lugar, se evalúan y analizan diferentes arquitecturas de filtrado digital integradas, manteniendo un equilibrio entre los requisitos de recursos y la preservación de la información fisiológica. En tercer lugar, se detalla una nueva propuesta de sistema SQA para señales PPG, implementada y evaluada mediante el uso de conjuntos de datos públicos y propios. Este sistema SQA también informa de las métricas de tiempo y consumo de energía para diferentes características extraídas. Posteriormente, se presenta una exploración completa del espacio de diseño de la extracción de características embebidas para un caso de uso HRV. Aquí se analizan y discuten las técnicas de procesamiento de datos temporales y de frecuencia. Esto se hace para proporcionar una perspectiva en profundidad sobre las consideraciones y limitaciones del diseño de la extracción de características. Además, se presenta y discute una comparación entre las características basadas en HRV obtenidas con la pulsera y las obtenidas con un kit de herramientas de investigación. Por último, se presentan las métricas de consumo de

energía, que proporcionan un análisis exhaustivo de la duración de la batería de la pulsera.

5.2.1 Arquitectura del sistema

Como se muestra en la Figura 5-4, el brazalete se compone de diferentes elementos de hardware y software. Estos se pueden clasificar en cuatro grupos: el SoC, los actuadores, los elementos de gestión de la energía y los sensores fisiológicos. Se describen como sigue:

- Unidad de microprocesador. Bindi 1.0 está equipado con el SoC nRF52832 que incluye ARM® Cortex®-M4, una unidad de microcontrolador de ultra bajo consumo de energía con 512KB de memoria flash y 64KB de RAM, unidad de punto flotante de precisión única, conjunto de instrucciones Thumb®-2, reloj de 64MHz, y algunos periféricos integrados (USB, UART, SPI, I2C, I2S, ADC, PDM, y AES) [224]. Cabe destacar que el módulo de radiofrecuencia a través de la comunicación Bluetooth Low Energy®(BLE) también está integrado dentro de esta unidad host. Además, los diferentes procesos de señalización digital empleados fueron embebidos en este SoC.
- Actuadores. La pulsera está equipada con un botón electromecánico convencional para la activación manual del usuario, que actúa como botón de pánico. Adicionalmente, se incluye un zumbador para dar respuesta física a las diferentes alarmas del sistema [225].
- Elementos de gestión de la energía. En este caso se utilizan los componentes BQ2019 y MCP73831 de Texas Instruments y Microchip [226, 227]. Estos dos circuitos integrados se encargan de controlar y cargar la batería, respectivamente. Para el Bindi 1.0 se ha empleado una batería de polímero de iones de litio de 500 *mAh* de 3,7V.
- Sensores fisiológicos. Tres sensores fisiológicos diferentes están presentes en el Brazalete: PPG, GSR, y SKT. En la siguiente sección se ofrecen detalles específicos, junto con las limitaciones encontradas en cuanto a la implementación del hardware. Tenga en cuenta que estas últimas se abordaron en las siguientes versiones de Bindi (Bindi 2.0).

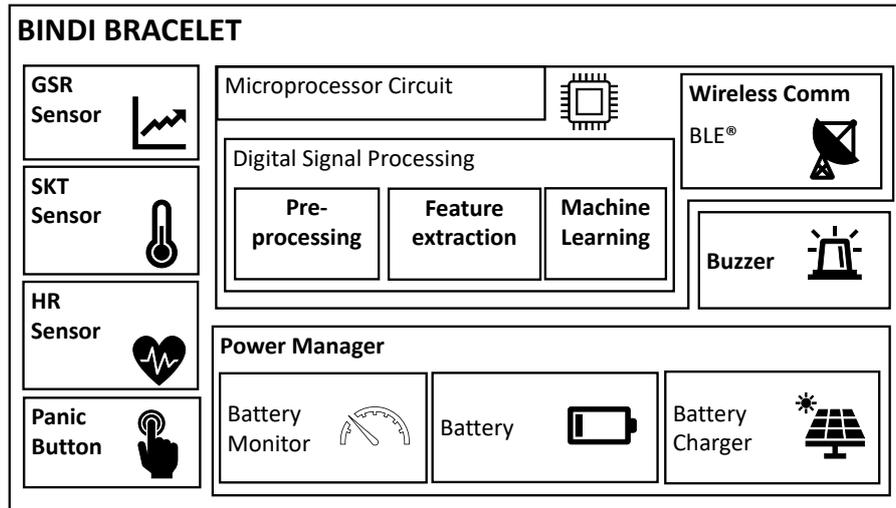


Figure 5-4: Arquitectura simplificada del brazalete.

Desde el punto de vista del hardware, la mayor parte de los diferentes elementos del brazalete se basan en sensores inteligentes, microcontroladores y actuadores disponibles en el mercado. Esta decisión se basó en tres hechos principales 1) facilitar todos los procesos de diseño e integración, 2) crear la primera versión wearable de Bindi con piezas disponibles en el mercado siempre que fuera posible, y 3) reducir costes al no tener que diseñar muchos de los elementos desde cero. Posteriormente, esta decisión de diseño nos permitió identificar los inconvenientes y limitaciones actuales del Commercial-Off-The-Shell (COTS) empleado. Estos aparecerán a lo largo de las siguientes subsecciones.

5.2.1.1 Diseño e integración de sensores fisiológicos

Esta sección ofrece un análisis en profundidad de los sensores integrados en Bindi 1.0, así como las limitaciones encontradas durante este proceso. Hay que tener en cuenta que la ubicación corporal de los sensores se vio directamente afectada por la forma del factor del brazalete, así como por la literatura anterior que demostró las diferencias fisiológicas [102, 228, 229].

Sensor cardiovascular

El sensor de frecuencia cardíaca integrado se basa en un sensor fotopletoislográfico que detecta los cambios de BVP mediante la medición de la absorción de la luz emitida a través de la piel, como se estudió en el capítulo 2. Este sensor es el MAX30101 Pulsioxímetro reflexivo de alta sensibilidad, con ADC de 18 bits, comunicación I2C, cancelación de ruido digital, y diferentes LEDs integrados (rojo -660nm-, verde -

527nm-, e infrarrojo -880nm-), [10]. Teniendo en cuenta la eficiencia cuántica del fotodiodo del sensor, figura 5-5, y el voltaje de avance requerido por los diferentes LEDs, se seleccionó finalmente el LED rojo. Nótese que la eficiencia cuántica de cualquier fotodiodo o fotodetector se refiere al porcentaje o fracción de fotones absorbidos o incidentes que contribuyen a la fotocorriente real, es decir, la sensibilidad esperada del fotodiodo dividida por la fotosensibilidad máxima en caso de que cada fotón genere un electrón. Además, decidimos utilizar sólo uno de los LEDs para reducir el consumo de energía y para abrir una nueva línea de investigación sobre la eliminación de artefactos de movimiento mediante técnicas de separación ciega de fuentes. Esto último dio lugar a una Tesis de Máster supervisada [230], en la que se establecieron las bases para el uso de algoritmos de eliminación de artefactos de movimiento. Cabe destacar que, aunque esto último no entra en el ámbito de este documento, servirá de base para futuras investigaciones. Entre las capacidades revisadas de este sensor inteligente, también ofrece una frecuencia de muestreo configurable desde 50 Hz hasta 3,24 kHz, y control de la corriente del LED programable. En nuestro caso, para las implementaciones embebidas presentadas en este capítulo, empleamos la corriente máxima del LED (50mA con un ancho de pulso de 411 μ s) y una frecuencia de muestreo de 100 Hz. La primera se decidió para proporcionar una penetración más profunda, lo que derivó en una mayor diferencia de ciclo entre las fases sistólica y diastólica. La frecuencia de muestreo se eligió porque es la disponible en el sensor que permite una resolución temporal adecuada para extraer las características deseadas [231]. Una de las principales limitaciones de este sensor es en realidad su principal ventaja, ya que proporciona una solución integral mediante la integración de diferentes LEDs, pero esto no proporciona flexibilidad para probar otras configuraciones de LEDs. Esto llevó al equipo de UC3M4Safety a investigar en diferentes configuraciones de LEDs modificando el entrehierro del sensor de piel o incluso probando configuraciones de múltiples longitudes de onda [232]. Independientemente de estas últimas investigaciones, este sensor se ha mantenido para la integración de Bindi 2.0.

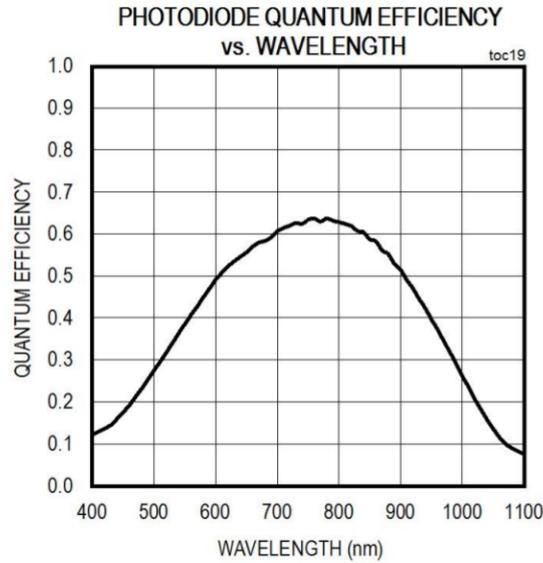


Figure 5-5: Fotodiodo de eficiencia cuántica MAX30101 [10]

Sensor Electrodermico (GSR)

Para el sensor GSR, se ha implementado una medición exosomática de corriente continua aplicando un voltaje constante a través de la piel. Además, se han utilizado electrodos de acero seco. Nótese que, en este caso, éste es el único sensor, de los tres integrados en la Brazalete, que se diseñó en lugar de adquirir un COTS analógico-front-end o smart-sensor. El diseño de este sensor se basó en el primer circuito integrado en el *iGlove* [221]. La figura 5-6 muestra el esquema del front-end analógico para el sensor de corriente GSR en Bindi 1.0. En concreto, los electrodos están conectados a *J7*, por lo que se realiza una medición del potencial de piel. Esto se realiza gracias al divisor de tensión entre la piel y *R14*. Para evitar las perturbaciones endosomáticas, se considera una referencia común a la salida y a la entrada para que la diferencia de tensión sea independiente de la posición del electrodo de referencia. En función de la tensión de salida del sensor que se va a medir, se aplica una tensión de referencia para evitar la saturación utilizando una resistencia variable (*R7*). Obsérvese que se aplican seguidores de tensión en ambas ramas como amortiguadores para evitar problemas relacionados con la impedancia. Por último, se emplea un amplificador diferencial para obtener la diferencia entre la referencia de tensión conocida y el divisor de tensión de la piel. La amplificación entre estas

dos tensiones viene dada por la ecuación 5.1:

$$V_{OUT3} = \frac{((R_{skin} - R7) * 2 * VCC_{1.8B} * 2 * 10^5)}{((R7 + 2 * 10^5) * (R_{skin} + 2 * 10^5))}. \quad (5.1)$$

Esta tensión de salida es seguida por un filtro de paso bajo ($R11$ y $C11$) para evitar el ruido de alta frecuencia con una frecuencia de corte de hasta 1,5 Hz. Hay que tener en cuenta que la información del GSR se mantiene por debajo de dicha frecuencia, tal y como se estudia en el capítulo 2. En cuanto al consumo de energía, el propio sensor consume alrededor de 0,7mA.

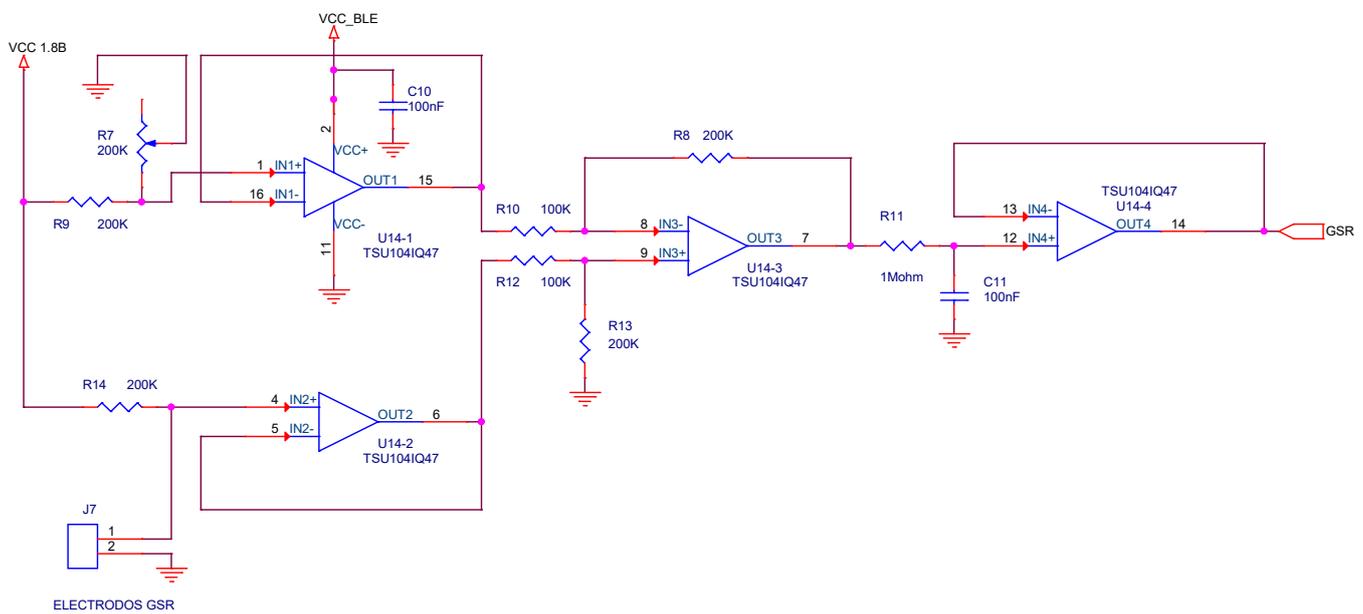


Figure 5-6: Implementación del frente analógico del sensor de GSR en el brazalete de Bindi.

Una de las principales limitaciones del circuito de adquisición GSR implementado es el comportamiento no lineal. Las figuras 5-7 y 5-8 muestran la tensión de salida (V_{OUT3}) y la corriente inyectada en la piel, respectivamente. Obsérvese que el voltaje se representa utilizando diferentes valores de $R7$, y $R14$ se fija en $200\text{ k}\Omega$. Este último se fijó en ese valor para limitar la corriente inyectada por debajo de los límites recomendados de $10\mu\text{A}/\text{cm}^2$ para los requisitos de seguridad [120]. Siguiendo un compromiso entre la sensibilidad y un rango deseado de hasta $0\text{-}20\ \mu\text{S}$, decidimos fijar la resistencia variable en $50\text{ k}\Omega$. Así, considerando una resolución de muestreo de 14 bits (ADC), el LSB es de hasta $219\mu\text{V}$, y la peor resolución de conductancia del sensor es de $0,007\mu\text{S}$. Esta resolución es suficiente para capturar cambios de $0,01\mu\text{S}$ para registrar correctamente todos los SCRs. Tenga en cuenta que, suponiendo un

error de cuantificación máximo de $LSB/2$, para este caso que lleva hasta $\pm 0,003\mu S$.

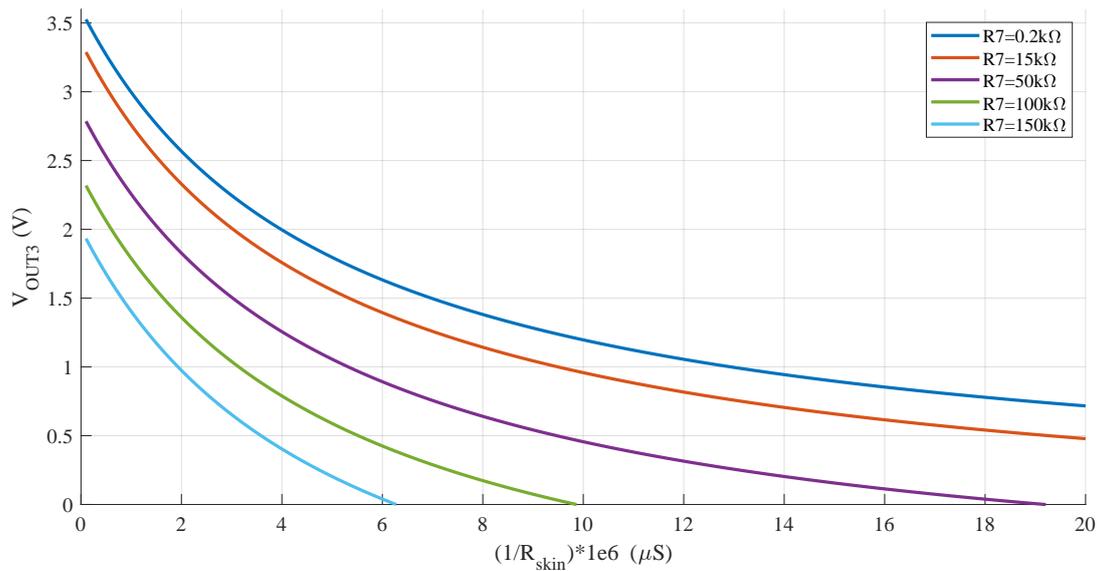


Figure 5-7: Respuesta GSR de Bindi considerando diferentes resistencias de la piel.

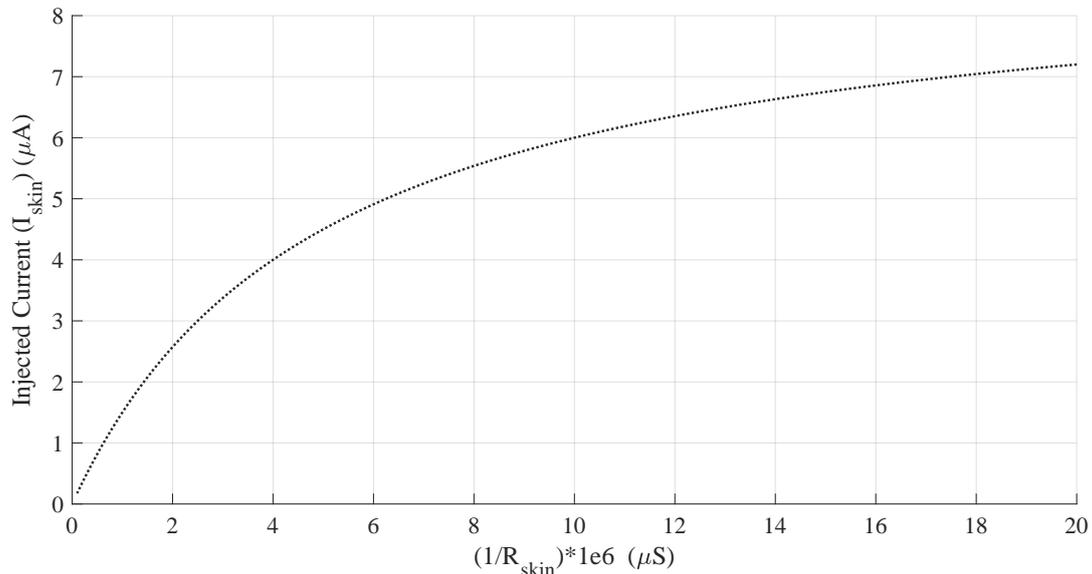


Figure 5-8: Respuesta no lineal de la corriente de piel dada por el divisor de tensión entre $R14$ y R_{skin} .

Este sensor fue validado empíricamente con éxito utilizando componentes pasivos (resistencias) en [221]. Además, debido a la complejidad para generar un modelo de piel adecuado [233], decidimos utilizar un sensor GSR de grado de investigación y un conjunto reducido de voluntarios para validar una medición GSR real [234]. En estos experimentos, nuestro sensor GSR se colocó en la parte distal del antebrazo debido al factor de forma Brazalete, mientras que el sensor GSR de validación se situó en la palma de la mano. Se sabe que en esta última localización se encuentra la mayor densidad de glándulas sudoríparas del cuerpo [112], lo que implica una señal más sensible a los afectos. Además, Bindi trabaja a base de electrodos secos,

mientras que el sensor de validación utiliza electrodos de hidrogel, véase la figura 2-15. Este hecho es clave a la hora de comparar las señales, ya que el hidrogel mejora la calidad de la señal al disminuir la impedancia que existe en la interfaz electrodo-piel. La figura 5-9 muestra las señales GSR normalizadas obtenidas por ambos dispositivos para un voluntario durante dos ensayos diferentes. La línea vertical de guiones en la figura marca la separación de estímulos, donde el primer y segundo estímulo son la alegría y el miedo, respectivamente. Al analizar la correlación de las señales adquiridas por ambos sensores, se obtiene una métrica de Pearson de 0,85, lo que denota una fuerte correlación positiva directa. Se obtuvieron coeficientes de correlación similares para el resto de los voluntarios. Se aprecian diferencias entre ambas señales, que pueden deberse a los artefactos de movimiento del sensor, a los efectos del hidrogel y a la ubicación del sensor. Además, la mayoría de los SCRs captados por el sensor de validación están presentes en la señal de Bindi. Por lo tanto, concluimos que la validación del sensor fue exitosa.

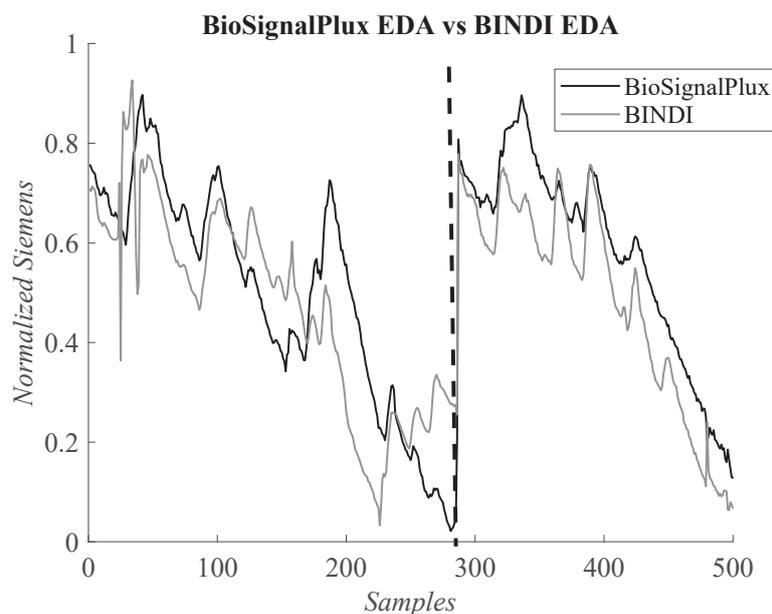


Figure 5-9: Señales filtradas normalizadas del sensor de GSR obtenidas por Bindi y el sensor de validación para un voluntario en dos estímulos. La línea vertical de guiones denota la separación de estímulos.

Las limitaciones encontradas durante el desarrollo e integración de este sensor GSR hicieron que el equipo de UC3M4Safety trabajara en un nuevo sensor que aborda la respuesta no lineal y posee un hardware ajustable e independiente del sujeto. Esto hace que el nuevo sistema de sensores sea capaz de ajustar su hardware en función de

la línea de base detectada o de cualquier otro parámetro individual basado en el GSR para asegurar la sensibilidad recomendada sin exceder los límites de recomendación de la densidad actual. Esto se está probando actualmente y se está publicando [235].

Sensor de temperatura de la piel

Por último, se propone el componente MAX30205 para adquirir una medición fiable de la temperatura de la piel [236]. Este circuito integrado se define como un sensor de grado clínico para aplicaciones vestibles, proporcionando una precisión de $\pm 0.1^\circ\text{C}$ en un rango de temperatura de 30°C a 50°C . Integra comunicación I2C y un ADC de alta resolución, sigma-delta, de 16 bits. Además, cuando está en modo activo, consume alrededor de $0,6\text{mA}$.

Como se indica en el capítulo 2, la medición de la temperatura de la piel es un indicador robusto para caracterizar el proceso de homeostasis del cuerpo. Aunque el uso de sensores de temperatura de contacto es sencillo siempre que se disponga de la superficie de contacto (piel), la adquisición de mediciones precisas de dicha variable es una tarea difícil debido a las diferentes variables y condiciones de configuración. Esto se refiere a consideraciones como la homogeneidad de la piel, la resistencia térmica de contacto y la eficacia de la fijación, entre otras [139]. En concreto, el MAX30205 mide la temperatura de su propia matriz por la vía térmica entre ésta y la PCB. Por lo tanto, la temperatura medida se adquiere a través de los conductores y la almohadilla expuesta. En este contexto, y teniendo en cuenta la forma factorial del , decidimos integrar este sensor dentro de la PCB, justo debajo del sensor PPG, Figura 5-10. A pesar de que el fabricante en la hoja de datos afirma que los errores de temperatura debidos al autocalentamiento son bajos debido a la baja corriente de alimentación mínima, también se especifica que se requiere un período de muestreo ≥ 10 -segundos para evitar completamente estos efectos. Así pues, el principio de medición del sensor junto con la implementación de la PCB no resultó ser el más preciso para adquirir la temperatura corporal de la piel ni el más eficiente para evitar los problemas de autocalentamiento, masa térmica y/o conductividad térmica. La consecuencia de este problema fue un gradiente térmico inicial que dura alrededor de 200 segundos hasta que la masa térmica de la PCB está en equilibrio. Por ejemplo, la figura 5-11 muestra la salida filtrada del sensor tras colocar un dedo sobre el chip integrado en condiciones de temperatura ambiente controlada. Este

problema se resolvió para las siguientes versiones de Bindi (Bindi 2.0), así como para los experimentos realizados y explicados en el 6, considerando la integración del sensor de temperatura MAX30208 [237]. Este sensor fue la siguiente versión del MAX30205, incluyendo las mismas capacidades digitales, pero cambiando el principio de medición y el consumo de energía en modo de funcionamiento. En concreto, mide a lo largo del contacto superior del paquete en lugar de utilizar una almohadilla térmica, y consume alrededor de $70\mu A$ cuando adquiere. Obsérvese que el consumo de energía es considerablemente menor en comparación con el sensor anterior. La figura 5-12 muestra la modificación realizada al Brazalete para incluir el nuevo sensor de temperatura y una comparación del experimento para ambos. Hay que tener en cuenta que hemos utilizado parte de la placa de evaluación del MAX30208 [238]. El experimento realizado consistió en tres fases: 1) los sensores se dejaron en el exterior durante 1 hora (noviembre, $14^{\circ}C$), 2) el sistema se encendió y comenzó a medir justo después de entrar en la habitación, 3) se realizó el contacto con la piel para ambos sensores después de estar tres minutos midiendo a temperatura ambiente, y 4) se liberó el contacto con la piel después de un minuto. Así, podemos observar como la respuesta del MAX30208 es más rápida que la del MAX30205, concretamente dos veces más rápida, y como el principio de medición y la masa térmica de la PCB están afectando a la hora de alcanzar una medición precisa. Además, también se observa un desfase entre ambos, que también se debe a los factores comentados. A pesar de los problemas encontrados con la integración del MAX30205 y aunque existe un desfase con respecto al MAX30208, las medidas obtenidas del primero son válidas una vez que el transitorio térmico inicial de la PCB ha finalizado.

5.2.1.2 Diseño de procesamiento digital de señales

El firmware diseñado para Bindi aprovechó las funcionalidades o Software-Development-Kit (SDK) proporcionadas por el fabricante del microcontrolador, en este caso Nordic Semiconductors®. La figura 5-13 representa una estructura simplificada para la pila incrustada en el Brazalete. Cada parte se describe como sigue:

- Elemento nRF HAL. Forma parte del SDK de Nordic. En concreto, es el Hardware Abstraction Layer (HAL) para las diferentes funcionalidades de bajo nivel del sistema, incluyendo la interfaz directa con el núcleo ARM® los per-

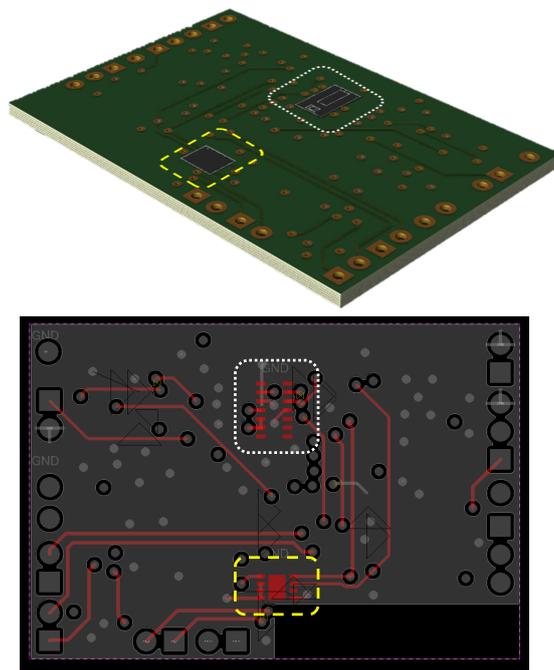


Figure 5-10: Integración de los sensores de temperatura de la piel (círculo amarillo) y de frecuencia cardíaca en la pulsera. La zona gris determina el plano de tierra.

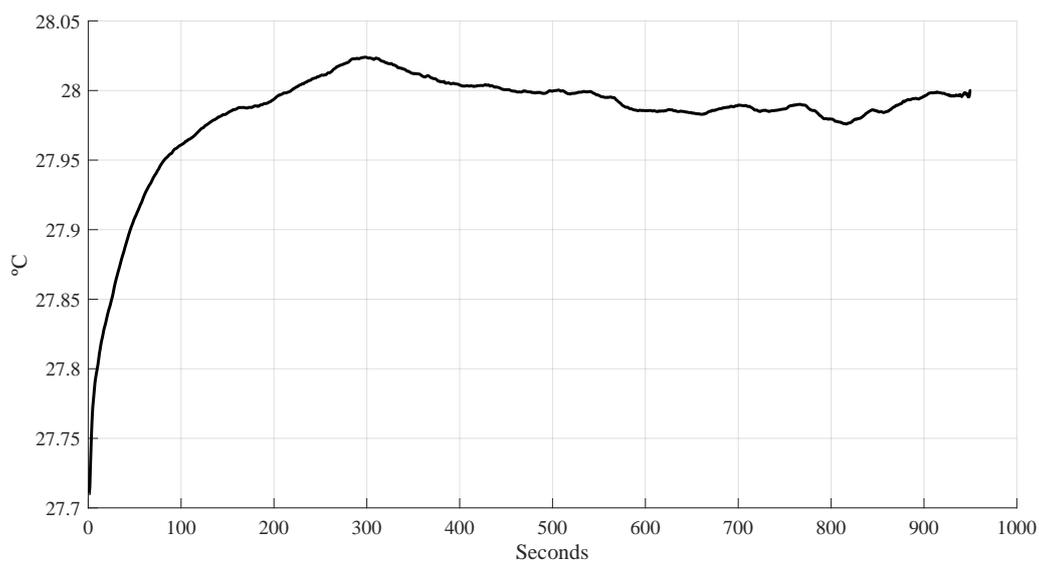


Figure 5-11: Salida filtrada del MAX30205 tras colocar un dedo sobre el chip integrado en condiciones de temperatura ambiente controlada.

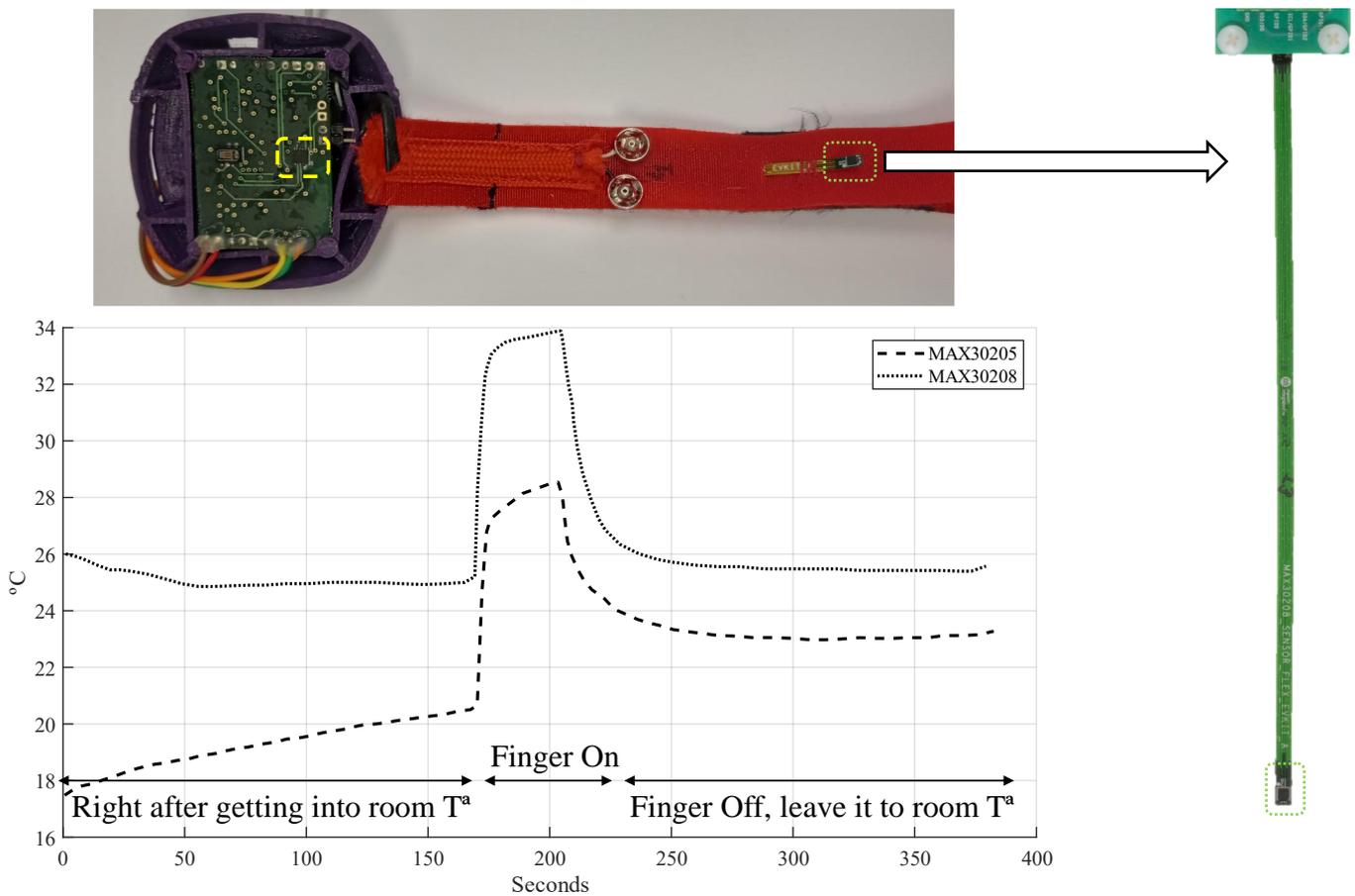


Figure 5-12: Modificación realizada al Brazalete para incluir el MAX30208 y comparación del experimento para ambos sensores de temperatura. A la derecha parte de la placa de evaluación del MAX30208.

iféricos y la radio, entre otros.

- BINDI BLE. Se trata de un sistema gestor ad-hoc que maneja las diferentes colas de transmisión y recepción de radio, realiza el formateo de los paquetes BLE, y gestiona la interacción directa con el softdevice. Nótese que esta última es la pila BLE que se está empleando, que en Bindi 1.0 es la S132 [239] que se basa en BLE 5.1 calificado.
- SYSTEM INIT. Es la parte encargada de gestionar todos los procesos de inicialización relativos a la solicitud de configuración de los periféricos necesarios, así como de la configuración inicial general de los GPIO.
- BINDI HAL. Es una de las partes principales de la pila. Se trata de un HAL ad-hoc a nivel de periféricos, que está específicamente destinado a gestionar todas las diferentes interacciones Bindirelacionadas con los periféricos, realiza la gestión de la adquisición en bruto, lleva a cabo las primeras etapas de filtrado inicial, y procede a segmentar los datos y almacenar los búferes procesados para ser procesados posteriormente por la capa BINDI APP. Además, también se ocupa de la interacción de los actuadores, es decir, de encender y apagar el motor vibrador y de recibir las interrupciones del botón de pánico.
- BINDI APP. Esta capa se encarga de las principales funcionalidades a nivel de sistema, como la gestión de los datos fisiológicos procesados, la extracción de características, el procesamiento principal de la señal digital (DSP) y la clasificación.
- Aplicación de Usuario. Las funcionalidades anteriores a nivel de sistema son manejadas y sincronizadas por una máquina de estado finito (FSM) que reside en esta capa y es modificada de acuerdo a la aplicación específica del usuario.
- CMS Task Handler. Se trata de una funcionalidad cruzada que puede interactuar con toda la pila. Se utiliza sobre todo para decodificar todos los paquetes recibidos (BLE) y desencadenar la respectiva acción requerida en relación con partes específicas de la pila. Esta herramienta también se utiliza para depurar cuando se está en modo de desarrollo.

Teniendo en cuenta las cadenas de procesamiento de datos y siguiendo la segmentación de datos adoptada al tratar el último sistema de detección de miedo propuesto en la Sección 4.2, la Figura 5-14 muestra los diferentes procesos de tem-

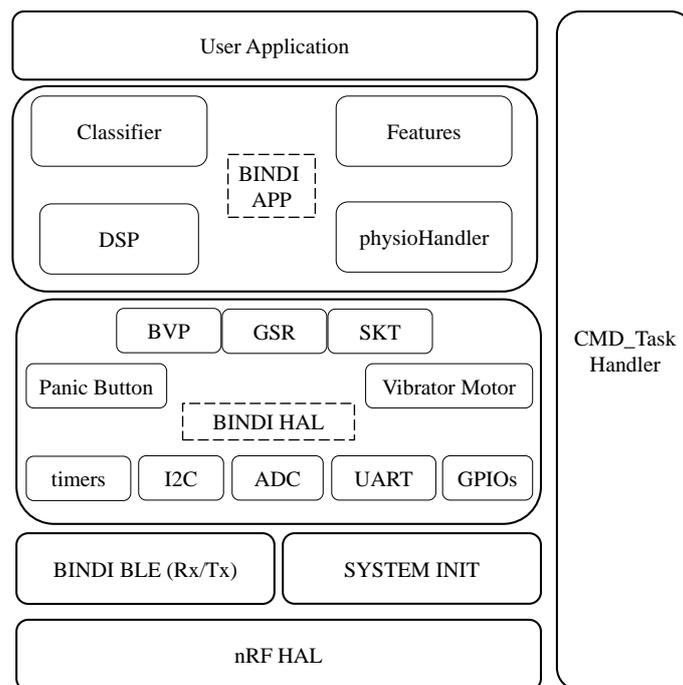


Figure 5-13: Pila actual del firmware del brazalete de Bindi.

porización que se realizan dentro del Brazalete. Con el fin de reducir las operaciones del host y el uso de periféricos internos, en lugar de emplear temporizadores independientes para cada una de las señales fisiológicas, hacemos uso de los tiempos proporcionados por el sensor inteligente PPG. Esto se hace porque este sensor es el que tiene la mayor frecuencia de muestreo, $100Hz$, mientras que el GSR y el SKT trabajan a $10Hz$ y $5Hz$ respectivamente. Así, cada vez que se escribe una nueva muestra del sensor PPG en el buffer BVP, se comprueba si es el momento de muestrear el resto del sensor de forma sincronizada. Este esquema de adquisición fisiológica se repite cada segundo y permite evitar cualquier cálculo complejo de sincronización o deriva temporal. Además, las muestras adquiridas para cada señal están separadas uniformemente. Nótese que esto último es crucial para aplicar correctamente diferentes procesos DSP, como las FFT. Los datos adquiridos se filtran y se almacenan en búferes de 20 segundos, que luego son alimentados a los módulos de extracción y clasificación de características. En la figura 5-14 también se muestra el esquema del proceso de superposición. Una de las principales limitaciones de este esquema de adquisición es la total dependencia del PPGsensor inteligente, ya que, en caso de fallo del sensor, todo el sistema se ve comprometido. Actualmente se están realizando diferentes trabajos para implementar y proporcionar un esquema de medición flexible capaz de hacer frente a eventos de mal funcionamiento. Además,

se está investigando la integración de pruebas en línea dentro del Brazalete para evaluar estos casos [240].

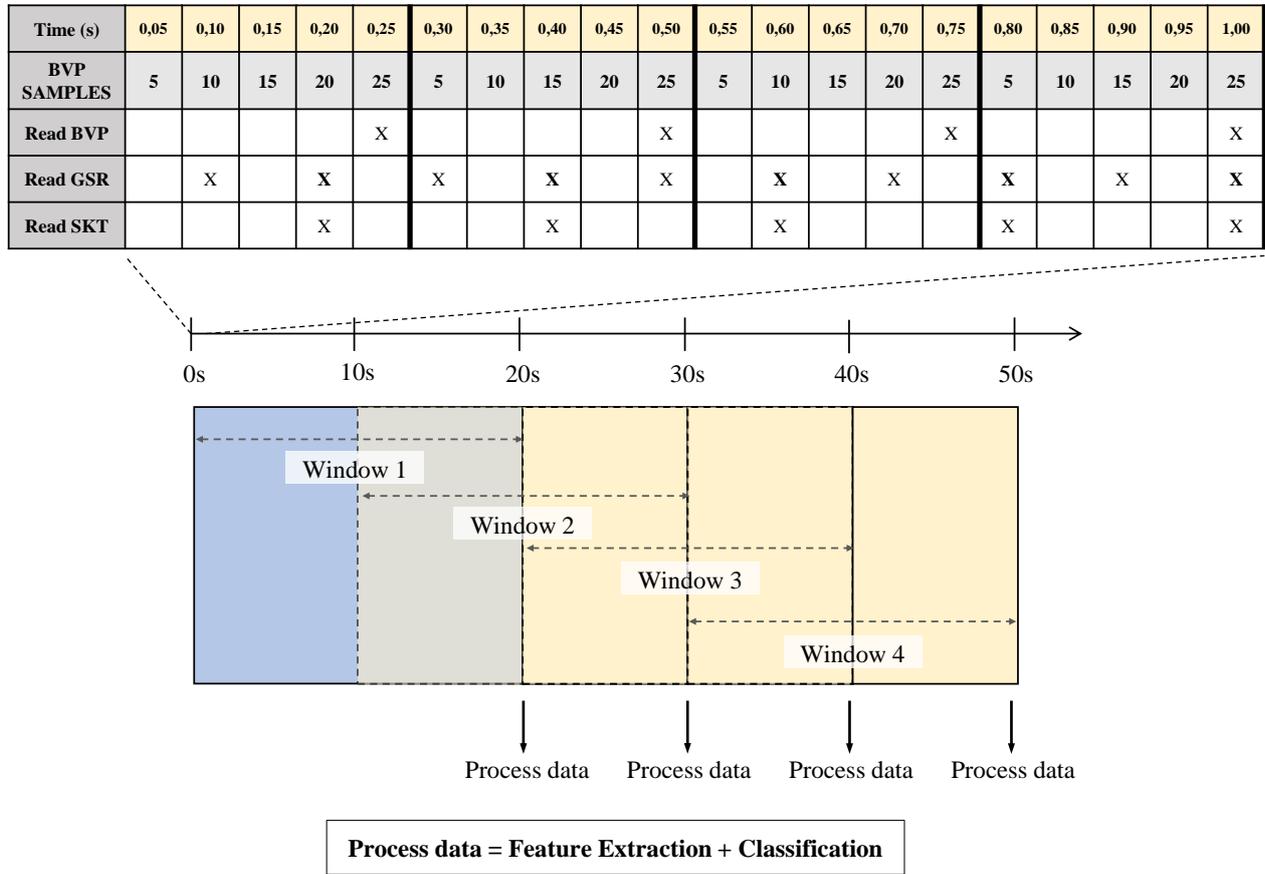


Figure 5-14: Sincronización fisiológica actual y tiempos de procesamiento de datos en el Brazalete.

En el caso de especificar cada uno de los diferentes procesos digitales embebidos que se realizan dentro de cada ventana temporal (20 segundos) y para cada sensor, la Figura 5-15 muestra parte de la arquitectura actual del sistema Brazalete centrada en el flujo de datos a lo largo de dichos procesos principales. Como ya se ha dicho, cada cadena de procesamiento de datos comienza con la recopilación de los datos del sensor respectivo utilizando la adquisición I2C o ADC. Después, los datos son filtrados y se realiza la segmentación (windowing). Algunas de las arquitecturas de filtrado integradas evaluadas e implementadas se explican en la sección 5.2.2. En este punto, los sensores siguen diferentes caminos. Por ejemplo, la implementación actual con respecto a los datos BVP se somete a un proceso de evaluación de la calidad empleando un sistema SQA, que se detalla en la Sección 5.2.3. En este punto, se aplican diferentes algoritmos de eliminación de artefactos de movimiento

para recuperar la mayor parte de la información de la señal si es necesario. Hay que tener en cuenta que estos algoritmos se están desarrollando en la actualidad y, aunque se representan en esta arquitectura, aún no están totalmente implementados. A continuación, se extraen las características de los datos filtrados y segmentados. Centrándonos en la cadena de procesamiento de datos PPG, la sección 5.2.4 detalla algunos de los procesos implicados durante la extracción de características para BVP métricas relacionadas. Por último, las características obtenidas se introducen en el motor de inferencia y la etiqueta resultante se transmite de forma inalámbrica a la APP. Cabe señalar aquí que, aunque en las siguientes secciones se analizan en profundidad algunos de estos procesos digitales, la implementación integrada de todas las cadenas de procesamiento de datos, incluido el bloque de inferencia, es un trabajo actualmente en curso. Por ejemplo, en [183], propusimos una cadena de procesamiento de datos totalmente embebida, desde la adquisición hasta la clasificación embebida, considerando el valor medio de cada variable para una ventana temporal de 10 segundos como etapas de filtrado y extracción de características. Implementamos un KNN ligero y aplicamos el aprendizaje sensible a los costes para entrenar y desplegar un sistema dependiente del sujeto. Este sistema fue una prueba de concepto inicial y sirvió de base para empezar a diseñar y mejorar la siguiente versión. Por este motivo, la discusión de la integración de cualquier clasificador integrado queda fuera del alcance de este documento y será objeto de la investigación que surja de este trabajo. Del mismo modo, la integración de la eliminación de artefactos de movimiento también queda fuera del ámbito de esta investigación.

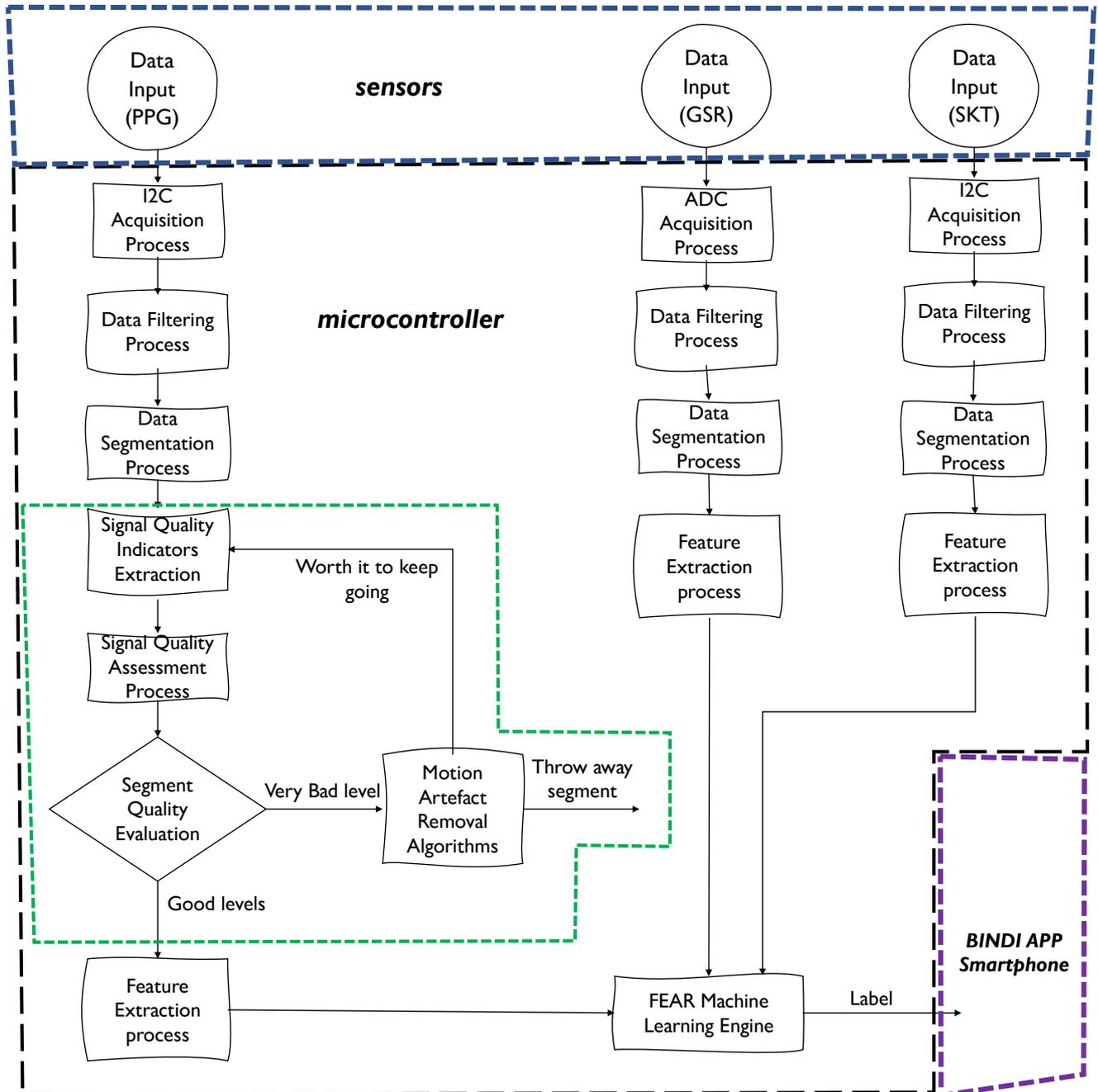


Figure 5-15: Arquitectura actual del sistema para las principales tareas de procesamiento digital del Brazalete.

5.2.2 Evaluación del filtrado integrado

En esta sección, se realiza una evaluación de filtrado incrustado para una etapa de filtrado basada en PPG. Este análisis se extrae de [159]. Teniendo en cuenta la limitación de recursos de Bindi, se han evaluado cinco parámetros cruciales para las diferentes arquitecturas de filtrado consideradas: el uso de memoria, el tiempo de cálculo de la ventana, el tiempo de asentamiento, la atenuación media en la banda de parada y el ripple de paso de banda. Los dos primeros parámetros son los que están directamente relacionados con la implementación del filtro integrado. El tiempo de cálculo también está limitado por los tiempos definidos por la aplicación. El resto de los parámetros están relacionados con las características de filtrado. Por ejemplo, el tiempo de asentamiento es especialmente relevante y denota el tiempo de estabilización del filtro, lo que podría estar relacionado con la pérdida de tiempo y de memoria. La atenuación media en la banda de parada está relacionada con el nivel medio de atenuación con la banda de rechazo designada, mientras que la ondulación de paso de banda es la cantidad de variación en la ganancia dentro del ancho de banda designado del filtro.

Desde una perspectiva embebida o digital, como ya se revisó en el capítulo 4, hay dos técnicas de filtrado comúnmente aplicadas: IIRs y FIRs. Los IIR son computacionalmente rápidos, aunque no tienen una respuesta de fase lineal, lo que podría llevar a no preservar la forma de la onda o la morfología fisiológica. Por ejemplo, este hecho puede dar lugar a que los algoritmos de detección de picos BVP identifiquen puntos erróneos. Esta desventaja se alivia utilizando una técnica de filtrado IIR hacia delante y hacia atrás, que requiere un doble filtrado y una doble inversión temporal de la señal. Esta última técnica conlleva un elevado tiempo de cálculo a costa de obtener una función de transferencia de fase cero. Por el contrario, los filtros FIR pueden diseñarse para tener una respuesta de fase lineal, preservando así la morfología fisiológica y no afectando a posibles patrones. Sin embargo, requieren más coeficientes y memoria que los IIR. Estas y otras técnicas digitales se utilizan para hacer frente a los ruidos fuera de banda, como la oscilación de la línea de base y el ruido de alta frecuencia. En el caso de una señal BVP, el rechazo de estos ruidos es clave para minimizar adecuadamente los cambios en su morfología que no tienen un origen cardíaco.

Las cuatro opciones de diseño de filtros consideradas son: tres filtros FIR pasa-banda con diferentes órdenes y un filtro de dos etapas basado en el promedio móvil. Por un lado, los coeficientes resultantes se cuantificaron a un entero de 14 bits para reducir el uso de memoria y aumentar el tiempo de procesamiento. Este número de bits es la máxima precisión que garantiza que no haya desbordamiento en nuestro sistema, con señales BVP de 18 bit/muestra y registros de 32 bits. El impacto de la respuesta en frecuencia es mínimo y la desviación cuadrática de la raíz de la salida en comparación con los coeficientes de punto flotante de 64 bits es insignificante. Por otro lado, el filtro de dos etapas se compone de dos pasos de promedio móvil. El primero es un filtro de paso bajo de 4 muestras, mientras que el segundo es la sustracción de la señal de los 100 valores, media móvil centrada.

En la tabla 5.1 se muestran los resultados obtenidos para las arquitecturas de filtrado embebidas evaluadas. Analizando esta tabla, podemos observar que, para los filtros pasa-banda, el aumento del orden del filtro (el número de coeficientes) aumenta la atenuación media de la banda de parada, pero también el uso de memoria ROM, el tiempo de cálculo y el tiempo de establecimiento. El tiempo de cálculo de estos filtros pasa-banda podría reducirse si los coeficientes se almacenan en la RAM a expensas del uso de la memoria. Obsérvese que el tiempo de cálculo del filtro de dos etapas es significativamente menor que el de los filtros pasa-banda.

En cuanto a la atenuación de la banda de parada, este parámetro beneficia a los filtros pasa-banda, proporcionando una mayor atenuación en toda la banda de parada. Obsérvese que la relación más eficiente de atenuación en la banda de parada frente al uso de memoria de la ROM se alcanza con el filtro pasa-banda de coeficiente 400, debido al efecto de tamaño de código constante. Centrándonos en el ripple de paso de banda, se desea un valor bajo para evitar la deformación de la señal. En el caso de las arquitecturas de filtrado propuestas, el ripple de paso de banda de los filtros no provoca ninguna distorsión en la señal. En cuanto al tiempo de asentamiento, la diferencia entre los filtros de dos etapas y los de paso de banda es grande, lo que beneficia a estos últimos.

En general, a partir de todo este análisis, se recomienda el filtro de dos etapas. Este posee un buen equilibrio entre el tiempo de cálculo, la atenuación y el uso de memoria para un sistema portátil limitado como Bindi. Aparte de estas decisiones

Table 5.1: Resultados obtenidos para las arquitecturas de filtrado integradas evaluadas.

Diseño Opciones	Tiempo. Cómputo [ms]	RAM [bytes]	ROM [bytes]	Set. Time [samples]	Media Banda parada att. [dB]	Paso banda ripple [dB]
400-coef	0.2474	10	626	400	-38.8	0.09
200-coef	0.1240	10	426	200	-25	0.64
100-coef	0.0623	10	326	100	-14.9	3.09
2-stage	0.0048	20	470	4	-9.3	1.93

de diseño respaldadas por estas métricas, cabe señalar que, desde un punto de vista fisiológico, el número de coeficientes asociados al tiempo de asentamiento puede afectar negativamente a la monitorización fisiológica. Este hecho está motivado por el número de muestras que hay que extraer para cada arquitectura considerada, lo que afecta negativamente a la cantidad final de información fisiológica de la que extraer las diferentes características. Nótese que, aunque nos centramos únicamente en esta señal, algunas de las conclusiones extraídas pueden ser extrapoladas para las otras dos etapas de filtrado que se abordan dentro del sistema (GSR y SKT).

5.2.3 Signal Quality Assessment

SQA es un proceso clave para la monitorización fisiológica continua y fiable [241]. En concreto, este tipo de procesos benefician en gran medida a Bindi ya que se centran en la evaluación de la calidad de la señal mediante diferentes características extraídas de la misma y la regla de decisión. Así, estos sistemas proporcionan una medida de la calidad de la señal segmentada que se procesa. Nótese que este sistema no se ocupa de ninguna tarea de eliminación de artefactos de movimiento o similares. Esta salida de calidad de la señal puede ser utilizada posteriormente por los diferentes algoritmos de extracción de características o incluso por el aprendizaje automático del miedo para ajustar o ponderar adecuadamente la calidad de dicha instancia temporal. En cuanto a sus diferentes etapas, está formado por hasta tres procesos principales:

- La primera es la etapa de extracción de características o indicadores de calidad de la señal (SQI). Se extraen diferentes SQI del segmento de la señal para caracterizarlo adecuadamente. Hay que tener en cuenta que las características o SQIs adecuados son los que cambian entre los segmentos limpios y los ruidosos de la señal.
- Siguiendo el proceso anterior, las características extraídas se evalúan en base

a diferentes reglas de decisión para cuantificar el nivel de ruido.

- El resultado de esta última etapa es el índice de calidad de la señal (SQ_i), que en la mayoría de los casos es de base binaria. Cuando se utilizan diferentes fuentes de la misma señal o incluso diferentes señales, se realiza una tercera etapa de fusión de datos. En dicha etapa, los SQ_is individuales se combinan para obtener la métrica de calidad final.

Hay que tener en cuenta que el proceso de filtrado y la segmentación de datos de la señal no están dentro del ámbito de las tareas del SQA; sin embargo, es necesario filtrar y segmentar la señal antes de la aplicación del SQA. Al igual que en la anterior evaluación de filtrado embebido, el sistema SQA presentado, [212], también se centra en las señales PPG debido a su relevancia e importancia dentro de Bindi. Cabe destacar que el trabajo presentado en esta sección es el resultado de una colaboración internacional con la Universidad de Essex [212].

En la literatura, la mayoría de los métodos de incrustación de recursos bajos propuestos por PPG SQA comparten las siguientes características:

- Se basan en reglas de decisión con umbrales duros para evaluar el SQ_i. Esta metodología obvia la alta incertidumbre como resultado de las diferencias entre sujetos o intra-sujetos, como los niveles de ruido variables a través del tiempo.
- Consideran una gran cantidad de datos de entrenamiento o de ajuste de umbral utilizando una combinación de diferentes conjuntos de datos. Sin embargo, el número real de conjuntos de datos públicos que contienen anotaciones sobre la calidad de la señal es escaso, lo que obliga a los investigadores a etiquetar los datos utilizados.
- Los sistemas propuestos se adaptan al conjunto de datos específicamente etiquetados, lo que da lugar a un sistema dependiente del experimento que dificulta la consecución de una generalización suficiente para hacer frente a diferentes escenarios experimentales.

Siendo conscientes de que la generación de conjuntos de datos anotados es una tarea difícil, se podría aprovechar una validación o un ajuste de consideración de pocos disparos junto con un autoajuste posterior en línea para el diseño de sistemas heterogéneos que puedan hacer frente a la baja cantidad de datos anotados disponibles. Hay que tener en cuenta que este tipo de perspectiva de diseño también puede apli-

carse a los sistemas que se espera que sean entrenados o ajustados en base a los datos en la naturaleza y a las anotaciones diarias de los voluntarios, ya que en estos experimentos se espera que las anotaciones recogidas sean escasas. Además, son escasas las investigaciones anteriores que hayan realizado una implementación integrada de SQA y hayan presentado diferentes compensaciones a tener en cuenta en el diseño. Sobre esta base, en esta sección se presenta un novedoso sistema de SQA integrado e invariante del sujeto que utiliza un conjunto reducido de características combinado con un sistema basado en reglas difusas de intervalo (FRBS). Este sistema es el actual SQA que se ejecuta en el Brazalete. En concreto, para hacer frente a la generalización y el ajuste del SQA procedentes de la amplia casuística de la señal PPG, se implementa un sistema difuso de tipo 2, ya que proporciona un mejor marco de incertidumbre para su aprovechamiento. Además, se propone y aplica una etapa de ajuste fino adaptativo para autoajustar el FRBS de forma *online*, lo que proporciona una adaptación agnóstica al usuario.

Centrándonos en la etapa de extracción del SQI para los sensores PPG, existe una clara división entre las metodologías en el dominio del tiempo y en el dominio de la frecuencia. Las primeras representan las técnicas más comunes utilizadas en los sistemas PPG-SQA en la literatura. Por ejemplo, en [242] se estudió el comportamiento estadístico de diferentes SQIs basados en la tendencia. En concreto, se probaron siete indicadores (perfusión, kurtosis, asimetría, potencia relativa, relación señal/ruido, cruces de cero y entropía) utilizando 160 registros de 60 segundos cada uno, un total de 9600 segundos. En los resultados presentados, skewness superó a los demás SQIs al conseguir una puntuación F1 de hasta el 87,20% en la detección de pulsos aceptables y no aceptables. Esta publicación definió tres niveles diferentes de calidad en lugar de la habitual clasificación binaria. Independientemente de la ventaja que supone la baja complejidad computacional de estos SQIs basados en tendencias, el diseño de un SQA basado pura y exclusivamente en estas métricas queda expuesto a las reglas de decisión heurísticas con umbrales duros. En cuanto a los sistemas de SQA basados en la extracción de características en el dominio de la frecuencia, Krishnan et al. en [243] utilizaron el espectro de la asimetría de la señal (bi-espectro) para explotar las relaciones de fase que existen en una señal PPG limpia. Estos métodos implican un alto esfuerzo computacional en compara-

ción con otros basados en el dominio del tiempo que no requieren realizar algoritmos de Transformada Rápida de Fourier (FFT) ni ninguna transformación de bases. Por otra parte, el desarrollo de algoritmos de aprendizaje profundo y de máquina ha dado lugar a sistemas de clasificación que detectan automáticamente las diferentes anomalías dentro de la señal PPG de una manera más robusta [244]. Sin embargo, no están libres de reglas de decisión determinadas empíricamente, y el enfoque de aprendizaje profundo dificulta una implementación óptima integrada.

Centrándonos en los sistemas de SQA propuestos en la literatura que fueron embebidos, podemos destacar tres trabajos recientes. En [245], Vadrevu et al. propusieron uno de los primeros sistemas PPG SQA en tiempo real mediante la extracción de características en el dominio del tiempo. Aplicaron seis reglas heurísticas predefinidas para evaluar la calidad de la señal y utilizaron un microcontrolador ARM Cortex-M3 de 32 bits. Combinaron dos bases de datos públicas diferentes de PPG de referencia con su propio conjunto de datos. Esta combinación de datos se utilizó tanto para el ajuste del umbral como para la validación del rendimiento. Finalmente, lograron hasta un 95,93% de precisión global. Aunque mostraron datos de consumo de energía competitivos en relación con el efecto de la disminución de la retención de datos y la implementación integrada de SQA, su sistema seguía estando sujeto al ajuste empírico del umbral. Este hecho adaptó el sistema propuesto a ese conjunto específico de umbrales estimados. Además, no realizaron ninguna prueba ciega. De forma similar, en [246], Reddy et al. propusieron el uso de características en el dominio del tiempo con un conjunto de reglas y umbrales empíricos. También combinaron diferentes bases de datos públicas de PPG de referencia, pero las dividieron en dos conjuntos de datos. Uno de ellos se utilizó para el ajuste de umbrales y el otro para las pruebas. Implementaron el sistema en el mismo microcontrolador que Vadrevu et al. y lograron una precisión global del 93,21%. Por último, en [247], Samiul Alam et al. emplearon la función de Kurtosis y autocorrelación con umbrales empíricos también. Siguieron la misma disposición del conjunto de datos que Reddy et al., y lograron hasta un 96,50% de precisión global. Además de verse afectados por la misma consideración de adaptación empírica comentada, utilizaron una plataforma integrada de alto rendimiento (ARM Cortex-A53 de cuatro núcleos). Esto último dificulta la tarea de comparación con un contexto extremo de edge-computing que

tiene que ver con los dispositivos wearables. Entre las ventajas y desventajas comentadas de estos sistemas, cabe destacar dos factores. En primer lugar, el conjunto completo de características utilizadas en estos trabajos era específico del dominio, lo que requiere un cierto conocimiento previo de la naturaleza del tipo de ruido que debe detectarse. En segundo lugar, todos los sistemas propuestos se ajustaron o entrenaron utilizando el mismo conjunto de datos o parte de una combinación de diferentes conjuntos de datos. Este último hecho es especialmente relevante debido al reto de heterogeneidad detallado anteriormente, ya que conseguir un sistema SQA aplicable a un amplio abanico de situaciones y actividades de la vida real requiere no sólo tener en cuenta a diferentes voluntarios sino también realizar pruebas ciegas con diferentes bases de datos.

Después de haber revisado los sistemas de SQA para la monitorización de PPG, podemos concluir que no existe un conjunto común general de técnicas para tratar este problema, sino diferentes metodologías de dominio e incluso la combinación de las mismas. Además, independientemente de la naturaleza de dichas técnicas de extracción de características o de los algoritmos de clasificación, los sistemas presentados en la literatura recurren a enfoques de umbral duro. Esto hace que el sistema se adapte al conjunto de datos de entrenamiento debido a esas decisiones heurísticas. Cuando se buscan otros tipos de SQAs que traten de superar tales limitaciones y se ocupen de la generalización, se encuentran algunas investigaciones que aplican un sistema de lógica difusa (FLS) de tipo I [248]. Sin embargo, el SQA fisiológico se convierte en un reto cuando se tienen escenarios heterogéneos. Por lo tanto, el FLS de tipo I está limitado en cuanto a la cantidad de incertidumbre a la que puede hacer frente.

Por esta razón, y con vistas a la aplicación de SQA heterogéneos, en esta investigación se explota un conjunto reducido de características específicas y agnósticas del dominio con un FLS de tipo II de intervalo, específicamente un clasificador basado en reglas difusas (FRBC). Obsérvese que la técnica de tipo II está pensada específicamente para tratar la incertidumbre comentada, ya que cada nivel de las características se fuzzifica basándose en un conjunto difuso de intervalo denominado Huella de Incertidumbre (FOU) [249].

5.2.3.1 Diseño, formación y validación de SQA

La figura 5-16 muestra la arquitectura de formación SQA utilizada en esta investigación. En concreto, esta arquitectura se compone de siete procesos diferentes. Los siguientes subapartados ofrecen una visión general de cada una de las etapas de esta arquitectura.

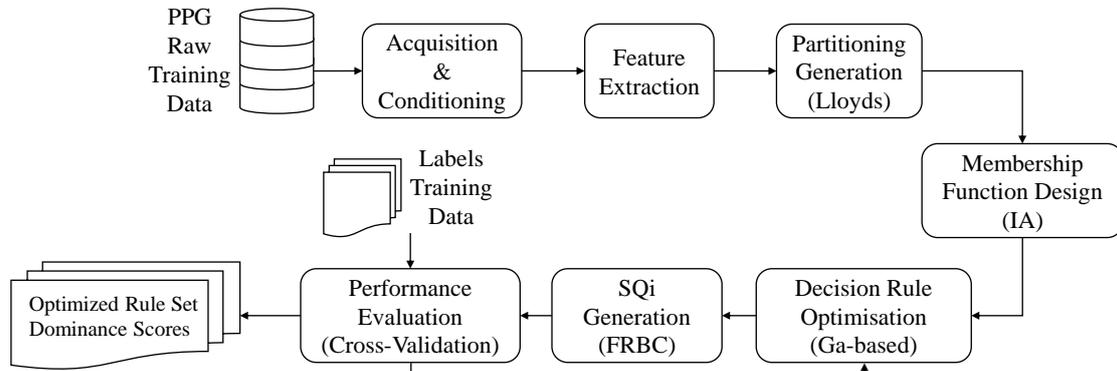


Figure 5-16: Arquitectura de formación SQA propuesta.

Adquisición y acondicionamiento

Una vez filtrada la señal siguiendo la arquitectura de filtrado integrada seleccionada anteriormente, se procede a la segmentación. Este proceso se basa en el hecho de que realizar la extracción de características en pequeños trozos de datos aliviará los diferentes procesos estadísticos a realizar (por ejemplo, los cálculos de complejidad temporal de la media o desviación estándar se basan en la cantidad de datos o muestras, es decir, $\mathcal{O}(n)$). En nuestro caso, para el SQA propuesto, la longitud de la ventana segmentada se establece en 3 segundo. Esta duración específica puede proporcionar dos períodos de frecuencia cardíaca (HR) para un mínimo de 40 latidos por minuto (BPM). Además, dentro de este corto periodo de tiempo, podemos considerar incluso un comportamiento cuasi-estacionario de esta señal fisiológica. Obsérvese que, a medida que disminuimos la ventana de procesamiento, también disminuye el uso de recursos dentro de un sistema integrado, pero aumenta el mínimo de BPM al que podemos asegurar dos períodos de la señal. Este hecho nos lleva a una decisión de compromiso que en nuestro caso viene dada por los hechos fisiológicos comentados y por trabajos anteriores que utilizaban longitudes de ventana temporal iguales o similares [247].

Feature extraction

A continuación, se aplican diferentes técnicas de extracción de rasgos para caracterizar el procesamiento de la ventana actual. En concreto, se extraen cuatro características. Obsérvese que todas las características aplicadas se basan en el tiempo. Esta decisión se debe a su menor complejidad computacional y a su robusto rendimiento, demostrado en publicaciones recientes [242]. Las diferentes características extraídas se detallan a continuación:

- **Curtosis.** Es la métrica estadística relacionada con la forma de una distribución de probabilidad que mide el grado de concentración que se presenta alrededor de la media de la distribución de frecuencias para una variable de valor real. También se describe como la medida de la cola. Esta medida estadística de orden superior viene dada por la ecuación 4.18.
- **Entropía.** La Entropía de Shannon proporciona una medida cuantitativa con respecto a la incertidumbre o aleatoriedad de la señal. Esta característica se define como:

$$e = - \sum_{i=1}^N (x_i^2) \log(x_i^2), \quad (5.2)$$

donde N es el tamaño de la muestra, y x_i es cada una de las muestras de datos filtrados.

- **Relación señal/ruido (SNR).** Se trata de una de las características más utilizadas en los sistemas SQA. Compara la potencia de una señal deseada con respecto al ruido observado. En este caso, se realiza el siguiente cálculo:

$$snr = \frac{\sigma_{abs(x)}}{\sigma_x}, \quad (5.3)$$

donde $\sigma_{abs(x)}$ es la desviación estándar del valor absoluto de la señal, mientras que σ_x es la desviación estándar de la señal.

- **Perfil de la matriz.** Hasta donde yo sé, esta característica no se ha utilizado para ningún sistema PPG SQA en la literatura, aunque se utiliza ampliamente en la detección de anomalías de series temporales [250,251]. Esta métrica ofrece diferentes ventajas que pueden proporcionar un SQI robusto y fiable, como el agnosticismo de dominio, el tiempo determinista y la ausencia de parámetros. La ecuación de trabajo para el perfil de la matriz se basa en el perfil de distancia

dado por las distancias euclidianas normalizadas de puntuación Z de diferentes subsecuencias dentro de la serie temporal:

$$d_{i,j} = \sqrt{2m \left(1 - \frac{Q_{i,j} - \mu_i \mu_j}{m \sigma_i \sigma_j}\right)}, \quad (5.4)$$

donde $Q_{i,j}$ es el producto punto de las dos subsecuencias con longitud m ($T_{i,m}$ y $T_{j,m}$) de la serie temporal, y μ y σ son la media y la desviación estándar de la subsecuencia respectiva. Obsérvese que para este trabajo de investigación se calcula la media sobre el conjunto de valores almacenados en $d_{i,j}$ y se asigna a cada ventana de procesamiento de 3 segundos. En cuanto al algoritmo específico, utilizamos SCRIMP++, que ofrece la menor complejidad temporal entre las diferentes implementaciones posibles [252].

Después de extraer el conjunto completo de características para todos los diferentes sujetos, se aplica un controlador automático de ganancia (AGC) para limitar la amplitud y escalar la información extraída. En este caso, utilizamos un AGC de 0 – 10.

Cuantificación y generación de particiones

Siguiendo un enfoque totalmente orientado a los datos, este trabajo de investigación utiliza la cuantificación y la partición de datos sobre los datos de entrenamiento considerados para generar los diferentes conjuntos difusos de forma no supervisada. Así, estos procesos son esenciales para evaluar los límites de las representaciones lingüísticas conceptuales definidas para cada característica y modelar adecuadamente las diferentes funciones de pertenencia. Esto se hace para evaluar si existe una partición o separación de los valores de las características en función de su distribución.

En concreto, en este caso, aplicamos el algoritmo cíclico de Lloyd [253] para optimizar las diferentes particiones utilizando las características revisadas y apuntando a la extracción de tres variables lingüísticas: Baja (L), Media (M) y Alta (H). El algoritmo de LLoyd se ejecuta en un proceso iterativo para cada secuencia o rasgo entrante, $A_1, A_2, A_3, \dots, A_m$, dirigiéndose a una mínima distorsión cuadrática media o error cuadrático medio para las particiones generadas $B_1, B_2, B_3, \dots, B_m$. Debido al sistema de lógica difusa que se aplicará dentro del sistema propuesto, la salida de esta etapa debe ser las particiones o intervalos cuantificados para cada secuencia o

característica entrante. Obsérvese que, al tratarse del diseño de un sistema SQA invariante de sujeto, la cuantificación y optimización de la partición se aplica de forma independiente para cada sujeto, lo que da un conjunto de m particiones o intervalos individuales (ν_i) con $m - 1$ puntos finales (τ). Esto se describe en la figura 5-17. Las particiones generadas son consideradas posteriormente por la siguiente etapa

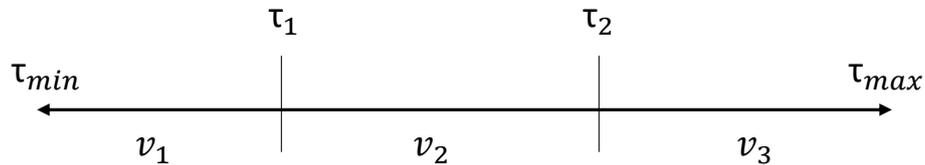


Figure 5-17: Representación de intervalo con tres (m) particiones (ν) y dos puntos finales encontrados (τ). Los valores τ_{min} y τ_{max} son el mínimo y el máximo de la secuencia entrante que se evalúa o los puntos finales izquierdo y derecho.

de la arquitectura de entrenamiento de SQA propuesta para diseñar las diferentes funciones de membresía a implementar.

Diseño de funciones de pertenencia de tipo II

A partir de los intervalos o particiones generados y de los puntos finales, utilizamos la metodología de Aproximación a los Intervalos (IA) [254] para diseñar las funciones de pertenencia iniciales así como las FOU's. Esta técnica se aplica sobre el conjunto completo de particiones de características $[\nu_1, \nu_m]$ y se basa en dos procesos diferentes. En primer lugar, se aplica un paso de preprocesamiento para el conjunto completo de intervalos. Este paso se basa en cuatro etapas. La primera etapa aplica una comprobación de saturación para garantizar que todas las características están dentro del intervalo. A continuación, las dos etapas siguientes se ocupan de la detección de valores atípicos. Por un lado, se utiliza una prueba de Box y Whisker para eliminar posibles valores atípicos fuera de ciertos criterios de límite intercuartil, mientras que, por otro lado, se aplica un tratamiento de límites de tolerancia para comprobar que cada punto está contenido dentro de un rango específico con respecto a la media y las desviaciones estándar de los puntos finales izquierdo y derecho y los intervalos. En la última etapa, se realiza un tratamiento de intervalos razonables. Éste se basa en definiciones o requisitos específicos que deben cumplir los distintos intervalos. Por ejemplo, un intervalo no válido o no razonable es aquel que no se solapa con otro intervalo de datos. Por último, tras las etapas de preprocesamiento, se lleva a cabo la segunda etapa de la técnica de IA. Comprende también diferentes etapas, desde la asignación de un intervalo a una función de pertenencia inicial de

tipo I hasta el cálculo de un modelo matemático para las FOU finales propuestas. Por ejemplo, la Figura 5-18 muestra el resultado final de las funciones de pertenencia y FOUs generadas utilizando los datos de entrenamiento de la característica del perfil de la matriz.

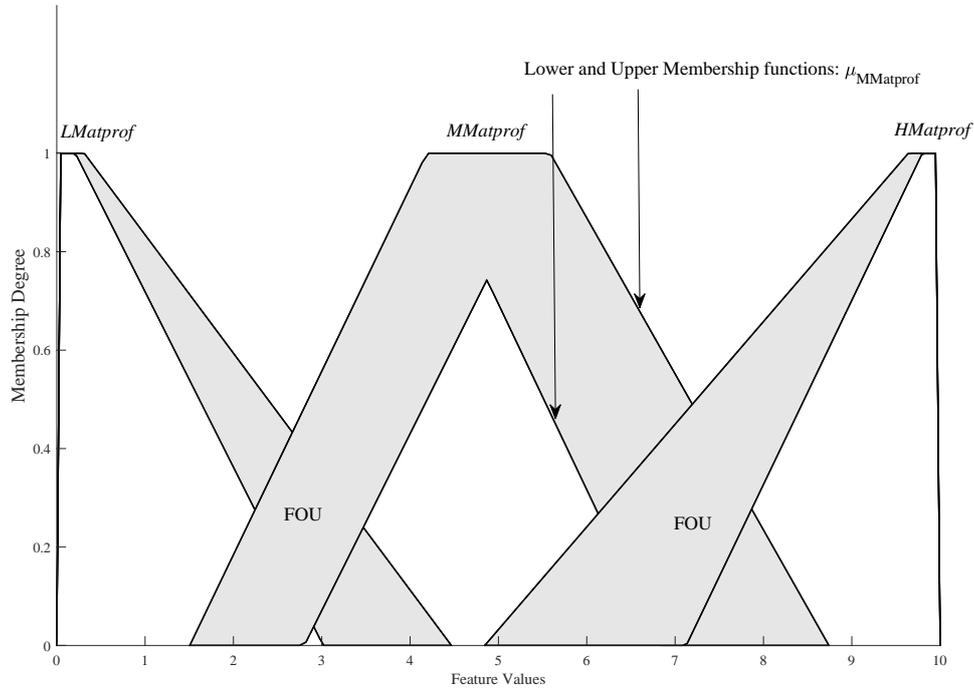


Figure 5-18: Funciones de pertenencia de tipo II generadas a partir de los datos de características del perfil de la matriz aplicando IA para todos los sujetos de entrenamiento. Tres variables lingüísticas: Baja (L), Media (M), Alta (H). El área sombreada en gris es la FOU obtenida.

Formalmente, todo conjunto difuso de tipo II o concepto lingüístico está definido por una función de pertenencia que viene dada por la siguiente ecuación 5.5:

$$\tilde{A} = \{(x, u, f_x(u)) | \forall x \in X, \forall u \in [\underline{\mu}_{\tilde{A}}(x), \bar{\mu}_{\tilde{A}}(x)] \subseteq [0, 1]\}, \quad (5.5)$$

donde x es el universo del discurso contenido en X , u es el valor de pertenencia primario, $f_x(u)$ es el valor de pertenencia secundario, y $\mu_{\tilde{A}}$ representa las respectivas funciones de grado de pertenencia inferior ($\underline{\mu}_{\tilde{A}}(x)$) y superior ($\bar{\mu}_{\tilde{A}}(x)$) del concepto lingüístico \tilde{A} . En concreto, para un sistema difuso de tipo intervalo II, $f_x(u)$ se simplifica como

$$f_x(u) = 1, \forall x \in X, \forall u \in [\underline{\mu}_{\tilde{A}}(x), \bar{\mu}_{\tilde{A}}(x)] \subseteq [0, 1]. \quad (5.6)$$

Así, dado un universo discreto de discurso e independientemente de la forma de

pertenencia, se almacenan ocho puntos para cada conjunto difuso, es decir, cuatro puntos x por funciones de pertenencia (inferior y superior) que delimitan dicho universo. Nótese que en nuestro caso, debido al factor de escalado de características anterior, $X \subseteq [0, 10]$.

Optimización de las reglas de decisión

Uno de los objetivos finales de esta arquitectura de entrenamiento es generar el conjunto de reglas óptimas para integrar en el FRBC integrado. Nótese que cada regla se conceptualiza con la siguiente nomenclatura:

$$R_j : IF \phi_a \text{ is } \lambda_b \text{ and } \dots \text{ and } \phi_c \text{ is } \lambda_d \text{ then } Y \text{ is } \gamma_{n/c} \quad (5.7)$$

donde $a \neq c$, ϕ son los diferentes antecedentes o características contenidas en la regla j , λ son las variables lingüísticas activadas para cada antecedente, y γ es el respectivo consecuente de la regla. Obsérvese que para este trabajo de investigación, el FRBC implementado se basa en una salida binaria, que conduce a dos clases o consecuentes diferentes, es decir, clase positiva para los segmentos ruidosos (γ_n) y clase negativa para los segmentos limpios (γ_c).

Para lograr el propósito de esta etapa, se integra un algoritmo genético evolutivo (GA) y se utiliza para identificar las reglas que en conjunto dan los mejores resultados de clasificación. Nótese que esta optimización del entrenamiento se ha utilizado en trabajos anteriores para diferentes aplicaciones [255]. En concreto, en este caso, el AG se fija en un máximo de generaciones (es decir, el número máximo de iteraciones antes de que el algoritmo se detenga) y el tamaño de la población (es decir, el número de soluciones factibles) hasta 50, utiliza una función de selección de torneo, y emplea un cruce de un punto para la combinación de cromosomas. Nótese que la tolerancia del AG se fija en $1 * 10^{-5}$. Así, la estructura de cada fenotipo viene dada por

$$\begin{aligned} \rho^j = \{ & \phi_1^1, \phi_2^1, \phi_3^1, \phi_1^2, \phi_2^2, \phi_3^2, \dots, \phi_i^j, \\ & \lambda_1^1, \lambda_2^1, \lambda_3^1, \lambda_1^2, \lambda_2^2, \lambda_3^2, \dots, \lambda_i^j, \\ & \gamma_n, \gamma_c, \dots, \gamma_i \}. \end{aligned} \quad (5.8)$$

Obsérvese que, inicialmente, las reglas se generan de forma aleatoria, el número máximo de antecedentes permitido para cada regla (A_{max}) se fija en tres, y el número

máximo de reglas totales (M) se establece en diez. Estas últimas consideraciones se hacen para garantizar que el conjunto final de reglas sea lo suficientemente completo e interpretable.

Además, dentro de esta etapa, se asigna un Peso de la Regla (RW) a cada regla generada, tanto para los miembros superiores como para los inferiores. Esta puntuación se calcula como se indica en [256], a continuación:

$$\begin{aligned}\overline{RW}_j &= \bar{c}_j \cdot \bar{s}_j \\ \underline{RW}_j &= \underline{c}_j \cdot \underline{s}_j\end{aligned}\tag{5.9}$$

donde c_j y s_j son la confianza de la regla y el soporte de la regla j respectivamente. La primera representa la probabilidad condicional de que un patrón clasifique correctamente una instancia de datos, mientras que la segunda es una medida para cuantificar la cobertura de la regla en el conjunto de datos de entrenamiento. Vienen dados por,

$$\begin{aligned}c_j(\phi_j \Rightarrow \gamma) &= \frac{\sum_{x_t \in \gamma} w_j^s(x_t)}{\sum_{j=1}^M w_j^s(x_t)}, \\ s_j(\phi_j \Rightarrow \gamma) &= \frac{\sum_{x_t \in \gamma} w_j^s(x_t)}{M},\end{aligned}\tag{5.10}$$

donde x_t es cada instancia de datos contenida en el conjunto de entrenamiento, y w_j^s es la fuerza escalada de activación de dichos datos con respecto a cada regla, es decir, el grado de coincidencia de la regla j con la entrada x_t . La fuerza de activación escalada se calcula como:

$$w_j^s(x_t) = \frac{w_m(x_t)}{\sum_{k, Y=\gamma} w_k(x_t)},\tag{5.11}$$

donde $w_m(x_t)$ es la fuerza de activación, y $w_k(x_t)$ es la suma de todas las fuerzas de activación que tienen la misma clase que el consecuente de la regla j . Finalmente, la fuerza de activación se calcula como se indica en la siguiente ecuación:

$$w_j(x_t) = \prod_{z=1}^{A_{max}} \mu_A^z(x_t),\tag{5.12}$$

donde $\mu_{widetilde{de}A}^z(x_t)$ representa el valor del grado de pertenencia de la instancia de

datos x_t para las funciones de grado de pertenencia inferior y superior difusas de tipo II, como se denota en la ecuación 5.5.

Generación del nivel de calidad (SQi)

Durante la evaluación de cada iteración del AG, la aptitud se calcula en base a un conjunto de validación específico. La división entre los conjuntos de entrenamiento y de validación se realizó utilizando diferentes técnicas de CV, hold-out y k-fold. Por un lado, se emplearon diferentes porcentajes de conjuntos de validación para la validación hold-out. En concreto, el sistema se ha entrenado utilizando un hold-out aleatorio y estratificado del 40%, 30%, 20% y 10%. Por otra parte, se utilizaron conjuntos de datos de entrenamiento y validación disjuntos de 5 veces. Estos procesos garantizan que no haya ningún sesgo en la selección de los conjuntos de datos de entrenamiento y validación. Obsérvese que, como la adquisición de segmentos de la señal y la extracción de características no se someten a ningún proceso de solapamiento, no hay flujo de información del aprendizaje de reglas de un conjunto de entrenamiento o pliegue a otros.

En cuanto a la generación de SQi específicas para cada una de las instancias contenidas en el conjunto de validación, se aplican dos métodos. Ambos se basan en el cálculo del grado de asociación con respecto a la regla j que se está evaluando, que viene dado por

$$\bar{h}_j(x_t) = \bar{w}_j^s(x_t) \cdot \overline{RW}_j, \quad \underline{h}_j(x_t) = \underline{w}_j^s(x_t) \cdot \underline{RW}_j, \quad (5.13)$$

donde la fuerza de activación y el RW se obtienen utilizando las ecuaciones 5.12 y 5.9 respectivamente. Así, el grado de asociación global considerando la contribución de las funciones de pertenencia de tipo II superior e inferior para una regla j se calcula como

$$h_j(x_t) = \frac{\bar{h}_j(x_t) + \underline{h}_j(x_t)}{2}. \quad (5.14)$$

Basándose en esta puntuación de clasificación final, el primer método de razonamiento (α) empleado para asignar la clase predicha se basa en el método de máxima coincidencia seleccionando el consecuente de la regla con el máximo grado de asociación. El segundo método (β) se basa en el grado de asociación máximo de la agregación de todos los grados de asociación que tienen el mismo consecuente.

Obsérvese que, como el resultado del sistema es binario, este último se traduce en el máximo entre dos grados de asociación acumulados. En caso de empate, clasificamos aleatoriamente la clase predicha para ambos métodos. Por lo tanto, estos procesos pueden expresarse como

$$Y_\alpha = \gamma_j \Rightarrow \max_{j \in [1, M]} (h_j(x_t)), \quad (5.15)$$

$$Y_\beta = \gamma_j \Rightarrow \max_{\forall k \in j} \left(\sum_{k, Y=\gamma_n} h_k(x_t), \sum_{k, Y=\gamma_c} h_k(x_t) \right), \quad (5.16)$$

donde Y_{alfa} y Y_{beta} son la clase predicha obtenida con el primer y segundo método de razonamiento respectivamente. Hasta donde yo sé, es la primera vez que se propone, valida y aplica el segundo método de razonamiento.

Métodos de evaluación del rendimiento

Por último, la evaluación del rendimiento de cada iteración validada de forma cruzada se realiza a lo largo del coste calculado mediante el Coeficiente de Correlación de Mathew (MCC) como sigue:

$$cost = 1 - \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (5.17)$$

donde TP , TN , FP y FN son los verdaderos positivos, los verdaderos negativos, los falsos positivos y los falsos negativos obtenidos de la matriz de confusión utilizando las etiquetas predichas comparadas con respecto a las etiquetas doradas. Tenga en cuenta que para la validación cruzada de 5 pliegues, el coste se calcula como la media de todos los costes del conjunto de datos de validación de pliegues. Una vez recuperado el coste, el AG compara dicho valor con un criterio de tolerancia predefinido. Si el coste es mayor que la tolerancia del AG, éste rellena un nuevo conjunto de reglas, y el proceso se repite hasta que se cumpla el criterio de tolerancia del AG. Además del MCC, también se utilizan otras métricas para seguir comparando las distintas validaciones cruzadas. Estas métricas son: la sensibilidad, la especificidad, la media geométrica entre la sensibilidad y la especificidad (Gmean) y la precisión (ACC).

5.2.3.2 Implantación de SQA y autoajuste

En cuanto a las diferencias entre las arquitecturas online (embebidas) y offline, cabe destacar que en la primera no se realiza ningún tipo de partición, generación

de miembros ni optimización de reglas, ya que estos procesos ya se realizaron durante el proceso de entrenamiento. La figura 5-19 representa la arquitectura incrustada completa. En primer lugar, la adquisición, el acondicionamiento y la extracción de características siguen el mismo esquema que en la sección 5.2.2. En segundo lugar, con respecto al FRBC, el conjunto optimizado de reglas, los RW y los valores de las funciones de pertenencia, que se obtuvieron tras el entrenamiento y la validación, son parámetros codificados. Estos se utilizan específicamente para los cálculos en línea de la fuerza escalada de activación y los grados de asociación para cada nueva instancia de datos. Obsérvese que dichos cálculos siguen las ecuaciones 5.11 y 5.14 respectivamente.

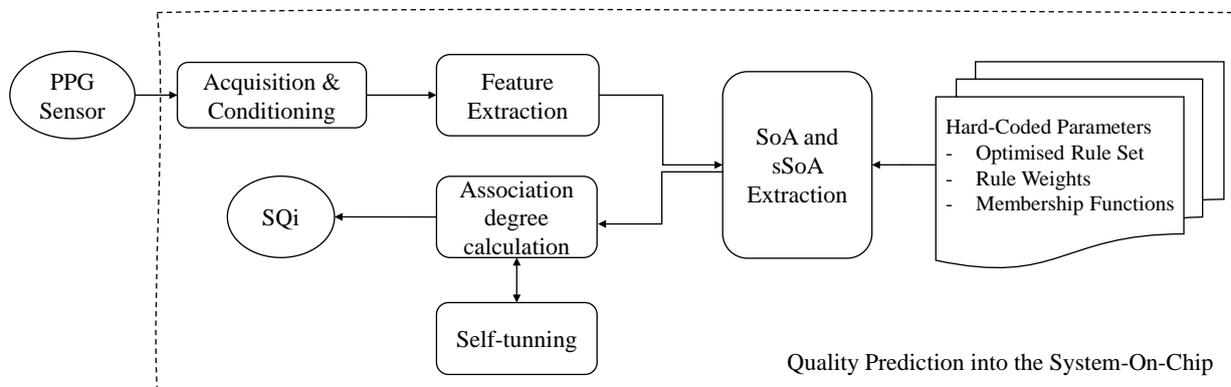


Figure 5-19: Arquitectura integrada SQA implementada. SoA: Fuerza de Activación. sSoA: Fuerza de activación escalada.

La pequeña base de reglas considerada en este trabajo de investigación como requisito para facilitar la interpretabilidad del modelo puede llevar, a largo plazo, a afrontar incertidumbres acumuladas. Por ello, proponemos un FRBC de tipo II autoajutable en línea. En concreto y teniendo en cuenta trabajos anteriores relacionados con los sistemas difusos adaptativos de tipo II [257], el autoajuste implementado se basa en la generación online de nuevas reglas en caso de que la nueva instancia de datos muestre un grado de asociación nulo teniendo en cuenta las diez reglas iniciales, es decir, que ninguna de las reglas existentes se dispare. El algoritmo 1 muestra el proceso implementado para este autoajuste en línea. En primer lugar, los antecedentes y los conceptos lingüísticos activados que conforman la nueva regla se estiman en función de la fuerza máxima de activación de tres, basada en el grado de pertenencia de la nueva instancia de datos, línea 1. Obsérvese que la selección se limita a tres debido al requisito de antecedente máximo impuesto, como se ha indicado en la sección anterior. Así, la nueva regla estará siempre formada por tres

antecedentes con sus respectivos conceptos lingüísticos activados. A continuación, se calcula el grado de similitud mediante el algoritmo de Nguyen [258], línea 2. Este proceso no requiere ejecutar el cálculo de las similitudes en línea, ya que los diferentes grados de similitud entre las diferentes membresías para cada antecedente ya han sido calculados fuera de línea. Pueden existir reglas, dentro de la base de reglas original, que posean menos de tres antecedentes, por lo que, para considerar tales particularidades, las similitudes devueltas se escalan multiplicándose por el número de antecedentes de la regla con la que se compara. A continuación, las RWs para la nueva regla se derivan considerando la matriz de similitudes obtenida, es decir, hay un grado de similitud de la nueva regla con cada regla existente, y las RWs existentes dentro de la base de reglas, líneas 3-4. Finalmente, para estimar el consecuente (γ) o clase de la nueva regla, se selecciona el consecuente de la regla que tiene la máxima similitud con la nueva regla, líneas 7-11. Después de ejecutar este algoritmo, la base de reglas y las RW se actualizan con la nueva información. Además, se repite el mismo proceso de inferencia basado en los grados de asociación, que se realizó antes de la ejecución del algoritmo, asignando la etiqueta correspondiente a la nueva instancia de datos. Para limitar el impacto de este proceso en el sistema, el número máximo de reglas nuevas se fijó en cinco, lo que puede llevar a un máximo de 15 reglas considerando las 10 iniciales.

Algorithm 1: Autoajuste en línea mediante el uso de similitudes de Nguyen.

Input : Dato nuevo (d_{new}); Valores de membresía (MFs);
Reglas actuales, pesos para la parte baja y alta (R, RW_{low}, RW_{up});

Output: Nueva base de reglas, nuevos pesos inferiores y superiores
($R_{new}, RW_{lnew}, RW_{unew}$);

Data: Similitud para la afiliación inferior y superior (s_{low}, s_{up});

Similitud media y posición máxima (s_{mean}, s_{maxPos});

- 1 $R_{new} \leftarrow GetMaxSoA(d_{new}, MFs)$ basado en (5.12);
 - 2 $s_{low}, s_{up} \leftarrow Nguyen(d_{new}, MFs, R, R_{new})$;
 - 3 $RW_{lnew} = \sum_{j=1}^M (s_{low_j} * RW_{low_j}) / \sum_{j=1}^M RW_{low_j}$;
 - 4 $RW_{unew} = \sum_{j=1}^M (s_{up_j} * RW_{up_j}) / \sum_{j=1}^M RW_{up_j}$;
 - 5 Update $RW_{low,up}$ with $RW_{lnew,unew}$;
 - 6 **for** $i \leftarrow 1$ **to** M **do**
 - 7 | $s_{mean_i} = (s_{up_i} + s_{low_i}) / 2$;
 - 8 **end**
 - 9 $s_{maxPos} \leftarrow FindMaxPos(s_{mean})$;
 - 10 $R_{new}(\gamma) = R(s_{maxPos}, \gamma)$;
 - 11 Actualizar R con la nueva regla R_{new} ;
-

5.2.3.3 Herramientas y métodos

Se utilizaron tres conjuntos de datos diferentes para entrenar, validar y probar el SQA propuesto. En primer lugar, llevamos a cabo nuestro propio experimento mediante el cual se recopilaban los datos de entrenamiento y validación de unos pocos disparos. Los datos se extrajeron del experimento explicado en el capítulo 6. En concreto, para el sistema SQA propuesto, se utilizaron unos pocos disparos de 993 segundos de señal PPG registrados a 200 Hz de 10 voluntarios diferentes. Hay que tener en cuenta que los estímulos eran dinámicos en cuanto al movimiento del voluntario, es decir, el voluntario podía moverse sin más restricción que estar sentado. A partir de estos datos, un experto familiarizado con la PPG y los artefactos realizó anotaciones manuales para etiquetar la calidad aceptable e inaceptable de los segmentos de PPG. El etiquetado se evaluó para cada ventana de PPG de 3 segundos no solapada, lo que dio lugar a 331 ventanas con 269 segmentos de PPG aceptables y 62 inaceptables. A continuación, se utilizaron dos conjuntos de datos públicos de referencia para realizar una prueba totalmente ciega, es decir, no se proporciona información sobre los datos de prueba durante el entrenamiento y la validación. El primer conjunto de datos es Capnabase [259], del que se obtuvieron 9120 segundos de señal PPG, mientras que el segundo conjunto de datos fue el Complex Systems Laboratory (CSL) [260] con un total de 7200 segundos de señal PPG. Capnabase fue el primer punto de referencia público para el análisis de la calidad respiratoria y de la PPG y originalmente contiene 42 casos (voluntarios) con grabaciones de PPG de 8 minutos de duración a una frecuencia de muestreo de 300 Hz. Sin embargo, sólo 19 de los 42 casos presentan segmentos de PPG aceptables e inaceptables. Este conjunto de datos proporciona etiquetas de artefactos sin restricción temporal de ventanas. CSL recoge señales de PPG de dos horas de duración de dos voluntarios diferentes, niños de una unidad de cuidados intensivos pediátricos, que se registraron a una frecuencia de muestreo de 125 Hz. Este conjunto de datos también proporciona anotaciones de artefactos sin restricción temporal de ventanas, que se publicaron recientemente en [261]. Cabe destacar dos consideraciones principales en relación con esta colección de bases de datos específica. Por un lado, se utilizó un total de 17313 segundos de señal PPG, de los cuales aproximadamente sólo el 5% se utiliza para la validación y el 95% para la prueba ciega. Por otro lado, la selección específica

de los dos conjuntos de datos de prueba detallados se basó en la disponibilidad de anotaciones. Por lo tanto, estas consideraciones apuntaban al hecho de que el diseño de los sistemas SQA debe proporcionar una generalización suficiente para hacer frente a entornos heterogéneos, como se ha dicho anteriormente.

Para ajustar las etiquetas de los segmentos de PPG proporcionadas por los conjuntos de datos de prueba a la ventana de procesamiento de 3 segundos del sistema propuesto, los datos de prueba se segmentaron en dicha longitud de ventana y la etiqueta de cada ventana fue positiva (segmento inaceptable) en caso de estar dentro de las etiquetas originales o solaparse con ellas. Así, tras este proceso, la cantidad total de segmentos PPG aceptables e inaceptables obtenidos de Capnabase fue de 2909 y 131 respectivamente, y de 2131 y 269 de CSL.

5.2.3.4 Resultados

Esta sección presenta los resultados experimentales relativos a la validación, las pruebas y el funcionamiento en tiempo real del sistema PPG SQA propuesto.

Validación y pruebas

Antes de realizar la validación y los procesos posteriores, se obtuvo la base de reglas inicial mediante la optimización de AG, tal y como se detalla en la sección 5.2.3.1. En la tabla 5.2 se presentan los resultados obtenidos para ambos métodos de clasificación de razonamiento y las técnicas de validación empleadas. Antes de explicar estos resultados cabe hacer varias consideraciones. En primer lugar, no se aplica ningún autoajuste durante la validación. En segundo lugar, se ejecuta un total de 30 iteraciones independientes para cada método de validación cruzada, es decir, las particiones de entrenamiento y validación se seleccionan aleatoriamente para cada ejecución. Esto se hace para proporcionar un valor estadístico. Por último, se obtienen los RW para cada partición de entrenamiento de forma independiente.

En cuanto a los diferentes métodos de razonamiento, se comparan α , como el que considera el máximo grado de asociación entre todas las reglas, y β , como el que considera el máximo grado de asociación entre los grados de asociación agregados para la clase positiva y los de la clase negativa. Se puede observar que τ supera a α alcanzando métricas medias más altas con menos desviación en la mayoría de los casos. Esto se debe al equilibrio o distribución del RW entre las reglas, ya que en este caso las reglas que tienen un consecuente de clase negativo poseen un RW

Reasoning Method	Cross-Validation Method	Validation Performance Metrics				
		Sensitivity $\mu(\sigma)$	Specificity $\mu(\sigma)$	Gmean $\mu(\sigma)$	MCC $\mu(\sigma)$	ACC $\mu(\sigma)$
α	40% Hold-Out	45.19 (7.17)	99.54 (0.47)	66.87 (5.21)	0.61 (0.06)	89.47 (1.37)
	30% Hold-Out	55.78 (16.68)	97.81 (2.29)	72.92 (11.03)	0.64 (0.11)	89.97 (2.67)
	20% Hold-Out	72.91 (11.00)	90.98 (3.98)	81.25 (7.23)	0.62 (0.13)	87.58 (4.56)
	10% Hold-Out	71.27 (14.93)	93.50 (5.10)	81.20 (9.80)	0.66 (0.18)	89.19 (5.67)
	5 k-fold	83.71 (1.34)	92.63 (0.44)	88.05 (0.68)	0.73 (0.01)	90.95 (0.38)
β	40% Hold-Out	86.46 (5.55)	87.57 (3.10)	86.96 (3.33)	0.66 (0.06)	87.37 (2.80)
	30% Hold-Out	87.68 (6.96)	87.59 (4.26)	87.56 (4.64)	0.67 (0.09)	87.60 (4.07)
	20% Hold-Out	87.09 (8.60)	88.24 (4.31)	87.57 (5.41)	0.68 (0.11)	88.03 (4.30)
	10% Hold-Out	88.41 (10.43)	90.10 (10.44)	89.07 (6.46)	0.72 (0.13)	89.80 (5.54)
	5 k-fold	87.37 (1.30)	88.54 (0.45)	87.95 (0.66)	0.68 (0.01)	88.31 (0.40)

Table 5.2: Métricas de rendimiento de validación utilizando los métodos de razonamiento α y τ y nuestro propio conjunto de datos.

Dataset	Método Clasificación	Comprobación de las métricas de rendimiento				
		Sensibilidad	Especificidad	Gmean	MCC	ACC
[259]	α w/o s-T	79.39	93.92	86.34	0.51	93.29
	α w/ s-T	82.44	92.05	87.11	0.48	91.64
	β w/o s-T	80.91	93.81	87.12	0.52	93.25
	β w/ s-T	84.73	90.82	87.72	0.47	90.55
[261]	α w/o s-T	71.75	99.48	84.48	0.81	96.38
	α w/ s-T	75.47	99.06	86.46	0.82	96.41
	β w/o s-T	73.60	99.48	85.56	0.82	96.58
	β w/ s-T	81.41	98.82	89.69	0.84	96.88

Table 5.3: Métricas de rendimiento de las pruebas para los diferentes conjuntos de datos de prueba utilizando los métodos de razonamiento α y β , y el autoajuste (s-T).

más alto en comparación con las reglas que tienen un consecuente de clase positivo. Además, en general, se puede observar que α presenta dependencia sobre la cantidad de datos de entrenamiento, mientras que β proporciona una validación del sistema más robusta independientemente de este hecho. Nótese que, aunque esto se ha observado específicamente para este conjunto de datos de entrenamiento, podría ser aplicable a otros conjuntos de datos, así como a otros problemas. El mejor resultado para el método α se consigue utilizando 5 k-fold, lo que lleva a un 88,05% y 0,73 de los valores promediados de Gmean y MCC respectivamente. Cuando se comparan los resultados del método β , se observa que la validación cruzada del 10% obtiene los mejores resultados promediados. Sin embargo, en el caso de considerar el equilibrio entre las métricas promediadas y sus desviaciones, la configuración de 5 k-fold muestra el mejor rendimiento con 87,95%, 0,68, 0,66 y 0,01 de los valores promediados y de desviación estándar de Gmean y MCC respectivamente. Este análisis se completa comparando los resultados de la validación k-fold para ambos métodos

de razonamiento. En ese caso, α presenta unas métricas ligeramente mejores que β para todas las métricas de rendimiento excepto para la sensibilidad. Este último hecho es una indicación del comportamiento real del sistema para ambos métodos de razonamiento sobre los futuros datos no vistos, ya que τ proporciona una mejor sensibilidad a expensas de la disminución de la especificidad. A pesar de estas últimas diferencias entre ambos métodos, se puede concluir que la validación cruzada k-fold supera al resto de los métodos y, por tanto, los RWs optimizados a utilizar para las pruebas se obtienen promediando los RWs obtenidos durante la validación k-fold considerando los 5 pliegues y las 30 iteraciones independientes.

Una vez realizada la validación del sistema y obtenida la base de reglas inicial optimizada y sus respectivos RWs, se procede a la recogida del conjunto de datos de prueba tal y como se detalla en la Sección anterior. La tabla 5.3 muestra los resultados obtenidos para los dos conjuntos de datos de referencia considerados y los métodos de razonamiento. En este caso, también proporcionamos los resultados relativos a la aplicación de autoajuste. Obsérvese que en negrita están las métricas que aumentaron tras la integración del autoajuste. Por un lado, el método β consigue generalmente métricas más altas que el método α para ambos conjuntos de datos cuando no se aplica el autoajuste. De hecho, la diferencia entre cualquiera de las métricas que son peores para el método β que para el método α no supera siquiera el 0,2%. Por otra parte, la aplicación del proceso de autoajuste condujo a la adición de una nueva regla por conjunto de datos con consecuente de clase positivo. Se puede destacar una clara diferencia entre los resultados para este caso de uso, ya que mientras que para CapnoBase sólo hay una mejora de la sensibilidad, para CSL se puede observar un aumento para la mayoría de las métricas de rendimiento excepto para la especificidad, cuyo empeoramiento no supera el 0,7%. Esta diferencia en la tendencia de los resultados para los distintos conjuntos de datos puede atribuirse a la naturaleza de los mismos. Hay que tener en cuenta que la heterogeneidad está presente ya que estos conjuntos de datos contienen datos de diferentes voluntarios, pero también que los artefactos dentro de los segmentos de PPG extraídos pueden tener diferentes dinámicas caracterizadas. De hecho, aunque CapnoBase es mayor que CSL, este último contiene el doble de segmentos PPG etiquetados inaceptables. Por lo tanto, los resultados y también los efectos de la aplicación de autoajuste vari-

Método Clasificación	Métricas de rendimiento de las pruebas				
	Sensibilidad	Especificidad	Gmean	MCC	ACC
α w/o s-T	75.57	96.70	85.41	0.66	94.84
α w/ s-T	78.96	95.56	86.79	0.66	94.03
β w/o s-T	77.25	96.64	86.40	0.66	94.92
β w/ s-T	83.07	94.82	88.75	0.66	93.72

Table 5.4: Métricas de rendimiento promedio de las pruebas utilizando los métodos de razonamiento α y β , y el autoajuste.

arán en función de lo bien caracterizado que esté el objetivo a detectar (segmentos PPG inaceptables). En resumen, la tabla 5.4 presenta los resultados promediados de las pruebas tras combinar ambos conjuntos de datos de prueba. Los mejores resultados se obtienen utilizando el método de razonamiento β y el ajuste de autoajuste. Esto lleva a un 88,75% y 0,66 Gmean y MCC promediados, lo que es comparable a los resultados de validación obtenidos.

Funcionamiento en tiempo real

Para la implementación embebida del SQA propuesto, la base de reglas inicial optimizada, los RWs y las funciones de membresía se codifican en el SoC. Para ello se cuantifican dichos parámetros mediante registros de 32 bits. Para validar la integración embebida, se ejecutan en el SoC 33 segmentos de PPG (24 aceptables y 9 inaceptables de 3 segundos) obtenidos del primer paciente del conjunto de datos de validación. Por ejemplo, la Figura 5-20 representa un extracto de la señal PPG filtrada, así como los diferentes valores de características para cada ventana de procesamiento de 3 segundos evaluada. Tenga en cuenta que el método de razonamiento incrustado es β , ya que logró mejores resultados durante la validación fuera de línea y las pruebas.

La tabla 5.5 informa de la comparación de las métricas de rendimiento entre la implementación incrustada y la offline. Como era de esperar, esta última consigue un mejor rendimiento. En concreto, alcanza hasta un 92,29% y un 0,85 de los valores de Gmean y MCC respectivamente, mientras que la implementación embebida empeora esos resultados bajándolos hasta un 91,67% y un 0,78. Esta diferencia de rendimiento se debe a la pérdida de precisión de los diferentes procesos dentro del SoC. De hecho, la Tabla 5.6 muestra el coeficiente de determinación (R^2) entre los resultados embebidos y los registrados en MATLAB para los principales procesos del sistema.

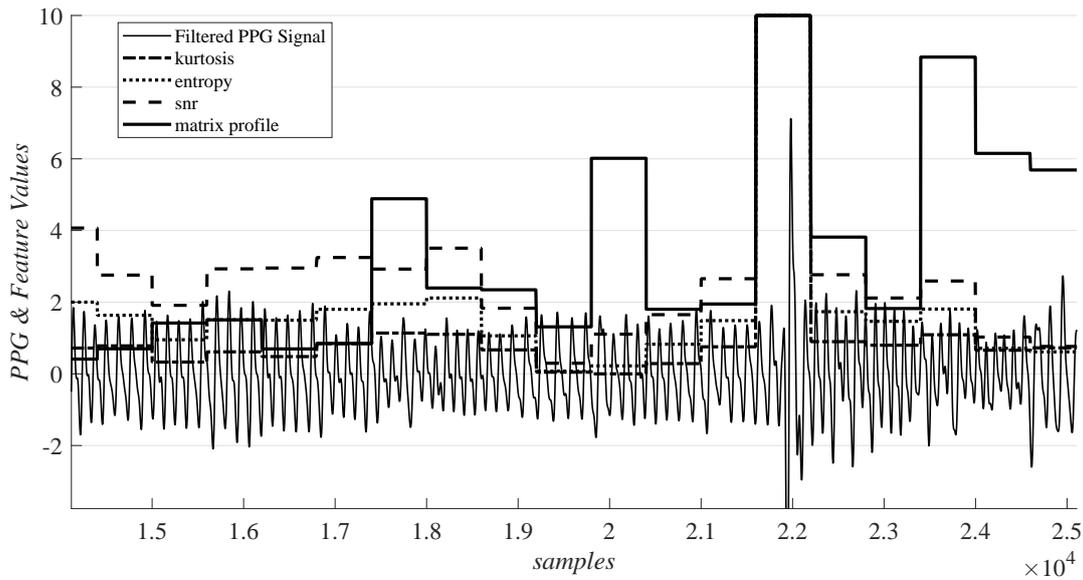


Figure 5-20: Captura en tiempo real de la implementación de SQA integrada que muestra los diferentes valores de las características en cada ventana de procesamiento (3 segundos).

Plataforma	Sensibilidad	Especificidad	Gmean	MCC	ACC
MATLAB®	88.89	95.83	92.29	0.85	93.94
SoC	88.89	91.67	90.27	0.78	90.90

Table 5.5: Comparación entre las métricas de rendimiento registradas por MATLAB para los 33 segmentos de PPG evaluados en el SoC utilizando el método de razonamiento *beta*.

Teniendo en cuenta que se trata de la primera implementación del sistema propuesto, la mayoría de los procesos manejan números de punto flotante de 32 bits en sus operaciones (IEEE 754). Así, las etapas menos exigentes desde el punto de vista computacional y más sencillas, como los cálculos de fuerza de activación, Kurtosis y SNR, obtienen un R^2 de 0,99. Sin embargo, al enfrentarse a operaciones matemáticas más complejas como el logaritmo dentro del cálculo de la Entropía y los múltiples algoritmos FFT realizados durante el cálculo del perfil de la matriz, el R^2 baja a 0,98 y 0,97 respectivamente. Aunque estos errores de precisión son bajos, su acumulación a lo largo de toda la predicción hace que el proceso final, es decir, el grado de asociación para las clases negativas y positivas, tenga un R^2 de 0,94 y 0,80 respectivamente. Esta es la razón de la caída de la métrica de rendimiento dentro del SoC. Hay que tener en cuenta que la implementación de MATLAB® opera con tipos de datos dobles, que corresponden a números de 64 bits en coma flotante.

En cuanto al análisis del ahorro de energía con y sin el método SQA, la tabla 5.7

	SoA	Kurtosis	Entropy	SNR	MP	-AD	+AD
R^2	0.99	0.99	0.98	0.99	0.97	0.94	0.80

Table 5.6: Coeficiente de determinación (R^2) para los principales procesos realizados dentro del SoC. SoA: Fuerza de Activación. MP: Perfil de la matriz. - y + AD: Grados de Asociación de Clase Negativa.

informa de cuatro escenarios diferentes de señales de prueba. Estos últimos fueron elegidos para facilitar la comparación con [245] y [246]. Independientemente de la aplicación del método SQA propuesto, el consumo de energía del sensor PPG es inevitable para un sistema o aplicación de monitorización fisiológica continua. Así, la línea base de consumo de energía del sistema es de $1053,36 \text{ mJ}$, que corresponde al funcionamiento normal del sensor así como a la comunicación I2C implicada para recoger los datos del mismo. Para la transmisión de cada ventana de procesamiento de 3 segundos, la energía consumida por el BLE es de unos $15,50 \text{ mJ}$, lo que lleva a $318,60 \text{ mJ}$ consumidos para la transmisión de la señal sin ruido durante 60 segundos. El consumo de energía debido a la ejecución de todas las etapas implicadas en el sistema SQA propuesto es de $59,40 \text{ mJ}$. Por lo tanto, teniendo en cuenta los diferentes escenarios de la señal de prueba, podemos concluir que el método SQA propuesto puede ahorrar un consumo de energía global del 1,5% al 20,7% para señales PPG ruidosas con una duración de 12 a 60 segundos. Por el contrario, el consumo de energía extra debido a la ejecución de SQA alcanza hasta el 4,3% para una señal completa de 60 segundos sin ruido. Por último, también se cuantificaron las complejidades de tiempo y memoria para el SQA propuesto. El tiempo medio obtenido para ejecutar el método SQA propuesto fue de $53,07 \text{ ms}$. La memoria total requerida para manejar las variables globales y temporales, así como los buffers de adquisición y procesamiento es de 15kB.

Para contextualizar algunos de los resultados obtenidos con respecto a otros trabajos reportados sobre SQA, la Tabla 5.8 presenta las métricas clave para el sistema propuesto y tres trabajos recientes [245–247] que también fueron revisados. El trabajo propuesto proporciona unas métricas de rendimiento comparables a las del estado de la técnica. Cabe señalar que los otros trabajos no utilizaron las etiquetas de artefactos proporcionadas por los conjuntos de datos de referencia. En su lugar, etiquetaron de nuevo esos datos. Además, ninguno de ellos utilizó exactamente

Escenarios de señales de prueba	Sistema sin SQA				Sistema con SQA			Overall Energía (with SQA) saving/extra
	EC_{Sensor} (mJ)	EC_{SQA} (mJ)	EC_{TR} (mJ)	Total (mJ)	EC_{SQA} (mJ)	EC_{TR} (mJ)	Total (mJ)	
60-segundos Señal sin ruido	1053.36	NE	318.60	1371.96	59.40	318.60	1431.36	4.3% Extra
60-segundos Señal ruidos	1053.36	NE	318.60	1371.96	59.40	NE	1112.76	20.7% Saving
6-seg. Señal ruidosa de 60 segundos	1053.36	NE	318.60	1371.96	59.40	292.05	1404.81	2.3% Extra
12-seg. Señal ruidosa de 60 segundos	1053.36	NE	318.60	1371.96	59.40	238.95	1351.71	1.5% Saving

Table 5.7: Análisis de ahorro de energía en tiempo real con y sin el método SQA. EC_{SQA} : Consumo de energía para el sistema implementado con SQA. EC_{Sensor} : Consumo de energía del sensor PPG. EC_{TR} : Consumo de energía para la transmisión BLE. *NE*: No ejecutado.

Trabajo	Validación Métricas	Experimento Independiente	Few Shot	Experimento Observaciones	ACC (%)	Clock (MHz)	Memoria (kB)	Energía (mJ)	Plataforma
Vadrevu (2019) [245]	amplitud absoluta, ratio cruzado y autocorrelación	×	×	38620	95.93	84	13	210	SAM3X8E ARM Cortex-M3
G.N.K. Reddy (2020) [246]	FOPC-DC	×	×	15000	93.21	84	29.56	–	SAM3X8E ARM Cortex-M3
Samiul Alam (2021) [247]	Curtosis, y autocorrelación, límites empíricos	×	×	8000	96.50	1200	88	63.1	Quad-core ARM Cortex-A53
Propuesto	Type-2 Fuzzy Independiente del sujeto (β w/ autoajuste)	✓	✓	331	93.72	64	15	59.40	nrf52832 ARM Cortex-M4

Table 5.8: Comparación con los trabajos publicados sobre SQA.

los mismos conjuntos de datos, tanto si los usaban para ajustar el umbral como si los probaban. En cuanto al consumo de memoria y energía, ofrecemos una de las métricas más bajas. En concreto, para el consumo de energía, [245] y [246] también proporcionaron un análisis de ahorro de energía en tiempo real comparable al que se indica en la tabla 5.7. Obtuvieron un ahorro energético global superior al 90,00% para el segundo escenario de señales de prueba. Sin embargo, no consideraron ni informaron del consumo de energía del sensor, que se supone que está funcionando continuamente. También hay que destacar la diferencia de plataforma, ya que mientras [245] y [246] utilizaron un dispositivo embebido (microcontrolador) comparable al utilizado en este trabajo de investigación, [247] empleó un Cortex-A53 que está lejos de poder ser comparado adecuadamente con el nuestro. Por último, se destacan los enfoques de pocos y experimentos hacia la aplicación y la adaptación del sistema a entornos heterogéneos, así como la usabilidad en la naturaleza.

El SQA propuesto proporciona un modelo de inferencia Mamdani simplificado y de baja complejidad basado en reglas difusas que se despliega en dispositivos de borde con pocos recursos. La principal novedad de esta investigación son los aspectos

no heurísticos, adaptativos, orientados a los dispositivos portátiles e invariantes del sistema SQA propuesto. En primer lugar, la característica no heurística se obtiene utilizando un novedoso método no supervisado para generar conjuntos difusos de tipo intervalo-II a partir de señales PPG basadas en la cuantificación. En segundo lugar, la adaptación del sistema se consigue definiendo e implementando un novedoso ajuste fino no supervisado en línea basado en la similitud escalada entre conjuntos difusos de tipo intervalo-II para las actualizaciones autoadaptativas del modelo. Por último, se logra el aspecto de la invariabilidad del sujeto y la heterogeneidad, ya que todos los conjuntos de datos empleados contienen datos de diferentes voluntarios. Este hecho hace que los artefactos dentro de los segmentos de PPG extraídos tengan diferentes dinámicas. Para demostrar la implementación del PPG SQA en línea, se realiza un análisis detallado del rendimiento integrado de los métodos propuestos en el Brazalete y se compara con el estado del arte. En general, el trabajo propuesto proporciona métricas comparables con el estado del arte comparado. Se logró una precisión global en las pruebas a ciegas de hasta el 93,72%. La evaluación en tiempo real mostró un consumo de energía de hasta 59,40 *mJ* para el SQA propuesto, lo que supuso un ahorro global de energía del 20,7%. Dentro de este contexto y comparación, también hay que tener en cuenta ciertas limitaciones del sistema propuesto. En primer lugar, pueden aplicarse más optimizaciones de procesamiento de señales digitales, como el escalado de cálculos enteros más pequeños y la instrucción única de datos múltiples. En segundo lugar, se están realizando más experimentos y recopilaciones de datos para aumentar los datos de entrenamiento y explorar el espacio de diseño. Algunas de las ventajas y limitaciones identificadas durante la realización de este sistema confirman la necesidad de contar con sistemas de SQA centrados en proporcionar una generalización suficiente para tratar con entornos heterogéneos, así como con implementaciones de SQA integradas en dispositivos de borde extremo. Hay que tener en cuenta que, aunque nos centramos únicamente en la señal PPG, se puede seguir investigando hacia enfoques similares para las demás señales considerando este sistema como referencia.

5.2.4 Extracción de características Exploración del espacio de diseño

En esta sección se presenta una exploración del espacio de diseño de extracción de características, que se centra en la extracción de información relacionada con BVP. El análisis presentado se divide en las diferentes etapas de la arquitectura del software, como muestra la Fig. 5-21. Así, se asume que las señales consideradas están debidamente filtradas y segmentadas antes de la aplicación de cualquier técnica de extracción de características. En cada etapa explorada, se evalúan y recomiendan los parámetros de interés. En primer lugar, se discute la delineación morfológica (bloque de detección de picos) mediante una comparación detallada de diferentes algoritmos de detección de picos. En segundo lugar, se presentan las técnicas habituales aplicadas para extraer la información de frecuencia cuando se trata de los puntos delineados de forma desigual o no uniforme, es decir, los bloques de interpolación o de recuento de latidos. En el caso de la aplicación de Bindi, se emplea, implementa y discute la técnica de interpolación. Por último, las recomendaciones específicas de las diferentes compensaciones discutidas en estas secciones se aplican en la sección 5.2.4.3 para un caso de uso particular de detección de estrés de 4 segundos, en el que una validación de los procesos de extracción de características incrustadas en el Brazalete se validan contra una herramienta de grado de investigación. Todas las métricas consideradas en esta sección para caracterizar las diferentes etapas se obtienen a partir de la implementación incorporada dentro de la exploración del espacio de diseño en la pulsera. Nótese que la selección final de los diferentes parámetros evaluados dependerá de los requisitos y necesidades de la aplicación. En cuanto a la sección 5.2.2, este análisis se extrae de [159]. Así, aunque los algoritmos evaluados se centran en un caso de uso PPG, algunas de las técnicas de extracción de características, como el FFT, son comunes al resto de señales.

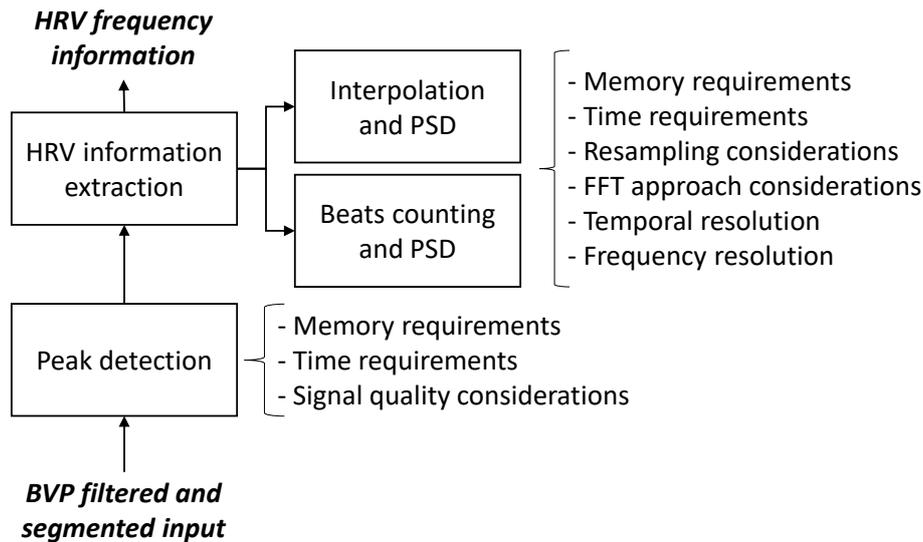


Figure 5-21: Parameters and processes involved in the BVP-based DSE.

5.2.4.1 Extracción de características: Detección de picos

Centrándonos específicamente en la casuística del PPG, se pueden utilizar diferentes enfoques para delinear las series temporales del PPG. La robustez de este proceso de delineación es clave para detectar correctamente los parámetros morfológicos de PPG deseados. Este hecho viene determinado no sólo por los pasos de filtrado anteriores, sino también por la diferente morfología de las ondas PPG, que puede verse directamente afectada por factores como la edad y las emociones [262,263]. La figura 5-22 muestra la diferencia morfológica entre tres grupos de edad diferentes medida con nuestro sensor PPG. Las diferencias observadas coinciden con las publicadas en la literatura [264]. Por ejemplo, la parte diastólica de la onda es la más afectada. Este hecho se debe principalmente a la variación del tono vascular con la edad, que se traduce directamente en una mayor o menor vasoconstricción y vasodilatación. Esta situación produce diferencias en la presión arterial que conducen a una distorsión del escurrimiento diastólico. Dentro de este contexto fisiológico variable, diferentes algoritmos de delineación podrían proporcionar resultados diferentes, es decir, diferentes puntos morfológicos identificados de la señal PPG. Este hecho puede incluso empeorar cuando se restringen las arquitecturas de filtrado de la etapa inicial aplicadas, lo que solía ocurrir en las aplicaciones integradas de monitorización fisiológica. Por ejemplo, si la aplicación no utiliza ningún filtro de eliminación de la línea de base, por ejemplo, un filtro de muesca por debajo de 0,5Hz, los diferentes puntos extraídos por el algoritmo de delineación empleado deben ser lo suficientemente robustos como para no verse afectados por las tendencias de baja frecuencia fuera de banda.

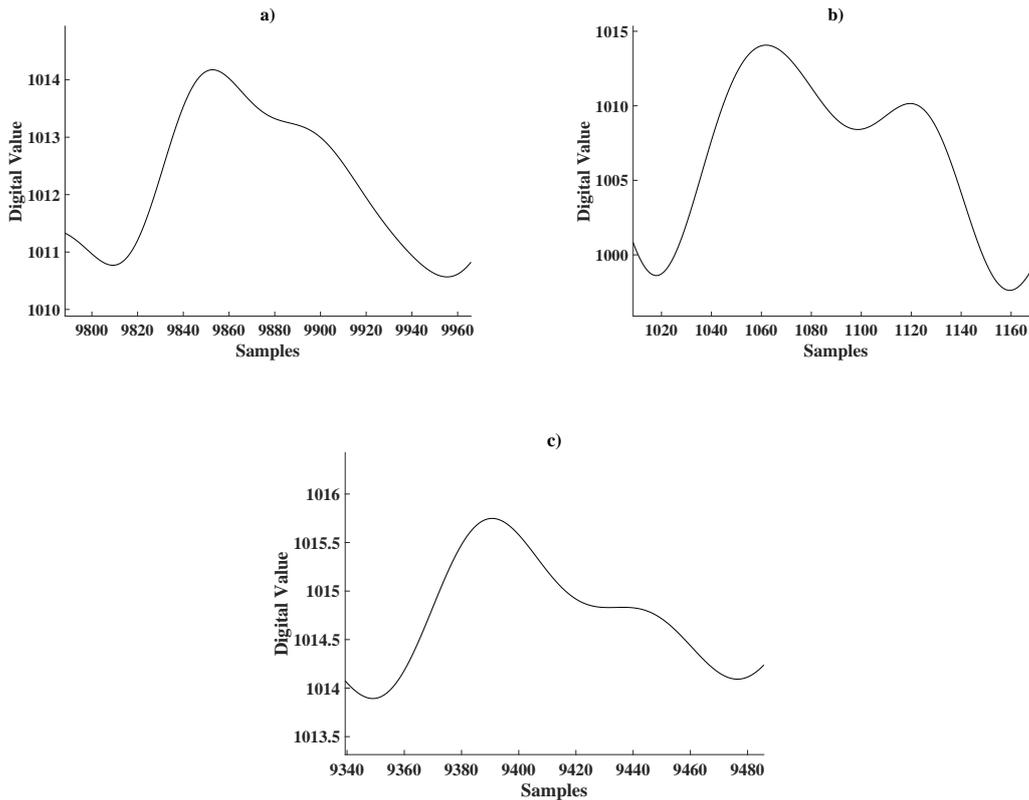


Figure 5-22: Diferencias morfológicas del PPG entre tres grupos de edad. (a) Persona de 18-24 años. (b) Persona de 35-44 años. (c) Persona de 55-65 años. Las señales mostradas fueron adquiridas por el brazalete Bindi.

Se pueden aplicar diferentes técnicas para asegurar una detección robusta de pico a pico. Sin embargo, algunas de ellas requieren la aplicación de cruces de cero a lo largo de la primera y segunda derivadas de la señal [265, 266]. Este hecho afecta directamente al tiempo de cálculo dentro de la cadena de procesamiento de datos.

Teniendo en cuenta las particularidades comentadas de esta etapa para un caso de uso PPG, se presenta una comparación entre dos enfoques ligeros conocidos. Por un lado, el primero se basa en un método de máximos/mínimos locales (LCM) desarrollado por el equipo UC3M4Safety que utiliza la evolución de la pendiente y la media local en periodos cortos de muestras junto con la ventana de procesamiento de datos [159]. Los métodos LCM son bien conocidos dentro de los algoritmos de detección de picos de PPG ya que solían ser menos exigentes computacionalmente a costa de un menor rendimiento. Por otro lado, el segundo algoritmo está tomado de [267] que se basa en un método de detección de umbral adaptativo (ADT) que utiliza una pendiente variable calculada iterativamente en función de la desviación estándar de la señal. Obsérvese que el segundo algoritmo fue validado con conjuntos

de datos disponibles públicamente y superó a las técnicas de LCM sin requerir operaciones de primera y/o segunda derivada de la señal.

Para el primer método, el Algoritmo 2 describe las operaciones realizadas para cada ventana BVP. En concreto, el algoritmo comienza asumiendo que el valor máximo es la primera muestra de la señal. A continuación, se calculan y evalúan los valores de la media y la pendiente sobre un número determinado de muestras a comparar (*stc*). Este paso se realiza si se cumple una de las dos condiciones de nivel de señal umbral (máximo o mínimo). Por lo tanto, es esencial un equilibrio para ajustar *stc*. Por ejemplo, si la señal BVP se muestrea a 100Hz, la evaluación media sobre diez muestras supone una atenuación de -6dB para 6Hz y una atenuación de -3dB para 4,5Hz, siendo esta última cercana a 4Hz que es una frecuencia de interés para apuntar a las frecuencias cardíacas máximas ($\text{BPM} \geq 240$). Por lo tanto, en función de las altas frecuencias de ruido residuales esperadas de la señal filtrada, se puede ajustar este parámetro. Para este algoritmo en particular, y guiado por el compromiso adoptado en la etapa de filtrado en la Sección 5.2.2, se puede elegir un *stc* igual a diez, lo que aumenta la capacidad de detección de picos a expensas de la complejidad del tiempo. Otro parámetro clave incluido en el algoritmo es $dist_{min}$, que se asigna inicialmente a un número específico de muestras k . Esta variable se refiere a la distancia mínima permitida entre dos picos sistólicos identificados y se fija en el número de muestras para la frecuencia más alta dentro del ancho de banda del BVP. Como en el ejemplo anterior, si la señal BVP se muestrea a 100Hz y la frecuencia de FC más alta aceptable es de 3,5 Hz (210 BPM), entonces k se fija en 28 (muestras). Este parámetro no afecta a la complejidad del tiempo del algoritmo, sino que proporciona un manejo robusto para los posibles transitorios que todavía están en la señal y que podrían estar afectando a la detección del pico. Además, hemos introducido diferentes condiciones en el algoritmo desarrollado para cubrir casos morfológicos especiales. Por un lado, en el caso de tratar con crestas sistólicas amplias, sólo el último punto de las mismas se considera el pico sistólico, líneas 11-17 del Algoritmo 2. Por otro lado, en caso de tener crestas sistólicas cortas, podría ocurrir que se dejara un pico potencial, lo que se tiene en cuenta en las líneas 19-21 del Algoritmo 2. Aunque el algoritmo se utiliza para extraer la información de los picos sistólicos, el procesamiento de los valles también se realiza dentro de

él. Este último se realiza a través de operaciones opuestas a las del procesamiento de los picos. Nótese que, en nuestro caso, la señal está centrada sin ninguna deriva o tendencia de DC antes de la aplicación de este algoritmo gracias a las etapas de filtrado previas.

Algorithm 2: Algoritmo de detección de picos BVP

```

1 function getPeaks ( $bvp_{signal}, bvp_{len}$ );
   Input :
   Señal limpia de BVP  $bvp_{signal}$ ;
   Número total de muestras  $bvp_{len}$ ;
   Output:
   Posición de los picos detectados  $peaks_{index}$ ;
   Número total de picos  $peaks_{total}$ ;
   Data:
   Máximo y mínimo para cada búsqueda  $peaks_{max}, peaks_{min}$ 
   Contador de picos detectados  $peaks_{count}$ 
   Separación mínima entre los picos detectados  $dist_{min}$ 
2  $peaks_{max} \leftarrow bvp_{signal}(0)$ ;
3  $peaks_{min}, peaks_{count} \leftarrow 0$ ;  $dist_{min} \leftarrow k$ ;
4 for  $i \leftarrow 1$  to ( $bvp_{len} - stc$ ) do
5     if  $bvp_{signal}(i) > peak_{max}$  then
6          $peak_{max} \leftarrow bvp_{signal}(i)$ ;
7          $peaks_{min} \leftarrow peak_{max} - stc$ ;
8         Get  $vtc_{mean}$  for  $[i, stc]$ ;
9         Get  $vtc_{slope}$  for  $[i, stc]$ ;
10        if  $vtc_{mean} \geq peak_{max}$  then
11            if  $peaks_{count} \&\& i - peaks_{index}(peak_{count} - 1) < dist_{min}$  then
12                 $peaks_{index}(peak_{count} - 1) \leftarrow i$ ;
13            else
14                 $peaks_{index}(peaks_{count}) \leftarrow i$ ;
15                 $peaks_{count} \leftarrow peaks_{count} + 1$ ;
16                 $peaks_{min} \leftarrow 0$ ;
17            end
18        else
19            if  $peaks_{min} \&\& vtc_{slope} < 0$  then
20                Actualizar contador, índice y  $peaks_{min}$ ;
21            end
22        end
23    end
24    if  $bvp_{signal}(i) < peak_{min}$  then
25        Realiza la operación inversa para detectar los valles;
26    end
27 end
28  $peaks_{total} \leftarrow peaks_{count}$ 

```

La figura 5-23 muestra un análisis del impacto en el tiempo de los dos algoritmos diferentes de detección de picos, considerando el número de muestras en la ventana de procesamiento. En relación con esta complejidad temporal, se observa un rendimiento lineal para nuestro método LCM, debido a la evaluación de vecinos realizada con cada muestra. En cuanto al algoritmo ADT, se observa un aumento del tiempo de cálculo entre el 30 % y el 50 % en comparación con LCM con el mayor

stc. Esta diferencia se debe principalmente al cálculo obligatorio de la desviación estándar para toda la señal de la ventana de procesamiento, que es necesaria para obtener la pendiente variable que utilizará el algoritmo ADT. En relación con las consideraciones de almacenamiento de memoria en esta etapa, el algoritmo LCM implementado necesita 2KB de ROM y 32B de RAM. En el caso del ADT implementado, este algoritmo utiliza 2,5KB de ROM y 64B de RAM.

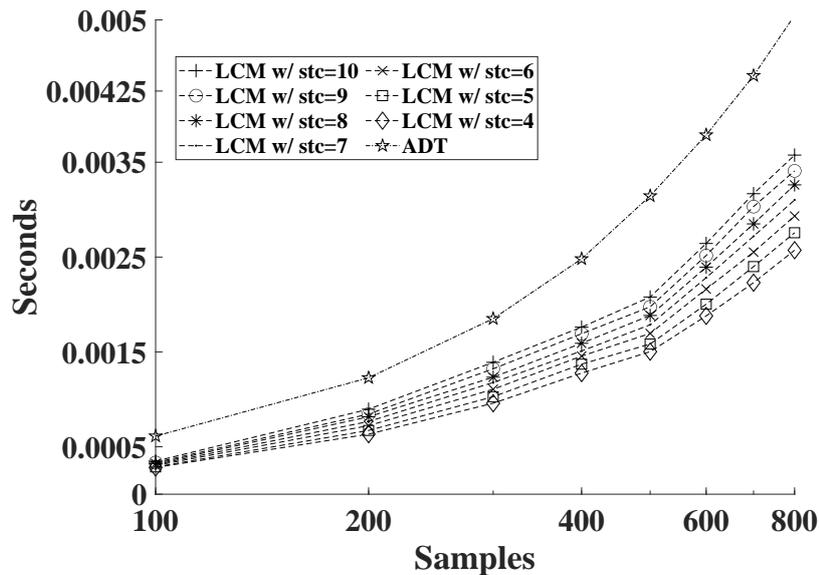


Figure 5-23: Análisis del impacto temporal de los algoritmos de detección de picos considerados.

A partir de esta etapa y teniendo en cuenta los resultados de complejidad temporal y uso de recursos por parte de los algoritmos evaluados, concluimos que, aunque los métodos LCM son de los más sencillos, pueden ajustarse a los requisitos de muchas aplicaciones. Por tanto, se recomienda el uso de estos algoritmos ligeros. Con el fin de proporcionar una validación de la aplicación real siguiendo estas compensaciones de extracción de características, la sección 5.2.4.3 presenta los resultados de un caso de uso real de monitorización de la activación de la VFC.

5.2.4.2 Extracción de características: Información de la VFC

En los sistemas digitales embebidos con restricciones, los análisis de frecuencia se realizan mediante la DFT. Uno de los algoritmos habituales es la transformada rápida de Fourier (FFT). Sin embargo, este algoritmo se basa en la suposición de una entrada muestreada equidistante. En este punto, surgen dos posibilidades basadas en las necesidades de la aplicación. Si la aplicación no está limitada por ninguna

restricción de tiempo de inferencia, el sistema puede esperar hasta que se extraigan suficientes puntos de VFC y sea posible la resolución de frecuencia deseada. Por el contrario, cuando se necesita una inferencia rápida continua dentro de una ventana temporal fija, se aplica la interpolación entre las muestras de VFC para restablecer la coherencia temporal.

Centrándonos en los casos de uso rápido continuo para aumentar el tiempo de respuesta de Bindi, consideramos dos parámetros principales para la evaluación de esta etapa de extracción de características: el tipo de interpolación y la longitud de la FFT. En la Fig. 5-24 se muestra un análisis del impacto en el tiempo para las dos técnicas de interpolación (lineal y polinómica) y la FFT implementada y considerada para diferentes longitudes de ventana de procesamiento. Como era de esperar, los métodos polinómicos tienen una mayor complejidad temporal, aunque producen resultados más precisos si se necesita posteriormente la precisión espectral. Obsérvese que para esta comparación se considera la interpolación cuadrática polinómica de Lagrange. En cuanto a la FFT, se utiliza un algoritmo de FFT radix-2 de 32 bits de punto fijo, que proporciona una de las complejidades computacionales más bajas ($\mathcal{O}(n \log n)$) y es entonces adecuada para el dispositivo embebido. Cabe destacar que para todas las longitudes de FFT evaluadas, aplicar la interpolación polinómica implica duplicar el tiempo de procesamiento respecto a la interpolación lineal. Teniendo en cuenta este hecho y que se prefieren las interpolaciones cuadráticas en el dominio del tiempo para el [268] de la VFC, se puede asumir esta diferencia de complejidad temporal.

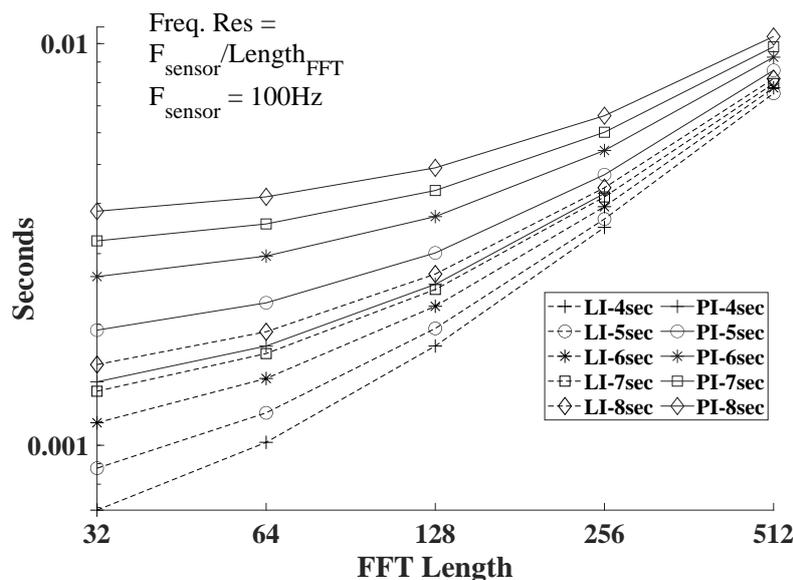


Figure 5-24: Análisis del impacto temporal basado en diferentes métodos de interpolación y en la FFT implementada y considerada.

En esta etapa, debe considerarse un compromiso entre la resolución temporal y la de frecuencia. Obsérvese que, independientemente de f_{sensor} , si la longitud de procesamiento de la ventana es fija, la resolución de la casilla de frecuencia para el FFT_{len} elegido no cambiará. Por lo tanto, para mejorar la resolución de frecuencia para una ventana temporal fija, se aplican técnicas de remuestreo después de la interpolación en estas situaciones. Por ejemplo, si la VFC se interpola a 100Hz para una ventana temporal fija de cuatro segundos, se obtiene una resolución de 0,39Hz/bin. Sin embargo, tras aplicar un remuestreo de 1Hz, la resolución de la frecuencia aumenta hasta 0,25Hz/bin. Tenga en cuenta que para esta última resolución sólo se toman 256 puntos disponibles. En caso de tomar más puntos que la longitud de la ventana, hay que aplicar un relleno de cero. Así pues, la resolución de tiempo y frecuencia, así como las técnicas de interpolación y remuestreo, dependen de la aplicación. Este es un aspecto clave cuando se trata de aplicaciones que requieren la extracción de información de la frecuencia de la VFC, ya que la banda más baja de interés se encuentra entre 0,01Hz y 0,04Hz. Por lo tanto, para lograr una capacidad completa de detección de la banda de frecuencia de la VFC se debe asegurar un mínimo de 0,04Hz/bin. Un valor de resolución de frecuencia superior a este disminuirá dicha capacidad de detección o separabilidad de las bandas espectrales. Hay que tener en cuenta que la resolución de la bandeja de frecuencias viene dada por la ecuación 4.7.

En cuanto a la resolución temporal, hay que tener en cuenta que la duración de la ventana de procesamiento debe seleccionarse para asegurar la presencia de al menos dos puntos de VFC. De lo contrario, la interpolación no es posible.

En relación con las consideraciones de almacenamiento de memoria en esta etapa, se debe tener especial cuidado con los requerimientos de recursos de la FFT, implementando propiedades in-place y un comportamiento no recursivo. Tenga en cuenta que los recursos utilizados durante la operación de remuestreo se consideran insignificantes. Para las interpolaciones, ambas consumen hasta 698B de ROM y 10B de RAM, mientras que la FFT necesita 3KB de ROM y 548B de RAM.

Este último paso es especialmente sensible. Por ejemplo, la interpolación lineal puede incluso introducir deformaciones en los espectros de potencia resultantes. En este caso, se recomienda una decisión de diseño basada en la calidad para que prevalezca la información fisiológica. Así, asumiendo el mismo uso de recursos para ambos métodos de interpolación y considerando todas las ventajas fisiológicas que proporciona la interpolación polinómica, se recomienda esta última sobre la lineal.

5.2.4.3 Implementación del caso de uso de VFC

Para dar un caso de uso real e implementar todas las diferentes recomendaciones concluidas para la última exploración de extracción de características, se presenta una aplicación específica de activación fisiológica de cuatro segundos de inferencia rápida. El brazalete está programado con toda la arquitectura detallada de procesamiento de señales y las compensaciones tomadas. En este caso, se utilizó un experimento con seis voluntarios y diez estímulos audiovisuales diferentes, estresados y no estresados, de un minuto de duración. Estos estímulos fueron previamente etiquetados y seleccionados por los autores. Después de cada estímulo, los voluntarios autoinformaron de su propio nivel de excitación o excitación que sentían al ver el vídeo. Para proporcionar una herramienta de validación o una medida de oro frente a las señales adquiridas por nuestra plataforma, se consideró un sistema sensorial de grado de investigación.

Para este experimento, se utilizó un f_{sensor} de 100Hz y se empleó una ventana de procesamiento temporal fija de cuatro segundos, que requería un buffer de 400 muestras (1,5KB). Hay que tener en cuenta que para FCs inferiores a 45BPMs, esta ventana no es aplicable, ya que sólo se pudo encontrar un punto de VFC. Cada

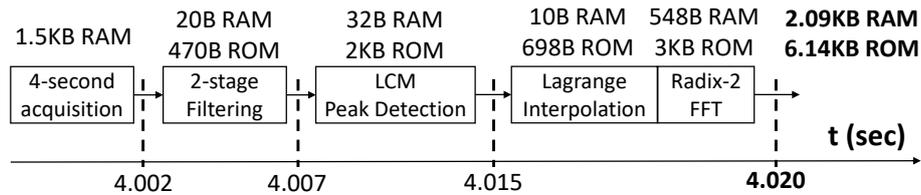


Figure 5-25: Cadena de datos completa para el procesamiento de una ventana de 4 segundos teniendo en cuenta las compensaciones comentadas.

cuatro segundos se extraen e interpolan los puntos de VFC, a lo que sigue un cálculo de la FFT y una estimación de la PSD dada por

$$PSD_i = \frac{2 * |fft|}{s}, \quad (5.18)$$

donde PSD_i es la densidad espectral de potencia para una casilla de frecuencia específica i , $|fft|$ es la magnitud del espectro al cuadrado y s es la suma de muestras al cuadrado de la función de ventana utilizada. En concreto, para hacer frente a la pérdida de festoneado y a los efectos de la valla, se aplica una ventana superior plana. Se utiliza un fijo FFT_{len} de 256 puntos, lo que lleva a una resolución de 0,39Hz/bin. Esta resolución es suficiente para observar la actividad de las bandas de frecuencia más bajas (hasta 0,4Hz) y las más altas (desde 0,4Hz hasta 1Hz). Hay que tener en cuenta que en caso de tener menos de 256 puntos después de la interpolación, se aplica el relleno cero. El mismo procedimiento digital se aplica para la herramienta de validación y Bindi. Teniendo en cuenta todos los datos proporcionados en las secciones anteriores, la implementación final de todas las diferentes etapas para esta aplicación específica requiere hasta 2KB de RAM, 6KB de ROM y tarda unos 20 milisegundos en proporcionar una estimación válida de la VFC desde la finalización de una ventana de procesamiento, Figura 5-25.

La tabla 5.9 muestra los resultados recopilados obtenidos para dos estímulos de estrés (H) y no estrés (L) seleccionados arbitrariamente para los seis voluntarios diferentes. \bar{P}_{Gf1} es el cociente promediado entre el primer bin de frecuencias (0,39Hz) y el segundo bin de frecuencias (0,78Hz) durante el estímulo utilizando la señal de la herramienta de validación, mientras que \bar{P}_{Gf2} es el observado para el cociente promediado entre el segundo bin de frecuencias y la suma del primero y el segundo. La cuarta y la quinta columnas son los valores análogos tomados de Bindi. Estos resultados muestran una disminución en el primer factor para todos

Table 5.9: Resultado de la medición del caso de uso específico del detector de estrés HRV.

<i>Type</i>	\bar{P}_{Gf1}	\bar{P}_{Gf2}	\bar{P}_{Bf1}	\bar{P}_{Bf2}	ε [%($\varepsilon_{f1}, \varepsilon_{f2}$)]
1_H	5.28	0.15	5.27	0.16	(0.18,6.66)
1_L	5.05	0.16	4.81	0.17	(4.75,6.25)
Δ	-0.23	+0.01	-0.46	+0.01	
2_H	4.09	0.24	4.29	0.20	(4.88,16.66)
2_L	3.83	0.27	3.45	0.29	(9.92,7.41)
Δ	-0.26	+0.03	-0.84	+0.09	
3_H	5.18	0.16	5.17	0.16	(0.19,0.00)
3_L	4.40	0.19	4.28	0.21	(2.72,10.52)
Δ	-0.78	+0.03	-0.90	+0.05	
4_H	5.27	0.15	5.32	0.15	(0.09,0.00)
4_L	5.16	0.16	5.12	0.17	(0.7,6.25)
Δ	-0.11	+0.01	-0.20	+0.03	
5_H	5.07	0.16	4.82	0.17	(4.93,6.25)
5_L	4.64	0.17	4.63	0.18	(0.21,5.88)
Δ	-0.43	+0.01	-0.19	+0.01	
6_H	4.98	0.16	4.96	0.17	(4.03,6.25)
6_L	4.84	0.17	4.46	0.20	(7.85,17.64)
Δ	-0.14	+0.01	-0.50	+0.03	

los pacientes desde el estímulo de estrés hasta el de no estrés. Por el contrario, hay un incremento en el segundo factor. Esto está en consonancia con la teoría del SNA. Como se comentó en los capítulos 2 y 4, las bandas de frecuencia más bajas están dominadas por el SNS, que se encarga de la respuesta de lucha o huida del cuerpo, mientras que el SNP está relacionado con las bandas más altas y es responsable de controlar las condiciones de relajación (descanso y digestión). Los errores entre los resultados de la validación y los resultados de la también se proporcionan en la tabla 5.9 . Estos errores son bajos ($\varepsilon < 10\%$), excepto en casos como 2_H o 6_L , en los que los fuertes artefactos de movimiento presentes en la señal de Bindino se limpiaron como se esperaba, dando lugar a segmentos de señales localmente contaminados, lo que afecta directamente al proceso de detección de picos y, por tanto, a la extracción de la VFC, véase la Figura 5-26.

Al llevar a cabo este caso de uso particular, se han implementado con éxito las diferentes compensaciones detalladas que se aplican a las aplicaciones de inferencia rápida. A pesar de que el caso de uso presentado no alcanza la capacidad completa de detección de la banda de frecuencia de la VFC, el objetivo de la detección rápida del estrés se logra utilizando una baja cantidad de recursos a expensas de la resolución

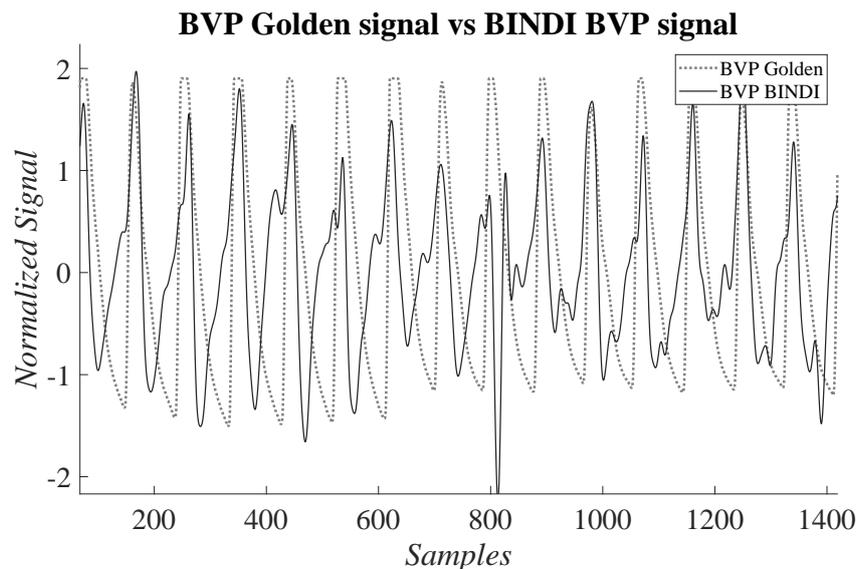


Figure 5-26: Efectos de artefactos de movimiento mostrados en un segmento del estímulo audiovisual de estrés del voluntario 2.

de la frecuencia.

5.2.5 Análisis del consumo de energía

La gestión del consumo de energía es un requisito en el diseño de un sistema wearable. En Bindi, una medición precisa del estado de carga de la batería y de la autonomía de los dos dispositivos wearables es esencial para garantizar que el sistema funcione cuando sea necesario. Esta sección proporciona un análisis cuantitativo del consumo de corriente para la Brazalete. Este análisis se realiza midiendo las acciones que más energía demandan a través de la parte de monitorización del dispositivo. Así, se mide por separado la corriente eléctrica consumida por la adquisición de datos a través de cada sensor fisiológico. Además, también se mide el consumo de energía que se produce al hacer uso del zumbador en intensidades suaves, medias y fuertes. Así, se optó por medir el consumo de energía debido a la comunicación y adquisición de datos de los sensores, que son esenciales para el sistema y están intrínsecamente relacionados con el diseño hardware específico de los dispositivos.

Los resultados obtenidos en el análisis del consumo de corriente para el Bindi 1.0 aparecen en la Figura 5-27. Los modos de vibración son las acciones que más corriente consumen, donde cuanto mayor es la vibración producida, mayor es la corriente requerida, como era de esperar. Sin embargo, el impacto del zumbido en la autonomía es reducido, ya que se activa durante poco tiempo en situaciones

de riesgo, por lo que su activación suele ser esporádica. Los sensores SKT y GSR también producen un pequeño incremento desde el estado de reposo. Sin embargo, el sensor PPG tiene un mayor impacto que los otros sensores. Así, podemos concluir que el cuello de botella actual del sistema, en términos de consumo de energía y tiempo de funcionamiento, es el sensor PPG. A pesar de este hecho, el bajo consumo de energía en el estado de reposo hace que la duración de la batería Brazalete sea de aproximadamente 40 horas cuando se utiliza una batería de 500 mAh. Tenga en cuenta que estos cálculos se basan en situaciones sin alarma.

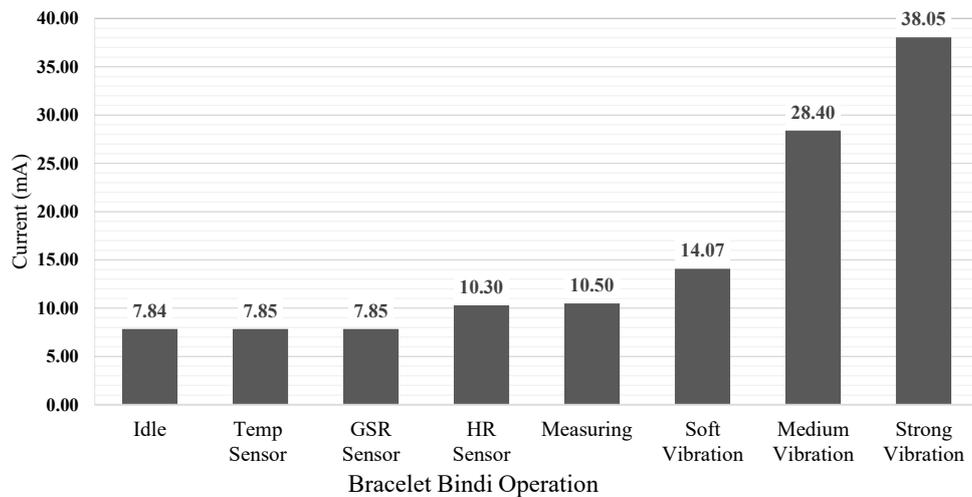


Figure 5-27: Consumo medio de corriente en el Brazalete [11].

5.3 Conclusión

En este capítulo se han expuesto las principales aportaciones de esta investigación en relación con el diseño de sistemas vestibles orientados a la adquisición y monitorización fisiológica continua. Ha quedado claro que el diseño de sistemas wearables para la monitorización continua de señales fisiológicas y el diseño e integración embebida de procesos relacionados con la computación afectiva es una tarea desafiante que requiere un cuidadoso equilibrio entre los recursos embebidos, el consumo de energía y el tiempo de respuesta del sistema. Además, presentamos diferentes trabajos que contribuyen a la implementación embebida de procesos relacionados con el aprendizaje automático en el borde, es decir, en nuestra umbrionada. Cabe destacar que la cadena completa de procesamiento de datos para todos los sensores, desde la adquisición hasta la clasificación, está siendo mejorada y ampliada para considerar y dimensionar el impacto embebido, así como sus limitaciones.

	BINDI	COMETA	ATENPRO	AlertCops	Safer Pro	EMPATIC A Embrace
Smart-watch/bracelet factor form	Yes	No	No	NA (App)	Yes	Yes
Automatic alarms	Sí	No (just GPS)	No	No	No	Sí
Technology	Cutting-edge	Obsolete	Obsolete	Mobile-based	Improvable	Cutting-edge
LEAs warning	Yes	Yes	Yes	Yes	No	No
Sensor Monitoring Goal	Emotional state	Distance (GPS)	No	No	No	Seizures
Use-Case	GBV	GBV	GBV	General (including GBV)	GBV	Epilepsy

Figure 5-28: La ventaja competitiva de Bindi sobre sus principales y más directos competidores.

En la sección 5.1 se han aportado los antecedentes técnicos necesarios para contextualizar los diferentes aspectos destacados que puede ofrecer la tecnología de Bindi. La figura 5-28 presenta un resumen compacto sobre las ventajas competitivas de Bindi respecto a sus principales y más directos competidores. Así, podemos concluir que las soluciones tecnológicas disponibles revisadas para combatir la violencia de género, las orientadas a un caso de uso general, o incluso las soluciones con objetivos diferentes pero tecnológicamente comparables a Bindi, no ofrecen las mismas funcionalidades ni se aprovechan de la tecnología de vanguardia. El Bindi tiene un gran potencial para ser una herramienta tecnológica efectiva para prevenir y combatir la Violencia de Género.

A lo largo de la Sección 5.2, se han diseccionado cuidadosamente las arquitecturas de hardware y firmware Brazalete para mostrar en detalle cada uno de los bloques de procesamiento digital, tanto los que están completamente cerrados a nivel de implementación e integración como los que aún están en fase de diseño y desarrollo. Dentro de este contexto y centrándonos en el aspecto wearable, en esta sección se ha presentado un sistema PPG SQA capaz de identificar segmentos de información fisiológica de baja calidad, sin embargo, no se ha presentado ningún trabajo rela-

cionado con la corrección de artefactos de movimiento o la eliminación de dichos segmentos. Al buscar este tipo de sistemas en la comunidad científica e independientemente de la proliferación de sistemas PPG SQA [241], existen multitud de sistemas de eliminación de artefactos de movimiento (MAR) propuestos en la literatura. Este hecho se basa en el anhelo de recuperar la señal PPG completa, sin tener en cuenta el tipo y la cantidad de ruido presente. Incluso se han creado diferentes iniciativas y retos a lo largo de la última década para fomentar el desarrollo de nuevos algoritmos y métodos de MAR. Por ejemplo, la Copa IEEE de Procesamiento de Señales de 2015 se basó en un conjunto de datos de PPG capturados en laboratorio utilizando una cinta de correr para generar diferentes tipos de artefactos de movimiento y pretendía ofrecer un marco general para tratar el MAR en la monitorización de la frecuencia cardíaca [269]. Sobre esta base, existen dos perspectivas diferentes. Por un lado, al centrarse en casos de uso offline, la aplicación de las técnicas MAR puede ser factible al considerar una gran cantidad de recursos informáticos disponibles. Sin embargo, estos algoritmos consideran la totalidad de la señal para sus líneas de procesamiento, y en algunos casos la reconstrucción o extracción de una medida válida a partir de una señal ruidosa no es posible. Por otro lado, las aplicaciones para llevar puestas que tienen como objetivo la monitorización continua de la PPG están sujetas al requisito de utilizar pocos recursos y consumir poca energía. Esto deja un fino vacío para implementar algunos de los mejores y más pesados algoritmos computacionales de SAM, como el análisis de componentes independientes, la descomposición de modo empírico y los métodos basados en el aprendizaje profundo [270]. Por lo tanto, la evaluación de la calidad de la señal a través de metodologías SQA antes de la aplicación de cualquier algoritmo MAR es esencial cuando se pretende la monitorización continua de PPG y otras señales relacionadas con el corazón [271]. A pesar de este último hecho, en nuestro grupo de investigación se ha iniciado una investigación relacionada con la propuesta de nuevas técnicas de MAR embebidas [230], con la intención de contribuir a este tema de investigación en particular.

En cuanto a los procesos detallados de extracción de características, hay que tener en cuenta ciertas limitaciones del sistema propuesto. En primer lugar, se pueden aplicar diferentes técnicas de procesamiento de señales. Por ejemplo, para tratar los

datos de VFC con espacios irregulares, se podría aplicar el método de periodograma de Lomb-Scargle en lugar de la FFT. En segundo lugar, el consumo de energía específico de cada una de las técnicas de extracción de características debe analizarse adecuadamente para identificar posibles cuellos de botella en la arquitectura de procesamiento de señales digitales. Esto último se está realizando actualmente y se están preparando más publicaciones.

Capítulo 6

Un nuevo conjunto de datos para el reconocimiento de emociones: WEMAC

En el capítulo 4 presentamos el trabajo realizado para el diseño de un sistema de detección de miedo utilizando bases de datos públicas. Durante el desarrollo de estos sistemas se identificaron diferentes limitaciones y se confirmó que, para llegar a un sistema óptimo de reconocimiento del miedo, se necesita una nueva base de datos centrada en la detección del miedo. Dicha base de datos debería incluir factores clave ya destacados en los capítulos anteriores, como:

- El uso de la tecnología de inmersión emocional.
- La modificación de la metodología de etiquetado para considerar la perspectiva de género.
- Una distribución de estímulos adecuadamente equilibrada con respecto a las emociones objetivo.
- Un mayor número de participantes.
- La integración de un proceso de recuperación basado en las señales fisiológicas de los voluntarios para cuantificar y aislar la activación emocional entre los estímulos.

Además, atendiendo a uno de los principales objetivos de esta investigación, como es la generación de nuevos mecanismos de prevención y lucha contra la Violencia de Género, esta base de datos debe ser concebida teniendo en cuenta las particularidades necesarias relacionadas con este perfil específico para llevar a cabo un diseño metodológico adecuado. En este contexto, este Capítulo presenta la base de

datos UC3M4Safety, cuyo objetivo final es el desentrañamiento de los mecanismos de activación de las Víctimas de Violencia de Género en situaciones de violencia. Este objetivo se pretende alcanzar mediante la generación y realización de diferentes experimentos:

- Experimento de preetiquetado. Esto generó los dos primeros conjuntos de datos, que están publicados en [272] y [273]. El objetivo era estudiar y validar la eficacia de un conjunto de estímulos audiovisuales a la hora de generar emociones discretas, concretas y únicas. Este experimento se centró en encontrar estímulos que fueran capaces de provocar la misma reacción emocional al mayor número de personas posible. Además, este estudio permitió analizar los métodos de clasificación de estos estados emocionales, la comprensión de los aspectos críticos por parte de los participantes y la influencia del género en la detección del miedo [82].
- Experimentos de laboratorio con víctimas de la violencia no de género. Estos experimentos generaron cuatro conjuntos de datos. El primero se denomina "Conjunto de datos de computación afectiva multimodal sobre mujeres y emociones" (WEMAC). Consisten en experimentos realizados en un entorno de laboratorio sólo con mujeres voluntarias que nunca han sufrido violencia de género. En concreto, se utiliza un conjunto reducido de estímulos, extraídos de los primeros conjuntos de datos, junto con la adquisición de información fisiológica y física (voz y audio). Aparte del indudable valor, para el área de la informática afectiva, de generar un conjunto de datos con reacciones emocionales en las mujeres mientras se registran sus variables fisiológicas y físicas, el desentrañamiento de las emociones de miedo mediante el seguimiento de las reacciones fisiológicas y físicas de las mujeres que no son Víctimas de Violencia de Género es también necesario para comprender estas variaciones y patrones bajo perfiles poblacionales no específicos. Esta investigación aborda estos conjuntos de datos centrándose en los datos fisiológicos y multimodales.
- Experimentos de laboratorio con víctimas de la violencia de género. En el momento de redactar este informe de doctorado, estos experimentos aún están en desarrollo. Generarán cuatro conjuntos de datos adicionales. Se basan en la misma metodología experimental seguida con las víctimas de la vio-

lencia no de género. Cabe destacar que se ha prestado especial atención a evitar la revictimización de las Víctimas de Violencia de Género, incluyendo el seguimiento psicológico y el trabajo en su recuperación emocional de la violencia. En estos experimentos, el objetivo es comparar las respuestas fisiológicas y multimodales con los experimentos anteriores con Víctimas de Violencia no de Género.

- Los experimentos Into-the-Wild con ambos perfiles, Víctimas de Violencia de Género y Víctimas de Violencia No de Género. Hasta la fecha, estos experimentos están en desarrollo. Generarán al menos diez conjuntos de datos más. Estos experimentos están pensados para ser realizados durante la vida cotidiana de algunos de los voluntarias que participaron en los experimentos de laboratorio. El objetivo es obtener un comportamiento fisiológico y multimodal real para seguir estudiando y caracterizando las emociones etiquetadas a lo largo de los diferentes días.

En cuanto a la estructura de este capítulo, comenzamos por ofrecer una explicación exhaustiva de la metodología y el desarrollo seguidos durante la generación de los experimentos de laboratorio para las Víctimas de la violencia no sexista. Al igual que en el Capítulo 4, también se presenta el análisis de la distribución del etiquetado autodeclarado para este conjunto de datos. A este análisis le sigue una exploración de la respuesta fisiológica para dar una visión adecuada de los patrones fisiológicos, las recuperaciones y otras particularidades observadas durante los experimentos. Esta exploración concluye con la presentación de los primeros resultados de detección del miedo basados en dicha información. Las métricas obtenidas se utilizan y se fusionan con los resultados del habla, lo que proporciona una perspectiva multimodal del problema. Esto último se ha realizado en un trabajo de investigación multidisciplinar con los miembros de la UC3M4Safety expertos en Teoría de la Señal y Comunicaciones. Cabe destacar que estos son los primeros resultados de reconocimiento de miedo utilizando información multimodal recogida en nuestra base de datos y que ya han sido presentados en [11].

6.1 Métodos, herramientas y estímulos

Como ya se ha mencionado, este capítulo utiliza la información recogida en el conjunto de datos WEMAC. El conjunto de datos contiene un total de 104 mujeres voluntarias que fueron expuestas a 14 estímulos audiovisuales validados relacionados con la emoción. Este conjunto de datos está destinado a estar disponible públicamente a través de diferentes versiones. En particular, esta investigación utiliza los datos contenidos en la primera versión, que representan un total de 47 de las 104 voluntarias. Se realizó un número total de 123 experimentos, de los cuales 104 grabaciones se consideraron válidas (sin mal funcionamiento de los sensores). Todos ellos se realizaron entre octubre de 2020 y julio de 2021.

En cuanto a la metodología diseñada para la experimentación, en la Figura 6-1 se muestra un esquema simplificado para cada voluntario y estímulo expuesto. El Comité de Ética de la Universidad Carlos III de Madrid aprobó este protocolo en cuanto a los aspectos éticos y de protección de datos. Antes del experimento, a los voluntarios reclutados se les explican los diferentes pasos a seguir y se les entrega una serie de documentos, como un consentimiento informado, el tratamiento de los datos personales y un cuestionario general. Como se especifica en la sección 2.3.3, este cuestionario puede proporcionar información adicional relacionada con la cognición, la valoración, la atención, los rasgos de personalidad, el género y la edad. Los datos recogidos fueron: grupo de edad, actividad física reciente o medicación que puede alterar la respuesta fisiológica del participante, cargas emocionales autoidentificadas debido a la situación laboral, económica y personal, y sesgo del estado de ánimo (miedos, fobias, experiencias traumáticas). Nótese que el análisis multivariante de estos factores está fuera del alcance de esta investigación.

Tras la lectura de la documentación, se realiza la puesta a punto del equipo, que consiste en:

- Oculus® Rift-S Headset¹ utilizado para maximizar la experiencia inmersiva y, en consecuencia, obtener una mejor elicitación de emociones. Esta es la plataforma a través de la cual se muestran los diferentes estímulos a las voluntarias tanto en formato 2D como 3D.
- Se empleó un sistema de medición estándar para adquirir diferentes señales

¹<https://www.oculus.com/rift-s/>

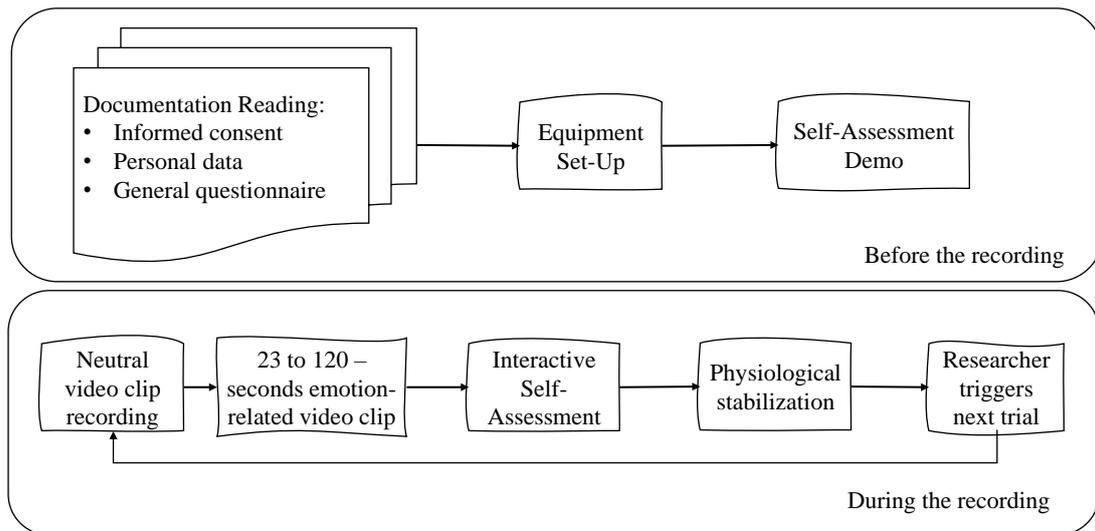


Figure 6-1: Metodología experimental seguida durante la elaboración del conjunto de datos WEMAC. Antes y durante la experimentación.

fisiológicas. En concreto, éstas fueron: PPG de dedo, GSR de muñeca ventral, SKT de antebrazo, electromiografía trapezoidal, respiración torácica y movimiento inercial de la muñeca mediante un acelerómetro. Proporciona una medida de oro que se puede comparar con el resto de los sensores, como la pulsera de Bindi. De hecho, las señales PPG y GSR obtenidas de BioSignalPlux y Bindi han sido comparadas previamente y correlacionadas con éxito en las publicaciones [159] y [234].

- Bindi con PPG de muñeca dorsal, GSR de muñeca ventral y SKT. La arquitectura de hardware y software y el diseño detallado de este elemento se detallan en el capítulo 5.
- Un sensor GSR adicional que se integrará en la siguiente iteración del brazalete (Bindi 2.0). El diseño del hardware y el software de este nuevo sensor están fuera del alcance de este documento, aunque está en marcha una publicación [235].

La sincronización de todos los diferentes sensores de adquisición junto con las etapas del experimento se lleva a cabo mediante un ordenador portátil que ejecuta un programa basado en el framework Unity®. Este trabajo fue realizado por el equipo UC3M4Safety. Hay que tener en cuenta que todos los dispositivos que detectan información fisiológica funcionaban a una frecuencia de muestreo de 200 Hz.

Finalmente, el último paso en la preparación del experimento es una demostración de etiquetado de autoevaluación, en la que las voluntarias se acostumbran al entorno

de realidad virtual y conocen mejor las diferentes categorías y particularidades del etiquetado. Los elementos recopilados durante el proceso de autoetiquetado son, por orden de aparición

- Muestra de audio grabada a través del micrófono de los auriculares Oculus, justo después de la visualización del videoclip relacionado con la emoción. Se pide a los voluntarias que revivan las emociones sentidas durante la visualización del estímulo relacionado con la emoción. Para esta investigación, se supone que la correspondencia es suficientemente sólida entre ambos instantes. Nótese que, aunque esta última suposición puede considerarse como una simplificación que se aplica para una primera manipulación de datos, necesitará una mayor validación en futuros trabajos.
- Maniqués modificados SAM para el mapeo de las dimensiones afectivas PAD, como se detalla en la sección 2.4 del capítulo 2.
- Nivel de familiaridad con respecto a la emoción sentida y a la situación mostrada en el video-clip. Ambas se preguntaron utilizando una escala Likert de 9 puntos, al igual que en el caso de los SAMs.
- Gusto por el vídeo con tres posibilidades: sí, neutro, no.
- Selección de una emoción discreta de un total de doce. Se obtuvieron a partir del estudio de preetiquetado realizado por el equipo UC3M4Safety, que utilizó los dos primeros conjuntos de datos y fue publicado y detallado en [82], [274] y [275].

Durante el experimento, como ya se ha detallado, cada voluntario visualizó un total de 14 estímulos audiovisuales en 2D o 360°. Estos estímulos se obtuvieron considerando 28 estímulos audiovisuales de un pool de estímulos mayor que contiene un total de 42 estímulos validados por más de 1332 personas (811 mujeres, 521 hombres) durante el primer experimento detallado de la base de datos UC3M4Safety [57,82]. Obsérvese que estos estímulos audiovisuales de 28 se seleccionaron basándose en tres premisas principales: el mayor acuerdo de etiquetado emocional discreto observado en las mujeres durante el experimento de preetiquetado, el objetivo de una duración adecuada del experimento de laboratorio y una distribución equilibrada de miedo frente a no-miedo considerando un modelo PA como el realizado para la selección de estímulos de la base de datos MAHNOB explicado en la sección 4.2.1

del capítulo 4. Así, se pueden aplicar dos lotes diferentes, con 14 estímulos por lote. Esta cantidad de estímulos audiovisuales, junto con la lectura de la documentación, la configuración del equipo y la demostración de autoevaluación, solía llevar de 1 a 1,5 horas por voluntario, mientras que el procesamiento de los datos implica de 3 a 8 horas. La tabla 6.1 informa de la estructura ordenada de estos lotes.

Estímulo	Emoción	Cuadrante (PA)	Longitud	Formato	Lote
1	Joy	1	1'26"	2D	1
2	Fear	2	1'20"	3D	1
3	Sadness	3	1'59"	2D	1
4	Anger	2	1'03"	3D	1
5	Fear	2	1'35"	2D	1
6	Calm	4	1'	3D	1
7	Anger	2	1'	2D	1
8	Fear	2	23"	2D	1
9	Disgust	3	40"	2D	1
10	Fear	2	2'	3D	1
11	Joy	1	1'41"	2D	1
12	Fear	2	1'20"	2D	1
13	Gratitude	4	1'40"	2D	1
14	Fear	2	1'27"	2D	1
15	Fear	2	1'52"	2D	2
16	Joy	1	1'28"	2D	2
17	Fear	2	46"	2D	2
18	Sadness	3	45"	2D	2
19	Fear	2	1'33"	3D	2
20	Calm	4	1'	2D	2
21	Anger	2	1'59"	2D	2
22	Fear	2	1'14"	2D	2
23	Disgust	3	1'36"	2D	2
24	Fear	2	2'	3D	2
25	Surprise	1	1'41"	2D	2
26	Fear	2	1'06"	2D	2
27	Gratitude	4	1'30"	2D	2
28	Fear	2	1'59"	3D	2

Table 6.1: Lista de los estímulos audiovisuales utilizados en el conjunto de datos WEMAC.

Para el primer lote, la duración de los estímulos está en $1'32'' \pm 46''$, mientras que para el segundo lote la duración está en $1'46'' \pm 44''$. Se puede observar que ambos lotes tienen 8 estímulos que pertenecen al segundo cuadrante PA, lo que se hizo a propósito para mantener un equilibrio adecuado entre las emociones parecidas al miedo y las que no lo son. Nótese que la premisa de equilibrio considera el modelo PA, en lugar del PAD, para facilitar y simplificar dicha tarea. Debido a este hecho, los estímulos preetiquetados como ira se consideraron también dentro del segundo cuadrante, estando así dentro de la clase positiva para el etiquetado de la verdad básica dimensional. Nótese que todos los voluntarios del mismo lote los visualizaron en el mismo orden.

Antes de la presentación de cada estímulo relacionado con la emoción, también se utiliza un clip neutral específico para facilitar la recuperación emocional. Estos se seleccionan aleatoriamente de un conjunto mayor proporcionado por el laboratorio de psicofisiología de Stanford [195]. Del mismo modo, pero al final de la autoevaluación, se presentan también escenas de recuperación de 360°. Estas son seleccionadas por consenso unánime del equipo de investigación. La principal diferencia entre los clips neutros y los de recuperación es que mientras los primeros son totalmente pasivos, es decir, no hay un seguimiento de la recuperación, los segundos implementan realmente un seguimiento fisiológico. Esto permite la evaluación en línea de la estabilización de las tres variables medidas. Este proceso se realiza utilizando las mediciones fisiológicas adquiridas por la pulsera de Bindi. Específicamente para estos primeros experimentos, implementamos un controlador de estabilización de la recuperación fisiológica en el SoC de la pulsera, que funcionó en base a ventanas de procesamiento de datos temporales segmentados. Dicho sistema realiza un proceso de filtrado básico online de las señales, extrae las BPMs de la frecuencia cardíaca calculada, y verifica la estabilización de las señales durante más de cuatro ventanas de procesamiento consecutivas. Una vez alcanzada la estabilización de al menos dos de las tres variables, la pulsera notificó mediante BLE al ordenador central que ejecutaba el marco de realidad virtual. Tenga en cuenta que la implementación de la recuperación fisiológica también fue seguido por una tesis de licenciatura bajo mi supervisión [276]. Su objetivo era implementar nuevos mecanismos de recuperación hacia la mejora de la actual durante los experimentos. Las nuevas características y mejoras implementadas se aplicaron incluso a otros proyectos que se están desarrollando dentro del grupo de investigación.

6.2 Exploración de respuestas de etiquetado autodirigidas

Como se ha detallado en la sección anterior, la primera versión del conjunto de datos WEMAC contiene datos de 47 voluntarias. En concreto, 32 y 15 voluntarias visualizaron el primer y segundo lote, respectivamente. Debido a las diferentes metodologías de etiquetado consideradas y basándose en los trabajos anteriores que utilizan los datos públicos de referencia explicados en el capítulo 4, se ha empleado

tanto el etiquetado discreto como el dimensional. Obsérvese que ambos son binarizados para proporcionar un problema de clasificación binaria similar al del miedo. Así, todas las etiquetas discretas que no fueron identificadas como miedo se codifican como la clase negativa, mientras que las evaluadas como miedo se establecen como la clase positiva. El mismo proceso se realiza para la metodología de etiquetado dimensional, pero siguiendo el método de binarización de miedo propuesto, véase la sección 2.3.4 del capítulo 2, tal y como se hace para las bases de datos públicas de referencia, véase el capítulo 4.

Las figuras 6-2 y 6-3 muestran el equilibrio de clases discreto y dimensional binarizado para las etiquetas autodeclaradas de los 47 voluntarias. Obsérvese que en estas figuras también se representa el balance de clases de la verdad terrestre por lote. En promedio, para la verdad básica, ambos lotes poseen 53,57% y 46,43% de clases negativas y positivas, respectivamente. El balance medio para las etiquetas de autoevaluación dimensional es de 55,80% y 44,20%, mientras que para las etiquetas de autoevaluación discreta es de 60,47% y 39,53% para las clases negativas y positivas, respectivamente. Sin embargo, la principal diferencia se obtiene al comparar la desviación estándar, que asciende a 15,22% para las etiquetas dimensionales y a 7,84% para las etiquetas discretas. Aunque el balance de clase promediado de los autoinformes dimensionales está más cerca del balance de clase dorado, su desviación es dos veces mayor que el balance de clase promediado de los autoinformes discretos, lo que está directamente relacionado con la concordancia de etiquetado de los diferentes 47 voluntarias. En este contexto, también definimos un umbral de 25,00% para identificar a los voluntarias cuyo balance de clase se vio afectado por una desviación de 1,5 (igual o superior) al balance de clase dorado, que identificamos como valores atípicos de etiquetado. Obsérvese que se trata de una simplificación de primera aproximación, ya que podrían realizarse más análisis de datos fisiológicos con dichos valores atípicos para caracterizar adecuadamente sus reacciones emocionales. Por lo tanto, los voluntarias que alcanzaron dicho umbral están marcados entre paréntesis. Para el etiquetado discreto se identificaron hasta cinco voluntarias (5, 6, 15, 33 y 40), y para el etiquetado dimensional se identificaron hasta nueve voluntarias (3, 5, 6, 13, 20, 21, 22, 40, 42). Cabe destacar las diferencias de equilibrio de clases observadas por voluntario al considerar ambas metodologías. Por ejemplo, el tercer voluntario

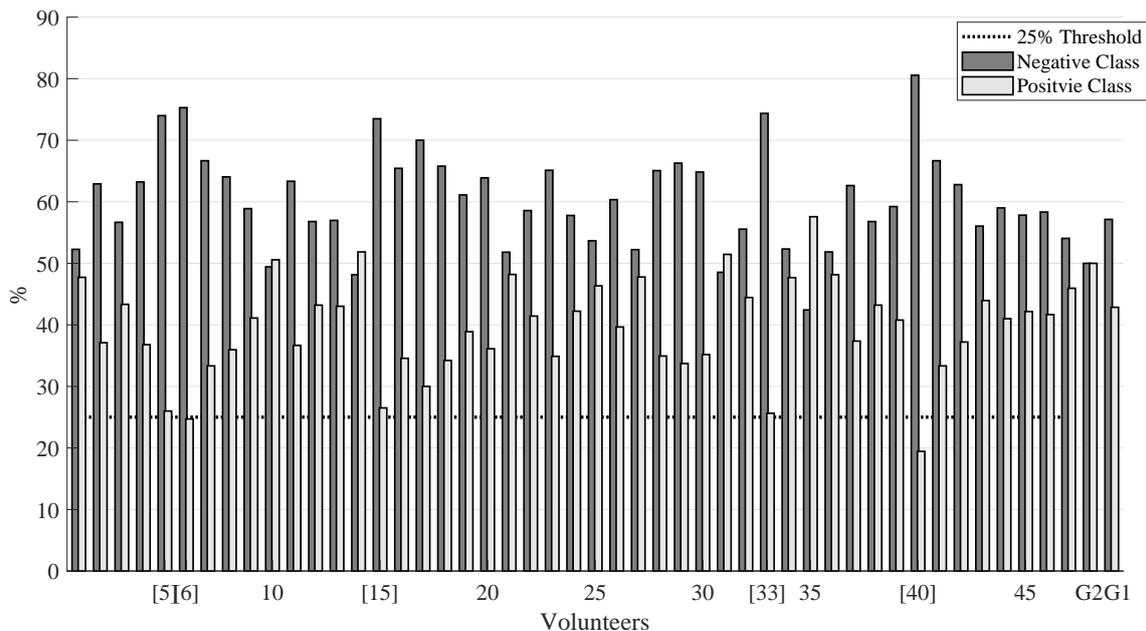


Figure 6-2: Distribución de clases para el mapeo de miedo binario sobre los autoinformes subjetivos discretos en WEMAC para todas las 47 voluntarias consideradas, y la distribución de clases original prevista en el experimento: G2 y G1 para el segundo y primer lote, respectivamente.

muestra un equilibrio de 57/43% aproximadamente para el etiquetado discreto, y un equilibrio de 85/25% para la evaluación dimensional. Este hecho sugiere una diferente comprensión y entendimiento de cada una de estas metodologías de etiquetado, lo que puede dar lugar a diferentes sistemas de aprendizaje automático cuando se utiliza una u otra. Esta es la razón principal que nos llevó a trabajar con ambos enfoques.

Siguiendo el mismo análisis de esquemas para este conjunto de datos que el aplicado a las bases de datos públicas de referencia en el capítulo 4, se han evaluado las correlaciones interindividuales de las etiquetas para comprobar si todos los voluntarias etiquetan cada estímulo relacionado con la emoción. En este caso, los resultados obtenidos tras una prueba de Levene y una prueba de Kruskal-Wallis para el etiquetado discreto binarizado proporcionaron resultados diferentes. La primera rechazó la hipótesis nula de que las varianzas son iguales en todos los voluntarias ($p < 0,001$), mientras que la segunda no la rechazó ($p > 0,001$). Por el contrario, los mismos métodos rechazaron la hipótesis nula para las etiquetas dimensionales binarizadas ($p < 0,001$). Esta diferencia es una consecuencia de la conclusión final extraída de las figuras anteriores, ya que indica que, al menos para la metodología de etiquetado discreto, no hay pruebas suficientes para afirmar que las varianzas sean diferentes

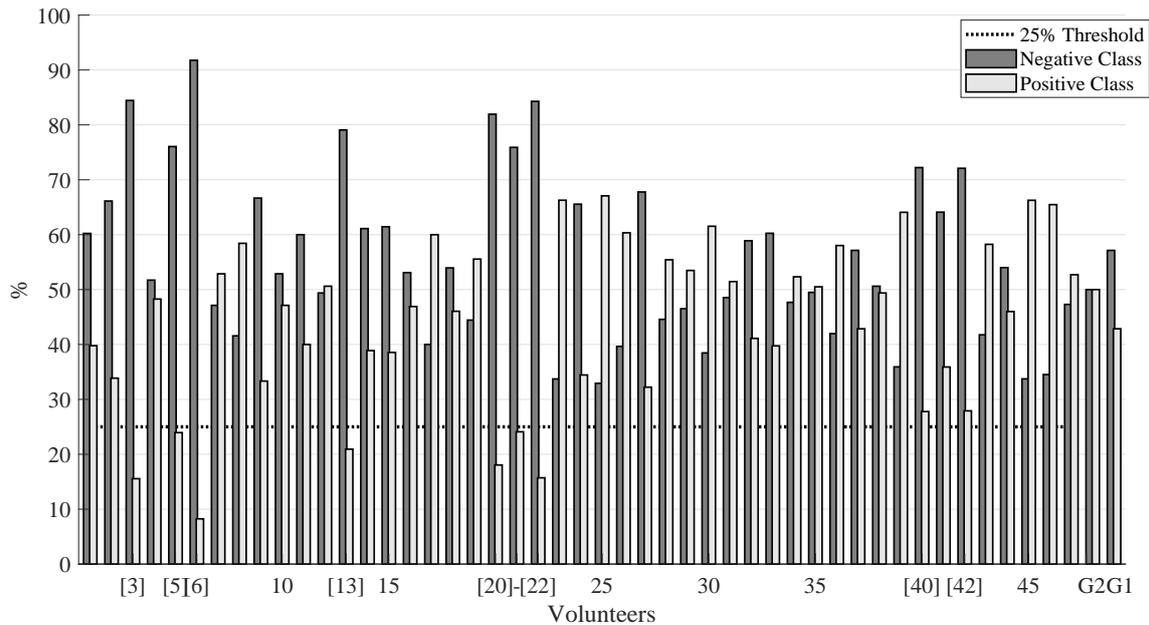


Figure 6-3: Distribución de clases para el mapeo binario del miedo sobre los autoinformes subjetivos dimensionales en WEMAC para todas las 47 voluntarias consideradas, y la distribución de clases original prevista en el experimento: G2 y G1 para el segundo y primer lote, respectivamente.

entre los voluntarias. Por tanto, este hecho sugiere que cada una de las metodologías está caracterizando aspectos diferentes de las emociones, lo que está en consonancia con la información aportada en el Capítulo 2 al afirmar que existen ambos modelos pero cada uno pretende explicar rasgos diferentes de las emociones [48]. Nótese que el conjunto de etiquetas binarizadas presenta una distribución no normal y que el nivel de significación se fijó en $p < 0,05$.

Tras el análisis de la varianza, también se aplica la correlación de Spearman para este conjunto de datos. Sin embargo, debido a la diferencia observada anteriormente, se muestran las matrices no promediadas para demostrar gráficamente el efecto y la consecuencia interindividual real. En concreto, la Figura 6-4 presenta la inter-correlación a través de los 47 voluntarias para ambas metodologías. Estas matrices proporcionan una información unitaria de los sujetos en relación con las diferencias de etiquetado. Al comparar las matrices de intercorrelación, podemos detectar algunas regiones comunes en ambas. Aunque se puede observar que la matriz de intercorrelación discreta posee un color gris más claro, lo que indica que las correlaciones son ligeramente más positivas, no hay una conclusión diferencial clara analizando únicamente estas matrices. Por lo tanto, los valores p correspondientes a dichas matrices de correlación se representan en la figura 6-5. En este caso, hay

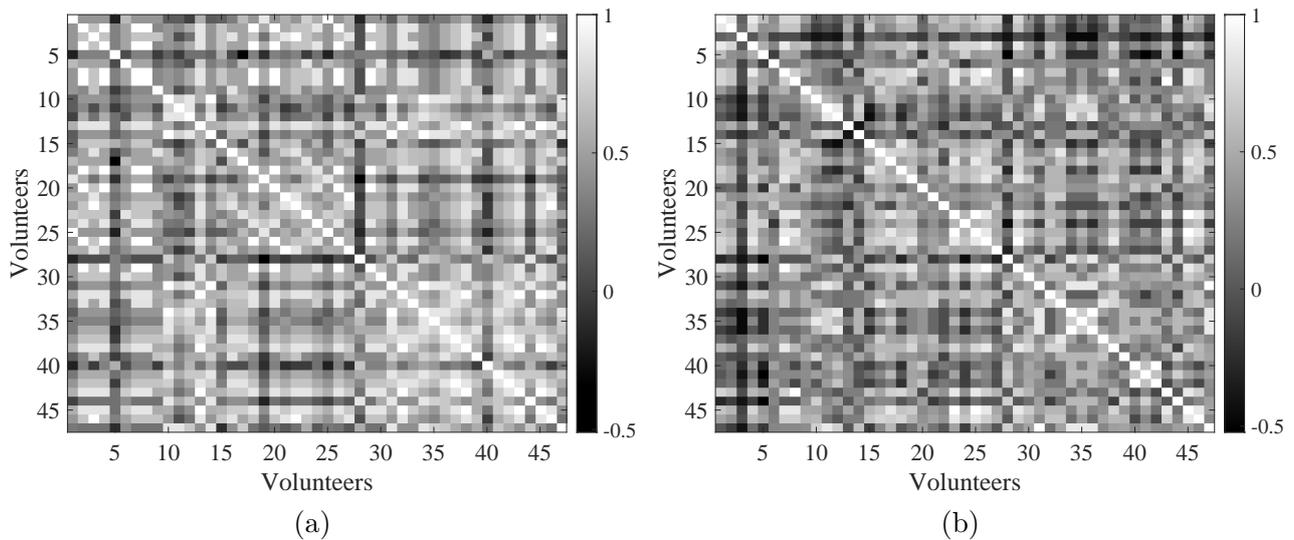


Figure 6-4: Intercorrelación de Spearman uno a uno entre los sujetos de los 47 voluntarias para ambas metodologías de etiquetado: a) discreta, y b) dimensional (PAD).

una clara distinción entre ambas metodologías. El etiquetado discreto muestra una identificación de color negro para la mayoría de los voluntarias, lo que indica un valor p inferior a 0,1. Por el contrario, la autoevaluación dimensional no informa de este comportamiento. Este hecho apoya las conclusiones anteriores y sugiere que la asociación o concordancia entre las etiquetas binarias de miedo de los voluntarias dentro del caso discreto es más fuerte que con la metodología dimensional. Por favor, para contextualizar este análisis, ténganse en cuenta las siguientes dos consideraciones: a) los voluntarias de ambos lotes fueron utilizados indistintamente, y b) este análisis sirve como estudio preliminar para evaluar la concordancia dentro de la misma metodología y las diferencias con respecto a ambas, sin embargo, se puede continuar profundizando en las diferencias específicas de los voluntarias uno a uno y/o incluso aplicando diferentes métodos estadísticos.

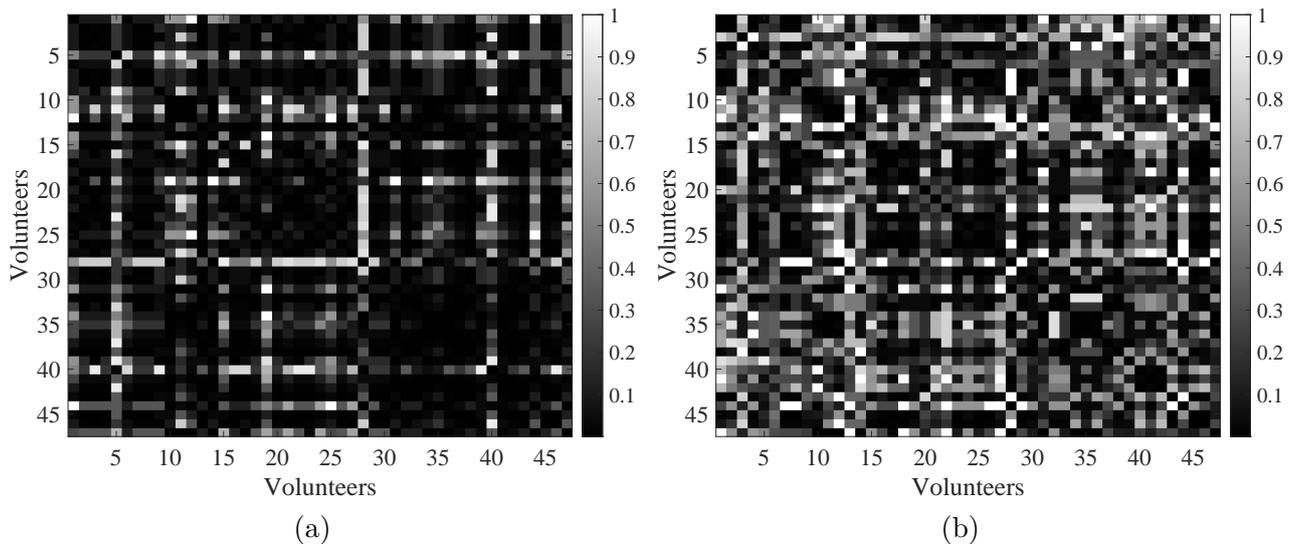


Figure 6-5: Valores P obtenidos a partir de la intercorrelación de sujetos de Spearman uno a uno a través de los 47 voluntarias para ambas metodologías de etiquetado: a) discreto, y b) dimensional (PAD).

Con el fin de proporcionar una perspectiva individual promediada para el acuerdo, se muestran las Figuras 6-6 y 6-7 promediadas. Los resultados obtenidos no rechazan la hipótesis nula en promedio para cada uno de los 47 voluntarias, lo que indica que no hay pruebas suficientemente sólidas para sugerir que existe una asociación entre las etiquetas binarias de miedo de los voluntarias de forma inequívoca. Por ejemplo, los valores p medios son de 0,15 y 0,31 para los casos discreto y dimensional, respectivamente. Así pues, aunque hay que señalar que se trata de un resultado promediado, la conclusión extraída está en consonancia con las anteriores. Obsérvese que los voluntarias que muestran los valores p promediados más altos son, en general, los que poseían las incoherencias de etiquetado indicadas en las figuras 6-2 y 6-3.

El análisis realizado en esta sección sugiere que la metodología de etiquetado discreto supera, en términos de acuerdo, a la evaluación dimensional. Esta conclusión no significa estrictamente que un sistema entrenado por separado con ambas etiquetas vaya a tener la misma diferencia de rendimiento, ya que las etiquetas autodeclaradas siempre se ven afectadas y sesgadas por procesos cognitivos, a diferencia de las respuestas fisiológicas, tal y como se detalla en el Capítulo 2. Por lo tanto, los diferentes resultados obtenidos en este estudio de equilibrio de estímulos y consideración de etiquetas deben contextualizarse también al evaluar los resultados obtenidos para los sistemas de aprendizaje automático basados también en ambas metodologías de etiquetado.

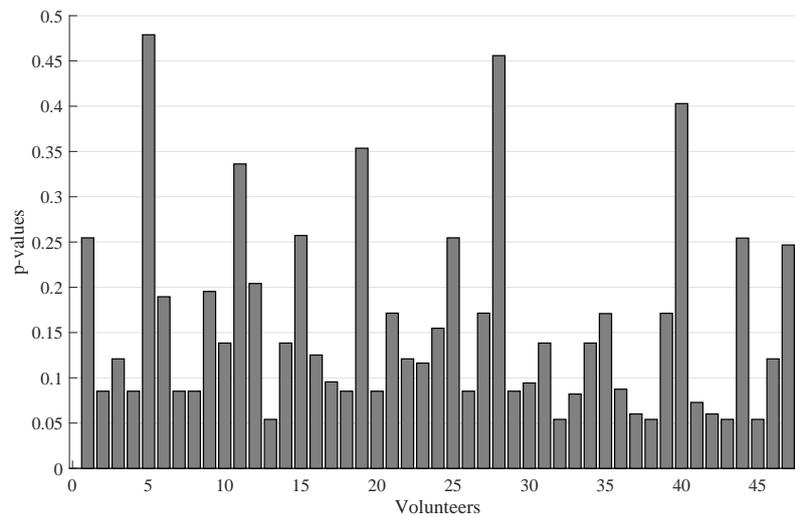


Figure 6-6: Valores p promediados para todos los voluntarias considerados y sus etiquetas aplicando la correlación de Spearman para sus etiquetas de mapeo binario de miedo basadas en PAD.

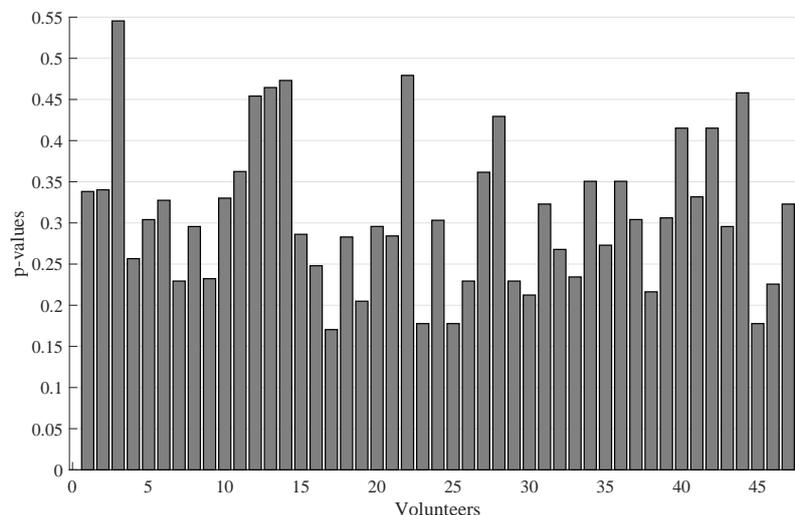


Figure 6-7: Valores p promediados para todos los voluntarias considerados y sus etiquetas aplicando la correlación de Spearman para sus etiquetas de mapeo binario del miedo basado en la discreción.

6.3 Exploración de la respuesta fisiológica

En esta sección, llevamos a cabo una exploración de la respuesta fisiológica para dar una visión adecuada de los patrones fisiológicos, las recuperaciones y otras particularidades observadas durante los experimentos. Además, se concluye presentando los primeros resultados de detección de miedo basados en dicha información utilizando las etiquetas discretas y dimensionales autodeclaradas para los 47 voluntarias. Las métricas obtenidas se utilizan y se fusionan con los resultados de audio en la siguiente sección.

Las señales utilizadas a lo largo de esta exploración de la respuesta fueron debida-

mente filtradas y denotadas. Las señales BVP se sometieron a la misma estrategia de filtrado descrita en la sección 4.1.2, y también a un filtro IIR de paso bajo hacia delante y hacia atrás para tratar el residuo de la línea de base. Para las señales GSR y SKT, se aplicó un filtro FIR básico con una frecuencia de corte de 2Hz. Después, esta salida filtrada se muestreó a 10 Hz y también se procesó con filtros de media móvil y mediana móvil. El primero utilizó una ventana de 1 segundo y ayudó a reducir el alto ruido residual después del FIR inicial, mientras que el segundo empleó una ventana de 0,5 segundos y se ocupó de los transitorios rápidos.

Todos los análisis y resultados fisiológicos presentados en esta sección se realizan utilizando las señales adquiridas por el sistema de herramientas de investigación BioSignalPlux®. Ésta decisión se considera para obtener resultados comparables para los diferentes sistemas de análisis fisiológico y de aprendizaje automático propuestos con respecto a la literatura. Este hecho es fundamental para poder replicar posteriormente el mismo análisis para los otros sistemas de sensores empleados en los experimentos y evaluar las diferencias. Aunque esta última tarea no está dentro del alcance de este trabajo de investigación, la verificación y validación de la adquisición de señales fisiológicas con Bindi y el sistema de herramientas de investigación BioSignalPlux® ya ha sido realizada y publicada en [159, 234].

6.3.1 Patrones fisiológicos y recuperaciones

El trabajo presentado en los capítulos 4 y 6, que trata de todas las etapas del procesamiento digital de la señal, como el filtrado y la extracción de características, nos permitió proporcionar un análisis más profundo de la respuesta fisiológica dentro de nuestro propio conjunto de datos (WEMAC). La exploración fisiológica es una tarea difícil cuando se considera este tipo de experimentos. Este hecho se ve afectado principalmente por la complejidad de las emociones y por las incertidumbres o variaciones fisiológicas intrínsecas debidas a las diferencias intra e inter individuales. Así, en aras de la simplicidad, en esta sección realizamos una exploración fisiológica preliminar considerando algunas de las señales revisadas y un conjunto reducido de características. En concreto, para el análisis del patrón fisiológico, utilizamos la señal GSR extraída durante la visualización de los estímulos relacionados con la emoción para determinar el grado de similitud considerando las señales entre el mismo y otros voluntarias. En cuanto al análisis de la recuperación fisiológica,

se utilizaron características específicas extraídas de las señales GSR y BVP para proporcionar una comparación detallada entre las etapas de recuperación y de visualización de los estímulos relacionados con la emoción. Estos análisis proporcionaron información útil sobre las respuestas fisiológicas esperadas y reales de los voluntarias. Además, también pueden ampliarse estudiando el conjunto completo de señales y características fisiológicas.

6.3.1.1 Análisis de patrones

Para la exploración del patrón fisiológico, como ya se ha comentado, se analizó el comportamiento de la señal GSR. La selección de esta señal fisiológica se basó en la relación directa que tiene con las respuestas emocionales, tal y como se ha revisado y estudiado previamente en los capítulos 2 y 4.

En este estudio, la exploración de patrones se realizó para todos los 47 voluntarias utilizando una técnica común de análisis de patrones de series temporales llamada Dynamic-Time-Warping (DTW) [194, 277–279]. Esta técnica permite cuantificar la similitud entre dos series temporales con características equivalentes aunque tengan velocidades o trayectorias de espacio de fase diferentes. Por ejemplo, las señales GSR entre los diferentes voluntarias muestran este comportamiento. La figura 6-8 muestra las señales GSR de los voluntarias 4, 15 y 27, extraídas durante la visualización de uno de los estímulos de miedo. Se pueden observar diferentes localizaciones de los picos fásicos, algunos dentro del mismo intervalo temporal para los tres voluntarias y otros en instantes totalmente diferentes. Obsérvese que el comportamiento y la dinámica de la señal dependen principalmente del tipo de estímulos relacionados con la emoción y del voluntario (factores intraindividuales como se detalla en el capítulo 2). Cada una de estas señales, que se adquieren a intervalos regulares, puede definirse como

$$S^{i,j} = (s_1^{i,j}, s_2^{i,j}, s_3^{i,j}, \dots, s_N^{i,j}), \quad (6.1)$$

donde i y j son el voluntario y el estímulo, respectivamente; y $s_k^{i,j}$, con $k \subseteq [1, N]$, son las diferentes muestras adquiridas para toda la duración del estímulo relacionado con la emoción. Así, DTW encuentra la medida de distancia óptima entre señales, siguiendo algunas reglas de restricción, destacando las similitudes entre las señales y proporcionando una medida de su similitud independientemente de las variaciones no lineales. En concreto, se utiliza una función de coste para evaluar la disimilitud

entre todas las muestras de las dos series temporales que se comparan. En nuestro caso, la función de coste viene dada por la distancia euclidiana siguiente

$$d_{mn}(S^{i,j}, S^{q,p}) = \sqrt{\sum_{m,n=1}^K (s_m^{i,j} - s_n^{q,p}) * (s_m^{i,j} - s_n^{q,p})}, \quad (6.2)$$

donde m y n son las muestras específicas de cada serie temporal. Obsérvese que j y q pueden ser el mismo o diferentes estímulos. Los resultados obtenidos con esta operación se ordenan en una matriz de costes, que se utiliza para encontrar la trayectoria óptima o warping. Una vez encontrado dicho camino, el resultado final es el coste total o distancia, que está directamente relacionado con la similitud entre ambas secuencias, dada por

$$d_{min}(S^{i,j}, S^{q,p}) = \sum_{m,n \in K} d_{mn}(S^{i,j}, S^{q,p}). \quad (6.3)$$

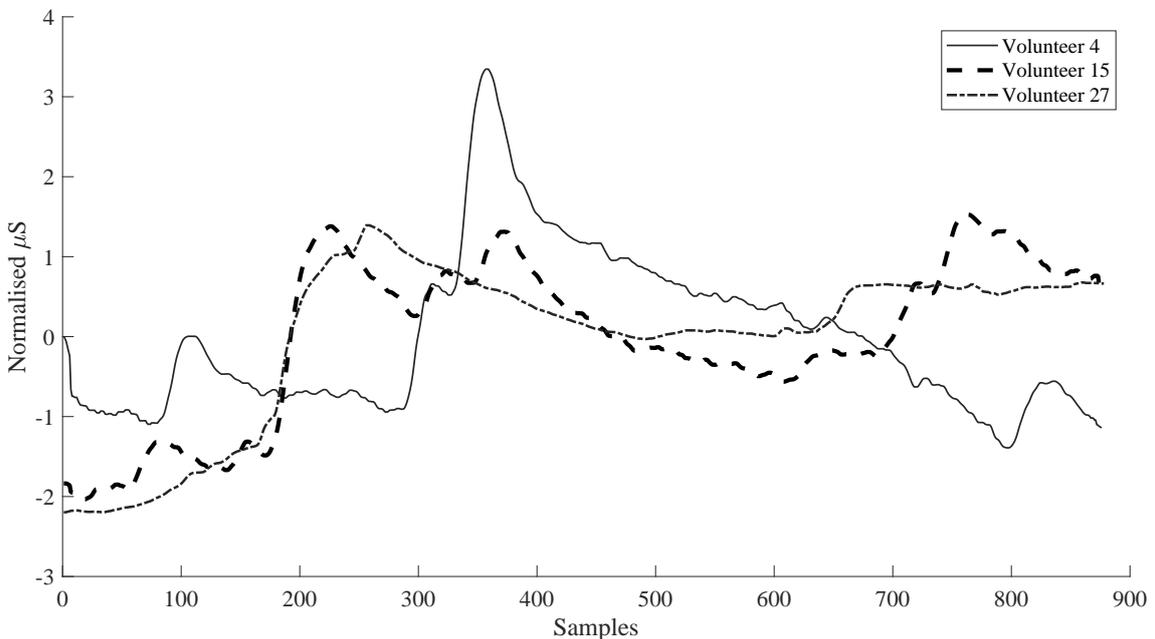


Figure 6-8: Señales GSR extraídas de la visualización completa del sexto estímulo de la primera tanda (último estímulo) de los voluntarios 4, 15 y 27.

Por lo tanto, en el caso de las señales GSR representadas para los tres voluntarios

en este ejemplo, las distancias obtenidas son:

$$\begin{aligned}d_{min}(S^{4,6}, S^{15,6}) &= 585.23 \\d_{min}(S^{4,6}, S^{27,6}) &= 549.13 \\d_{min}(S^{27,6}, S^{15,6}) &= 172.87,\end{aligned}\tag{6.4}$$

que indica que los voluntarias 15 y 27 poseen mayores similitudes que las otras dos combinaciones de voluntarias. Así, para este ejemplo, 2 de los 3 voluntarias examinados presentan un patrón fisiológico, un comportamiento o una dinámica similares con respecto a este estímulo específico.

Dentro de este contexto de análisis de patrones, se abordaron tres casos de uso diferentes de agrupación de patrones basados en la separación o combinación de los diferentes lotes. Para todos los casos de uso, los diferentes segmentos analizados GSR se normalizaron (puntuación Z) y se compararon entre sí. Nótese que en este caso, cada segmento se refiere a la señal GSR extraída para cada estímulo completo relacionado con la emoción. Antes de estudiar el análisis del patrón individual de segmento a segmento, generamos visualizaciones de matrices promediadas y gráficos de DTW agregados, como se muestra en las Figuras 6-9 y 6-10, respectivamente. El primero da una idea de la similitud de patrones promediada para todos los voluntarias y todo el experimento, es decir, la distancia total para cada voluntario se calcula promediando el conjunto de distancias obtenidas para cada comparación que consideró todo el experimento. Se puede observar cómo los puntos negros de la matriz son los que presentan mayor similitud (por ejemplo, la diagonal). Esta matriz puede servir como herramienta para evaluar una primera perspectiva gráfica respecto a la agrupación de patrones. En este caso, la Figura 6-10 representa los valores promediados de la matriz para los 32 voluntarias del primer lote y los 6 estímulos de miedo. Aunque se pueden ver algunos puntos oscuros dentro de la matriz (por ejemplo, los voluntarias 8-1, 15-2, 31-16, etc.), no podemos concluir que exista una formación de agrupación de patrones. En caso de agregar todas las distancias entre voluntarias dentro de la matriz y obtener la media y la desviación estándar (omitimos la parte diagonal), se puede informar de la figura 6-10. De este modo se obtiene una perspectiva macro del comportamiento de cada voluntario en comparación con los demás. Así, podemos concluir que, en promedio, no hay voluntarias

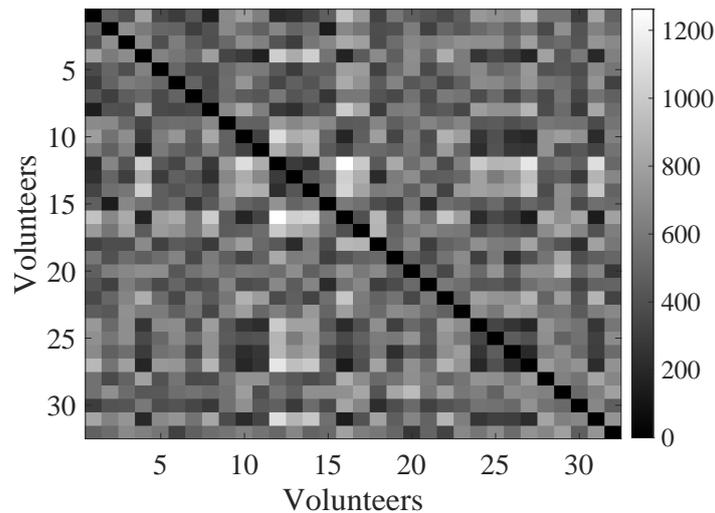


Figure 6-9: Matriz de distancia DTW promediada para los 32 voluntarias que visualizan los 6 estímulos de miedo del primer lote de estímulos relacionados con la emoción.

extremadamente desviados. Sin embargo, tampoco podemos afirmar la existencia de formaciones de patrones. Obsérvese que, en aras de la simplicidad, sólo se muestran esta matriz y el gráfico, pero el resto de los casos de uso también se analizaron y llevaron a la misma conclusión.

Después de realizar el análisis anterior y con el fin de cuantificar la agrupación de segmentos, se realizaron estudios de agrupación de segmentos independientes y dependientes del sujeto. En primer lugar, se extrajeron las similitudes (distancias) de cada segmento GSR con respecto al resto de segmentos de los mismos y diferentes voluntarias. En segundo lugar, considerando el conjunto de distancias reunidas, se encontró el mínimo. Por último, se asignó al segmento actual que se estaba procesando la misma etiqueta que la de dicho mínimo. Obsérvese que para este análisis se consideraron las etiquetas binarizadas de verdad básica como las esperadas (no autodeclaradas). Este hecho se basó en la evaluación de la viabilidad de encontrar agrupaciones de patrones fisiológicos sin ninguna información de sesgo cognitivo específica del sujeto. La tabla 6.2 informa de los resultados tras realizar los estudios de clustering para los diferentes casos de uso. Se puede observar que los resultados más altos para las tres métricas comparadas se obtienen siempre cuando se considera el enfoque de clustering independiente del sujeto. Nótese que este hecho también puede verse afectado por el tamaño del conjunto de datos, ya que la cantidad de datos para la agrupación dependiente del sujeto es considerablemente menor que para la independiente del sujeto. En general, hay más segmentos fisiológicos similares dentro de

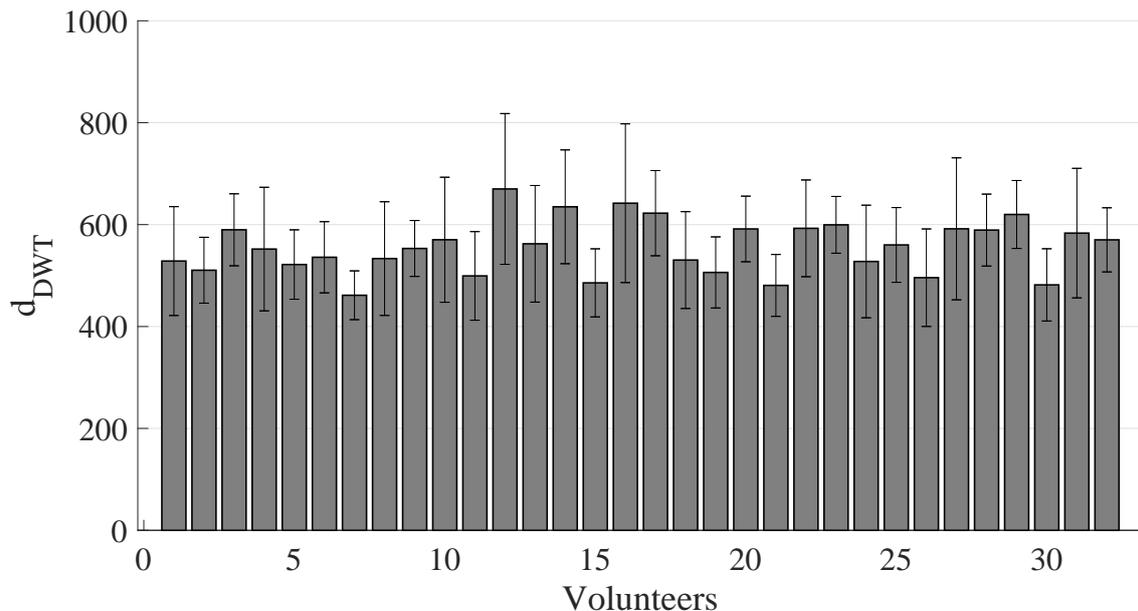


Figure 6-10: Resultados agregados obtenidos a partir de la matriz de distancia DTW promediada para los 32 voluntarias de la Figura 6-9.

los estímulos sin miedo (especificidad). De hecho, la agrupación de los segmentos de los estímulos de miedo no supera el umbral de detección del 50,00% (sensibilidad). La mayor diferencia entre ambos lotes se registra para el agrupamiento independiente del sujeto, con una sensibilidad de hasta el 47,92% y el 29,50% para el primer y el segundo lote, respectivamente. Una vez más, este hecho puede indicar una clara diferencia de respuesta fisiológica con respecto al efecto de los estímulos de miedo, sin embargo, es necesario contextualizarlo a la menor cantidad de voluntarias que se evalúan para el segundo lote. Por último, la mayor sensibilidad se consigue al considerar ambos lotes conjuntamente, aunque a costa de obtener la menor especificidad para la agrupación independiente del sujeto. A pesar de este último hecho, la Gmean para dicho caso de uso es una de las más altas con hasta un 53,45%. Por lo tanto, tras este estudio podemos concluir diferentes aspectos clave relativos a las respuestas fisiológicas debidas a los estímulos relacionados con el miedo y sin miedo, así como a la agrupación fisiológica global:

- El análisis de las similitudes de las señales GSR entre los distintos voluntarias no consigue distinguir adecuadamente entre los estímulos relacionados con el miedo y los que no lo están. Obsérvese que este análisis puede ampliarse para explorar otras señales, así como diferentes técnicas de exploración de patrones.
- La agrupación de estímulos no relacionados con el miedo está mejor caracterizada o identificada. Este hecho se da independientemente del lote considerado.

- La sensibilidad observada al considerar ambos lotes juntos es la más alta. Esto implica que existen estímulos relacionados con el miedo que evocan respuestas fisiológicas similares independientemente del lote y de su contenido audiovisual específico. Por lo tanto, en esta investigación aplicamos una perspectiva agnóstica de los lotes al considerar que ambos lotes pueden ser utilizados conjuntamente para seguir diseñando un sistema de aprendizaje automático de detección de miedo más eficiente.
- Las métricas de bajo rendimiento para los diferentes casos de uso del clustering indican que esta información no es suficiente para desentrañar y distinguir los mecanismos de activación fisiológica relacionados con el miedo. Por lo tanto, se podrían explotar más señales y/o características para lograr dicho objetivo.

Clustering Tipo	Lote Número	Métricas de identificación de segmentos		
		SPE	SEN	Gmean
Sujeto-dependiente	1	54.29%	43.23%	48.45%
	2	48.57%	48.57%	48.57%
Sujeto-independiente	1	61.33%	47.92%	54.16%
	2	64.76%	29.50%	43.70%
	1&2	57.34%	49.83%	53.45%

Table 6.2: Estudio de agrupación de un solo segmento para los sujetos dependientes e independientes. SPE: especificidad, SEN: sensibilidad, Gmean: media geométrica.

6.3.1.2 Análisis de recuperación

Para el análisis de la recuperación fisiológica, tal y como se ha explicado anteriormente en la sección 6.1, se realizó una evaluación de estabilización online de las tres señales fisiológicas diferentes que se adquirirían con la pulsera de Bindi durante los experimentos. Este proceso online operaba cada diez segundos realizando un filtrado básico online, extrayendo los BPMs de la señal BVP, y finalmente evaluando la estabilización de los BPMs, GSR, y SKTs para más de cuatro ventanas de procesamiento consecutivas. Este último proceso se realizó mediante el ajuste del umbral duro siguiendo un intervalo de confianza del 90% respecto al nivel de la primera ventana. Para analizar el efecto real de estas fases de recuperación, se realizó un estudio fisiológico posterior. En concreto, se utilizaron las características

extraídas de las señales GSR y BVP para proporcionar una comparación detallada entre las etapas de recuperación y de visualización de los estímulos. Por un lado, se compara el número de ERSCR o picos fásicos, la amplitud y el tiempo de subida. Por otro lado, se elaboran diferentes Poincaré-plots para valorar el estado simpático dentro de la etapa de recuperación [280]. Nótese que este análisis de recuperación se realiza considerando ambos baños.

La figura 6-11 muestra los resultados promediados de los picos detectados durante el experimento. Se hace una distinción dividiendo las respuestas fisiológicas relacionadas con el miedo y las no relacionadas con el miedo, que también se aplica para sus respectivas etapas de recuperación. Este proceso se realizó considerando toda la señal fisiológica adquirida para los diferentes estímulos y etapas de recuperación, es decir, no se aplicó ninguna segmentación. Esta última consideración se adopta porque el objetivo principal de este análisis es evaluar las respuestas fisiológicas. Por lo tanto, no es necesario aplicar restricciones reales como la segmentación de los datos. Nótese también que, para este análisis, se utilizó el algoritmo `cvxEDA`, que se detalla en la sección 2.5.2. Como se esperaba, los picos detectados para todos los estímulos de miedo superan a los detectados para las etapas de no-miedo y de recuperación. En concreto, durante la visualización de los estímulos de miedo se detectó una media de 2,30 picos por estímulo con una desviación estándar de 0,81, mientras que los estímulos de no-miedo produjeron 1,11 picos con una desviación estándar de 0,52. Uno de los aspectos clave de estos resultados es que las etapas de recuperación están por debajo de dichas métricas para ambos tipos de estímulos. Esto se obtiene tanto para los valores promediados como para la desviación estándar: la recuperación del miedo presenta una media de picos de 0,99 (0,37), y la recuperación del no-miedo llega a 1,03 (0,48). Para apoyar los resultados de la detección de picos, su amplitud y tiempo de recuperación también se extraen y se representan en las figuras 6-12 y 6-13. Obsérvese que la amplitud indicada se obtiene como la amplitud relativa desde el inicio del pico detectado, así como para el tiempo de recuperación, véase la Figura 2-14. En general, se observa el mismo comportamiento para estas métricas. Sin embargo, la media y la desviación estándar de la amplitud relativa para la recuperación sin miedo superan las métricas de los estímulos sin miedo. En concreto, los estímulos sin miedo alcanzan una amplitud

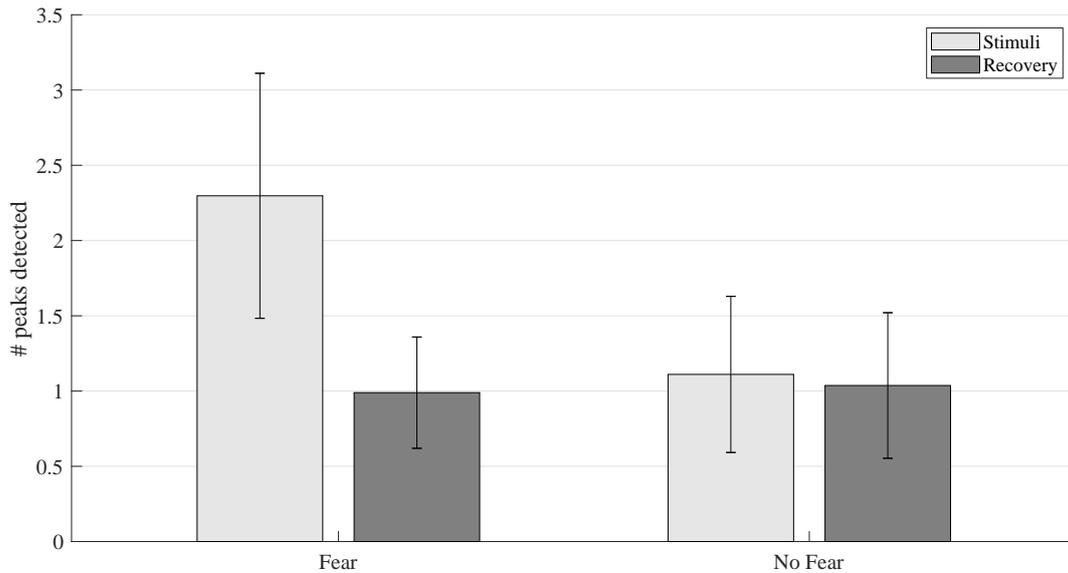


Figure 6-11: Comparación de resultados promediados obtenidos del proceso de extracción de picos GSR mediante el algoritmo *cvxEDA* para los voluntarios de 47 y ambos lotes.

relativa media de $0,01\mu\text{S}$ con una desviación estándar de $0,006\mu\text{S}$, y la recuperación sin miedo proporciona una amplitud relativa media de $0,02\mu\text{S}$ con una desviación estándar de $0,01\mu\text{S}$. Esta diferencia fisiológica debe contextualizarse junto con el tiempo de recuperación de los picos extraídos, en el que observamos exactamente el mismo comportamiento que para los picos detectados. Así, observando este comportamiento, podemos concluir que, en promedio, el nivel de excitación es el esperado para las etapas de recuperación en comparación con los estímulos de miedo y no miedo juntos. Por lo tanto, la aplicación del proceso de recuperación activa implementado reduce el sesgo emocional entre los estímulos.

Para el análisis del BVP, utilizamos una herramienta comúnmente aplicada para valorar la activación simpática, que se conoce como Poincaré-plot. Se trata de un gráfico de recurrencia en el que los IBIs consecutivos se trasladan a un diagrama de dispersión bidimensional para obtener una imagen gráfica del comportamiento del HRV para un intervalo de tiempo determinado, Figura 6-14. A partir de este gráfico concreto se obtienen diferentes métricas geométricas. En general, las dos más importantes son la desviación estándar a lo largo y perpendicular a la línea de identidad, SD_2 y SD_1 respectivamente. Se ha demostrado que estos rasgos pueden caracterizar la activación simpática y parasimpática. Por ejemplo, el hecho de que el cluster principal tenga una forma estrecha es una indicación del dominio de los

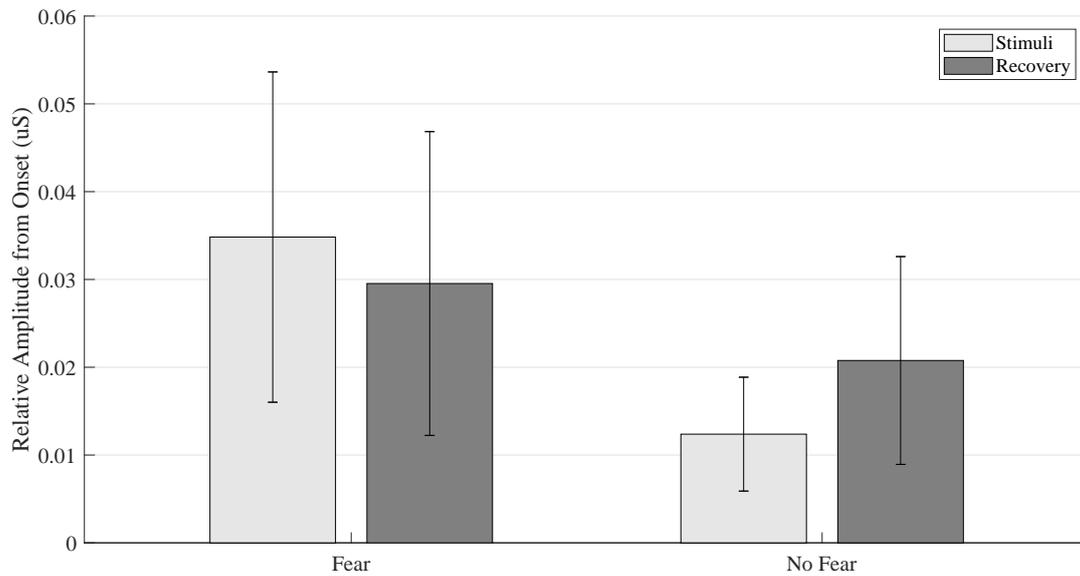


Figure 6-12: Comparación de los resultados promediados obtenidos del proceso de extracción de amplitudes relativas GSR mediante el algoritmo cvxEDA para los 47 voluntarias y ambos lotes.

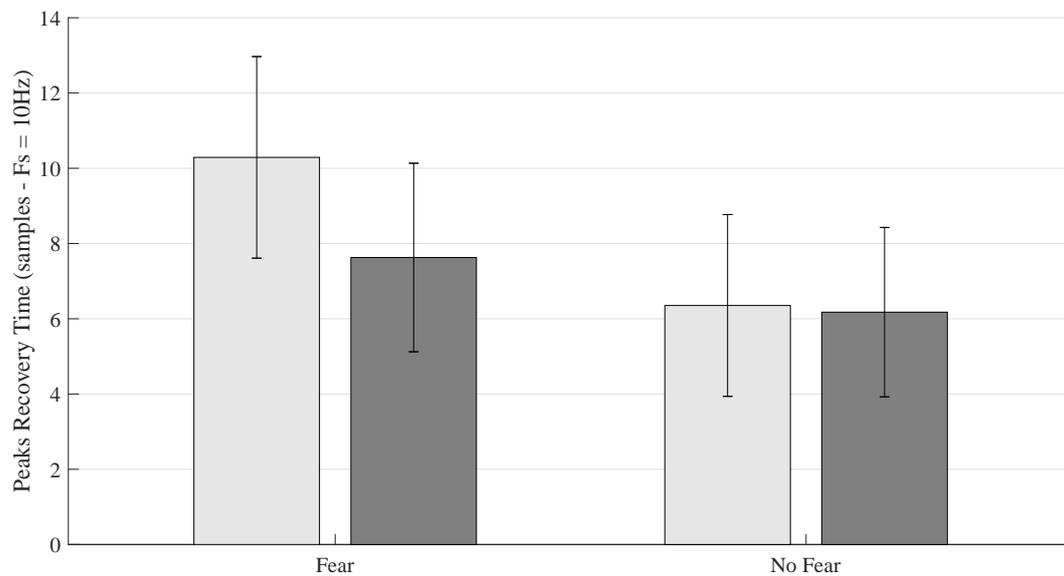


Figure 6-13: Comparación de los resultados promediados obtenidos del proceso de extracción del tiempo de recuperación de picos GSR mediante el algoritmo cvxEDA para los voluntarias 47 y ambos lotes.

componentes no respiratorios que regulan la frecuencia cardíaca, lo que está directamente relacionado con la activación simpática. Por el contrario, cuanto más amplio sea el clúster, mayor será la dominancia de los componentes respiratorios, lo que está relacionado con el predominio del parasimpático. Además, este tipo de gráfico permite estudiar la no linealidad de la información cardíaca, así como ser insensible a las tendencias de los IBI [281–283]. Obsérvese que la serie temporal del IBI se explica y representa en la sección 4.1.3.1 y en la ecuación 4.6. El cálculo de ambas características de desviación estándar se ha realizado siguiendo una simplificación considerando [284–286]. Así, estos se calculan como la desviación estándar de la serie temporal obtenida de la siguiente manera:

$$\begin{aligned} SD_2(i) &= \left(\frac{\sqrt{2}}{2}\right) * (IBI_i + IBI_{i+1}), \\ SD_1(i) &= \left(\frac{\sqrt{2}}{2}\right) * (IBI_i - IBI_{i+1}). \end{aligned} \quad (6.5)$$

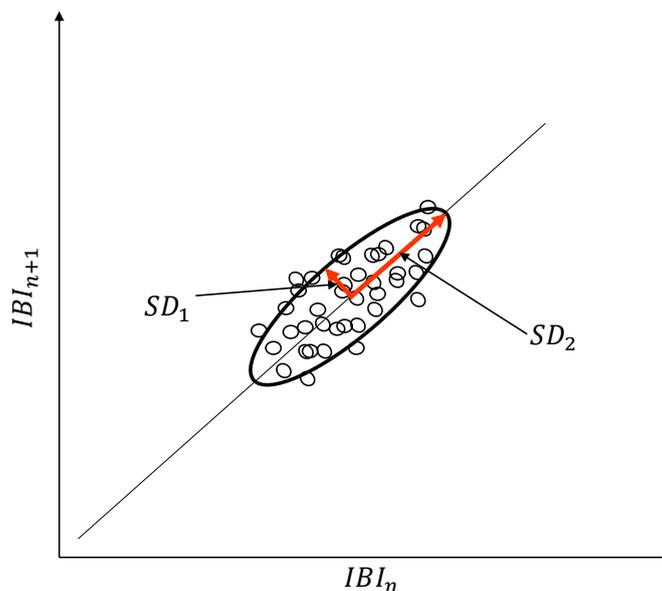


Figure 6-14: Ejemplificación de un gráfico de Poincaré recurrente y su métrica de desviación estándar a lo largo (SD_2) y perpendicular (SD_1) a la línea de identidad.

Para nuestro experimento, al igual que se hizo con la señal GSR, se consideró toda la señal BVP para los diferentes estímulos y etapas de recuperación, es decir, no se aplicó ninguna segmentación. La figura 6-17 muestra diferentes perspectivas para los tres Poincaré-plots obtenidos: estímulos con y sin miedo, y etapas de recuperación. Obsérvese que todas las series temporales del IBI de los 47 voluntarias están contenidas en estos Poincaré-plots. A primera vista, podemos observar que

los puntos de miedo tienden a estar ligeramente más cerca de la esquina inferior izquierda. Este hecho es una indicación de una menor variabilidad de la frecuencia cardíaca o de un mayor ritmo cardíaco. Además, los puntos de recuperación son los que presentan una mayor dispersión o una forma más amplia, lo que indica una dominancia parasimpática.

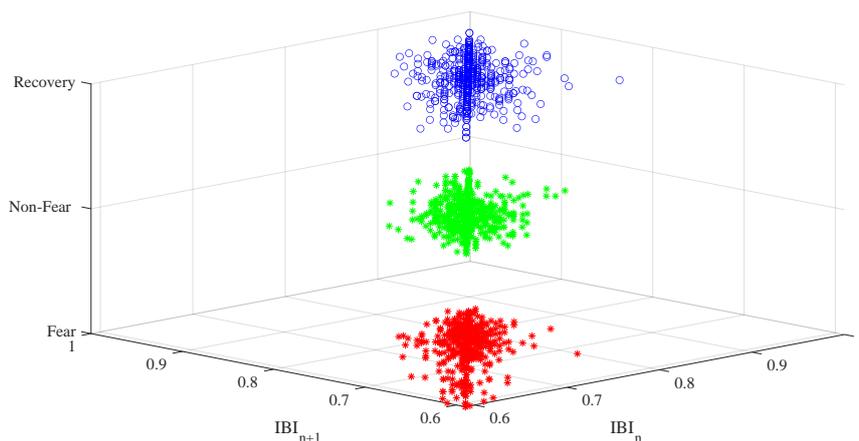


Figure 6-15: Diferentes perspectivas de Poincaré para todos los 47 voluntarios considerando los estímulos de miedo (rojo-abajo), los estímulos sin miedo (verde-medio) y las etapas de recuperación (azul-arriba). Vista frontal.

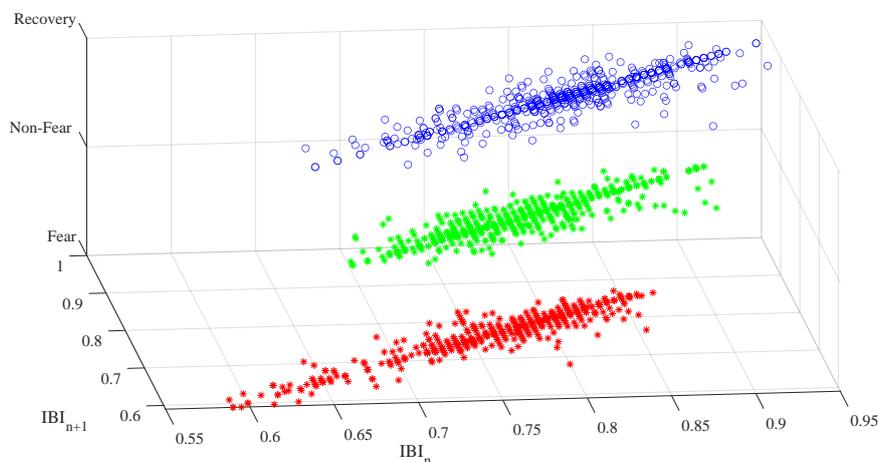


Figure 6-16: Diferentes perspectivas de Poincaré para todos los 47 voluntarios considerando los estímulos de miedo (rojo-abajo), los estímulos sin miedo (verde-medio) y las etapas de recuperación (azul-arriba). Vista longitudinal.

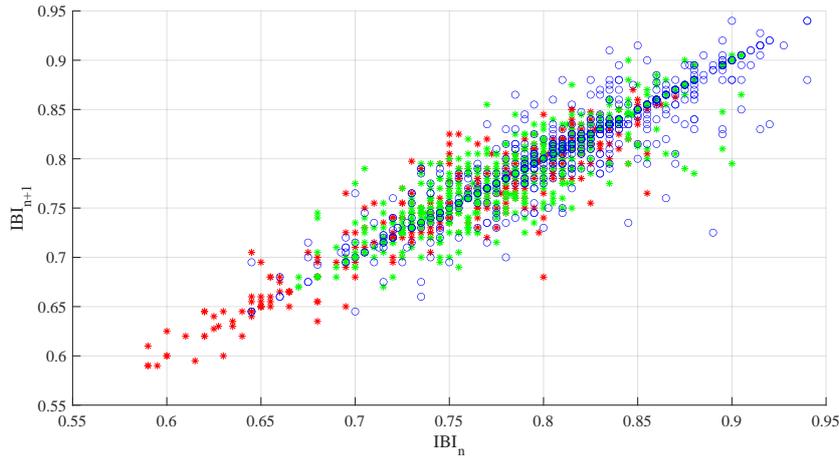


Figure 6-17: Diferentes perspectivas de Poincaré para todos los 47 voluntarias considerando los estímulos de miedo (rojo-abajo), los estímulos sin miedo (verde-medio) y las etapas de recuperación (azul-arriba). Vista 2D.

El análisis de este tipo de gráficos mediante la exploración visual es una tarea difícil. Por lo tanto, la tabla 6.3 informa de la media y la desviación estándar de los diferentes valores SD_2 y SD_1 obtenidos a partir de estos gráficos de Poincaré presentados. Nótese que en este caso las etapas de recuperación se dividen como se hizo para el análisis GSR.

Stimuli	SD_2 (ms)	SD_1 (ms)
Type	$\mu(\sigma)$	$\mu(\sigma)$
Miedo	62.78 (9.99)	14.28 (2.91)
Miedo en recuperación	72.37 (13.11)	17.19 (3.47)
No Miedo	60.96 (11.19)	14.42 (3.05)
No Miedo en recuperación	70.57 (11.84)	17.22 (3.34)

Table 6.3: Evaluación de las características del diagrama de Poincaré para los estímulos con y sin miedo, y sus respectivas etapas de recuperación. Estas métricas son la media y la desviación estándar promediadas para los 47 voluntarias.

Aunque la información del BVP no está directamente relacionada con una dimensión emocional, como lo está el GSR con el arousal, y su análisis es más complejo, los resultados obtenidos reafirman algunos de los efectos fisiológicos expuestos en la anterior exploración del GSR. Por ejemplo, al comparar las características obtenidas para los estímulos con y sin miedo, podemos observar cómo mientras el valor medio de SD_2 disminuye de 62,78 ms a 60,96 ms, el SD_1 aumenta de 14,28 ms a 14,42 ms. Para las etapas de recuperación, SD_1 aumenta en ambos tipos de estímulos, recuperación con miedo y sin miedo, lo que indica una formación de agrupación

más amplia y una activación parasimpática. Sin embargo, SD_2 también aumenta para ambas recuperaciones. Este hecho fisiológico implica que la agrupación de las fases de recuperación es más dispersa en ambas direcciones, lo que no es deseable. Idealmente, SD_2 debería tener el comportamiento opuesto para las etapas de recuperación. Así, a partir de este análisis, concluimos que el proceso de recuperación está teniendo un efecto en la parte parasimpática del ANS, pero no está disminuyendo la contribución simpática.

El proceso de recuperación fisiológica implementado y sus efectos fisiológicos se han verificado a lo largo de estos análisis. Una de las principales limitaciones de este proceso de recuperación en línea es que no está produciendo una respuesta fisiológica promedio cercana a cero, y en algunos casos no está suprimiendo la contribución simpática. Este hecho se debe principalmente a dos factores principales:

- Lograr una respuesta fisiológica plana (no activa), con predominio del parasimpático, cuando se está bajo un experimento de elicitación de emociones en realidad virtual es una tarea difícil. Teniendo en cuenta que este experimento, para la mayoría de los voluntarias, fue la primera experiencia de realidad virtual, las diferencias observadas entre la dinámica fisiológica en los estímulos sin miedo y en las etapas de recuperación son ligeramente notables para el análisis de la GSR.
- El intervalo de confianza implementado obvia la tendencia o dinámica fisiológica real dentro de cada ventana de procesamiento. Además, el gráfico de Poincaré es insensible a las tendencias de los IBI o de la frecuencia cardíaca. Por ejemplo, podría ocurrir que, dado un conjunto de diferentes ventanas de procesamiento temporal consecutivas y un intervalo de confianza codificado, la tendencia de la señal que se está evaluando sea positiva, lo que en el caso de la señal GSR significaría un incremento de la excitación.

Hasta donde yo sé, ninguna base de datos pública y abierta consideraba una monitorización de la recuperación basada en el biofeedback activo dentro de sus experimentos, lo que hace que esta parte de nuestra base de datos, así como el análisis presentado en esta sección, sean una contribución novedosa. Las limitaciones indicadas se utilizaron para seguir investigando en nuevas implementaciones de recuperación vestibles en línea. De hecho, se está desarrollando e implementando una versión

mejorada del proceso de recuperación presentado. Por ejemplo, una de las primeras etapas del nuevo seguimiento de recuperación en línea se ha implementado en [276]. En concreto, en esta primera etapa se ha utilizado la extracción de características en línea para la señal BVP y, mediante regresión lineal de mínimos cuadrados, se han analizado las diferentes tendencias de las características para asegurar la reducción de la activación simpática. Aunque este nuevo proceso de recuperación aún está en desarrollo, ya se ha probado con una pequeña muestra de voluntarias y ha demostrado superar el proceso inicial de control de la recuperación.

6.3.2 Resultados fisiológicos uni-modales

Como se ha indicado en la sección anterior, los resultados presentados en esta sección se han obtenido con las señales adquiridas por el sistema de herramientas de investigación BioSignalPlux®. Por lo tanto, estos resultados se basan en la implementación de un sistema de aprendizaje automático fuera de línea. Aunque dicho sistema no ha sido integrado, estos resultados representan los primeros resultados de detección de miedo de nuestro conjunto de datos. Hay que tener en cuenta que el diseño y la implementación del sistema de detección de miedos basado en nuestro conjunto de datos está motivado principalmente por las limitaciones encontradas en los anteriores sistemas de detección de miedos propuestos en el capítulo 4. En este caso, y con el objetivo de mejorar y profundizar en los modelos independientes del sujeto, el sistema presentado se centra en dicho enfoque. En cuanto a las etiquetas, se han utilizado tanto las discretas como las dimensionales siguiendo el mismo enfoque de miedo-binario expuesto en el capítulo 2 y aplicado en el capítulo 4. Sin embargo, debido a las incoherencias de etiquetado observadas para algunos de los voluntarias en la sección 6.2, decidimos excluir de la evaluación a los voluntarias número 5, 6, 15, 33 y 40 para el caso discreto y a los voluntarias número 3, 5, 6, 13, 20, 21, 22, 40 y 42 para el caso de uso dimensional, ya que sólo tenían alrededor del 25% de la clase positiva. Esto se tiene en cuenta aquí ya que las etiquetas utilizadas son las valoraciones autoinformadas, a diferencia de la sección anterior que utilizaba las etiquetas validadas o de destino. Se podrían realizar más investigaciones para analizar y cuantificar el efecto de los sujetos gravemente desequilibrados dentro de los sistemas de aprendizaje automático independientes del sujeto. Esto último no entra en el ámbito de esta tesis doctoral. Cabe destacar que durante el

desarrollo del sistema de detección de miedo presentado, dos tesis de licenciatura supervisadas [287,288] y una tesis de máster supervisada [289] proporcionaron apoyo en la exploración del espacio de diseño.

La arquitectura de procesamiento de datos fisiológicos implementada se muestra en la Figura 6-18. Las etapas iniciales se basan en los sistemas de prueba de concepto anteriores presentados en el capítulo 4. En primer lugar, las etapas de filtrado aplicadas siguen los mismos procesos que se detallan en la sección 6.3. Para la segmentación y la superposición de datos, se utiliza la estrategia de superposición de 20 segundos y 50%, como se hizo para el sistema MANHOB en la Sección 4.2.2. El proceso de extracción de características incluye características adicionales en comparación con los sistemas de detección de miedo presentados anteriormente. En concreto, se extraen 57 características: 31 de BVP, 20 de GSR, y 6 de SKT. Estas características se normalizan siguiendo una técnica de puntuación Z y posteriormente se introducen en la etapa de selección de características. Obsérvese que, antes del proceso de selección de características, la división tren-prueba se realiza de forma personalizada, mediante el uso de una técnica híbrida CV, LASO, como se detalla en la sección 3.1.7.3. Esta técnica tiene en cuenta tanto la variabilidad intra como inter de los voluntarios, a diferencia de LOSO y LOTO. En concreto, la partición LASO se realiza dejando fuera a la mitad de cada voluntario, es decir, las siete primeras respuestas de los estímulos audiovisuales se utilizan para el entrenamiento y las otras siete para la prueba. El conjunto de pruebas se emplea además para ejecutar una prueba ciega completa, que se utilizará para evaluar el rendimiento final del sistema. Hay que tener en cuenta que esta configuración de la partición de entrenamiento-prueba es una aproximación inicial y puede mejorarse y/o realizarse de forma diferente. Para el proceso de entrenamiento, se realiza una partición de entrenamiento y validación con un 5-kFold. Esta partición también se utiliza durante la optimización de hiperparámetros realizada a través de SMBO, como se detalla en la sección 4.2.4. Esta arquitectura se aplica, valida y prueba en base a los mismos tres clasificadores que para el sistema de detección de miedo presentado en la Sección 4.2.4: SVM, KNN, y ENS.

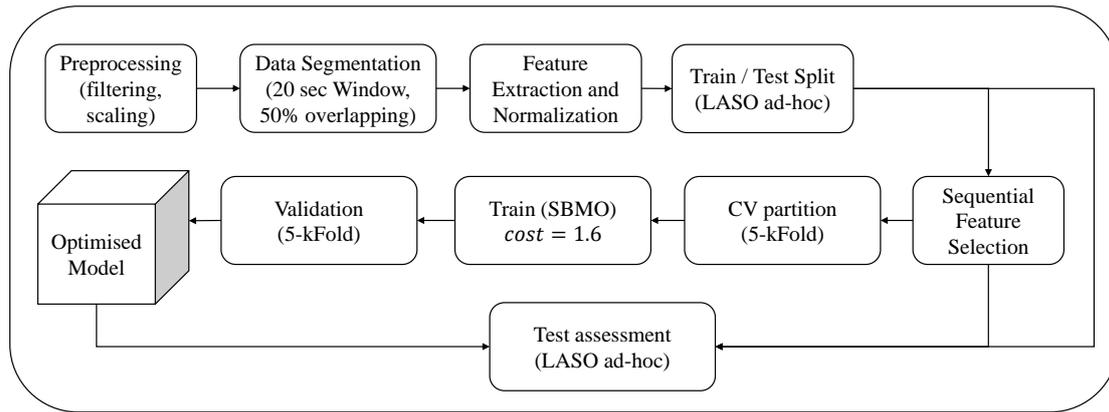


Figure 6-18: Arquitectura de procesamiento de datos fisiológicos para entrenar y probar los modelos de aprendizaje automático generados utilizando nuestro propio conjunto de datos.

6.3.2.1 Extracción de características

Dentro del procedimiento de extracción de características, también se realizan los diferentes procesos de delineación. En este caso, la señal BVP se somete a un apilamiento o a un cálculo aproximado mediante tres algoritmos diferentes. El primer algoritmo es el aplicado en la sección 4.1.3. El segundo algoritmo viene dado por [290] y se basa en un método de umbral adaptativo para PPG la detección de picos. El tercer algoritmo viene dado por [291] y se basa en una media móvil de diferencias valle-pico junto con filtros de umbral local para identificar los picos sistólicos de la señal PPG. Estos tres algoritmos pueden agruparse dentro de los métodos basados en la pendiente máxima y mínima local, que son los algoritmos de delineación BVP menos exigentes desde el punto de vista computacional. Obsérvese que obtuvieron una precisión de detección de picos superior al 90,00% para su respectiva validación. A diferencia de la utilización de un solo algoritmo de delineación como se hizo para los anteriores sistemas de detección de miedo presentado, este triple enfoque se considera en este caso para hacer frente a lo más posible con cualquier PPG deformación morfológica. Nótese que en nuestro conjunto de datos, los estímulos audiovisuales están basados en la RV, lo que introduce más movimientos corporales (artefactos de movimiento) que el resto de las bases de datos públicas abiertas basadas únicamente en estímulos 2D. La delimitación de la señal GSR se ha realizado como en la sección 4.1.3. Así, asumimos una combinación lineal de ambos componentes GSR, SCL y SCR, seguida de la ecuación 2.7.

En lo que respecta específicamente al número de nuevas características añadidas,

las tablas 6.4 y 6.5 detallan cada una de las características consideradas en este caso. Nótese que las características extraídas para la señal SKT no cambiaron y son las mismas que se especifican en la Tabla 4.14. Para la señal BVP, se incluyeron seis nuevas características. En el dominio del tiempo, se consideró el valor medio de la señal y la raíz cuadrada de la IBI (HRV-RMSSD). El primero se utiliza para tener en cuenta cualquier residuo de respiración que haya quedado lo suficientemente fuerte después de haber filtrado la señal. La segunda es una métrica adicional derivada de los IBI extraídos y proporciona otro indicador del control vagal cardíaco, es decir, cuanto más alta es la métrica, mayor es la activación parasimpática. En el dominio de la frecuencia, se reordenaron las distintas bandas de frecuencia y se obtuvo más información sobre la contribución energética de las distintas bandas, relativa y normalizada. Obsérvese que dicha reordenación de la definición de las bandas se ha asumido debido a las recientes publicaciones [179]. Además, se implementó un cambio importante para la estimación de la PSD teniendo en cuenta los IBIs adquiridos de forma desigual. Así, en lugar de interpolar y utilizar el estimador de promedio de segmentos superpuestos de Welch, se ha empleado el periodograma de Lomb-Scargle para estimar las PSDs para las bandas espectrales especificadas [292]. Esta técnica nos permite relajar las consideraciones de resolución de frecuencia. Una vez obtenida la contribución de la PSD para cada banda, se calcula su densidad espectral de energía, que posteriormente se normaliza para obtener la relación de energía entre las bandas LF y HF. Por último, las características no lineales se han ampliado con hasta siete métricas derivadas de Poincaré-plot. Éstas se basaron principalmente en [285, 286] y su cálculo utiliza SD_2 y SD_1 , que se detallan en la Sección anterior y vienen dadas por la ecuación 6.5. Así, la longitud longitudinal y transversal de la parcela se calcula como

$$\begin{aligned}L_{SD_2} &= 4 * SD_2, \\T_{SD_1} &= 4 * SD_1.\end{aligned}\tag{6.6}$$

Obsérvese que estas dos características basadas en la recurrencia están directamente relacionadas con las desviaciones estándar SD_2 y SD_1 . Por lo tanto, siguen la misma lógica fisiológica, pero de una manera mejorada debido al factor de multiplicación. Además, también se calcularon el índice simpático cardíaco (CSI), el CSI modificado

(MCSI) y el índice vago cardíaco (CVI). Se calculan de la siguiente manera

$$\begin{aligned}CSI &= L_{SD_2}/T_{SD_1} = SD_2/SD_1, \\MCSI &= L_{SD_2}^2/T_{SD_1}, \\CVI &= \log_{10}(L_{SD_2} * T_{SD_1}).\end{aligned}\tag{6.7}$$

Se puede observar que estas características están fuertemente relacionadas con la activación simpática y parasimpática dada una trama de Poincaré. Por ejemplo, el MCSI mejora la longitud longitudinal para enfatizar la respuesta simpática, lo que puede llevar a distinguir las activaciones simpáticas más débiles. Uno de los principales cambios dentro de las características no lineales del BVP es que el MSE no se considera en este sistema. Este hecho se debe a las limitaciones observadas al extraer esta característica para ventanas temporales cortas [293, 294]. Para el GSR, se incluyeron dos características más. Se trata del tiempo medio de recuperación relativa y del área media bajo los picos detectados. La primera ya se ha utilizado en la sección anterior. El área bajo los picos se calcula mediante una aproximación trapezoidal. Obsérvese que esta última puede mejorarse utilizando la regla de Simpson a costa de aumentar el tiempo de cálculo.

6.3.2.2 Selección de características

Para reducir la dimensionalidad del problema, se emplea SFS se emplea. Esta técnica permite seleccionar las características más relevantes y, por tanto, reducir la complejidad del tiempo de entrenamiento e inferencia y los requisitos de almacenamiento. Así, ejecutamos el SFS para cada uno de los tres clasificadores considerando cada conjunto de entrenamiento generado de voluntarias. En el caso del SVM, se utiliza un kernel RBF con $\gamma = 1$ y $C = 1$. El KNN se ajusta a la distancia euclidiana con 10- K vecinos más cercanos. Por último, el ENS utiliza un algoritmo AdaBoost con árboles de decisión potenciados como aprendices débiles y un número máximo de divisiones de hasta 10. La función de coste para cada iteración del SFS viene dada por $1 - MCC$. Obsérvese que este proceso de selección de características se realiza después de la división entrenamiento/prueba para evitar la fuga de información del conjunto de prueba.

En los siguientes puntos, proporcionamos la lista de características que se selec-

Table 6.4: Características extraídas para la señal BVP y la propuesta de reconocimiento de emociones binarias de miedo utilizando nuestro conjunto de datos.

Sensores	Dominio	Características
BVP (31)	Dominio Temporal: (4)	Filtered data mean value Mean of IBI HRV-SDNN HRV-RMSSD
	Dominio Frecuencial: (12)	Normalised IBI PSD contribution (summation) for: Low frequency (LF) (0.01–0.15 Hz) High frequency (HF) (0.15–0.40 Hz) Ultra-High frequency (UHF) (0.40–1.00 Hz) Energy contribution of those IBI PSD bands Relative energy of those IBI PSD bands Normalised energy ratio between LF and HF Normalised energy ratio for LF and HF
	No Lineal: (15)	From Poincaré-plot: SD2, SD1, LSD2, TSD1, CSI, MCSI, CVI Detrended fluctuation for the filtered signal Recurrence rate Determinism Laminarity Longest RP diagonal line Diagonal lines entropy Trapping time Correlation dimension

cionaron al menos una vez para todos los modelos generados. Cuando se utilizan las etiquetas discretas binarizadas, se obtienen las siguientes mejores características:

- Para el sistema basado en SVM (15 características seleccionadas en total):
 - BVP (9): valor medio de los datos filtrados, HRV-RMSSD, contribución energética de HF y UHF, relación LF/HF, SD_1 , T_{SD_1} , análisis de fluctuación de detrimento para la señal filtrada, laminaridad y entropía de las líneas diagonales.
 - GSR (5): valor medio de los datos filtrados, y su desviación estándar, área bajo el ERSCRs detectado, distribución del primer y tercer cuartil.
 - SKT (1): valor medio filtrado.
- Para el sistema basado en KNN (11 características seleccionadas en total):
 - BVP (2): valor medio de los datos filtrados, y laminaridad.
 - GSR (8): valor medio de los datos filtrados, número medio de picos ERSCR, amplitud relativa media, tiempo de subida y tiempo de recuperación de los picos ERSCR, área bajo los ERSCR, distribución del primer y tercer cuartil.
 - SKT (1): valor medio filtrado.

Table 6.5: Características extraídas para la señal GSR y la propuesta de reconocimiento binario de emociones de miedo utilizando nuestro conjunto de datos.

Sensores	Dominio	Características
GSR (20)	Dominio temporal: (9)	Filtered data mean value ERSCR including number of peaks ERSCR amplitude and rise time ERSCR recovery time and area under the peak Filtered data Standard deviation First quartile Third quartile
	Dominio frecuencial: (3)	Power spectral density of two bands for SCL and SCR components (0–0.05 Hz, 0.05–1.5 Hz) Spectral density ratio for 0–0.05 Hz
	No lineal: (8)	Detrended fluctuation for filtered data Recurrence rate Determinism Laminarity Longest RP diagonal line Diagonal lines entropy Trapping time Correlation dimension

- Para el sistema basado en ENS (13 características seleccionadas en total):
 - BVP (3): valor medio de los datos filtrados, entropía de las líneas diagonales y tiempo de captura.
 - GSR (8): valor medio de los datos filtrados, número medio de picos ERSCR, amplitud relativa media y tiempo de subida de los picos ERSCR, área bajo los ERSCR, distribución del primer y tercer cuartil, y laminaridad.
 - SKT (2): valor medio filtrado, y densidad espectral de potencia de la banda más baja (0–0,1 Hz).

Cuando se utilizan las etiquetas dimensionales binarizadas, se obtienen las siguientes mejores características:

- Para el sistema basado en SVM (11 características seleccionadas en total):
 - BVP (4): valor medio de los datos filtrados, media del IBI, tasa de recurrencia y entropía de las líneas diagonales.
 - GSR (5): valor medio de los datos filtrados y su desviación estándar, distribución del primer y tercer cuartil y tiempo de captura.
 - SKT (2): valor medio filtrado y densidad espectral de potencia de la banda más baja (0–0,1 Hz).

- Para el sistema basado en KNN (8 características seleccionadas en total):
 - BVP (2): valor medio de los datos filtrados, y laminaridad.
 - GSR (5): valor medio de los datos filtrados, número medio y amplitud relativa media de los picos ERSCR, distribución del primer y tercer cuartil.
 - SKT (1): valor medio filtrado.
- Para el sistema basado en ENS (8 características seleccionadas en total):
 - BVP (3): valor medio de los datos filtrados, entropía de las líneas diagonales y tiempo de captura.
 - GSR (4): valor medio de los datos filtrados, número medio de picos ERSCR, amplitud relativa media de los picos ERSCR y distribución del primer cuartil.
 - SKT (1): valor medio filtrado.

En general, la cantidad de características seleccionadas por cada modelo oscila entre 15 y 20. Así, la complejidad del problema se reduce a una cantidad relativamente baja de características para los diferentes clasificadores y para ambas metodologías de etiquetado. Nótese que este hecho afecta drásticamente a las diferentes etapas de la arquitectura fisiológica, como el entrenamiento y la prueba (inferencia). Además, en lo que respecta al número y la naturaleza específicos de las características más importantes enumeradas para cada caso de uso del etiquetado, aproximadamente el 50% son temporales y morfológicas, el 20% se basan en la frecuencia y el 30% son no lineales. Aunque cada clasificador no seleccionó exactamente las mismas características, dicha selección determina una primera aproximación para obtener las que proporcionan la información más valiosa. Para el caso del etiquetado discreto binarizado, cabe destacar que los tres clasificadores coincidieron en considerar el valor medio de los datos filtrados para los tres sensores, características relacionadas con los picos de ERSCR, y algunas características no lineales directamente relacionadas con la caracterización no periódica del sistema (laminaridad, tiempo de atrapamiento, entropía de las líneas diagonales). Para el caso del etiquetado dimensional binarizado, se repite el mismo comportamiento respecto a la consideración del valor medio de los datos filtrados, que también se acompaña de rasgos relacionados con los picos de ERSCR por parte de dos de los tres clasificadores y de los mismos

rasgos no lineales. Sin embargo, el SFS aplicado para el caso de uso de la dimensión binarizada considera menos características acordadas entre los modelos. Este hecho puede ser una consecuencia del mayor desacuerdo observado en dicho etiquetado.

6.3.2.3 Resultados de la validación y las pruebas

Además, tras aplicar el paso de selección de características, decidimos emplear un enfoque de aprendizaje sensible a los costes para tratar la situación de desequilibrio en el etiquetado. Para ello, se ajustó un parámetro de coste de clasificación errónea, tal y como se hizo con los sistemas de detección de miedo propuestos anteriormente en la sección 4.1.4. Para este caso concreto, se aplicó un coste de clasificación errónea de 1,6 sobre la clase positiva (miedo), que se fijó mediante un barrido experimental de parámetros tras la etapa de selección de características. Tenga en cuenta que esta consideración de diseño hace que el sistema sea menos propenso a omitir una situación peligrosa para el caso de uso que se aborda [183], es decir, aumenta la sensibilidad.

La salida del sistema de aprendizaje automático fisiológico es una etiqueta binaria cada 10segundo, como se indica en la sección 4.2.2. Por lo tanto, para esta primera aproximación, asumimos que la verdad básica de un estímulo específico es la etiqueta binarizada autodeclarada que se le asigna, independientemente de la cantidad total de instancias generadas, es decir, todas las instancias generadas dentro del mismo estímulo tienen la misma etiqueta. Por ejemplo, hay estímulos audiovisuales dentro de la misma clase que generan más instancias que otros. Esta aproximación puede ser crítica para estímulos de corta duración, como el estímulo número ocho de la primera tanda, cuya duración es de 23segundo. Dicha duración implica una instancia generada, que puede dañar seriamente el equilibrio del sistema o incluso ser insuficiente para caracterizar adecuadamente la emoción objetivo de ese estímulo. Esta limitación se aborda considerando el número completo de instancias para las clases de miedo y no miedo sin depender de la cantidad de información proporcionada por cada estímulo. La tabla 6.6 informa del número total de instancias basadas en las etiquetas discretas binarizadas y las autodeclaradas dimensionalmente, considerando una ventana de procesamiento de 20 segundos y un solapamiento del 50%. Nótese que estos valores se obtienen para todos los voluntarios independientemente de las etiquetas desequilibradas. Se puede observar que el equilibrio es cercano al

Class	Discrete Labels	Dimensional Labels
Fear	1496	1335
Non-Fear	2107	1942
Balance (Fear/Non-Fear)	42/58%	40/60%

Table 6.6: Número total de instancias para nuestro conjunto de datos basado en etiquetas discretas binarizadas y dimensionales autodeclaradas.

Table 6.7: Resultados de validación y prueba de los diferentes sistemas de aprendizaje automático fisiológico utilizando la primera versión de WEMAC. Se muestran los resultados de ambos enfoques binarizados discretos (Disc) y dimensionales (Dim).

Classifier	Partition Type	SEN (MAD)	SPE (MAD)	Gmean (MAD)	ACC (MAD)	AUC (MAD)	F1 (MAD)
SVM	Val-Disc	83.02(1.19)	78.79(0.99)	80.87(0.79)	80.72(0.79)	86.72(0.79)	78.16(0.16)
	Test-Disc	64.36(17.39)	67.88(13.77)	65.29(12.48)	65.70(11.44)	65.33(16.08)	62.84(13.79)
	Val-Dim	86.45(1.05)	75.36(2.91)	80.63(1.54)	82.26(1.07)	87.33(1.00)	82.23(0.99)
	Test-Dim	72.78(13.21)	53.86(11.24)	61.60(9.23)	62.51(9.03)	62.14(11.24)	65.01(9.64)
KNN	Val-Disc	81.15(4.52)	75.55(4.92)	78.28(4.71)	73.18(5.85)	86.84(4.26)	75.32(5.04)
	Test-Disc	65.53(14.63)	69.00(13.49)	66.08(10.67)	66.87(9.64)	66.45(14.30)	64.72(10.17)
	Val-Dim	84.27(4.49)	84.65(3.92)	84.46(4.17)	84.45(4.20)	92.34(3.44)	84.60(4.19)
	Test-Dim	61.08(14.32)	65.01(14.45)	61.37(7.68)	61.78(6.84)	61.43(10.46)	60.00(9.76)
ENS	Val-Disc	81.82(4.31)	75.40(5.29)	78.53(4.82)	68.17(3.70)	75.52(4.01)	64.19(3.74)
	Test-Disc	68.55(12.10)	61.61(16.81)	63.51(10.48)	64.50(9.54)	64.57(14.25)	65.11(8.31)
	Val-Dim	94.02(0.48)	93.18(0.53)	93.60(0.44)	93.71(0.44)	98.40(0.22)	93.72(0.44)
	Test-Dim	65.91(15.71)	64.98(14.69)	63.75(9.28)	64.23(8.15)	66.62(12.20)	63.37(10.63)

reportado en la Sección 6.2.

Los resultados de validación y prueba de los diferentes clasificadores se detallan en la Tabla 6.7. Obsérvese que se muestran los valores de desviación media y absoluta para los diferentes clasificadores, particiones y enfoques de etiquetado. Estos resultados proceden de los 42 y 38 modelos considerados para el modelo discreto y dimensional binarizado, respectivamente. En general, los resultados obtenidos están en línea con los obtenidos en la sección 4.2.4.2, en la que el SVM mostró el peor rendimiento, seguido del **KN!** (**KN!**), y siendo el ENS el mejor. De hecho, cuando se analizan conjuntamente las métricas promediadas proporcionadas y sus valores de dispersión tanto para el etiquetado discreto como dimensional, el clasificador Adaboost (ENS) es el que supera a los otros dos. Una de las principales diferencias entre ellos es que, mientras que el SVM y el KNN pierden especificidad para el caso de uso dimensional, el ENS mantiene el equilibrio entre sensibilidad y especificidad dando lugar a una Gmean muy similar para ambos casos.

Por un lado, específicamente para el etiquetado discreto, los mejores resultados promediados los obtiene el clasificador KNN tanto para las particiones de validación

como de prueba. Además, la mayor especificidad, Gmean, precisión y AUC se consiguen en este caso para este clasificador con hasta un 69,00%, 66,08%, 66,87% y 66,45%, respectivamente. Por otro lado, cuando se trata del etiquetado dimensional, los clasificadores no siguen exactamente el mismo comportamiento que para el discreto. De hecho, el mejor clasificador en este caso es el ENS alcanzando la mayor especificidad, Gmean, precisión y AUC con hasta 64,98%, 63,75%, 64,23% y 66,62%, respectivamente. Obsérvese que estos resultados se obtienen utilizando un conjunto reducido de características, como se detalla en la sección anterior. Este hecho, acompañado de los retos que supone tratar con un enfoque independiente del sujeto, confiere un gran valor y un gran potencial a estos primeros resultados iniciales del WEMAC.

Para contextualizar el comportamiento de los clasificadores para los diferentes modelos considerados, las figuras 6-19 y 6-20 muestran la métrica de rendimiento de MCC sobre sus particiones de prueba. Obsérvese que esta métrica utiliza toda la información de la matriz de confusión y proporciona un valor similar al de la correlación considerando las matrices de confusión reales y predichas. Para ambos casos de uso del etiquetado, observamos que los diferentes modelos siguen un comportamiento similar para algunos de los voluntarios considerados. Por ejemplo, la correlación de las métricas MCC es de hasta 0,48(0,17) y 0,44(0,13) para el etiquetado discreto y dimensional, respectivamente. Este hecho es destacable ya que es un indicador que demuestra la independencia del comportamiento específico de cada clasificador con respecto al conjunto de datos utilizado. En consonancia con los resultados reportados en la Tabla 6.7, el enfoque de etiquetado discreto supera al dimensional. Sin embargo, estos gráficos también muestran unos resultados muy dependientes del sujeto en algunos casos. Esto se refleja en una variabilidad muy alta, que también se refleja en la MAD de las diferentes métricas de rendimiento. En concreto, en estos gráficos, las métricas de MCC se sitúan aproximadamente entre -0,4 y 0,9 y entre -0,4 y 0,6 para los casos discretos y dimensionales, respectivamente. Dicha variabilidad y distribución se representa en la Figura 6-21, que muestra la distribución agregada para los diferentes clasificadores y casos de uso de etiquetado. Obsérvese que el valor mediano es la línea roja u horizontal dentro de los recuadros. Como ya se ha dicho anteriormente, el entrenamiento utilizando el etiquetado discreto bina-

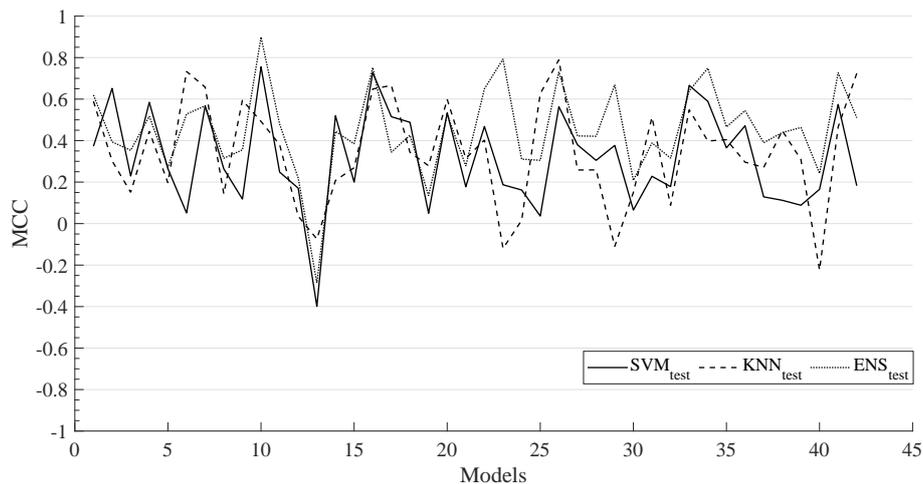


Figure 6-19: Evaluación de la métrica de prueba MCC para todos los 42 modelos considerados dentro del caso de uso de detección de miedo discreto binarizado.

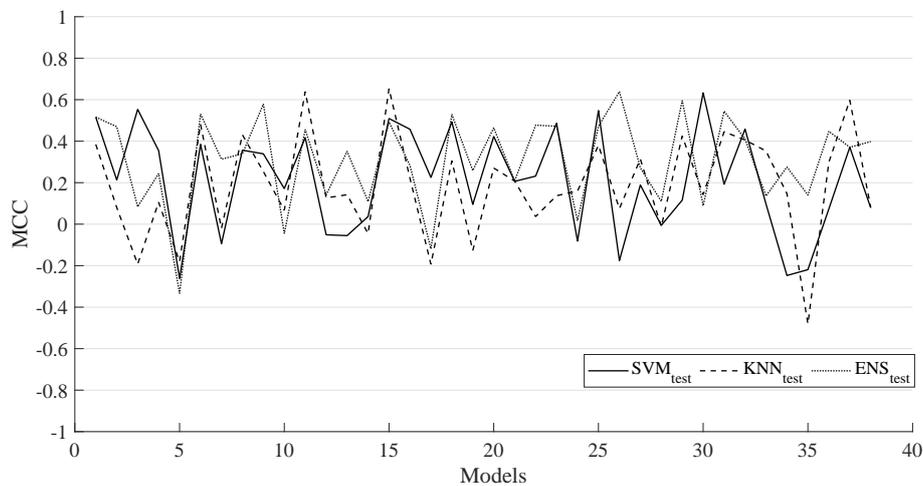


Figure 6-20: Evaluación de la métrica de prueba MCC para todos los 38 modelos considerados dentro del caso de uso de detección de miedo dimensional binarizado.

rizado consigue mejores modelos (medianas más altas) que el etiquetado dimensional binarizado. Además, la dispersión del rango intercuartil es siempre menor cuando se aplica el clasificador ENS.

Hasta donde yo sé, los sistemas generados son los primeros sistemas de detección del miedo que utilizan un conjunto reducido de señales fisiológicas y estímulos de realidad virtual. Por ejemplo, estos resultados constituyen la línea de base fisiológica para el conjunto de datos WEMAC. Se ha demostrado que, en general, los sistemas entrenados con el etiquetado discreto binarizado obtienen mejores modelos que los entrenados con el etiquetado dimensional binarizado. Este hecho está en consonancia con las conclusiones de la exploración de la respuesta del etiquetado propio expuestas en la sección 6.2.

La principal limitación de los modelos de aprendizaje automático generados es la

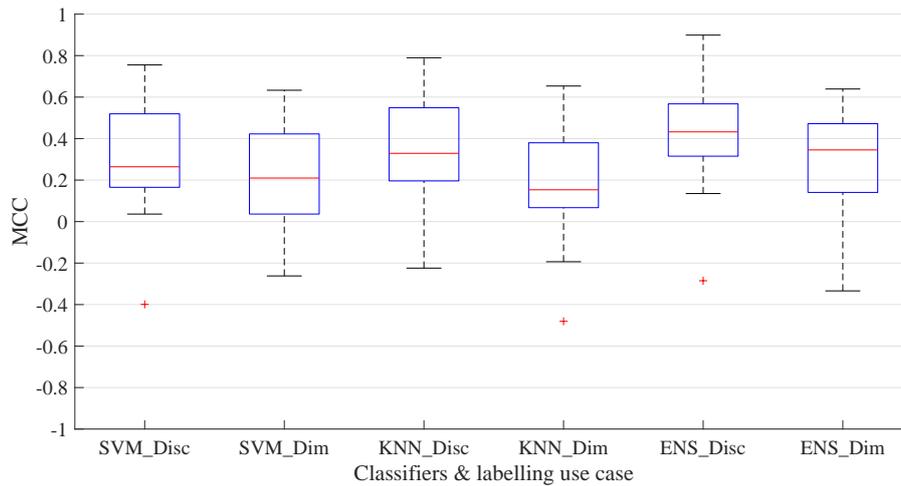


Figure 6-21: Distribución del diagrama de caja de la métrica de prueba MCC para todos los modelos de 42 y 38 considerados dentro de los casos de uso de detección de miedo discreto y dimensional binarizado.

alta variabilidad y dispersión observada. Este hecho podría deberse a un problema de sobreajuste, que podría estar provocando la generación de modelos de bajo sesgo y alta varianza. Sin embargo, las métricas de validación obtenidas no lo sugieren. De hecho, el clasificador que proporciona las mejores métricas de rendimiento de validación, es decir, AdaBoost con más del 90,00%, es el menos susceptible a este tipo de problemas, ver Sección 3.1.7. Por lo tanto, es necesario estudiar y analizar diferentes aspectos para dar una explicación adecuada a este problema:

- La técnica acLASO aplicada deja una pequeña cantidad de muestras en la partición de prueba. Aunque esta técnica está pensada para tratar la personalización de los sujetos, el pequeño tamaño del conjunto de datos de prueba puede dar lugar a un conjunto de datos poco representativo. Por lo tanto, se podrían aprovechar otras técnicas de partición para considerar un conjunto de datos de prueba más grande y/o incluir más inter e intravariabilidad. Este hecho también se ve afectado por la cantidad limitada de datos disponibles para cada voluntario cuando se trata de bases de datos de laboratorio.
- En vista de esta variabilidad entre los distintos voluntarios, algunos procesos de la arquitectura de entrenamiento propuesta pueden personalizarse en función del conjunto de datos específico, es decir, en función de la combinación concreta de voluntarios o de la distribución de clases. Por ejemplo, el enfoque sensible a los costes puede ajustarse para cada conjunto de entrenamiento diferente.
- Las técnicas de partición para la generación de los conjuntos de datos de

entrenamiento-validación podrían cambiarse por la misma técnica CV que se aplica para la primera partición (LASO).

- Se puede realizar un análisis fisiológico adicional para encontrar clusters y diferencias extremas entre los distintos voluntarios. Como primera aproximación, dichos clusters pueden basarse en filtros fisiológicos sencillos como: nivel de activación GSR (hipoactividad vs hiperactividad), rangos SKT, y nivel de ruido residual de la señal PPG tras el filtrado.
- También se pueden emplear diferentes tipos de técnicas de normalización y escalado para evaluar su efecto. Este hecho está directamente relacionado con la necesidad de descubrir la mejor forma de modelar el problema, es decir, la detección del miedo independiente del sujeto.

Entre estas consideraciones expuestas, desde mi punto de vista, la que más está afectando en la actualidad es el riesgo de falta de representatividad del conjunto de datos de prueba. En concreto, en este caso, el LASO aplicado dejó aproximadamente hasta un 1,3% de los datos totales para el conjunto de prueba. Esto supone una media de 48 muestras sobre un total de 3600 instancias disponibles. Con el fin de proporcionar un punto de partida en relación con esta discusión específica, la Tabla 6.8 informa de los resultados para la misma arquitectura fisiológica cuando se entrena el sistema KNN con el etiquetado discreto binarizado, pero utilizando una partición LOSO para la división de entrenamiento-prueba. Podemos observar que en el caso LOSO se obtienen resultados medios menores, que pueden verse afectados por la intravariabilidad del voluntario no visto que no se tiene en cuenta para el entrenamiento. Sin embargo, el hecho más importante se observa en la diferencia de dispersión. El sistema LOSO informa de una menor variabilidad, lo que puede ser una indicación de un conjunto de pruebas más representativo. Hay que tener en cuenta que es necesario realizar más estudios y análisis para caracterizar adecuadamente este hecho.

Table 6.8: Resultados de validación y prueba para los sistemas de aprendizaje automático KNN utilizando el etiquetado discreto binarizado y una técnica LOSO CV para la partición entrenamiento-prueba.

Partition Type	SEN (MAD)	SPE (MAD)	Gmean (MAD)	ACC (MAD)	AUC (MAD)	F1 (MAD)
Test-Disc-LASO	65.53(14.63)	69.00(13.49)	66.08(10.67)	66.87(9.64)	66.45(14.30)	64.72(10.17)
Test-Disc-LOSO	64.05(10.69)	60.93(9.26)	61.74(6.98)	61.90(6.93)	62.25(9.54)	58.48(7.18)

Dentro de esta casuística de variabilidad y además de los análisis recomendados para la mejora de los modelos, hay que considerar también que la latencia emocional y la dinámica fisiológica de cada voluntario para cada estímulo está afectando a la separación de clases binarias. Por ejemplo, la figura 6-22 muestra la relación LF/HF extraída de la señal IBI del voluntario número tres. En concreto, cada barra representa la característica extraída para una ventana de 20 segundos con un solapamiento de 18 segundos. Obsérvese que este solapamiento se aplica en este caso para reducir la resolución temporal y mejorar la visualización dinámica del rasgo a lo largo de cada estímulo. Podemos observar cómo varía la evolución de esta característica específica dentro de cada estímulo. A partir de esta información, podemos realizar un análisis como el proporcionado en la sección 6.3.1.1, en la que se estudiaron los patrones GSR. Sin embargo, el problema a destacar en este caso es que, independientemente de la evolución dinámica de las características dentro de los estímulos, se está asignando la misma etiqueta a todas las instancias generadas. Por lo tanto, habría que explotar diferentes técnicas de aprendizaje automático para considerar la evolución temporal de las características o la contextualización junto con un enfoque de etiquetado diferente. Esto último se refiere a la posibilidad de aplicar el aprendizaje automático semisupervisado para tratar las etiquetas duras actuales como etiquetas blandas. Esto puede incluso pensarse como modelos en los que las etiquetas son parámetros aprendibles. Además, la combinación de las dos metodologías de etiquetado actuales, la discreta y la dimensional, debería aprovecharse para sacar partido de cada una de ellas.

6.4 Marco de fusión de datos multimodales

Después de haber presentado y explicado el diseño del sistema de aprendizaje automático unimodal fisiológico, el otro enfoque principal a lo largo de este capítulo es la capacidad de fusión de datos multimodales que puede ofrecer la arquitectura del sistema Bindi². De hecho, los sistemas de reconocimiento de emociones basados en información multimodal están superando a los unimodales en la comunidad de la computación afectiva [155, 174]. La mayoría de los sistemas multimodales presentados en la literatura solían basarse en datos de audio y visuales [295], habla y

²la investigación presentada en esta sección se basa en un trabajo multidisciplinar con los miembros de la UC3M4Seguridad expertos en Teoría de la Señal y Comunicaciones

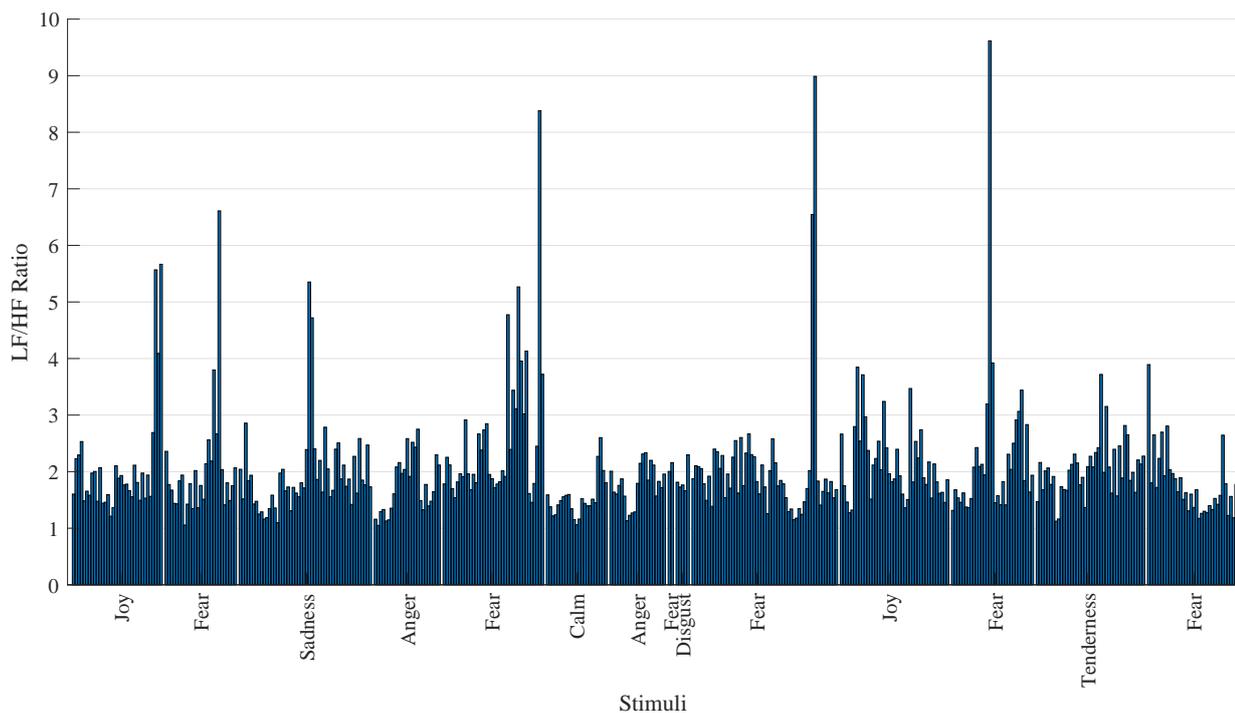


Figure 6-22: Relación LF/HF extraída del voluntario 3 del conjunto de datos WEMAC. Obsérvese que en las abscisas están representadas las emociones dirigidas para el primer lote.

gestos faciales [296], EEG y expresiones faciales [297]. Existen conjuntos de datos que recopilan información multimodal, que consideran la información fisiológica y del habla [298,299]. Se presta poca atención a la exploración del diseño multimodal utilizando información fisiológica y del habla. Sólo se ha encontrado un trabajo en la literatura que utiliza estos dos tipos de información para observaciones a corto plazo [300]. Emplearon una fusión híbrida por medio de la fusión de datos a nivel de características y de decisión, que arrojó hasta un 55,00% de precisión para un enfoque independiente del sujeto y una clasificación binaria de arousal-valencia. Por lo tanto, el marco y las metodologías multimodales propuestos sirven como un enfoque inicial para trabajar con el miedo real provocado en las mujeres y su procesamiento adecuado teniendo en cuenta tanto la información fisiológica como la del habla.

Antes de entrar en detalles sobre los diferentes sistemas diseñados y los resultados obtenidos a partir de la información recogida durante el desarrollo del conjunto de datos WEMAC, se podría explicar y detallar adecuadamente una contextualización sobre la casuística y las capacidades multimodales dentro de Bindi. Sobre esta base, se propusieron diferentes disposiciones de los componentes del sistema para explorar las posibilidades de dicho espacio de diseño multimodal. Esto proporcionó

el esbozo de una exploración del espacio de diseño para diferentes arquitecturas del sistema (fusión de datos de información fisiológica y de habla/audio procedentes del brazalete y del colgante, respectivamente). La figura 6-23 muestra dicho esquema y representa la relación potencial que se puede encontrar para los diferentes casos de uso presentados. Los casos de uso propuestos se detallan a continuación:

- Caso 1: Uni-modal. Esta disposición es la línea de base de la capacidad de detección del miedo para cada uno de los sistemas uni-modales, el fisiológico y los modelos de voz.
- Caso 2: Multimodal con prealarma. En este caso, la información fisiológica se evalúa continuamente a partir del sistema fisiológico unimodal. Cuando detecta que el usuario está experimentando miedo, activa una prealarma a la Bindi APP. Hay que tener en cuenta que esto se hace siguiendo un enfoque de computación en el borde, ya que es la propia pulsera la que ejecuta un motor ligero de aprendizaje automático. La detección del miedo hace que el Bindi APP comience a grabar audio durante un breve periodo de tiempo, lo que da lugar a una estrategia de bajo consumo de energía para el micrófono. La señal de audio se envía entonces a la capa for del sistema, es decir, a la Bindi APP, para que realice la detección del miedo utilizando también un motor de inteligencia unimodal basado en el habla.
- Caso 3: Multimodal con muestreo periódico de audio. Este caso sólo se diferencia del anterior en que no hay prealarma, sino que el habla/audio se muestrea de forma periódica.
- Caso 4: Multimodal con prealarma y muestreo periódico de audio. Esta configuración se basa en la conjunción de los dos casos de uso anteriores. Por lo tanto, representa una etapa intermedia entre tener información multimodal continua y los casos anteriores.
- Caso 5: Multimodal continuo. Esta es la última disposición propuesta del sistema y es la que requiere la mayor cantidad de recursos ya que los dos sistemas unimodales, fisiológico y habla/audio, están siempre activos, realizándose la fusión de datos de forma continua.

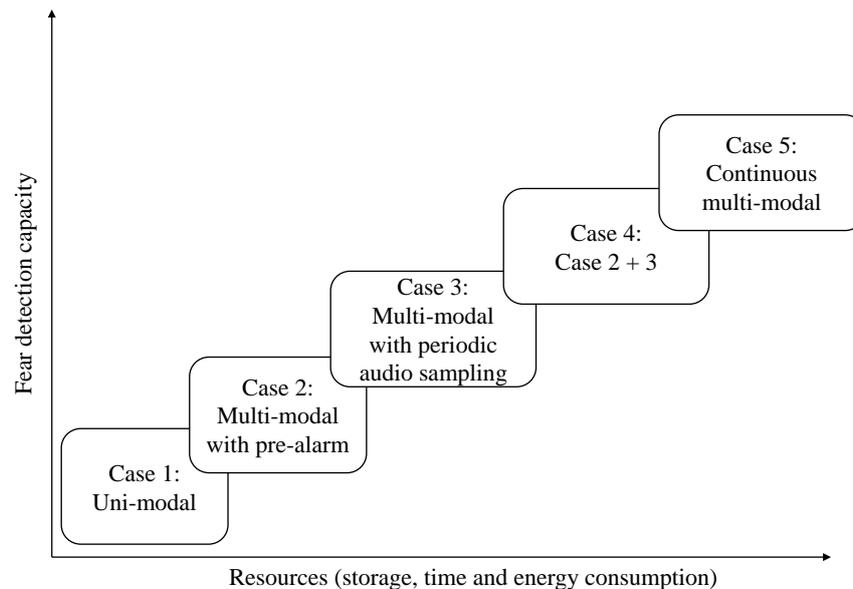


Figure 6-23: Esbozo de exploración del espacio de diseño para los diferentes arreglos de la modalidad a realizar con la arquitectura de Bindi.

Cabe señalar que estos casos de uso detallados definen la arquitectura del sistema en términos de disponibilidad de la información, en lugar de especificar la metodología de fusión de datos aplicada. Esta última se refiere a las técnicas utilizadas para realizar la fusión de datos fisiológicos y de audio. En realidad, pueden realizarse de diferentes maneras dentro del mismo caso de uso. Hay que tener en cuenta que las diferentes técnicas típicas de fusión de datos se describen y explican en el capítulo 3. De hecho, este trabajo de investigación, además de explorar el sistema fisiológico de detección de miedo unimodal, se centra en el análisis y la comparación de tres técnicas de fusión de datos propuestas mediante el uso de diferentes disposiciones de arquitectura del sistema, principalmente relacionadas con los casos 2 y 5. El resto de los casos de uso no están dentro del alcance de esta investigación. Por lo tanto, el análisis de otras disposiciones, así como el resto de los casos de uso, se dejan para los conjuntos de datos posteriores que se publicarán en la base de datos UC3M4Safety.

Dentro de este marco multimodal, he realizado la integración del sistema fisiológico detallado en el apartado anterior. En concreto, se ha empleado el clasificador KNN. El sistema unimodal de voz ha sido diseñado e implementado por los componentes del equipo de la UC3M4Safety con experiencia en el procesamiento de señales de audio. Este sistema incluye los siguientes módulos fundamentales Detección de Actividad Vocal (VAD), Sustracción Espectral (SS), extracción de características y un clasificador basado en redes neuronales [11]. Obsérvese que el caso de uso

del etiquetado discreto binarizado y el LASO CV se aplicaron para ambos sistemas uni-modales, y el sistema del habla también excluyó a los voluntarios altamente desequilibrados especificados en la sección anterior. Por último, los subsistemas unimodales fisiológico y del habla proporcionan una etiqueta binaria cada 10 y 1 segundos, respectivamente.

En cuanto a la exploración del espacio de diseño multimodal, el caso 2 fue el primero que se implementó; se explica en el capítulo 5 y se implementó en la primera versión de Bindi o [219], que se basa en una estrategia de fusión de datos jerárquica. En esta versión, la información fisiológica es recogida continuamente por el brazalete, que ejecuta un motor de inteligencia de detección de miedo fisiológico ligero y unimodal. Cuando detecta que el usuario está experimentando dicha emoción, lanza una prealarma a la Bindi APP. Esta acción hace que el Bindi APP comience a grabar audio durante un breve periodo, lo que supone una estrategia de bajo consumo de energía para el micrófono. A continuación, la señal de audio se envía a la Bindi APP para que realice la detección del miedo mediante un motor de inteligencia unimodal basado en el habla. Por último, si este último sistema confirma la detección, el Bindi APP inicia un procedimiento de seguridad para ayudar al usuario, activando una alarma a los respectivos respondedores. La segunda disposición del sistema analizado, Bindi 2.0a, también está relacionada con el caso 2 y se basa en las mismas dos tuberías de procesamiento de datos uni-modales de Bindi 1.0 pero aplicando, en la etapa de decisión final, una técnica de fusión tardía en lugar de una estrategia de acuerdo jerárquico o de confirmación [220], Figura 6-24. Hereda la funcionalidad de prealarma y casuística de Bindi 1.0 para tener un bajo consumo de energía para el micrófono. Finalmente, la última disposición del sistema, Bindi 2.0b, está relacionada con el caso 5. Este sistema es una variación de Bindi 2.0a pero basado en una adquisición continua de datos físicos y fisiológicos. No hay prealarma involucrada y este arreglo sigue el esquema de fusión tardía introducido en Bindi 2.0a.

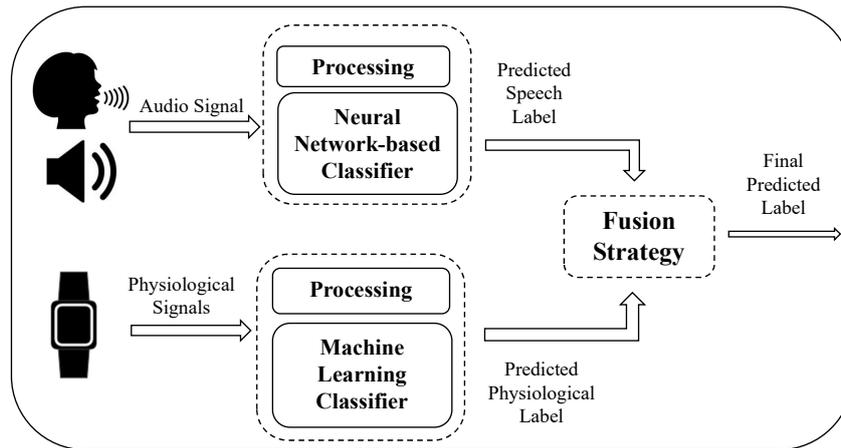


Figure 6-24: Diagrama de bloques de fusión de datos para Bindi 2.0a y Bindi 2.0b.

En las siguientes subsecciones se detallan las diferentes estrategias de fusión de datos consideradas y evaluadas con las tres disposiciones del sistema. Los resultados experimentales dan cuenta del proceso de validación realizado fuera de línea para evaluar la funcionalidad de los conductos de procesamiento de datos. Esto se hace para posteriormente integrar dichos módulos en la arquitectura, equilibrando las compensaciones observadas.

6.4.1 Métodos de fusión de datos multimodales

Antes de entrar en detalles sobre el marco multimodal propuesto, hay que tener en cuenta algunos puntos para la evaluación de las diferentes disposiciones del sistema multimodal. En primer lugar, de acuerdo con el diseño del conjunto de datos WEMAC, hay que señalar que los datos fisiológicos se recogen durante la elicitación del estímulo, mientras que la grabación del habla se registra durante la posterior anotación de audio. Esto significa que los datos fisiológicos y del habla no están alineados en el tiempo. Sin embargo, se requiere que ambos datos estén alineados para Bindi 2.0b, a diferencia de Bindi 1.0 y Bindi 2.0a. Dado que durante el etiquetado se pide a los voluntarias que revivan las emociones sentidas durante la elicitación del estímulo, se supone que la correspondencia es suficientemente sólida entre ambos instantes. Sin embargo, esta suposición deberá ser validada cuando se disponga del resto de subconjuntos del Base de datos UC3M4Safety.

Como ya se ha detallado en el apartado anterior, los subsistemas unimodales fisiológico y del habla estiman una etiqueta binaria, $y_k^m \in \{0, 1\}$, para cada ventana temporal k , donde $m \in \{textophy, sp\}$ son las dos modalidades, refiriéndose phy y sp

a los subsistemas fisiológico y del habla, respectivamente. Sin embargo, cada una de las modalidades utiliza una longitud de ventana temporal diferente, T_m , en segundos. Además, el sistema está pensado para emitir una respuesta por período de tiempo n (cada uno de los períodos de tiempo tiene una longitud L), en segundos. Así, se calcula una estimación de la probabilidad de miedo p_n^m para el periodo de tiempo n y la modalidad m , dada por

$$p_n^m = \frac{\sum_{k=1}^{K_m} y_{K_m \cdot n + k}^m}{K_m}, \quad (6.8)$$

donde $K_m = \lfloor \frac{L}{T_m} \rfloor$, es decir, el número de ventanas de tiempo que consideramos para cada modalidad para la estimación de probabilidades.

A partir de ahí, se puede calcular una única etiqueta binaria, Y_n^m , correspondiente a la probabilidad p_n^m como

$$Y_n^m = \begin{cases} 0 & \text{for } p_n^m < \text{th}_m \\ 1 & \text{otherwise} \end{cases}, \quad (6.9)$$

es decir, dará como resultado "1" (*miedo*) si p_n^m es mayor que el umbral predefinido relacionado con la modalidad, $\text{th}_m \in \{0, 1\}$, o "0" (*no-miedo*) en caso contrario. Obsérvese que los valores th_{phy} y th_{sp} se tratan en la sección 6.4.2.

Como métrica para representar el grado de confianza de cada sistema unimodal en la etiqueta de clase predicha en un periodo determinado, la entropía h_n^m para el periodo de tiempo n -th y la modalidad m -th se calcula como

$$h_n^m = -[p_n^m \cdot \log(p_n^m) + (1 - p_n^m) \cdot \log(1 - p_n^m)]. \quad (6.10)$$

Sobre esta base, se estudian tres estrategias de fusión tardía para producir la respuesta del sistema fusionado Y_n^{extf} para el período de tiempo n -th:

- Caso 1, Entropía más baja: La respuesta del sistema corresponde a la etiqueta binaria producida por el sistema unimodal con la menor entropía, es decir, la más segura. Para ello, la probabilidad de miedo fusionada p_n^f para el período

de tiempo n -th se calcula como

$$p_n^f = \begin{cases} p_n^{\text{phy}} & \text{if } h_n^{\text{phy}} < h_n^{\text{sp}} \\ p_n^{\text{sp}} & \text{otherwise} \end{cases}. \quad (6.11)$$

A continuación, aplicando el mismo razonamiento que en la ecuación (6.9), se obtiene una etiqueta binaria fusionada como

$$Y_n^f = \begin{cases} 0 & \text{for } p_n^f < \text{th}_f \\ 1 & \text{otherwise} \end{cases}, \quad (6.12)$$

donde, por ahora, th_f es el convencional 0,5.

- Caso 2, Combinación ponderada de la entropía inversa: La probabilidad de miedo fusionada p_n^f para el periodo de tiempo n -ésimo se calcula como una suma ponderada de probabilidades, tal y como viene dada por

$$p_n^f = \sum_m w_n^m \cdot p_n^m, \quad (6.13)$$

where

$$w_n^m = \frac{1/h_n^m}{\sum_m 1/h_n^m}. \quad (6.14)$$

A continuación, se obtiene una etiqueta binaria fusionada según la ecuación (6.12).

- Caso 3, OR lógico: La respuesta del sistema corresponde al cálculo del OR lógico sobre las etiquetas binarias de cada sistema unimodal. Es decir,

$$Y_n^f = Y_n^{\text{phy}} \vee Y_n^{\text{sp}}. \quad (6.15)$$

Las tres estrategias de fusión se basan en la literatura (por ejemplo, [301]) y se proponen como un compromiso entre la baja complejidad computacional y la robustez considerando la confianza del sistema en las predicciones. Al comparar teóricamente las tres estrategias de fusión, la OR lógica facilita la obtención de una predicción de clase de miedo sin comprobar la confianza del subsistema, lo que podría llevar a una falsa detección. Sin embargo, la estrategia de menor entropía confía en el modelo

de mayor confianza sin tener en cuenta las diferencias en las probabilidades. Por último, la combinación ponderada de la entropía inversa establece un equilibrio entre las probabilidades y las entropías para cada subsistema unimodal. Así, la confianza de esta última estrategia puede ser mayor que la de las demás.

En resumen, con respecto al procedimiento de prueba, las salidas del subsistema unimodal son matrices de etiquetas binarias. En concreto, para el "conjunto de datos", la longitud de las matrices es igual a la división de la duración de cada estímulo relacionado con la emoción por la respectiva ventana de respuesta unimodal, es decir, 10 y 1 segundo para los subsistemas fisiológico y del habla, respectivamente. Posteriormente, estas matrices recogidas se procesan calculando las probabilidades suaves y sus correspondientes etiquetas duras aplicando los umbrales fisiológicos (th_{phy}) y del habla (th_{sp}). Las estrategias de fusión de datos propuestas generarán también sus correspondientes etiquetas duras, tal y como se ha comentado anteriormente. Las métricas de evaluación seleccionadas, que son la precisión y la puntuación F1, se alimentan de las etiquetas duras finales obtenidas. La precisión puede representar de forma justa los índices de predicción, ya que el desequilibrio de clases es bajo. La puntuación F1 se considera para hacer frente al ligero desequilibrio observado. Aunque la puntuación F1 debería ser una buena métrica para un problema de detección como el que se aborda, en el que el número de positivos debería ser relativamente bajo en comparación con los negativos, el escenario experimental que se considera aquí está casi equilibrado y, por lo tanto, esta métrica no es tan significativa como se espera que sea cuando se prueba con datos capturados en condiciones reales.

6.4.2 Resultados de la fusión de datos multimodales

El primer análisis que hay que hacer es el rendimiento de los subsistemas fisiológico y del habla trabajando de forma independiente en un entorno continuo, es decir, teniendo en cuenta todas las muestras. Este experimento es esencial para determinar los umbrales, th_{phy} y th_{sp} , que convierten las etiquetas binarias obtenidas para cada periodo en una única etiqueta dura. Este paso es relevante porque determina si la arquitectura es más o menos propensa a las falsas alarmas, independientemente de la versión binaria o la disposición del sistema multimodal que se considere. Así, cada parámetro fue barrido en el rango $[0, 3, 0, 6]$ con pasos de 0,1 mientras se gen-

eraban los correspondientes 42 subsistemas unimodales siguiendo el enfoque LASO y considerando cada longitud de vídeo como los diferentes periodos aplicables. En este sentido, las Figs. 6-25a y 6-25b muestran los valores th_{phy} y th_{sp} frente a las métricas de precisión y puntuación F1-media para los 42 grupos de prueba en los subsistemas fisiológico y del habla, respectivamente.

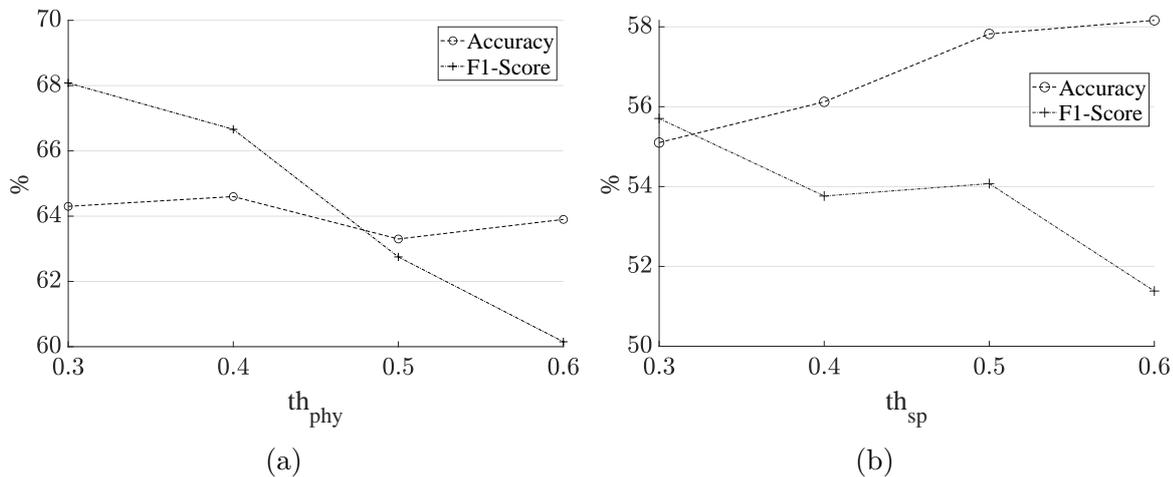


Figure 6-25: Barrido de parámetros para a) th_{phy} y b) th_{sp} en los subsistemas fisiológico y unimodal del habla, respectivamente.

Analizando la Fig. 6-25a, se puede observar cómo la puntuación F1 disminuye a medida que th_{phy} crece, mientras que la precisión se mantiene bastante estable. Obsérvese que la puntuación F1 depende en gran medida del número de Verdaderos Positivos (TP) predichos, pero no tiene en cuenta los Verdaderos Negativos (TN). Por lo tanto, si el TP aumenta y la suma de las tasas de falsos positivos (FP) y falsos negativos (FN) disminuye, la puntuación F1 aumenta. Esta compensación provoca el comportamiento observado, en el que cuanto más baja sea th_{phy} , mayor será la puntuación F1. Según este análisis, th_{phy} se fijó en 0,40, obteniendo un 66,66% y un 64,60% para la puntuación F1 y la precisión, respectivamente. Tenga en cuenta que estos valores son más altos que los reportados en la Tabla 6.7 debido al efecto de considerar un conjunto de salidas uni-modales para un período de tiempo determinado. La razón para elegir este valor de umbral específico es el buen compromiso observado entre ambas métricas y el hecho de que la falta de un TP podría ser dramática para el caso de uso de la violencia de género. Además, el sistema multi-modal combinado debe abstenerse de disparar falsas alarmas para evitar abrumar a las instituciones encargadas de protegerlas, y por eso se elige el subsistema de voz para ser más conservador en este sentido. Analizando la Fig. 6-25b para el

subsistema de voz, se puede observar cómo la F1 y la precisión empiezan a divergir a partir de 0,50. Por lo tanto, th_{sp} se fijó en este valor, obteniendo 54,07% y 57,82% para la puntuación F1 y la precisión, respectivamente. Obsérvese que la exactitud podría incluso aumentarse eligiendo un th_{sp} más alto.

Una vez fijados th_{phy} y th_{sp} , estudiamos la predicción del rendimiento medio en los 42 grupos de prueba para las diferentes configuraciones de la arquitectura, como se muestra en las figuras 6-27 y 6-26. Estas configuraciones son el subsistema unimodal fisiológico, el subsistema unimodal del habla, Bindi 1.0, Bindi 2.0a con fusión de datos de entropía mínima, Bindi 2.0a con fusión de datos de ponderación de entropía inversa, Bindi 2.0b con fusión de datos de menor entropía, Bindi 2.0b con fusión de datos de ponderación de entropía inversa, y Bindi 2.0b con fusión de datos de OR lógico. Nótese que Bindi 2.0a no se combinó con la fusión de datos OR lógica porque es equivalente a Bindi 1.0.

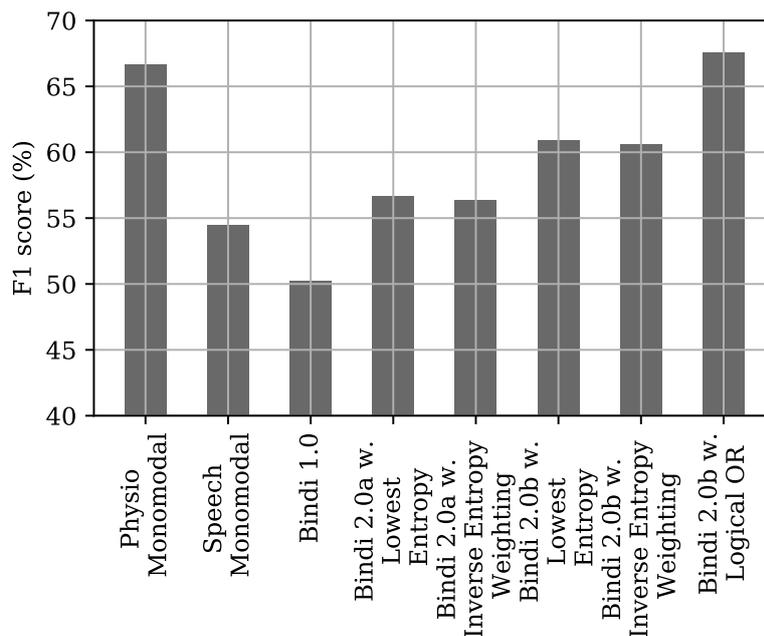


Figure 6-26: Análisis del rendimiento medio de las puntuaciones F1 prediciendo sobre los 42 voluntarias de prueba para las diferentes configuraciones de arquitectura.

Analizando la Figura 6-27, el subsistema uni-modal fisiológico consigue la mayor precisión proporcionando hasta un 64,63% y superando incluso a los esquemas de fusión. Para la métrica F1, este subsistema también proporciona la segunda tasa más alta con hasta un 66,67%. Este comportamiento puede estar relacionado con el sesgo introducido hacia la detección de la clase positiva, en primer lugar, con el enfoque de aprendizaje sensible al coste y en segundo lugar, con el barrido de

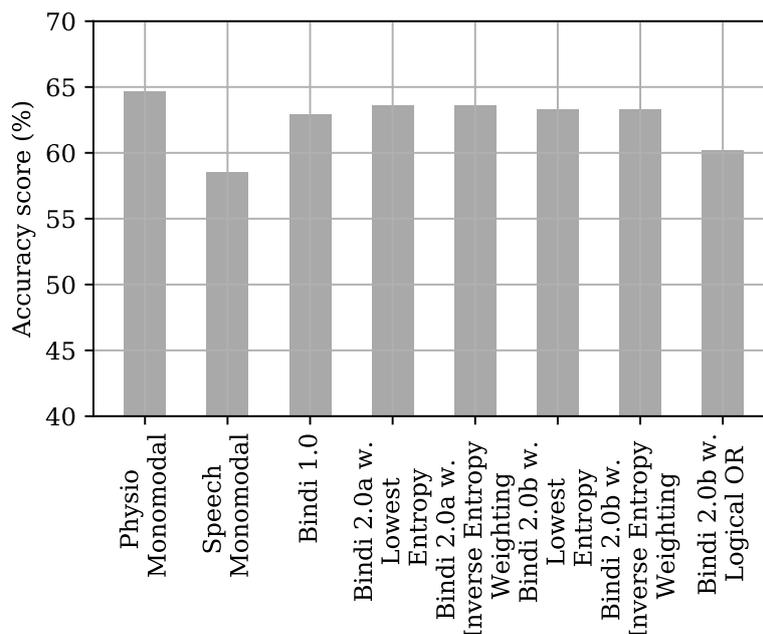


Figure 6-27: Análisis del rendimiento de la puntuación media de exactitud que predice sobre los 42 voluntarias de prueba para las diferentes configuraciones de arquitectura.

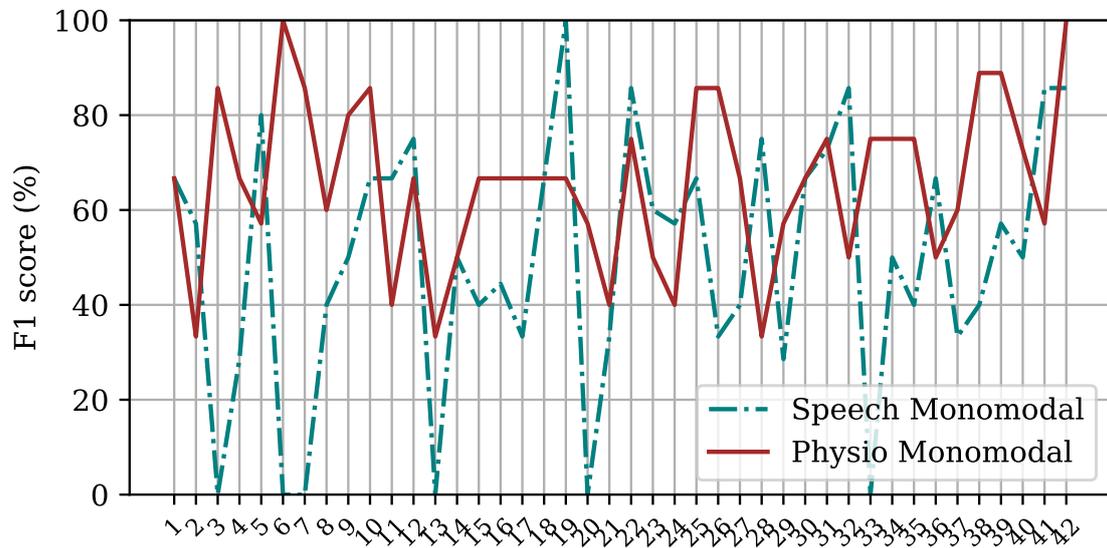
parámetros de th_{phy} . En la Figura 6-26, el subsistema unimodal del habla proporciona unas métricas significativamente inferiores a las del subsistema fisiológico. Este hecho podría estar relacionado con el limitado número de muestras para entrenar la red neuronal y la posible calidad limitada de las mismas debido a la acción de revivir la emoción sentida en la generación del conjunto de datos. Esta situación provoca que Bindi 1.0 proporcione la peor métrica en este análisis debido a que la respuesta final del sistema recae en el subsistema de habla. Bindi 2.0a y Bindi 2.0b ambos proporcionan una precisión similar cerca del subsistema fisiológico en la mayoría de los casos. Sin embargo, Bindi 2.0b logra la puntuación F1 más alta en todos los casos, especialmente para Bindi 2.0b con la fusión de datos lógica OR. Esta última estrategia proporciona la puntuación F1 más alta, un 67,59. En [11] se puede encontrar un breve avance de este análisis y una discusión de las matrices de confusión obtenidas para cada configuración.

Para elaborar los resultados mostrados en las Figuras 6-27 y 6-26, la Tabla 6.9 presenta resultados detallados para las diferentes configuraciones, incluyendo la desviación estándar media por voluntario probado. Los índices de desviación estándar bajos son buenos indicadores de una mejor capacidad de generalización, siempre que los resultados sean comparables. Obsérvese, por ejemplo, que aunque Bindi 1.0

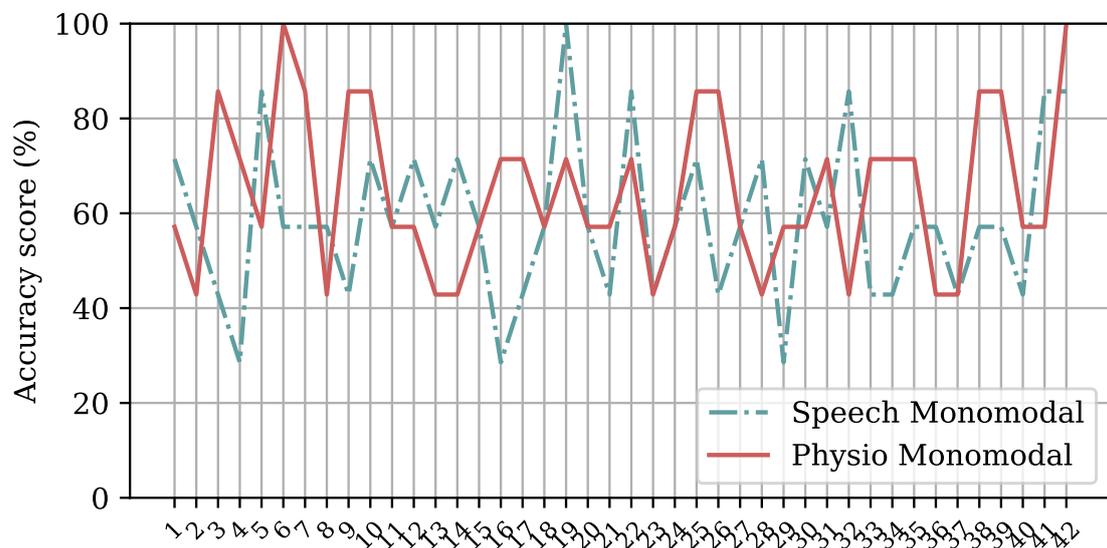
presenta la desviación estándar más baja, lo que puede considerarse una buena generalización, sus puntuaciones son superadas por la mayoría de las configuraciones, como ya se ha dicho. Además se puede observar que los valores de desviación estándar obtenidos son relativamente altos, especialmente para la puntuación F1. La causa se muestra en la Fig. 6-28, donde se proporciona la puntuación F1 y la precisión para cada una de las 42 pruebas y el subsistema uni-modal. Se observa que algunos voluntarios tienen una puntuación F1 de cero para el subsistema del habla en esta figura. Este hecho se produce porque la puntuación F1 depende del TP detectado, y no hubo predicciones positivas para algunos usuarios.

		Physiological uni-modal	Speech uni-modal	BINDI 1.0	Bindi 2.0a Lowest Entropy	Bindi 2.0a Inverse Entropy Weighting	Bindi 2.0b Lowest Entropy	Bindi 2.0b Inverse Entropy Weighting	Bindi 2.0b Logical OR
F1-score	Mean	66.67	54.48	50.23	56.68	56.33	60.87	60.58	67.59
	Std	17.31	26.73	27.64	23.91	24.05	26.63	26.98	14.27
Accuracy	Mean	64.63	58.5	62.93	63.61	63.61	63.27	63.27	60.2
	Std	16.56	16.73	14.30	14.35	14.35	17.94	18.21	15.75

Table 6.9: Análisis del rendimiento medio que predice los 42 voluntarios de las pruebas. Media y desviaciones estándar (Std).



(a)



(b)

Figure 6-28: Análisis del rendimiento individual de los dos subsistemas unimodales.

6.5 Discusión y conclusión

En este capítulo se han presentado diferentes contribuciones esenciales de esta investigación. En primer lugar, se detalla y explica el conjunto de datos WEMAC-nombre. En segundo lugar, se construye un sistema de detección del miedo basado en la fisiología a partir de los datos recogidos en dicho conjunto de datos. Por último, se proporciona un marco multimodal contextualizado en la tecnología Bindi. En concreto, estas contribuciones pueden detallarse como sigue:

- La generación de un nuevo conjunto de datos de detección de emociones que se ocupa de todas las limitaciones encontradas en las bases de datos públicas disponibles. Este conjunto de datos pertenece al Base de datos UC3M4Safety.
- El diseño e integración de un proceso de recuperación fisiológica activa dentro de los experimentos del conjunto de datos.
- El diseño, la implementación y la evaluación de un sistema de detección de miedo basado en la fisiología utilizando el conjunto de datos "Nombre del conjunto de datos".
- Un novedoso diseño de marco de fusión de datos multimodales utilizando información fisiológica y del habla.
- La aplicación de un modelo LASO que considera por primera vez el reconocimiento del miedo, la fusión de señales multisensoriales y los estímulos de realidad virtual.

Para el sistema de detección de miedo basado en la fisiología, el mejor resultado se obtiene utilizando un clasificador KNN y un clasificador AdaBoost (ENS) para el etiquetado discreto y dimensional del miedo, respectivamente. El primero alcanza hasta un 66,87% y un 66,45%, mientras que el segundo consigue hasta un 64,23% y un 66,62% para el ACC y el AUC promediados. Los resultados obtenidos están en consonancia con todos los sistemas Leave-One-Out (sujetos o ensayos) presentados en la literatura, véase la Tabla 4.19. Sin embargo, las recomendaciones proporcionadas al final de la sección 6.3.2.3 podrían aprovecharse e investigarse para mejorar estos resultados de referencia.

En lo que respecta específicamente al marco multimodal propuesto, el mejor resultado de fusión equilibrada se obtiene para la disposición Bindi 2.0b aplicando una estrategia de fusión de datos OR lógica. Este método reporta hasta un 60,20% y

un 67,59% para ACC y F1-score, respectivamente. Estos valores representan un resultado competitivo en comparación con el estado del arte que trata casos de uso multimodal similares [297, 300, 302]. Por otra parte, cabe destacar que las configuraciones descritas en este capítulo para la detección del miedo a través de datos fisiológicos y del habla son sólo una posible forma de caracterizar las situaciones y contextos en los que los usuarios pueden verse involucrados. Se trata de líneas de base iniciales para desarrollos posteriores que han permitido identificar importantes retos. En primer lugar, es crucial encontrar un equilibrio adecuado entre TP-TN y FP-FN, ya que el coste de no detectar una verdadera necesidad de ayuda es terrible, pero también hay que evitar interferir en la vida cotidiana de las víctimas de la violencia de género y la saturación de los servicios de protección con falsas alarmas. Por lo tanto, hemos tratado de reducir los FN en la medida de lo posible mientras los FP se mantienen en una tasa adecuada. Para ello, hemos considerado estrategias basadas en los costes de clasificación errónea y en la fijación de los parámetros de umbral. En concreto, hemos fijado el parámetro th_{phy} en el sistema fisiológico para obtener un mayor resultado de predicciones positivas con este sistema, de forma que en una fase posterior, las estrategias de habla (en Bindi 1.0) y de fusión de datos (en Bindi 2.0a y Bindi 2.0b) ayuden a corregir el sesgo mientras se intenta mantener la predicción TP. Se maximiza el Durante esta experimentación, el sistema unimodal del habla actual proporcionó tasas de rendimiento inferiores a las esperadas, lo que podría estar causado por la desalineación temporal de los datos fisiológicos y del habla en WEMAC. El desvanecimiento de la emoción elicitada en el momento en que se recoge la muestra de voz podría estar detrás de esta disminución del rendimiento.

En general, llegamos a la conclusión de que los sistemas de clasificación de miedo unimodal y multimodal obtenidos empleando el conjunto de datos WEMAC presentan resultados competitivos en comparación con el estado del arte. Sin embargo, es necesario seguir investigando para mejorar estos sistemas de cara a su aplicabilidad en la vida real. Así pues, el objetivo principal del marco multimodal propuesto y del conjunto de datos WEMAC es despertar el interés de la comunidad por este problema tan desafiante de la violencia de género y empezar a abordar la perspectiva de género en la inteligencia artificial.

Como trabajo futuro, el equipo de UC3M4Safety tiene previsto impulsar y desarrollar una serie de puntos clave y futuras líneas de actuación que se han identificado como limitaciones junto a la realización de este trabajo:

- Estudiar otras alternativas de fusión y modos de combinación para los subsistemas uni-modales.
- Aumentar el número de voluntarias y los datos disponibles de los sensores adquiridos con los dispositivos de borde.
- Incluir en la base de datos los datos de las víctimas de la violencia de género para comprender mejor sus mecanismos de activación en situaciones relacionadas con el miedo.
- Incorporar la arquitectura completa del sistema fisiológico unimodal y el procesamiento de datos en la pulsera Bindi y probar su eficacia en entornos y situaciones reales en experimentos en la naturaleza.
- Evaluar el uso de métricas de puntuación alternativas, como la información mutua y el área bajo la curva, para seguir encontrando un equilibrio adecuado entre las falsas alarmas y la probabilidad de fallo.
- Desarrollar y probar técnicas de adaptación del sujeto a los modelos unimodal y de fusión.

En el diseño de sistemas de detección de miedo para prevenir y combatir situaciones de violencia de género, pueden surgir varios problemas cuando el objetivo de un sistema es trabajar con datos de la vida real. En primer lugar, la dificultad de encontrar datos realistas, y en segundo lugar, la poca confianza en las arquitecturas desarrolladas si los datos utilizados son actuados o sintéticos. Esta situación lleva a la necesidad de generar bases de datos con emociones reales elicítadas, lo que supone un gran reto y requiere mucho tiempo. Sobre todo, trabajar con elicitación de emociones negativas fuertes, como las evocadas en WEMAC para la detección de miedo en mujeres en un entorno de laboratorio, puede llevar a problemas éticos. Por ello, hay que dedicar muchos recursos a salvaguardar el bienestar de los voluntarias que participan. Este problema en particular se agrava cuando el grupo objetivo de los voluntarias son mujeres que han sufrido violencia de género. Esto se debe a que los fallos del sistema o servicio de protección tienen consecuencias críticas para ellas. Por esta razón, la segunda versión del conjunto de datos que se está

recopilando actualmente en el laboratorio incluye únicamente a voluntarias víctimas de la violencia de género.

Parte IV

Conclusión

Conclusión

En este capítulo final, resumiremos las aportaciones de esta investigación doctoral en función de los objetivos propuestos. También aportaremos algunas sugerencias sobre posibles temas a estudiar en el futuro. Estas ideas provienen del último año de investigación y pueden suponer el punto de partida de nuevos proyectos de investigación.

Este doctorado comenzó con la creación de la UC3M4Safety, un equipo multidisciplinar que se creó al enfrentarse al problema de la Violencia de Género y reclamar que era necesario un enfoque multidisciplinar para fomentar nuevas y más innovadoras soluciones para prevenirla y combatirla. Impulsados por esta motivación, nos propusimos aportar nuevas herramientas para prevenir y combatir las situaciones de riesgo de Violencia de Género e, incluso, las agresiones, desde una perspectiva tecnológica, pero sin dejar de lado las diferentes consideraciones sociológicas relacionadas con el problema. En este contexto, y teniendo en cuenta el potencial tecnológico de la computación afectiva a través de la información fisiológica para generar esas nuevas herramientas, realizamos un análisis detallado sobre el desenrañamiento de la relación entre las señales fisiológicas y las emociones relacionadas con el miedo. Este estudio nos proporcionó el conocimiento para proponer un nuevo enfoque para detectar las emociones relacionadas con el miedo haciendo uso de las diferentes teorías emocionales y de los indicadores fisiológicos afectivos. Este estudio también se acompañó de una investigación exhaustiva sobre las herramientas de provocación de emociones, los informes de evaluación de emociones, las bases de datos de clasificación de emociones, el diseño de sistemas de computación afectiva.

tiva y las metodologías y herramientas relacionadas que nos permitieron construir una sólida base de conocimientos tecnológicos para cumplir con los retos de este doctorado.

Posteriormente, el enfoque de clasificación binaria del miedo se ha incluido en diferentes sistemas de computación afectiva contruidos sobre conjuntos de datos disponibles públicamente. En concreto, se han diseñado diferentes sistemas especializados de detección de miedo utilizando características de dominio temporal, frecuencial y no lineal. El valor añadido de las arquitecturas propuestas es la consideración de las restricciones de procesamiento digital para integrar adecuadamente dicho sistema en una plataforma de dispositivos de borde vestibles para permitir la protección de personas vulnerables. Durante el diseño de estos sistemas, se detectaron diferentes limitaciones en las bases de datos de libre acceso con las que estábamos trabajando. Por ejemplo, no se utilizaba tecnología de inmersión emocional, la metodología de etiquetado no tenía en cuenta la perspectiva de género, no se garantizaba una distribución de estímulos adecuadamente equilibrada con respecto a las emociones objetivo, y no se implementaba la integración de un proceso de recuperación basado en las señales fisiológicas de los voluntarios para cuantificar y aislar la activación emocional entre los estímulos. Sin embargo, los sistemas propuestos se compararon con éxito con el estado de la técnica.

Junto con el diseño y la validación de los diferentes sistemas de clasificación del miedo, se propuso una nueva solución de hardware vestible para desplegar las arquitecturas de los sistemas de detección relacionados con el miedo. Así, diseñamos Bindi, un sistema multimodal autónomo para la detección de situaciones de riesgo en contextos de violencia de género. La parte de computación de borde del sistema es una red ciberfísica inteligente. En concreto, esto se consigue mediante sensores inteligentes fisiológicos y físicos (audio y/o voz) que monitorizan continuamente al usuario. La capa del sistema basada en la niebla reside en una fusión de datos multimodal dentro de una aplicación ad-hoc para smartphones. Además, la información se envía a servidores informáticos específicos en la nube, que se encargan de almacenar los datos recogidos para posteriores acciones legales. El diseño de este sistema puede impulsar la generación de nuevos mecanismos de prevención y lucha contra la la violencia de género.

Finalmente, tras haber identificado la necesidad de generar una nueva base de datos y crear una nueva herramienta tecnológica, diseñamos y realizamos el conjunto de datos WEMAC. Consiste en de 104 mujeres que nunca han sufrido Violencia de Género que realizaron diferentes visualizaciones de estímulos relacionados con la emoción en un entorno de laboratorio. Los anteriores sistemas de clasificación binaria del miedo se mejoraron y se aplicaron a este novedoso conjunto de datos multimodal, lo que permitió obtener resultados competitivos en comparación con el estado del arte.

7.1 Contribuciones

Para ser más precisos, ordenaremos las contribuciones en función del capítulo en el que se realizan.

Las aportaciones sobre el capítulo 4 son las siguientes:

- La aplicación y validación de una nueva propuesta de clasificación binaria del miedo utilizando conjuntos de datos abiertos disponibles y un conjunto reducido de señales fisiológicas.
- El diseño y la evaluación de un sistema de clasificación del miedo que emplea la base de datos DEAP y el modelo PA. Logró un AUC de 81,60% y un Gmean de 81,55% en promedio para un enfoque independiente del sujeto y sólo dos señales fisiológicas (PPG y GSR).
- El diseño y la evaluación de un sistema de clasificación del miedo que emplea la base de datos MAHNOB y el modelo PAD. Logró un AUC de 86,00% y un Gmean de 73,78% en promedio para un enfoque independiente del sujeto y sólo tres señales fisiológicas (PPG, GSR y SKT). Hay que tener en cuenta que este sistema se probó con LOSO.

Las aportaciones sobre el capítulo 5 son las siguientes:

- El diseño, hardware y software, de un nuevo sistema smart-wearable basado en un conjunto reducido de señales fisiológicas y orientado a la generación de nuevos mecanismos y herramientas tecnológicas para prevenir y combatir la Violencia de Género.
- Diseño e implementación de un modelo de inferencia Mamdani de base de reglas difusas de baja complejidad para la evaluación de la calidad de las señales

en la pulsera de Bindi. Esto va acompañado de una propuesta de definición e implementación de un novedoso ajuste fino no supervisado en línea basado a través de la similitud escalada entre conjuntos difusos de tipo II de intervalo para las actualizaciones autoadaptativas del modelo. Los resultados muestran que el sistema alcanzó una precisión global del 93,72%. El sistema propuesto, que tiene en cuenta la calidad, presenta un consumo de energía de hasta 59,40 *mJ*, lo que repercute directamente en el consumo global de energía del 1,5% al 20,7% para la transmisión de una señal de fotopletiografía ruidosa de 12-60 segundos.

- Se implementaron diferentes estrategias de filtrado y técnicas de extracción de características en el brazalete de Bindi. Esto va acompañado de una medición exitosa y una comparación de resultados con un conjunto de herramientas específicas para la investigación.

Las aportaciones sobre el capítulo 6 son las siguientes:

- Para la generación de este conjunto de datos ha sido necesario un arduo trabajo en equipo. Por ejemplo, se ha empleado una cantidad global de 7000 horas. Se trata de una colección de experimentos capturados en laboratorio condiciones de laboratorio con mujeres voluntarias. Un conjunto de estímulos audiovisuales se emplean para provocar emociones realistas utilizando la realidad virtual y adquiriendo la información fisiológica y del habla de las voluntarias. Además, también se recogen anotaciones emocionales autoinformadas en escalas emocionales dimensionales y discretas. Los objetivos y contribuciones de este novedoso conjunto de datos multimodales son múltiples, como se muestra brevemente a continuación:

1. La integración de la tecnología inmersiva para provocar emociones. Se emplea la realidad virtual por ser la que más se asemeja a los escenarios del mundo real, ofreciendo un alto grado de correlación entre las condiciones de la investigación y el fenómeno emocional estudiado, es decir, con validez ecológica.
2. La consideración de un elevado número de voluntarios. El primer experimento contó con un total de 104 mujeres voluntarias que no eran víctimas de la violencia de género.

3. La aplicación de una distribución de estímulos debidamente equilibrada respecto a las emociones objetivo. Antes de generar el conjunto de datos, se aplicó una metodología mixta con jueces expertos y público en general para seleccionar los mejores estímulos audiovisuales para provocar reacciones emocionales. Se realizó una encuesta pública con 1.332 participantes para etiquetar los estímulos relacionados con las emociones preseleccionadas.
4. La modificación de la metodología de etiquetado para tener en cuenta la perspectiva de género. Este problema se abordó cambiando los maniqués de autoevaluación originales.
5. La implementación de un proceso de recuperación activa respecto a la estabilización fisiológica entre estímulos. Hasta donde sabemos, no existe ningún conjunto de datos público que haya implementado una evaluación de la estabilización en línea mediante la evaluación de la retroalimentación fisiológica durante los experimentos.

Entre estas contribuciones, he participado directamente en los objetivos 1, 2, 4,y5.

- Los primeros resultados experimentales multimodales con WEMAC. Éstos muestran una precisión media de la tasa de reconocimiento del miedo de hasta 63,61% con el método Leave-half-Subject-Out (LASO). Que yo sepa, es la primera vez que se presenta un modelo LASO que considera el reconocimiento del miedo, la fusión de señales multisensoriales y los estímulos de realidad virtual.

7.2 Trabajo futuro

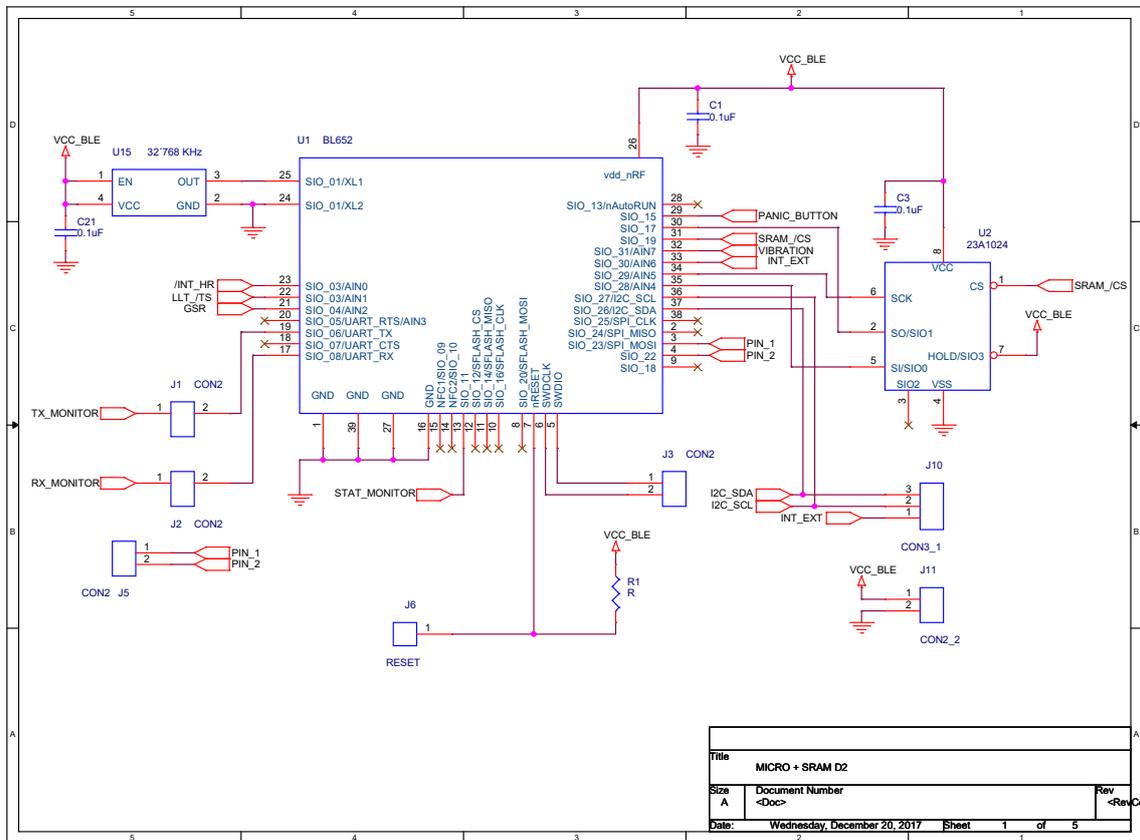
El autor desea aportar algunas sugerencias para futuras investigaciones:

- La consideración de más señales fisiológicas, o incluso biológicas, para ampliar el alcance de esta investigación y mejorar los resultados de la clasificación del miedo.
- Análisis multivariante considerando los cuestionarios iniciales recogidos al principio de los experimentos WEMAC junto con la información fisiológica y auditiva.

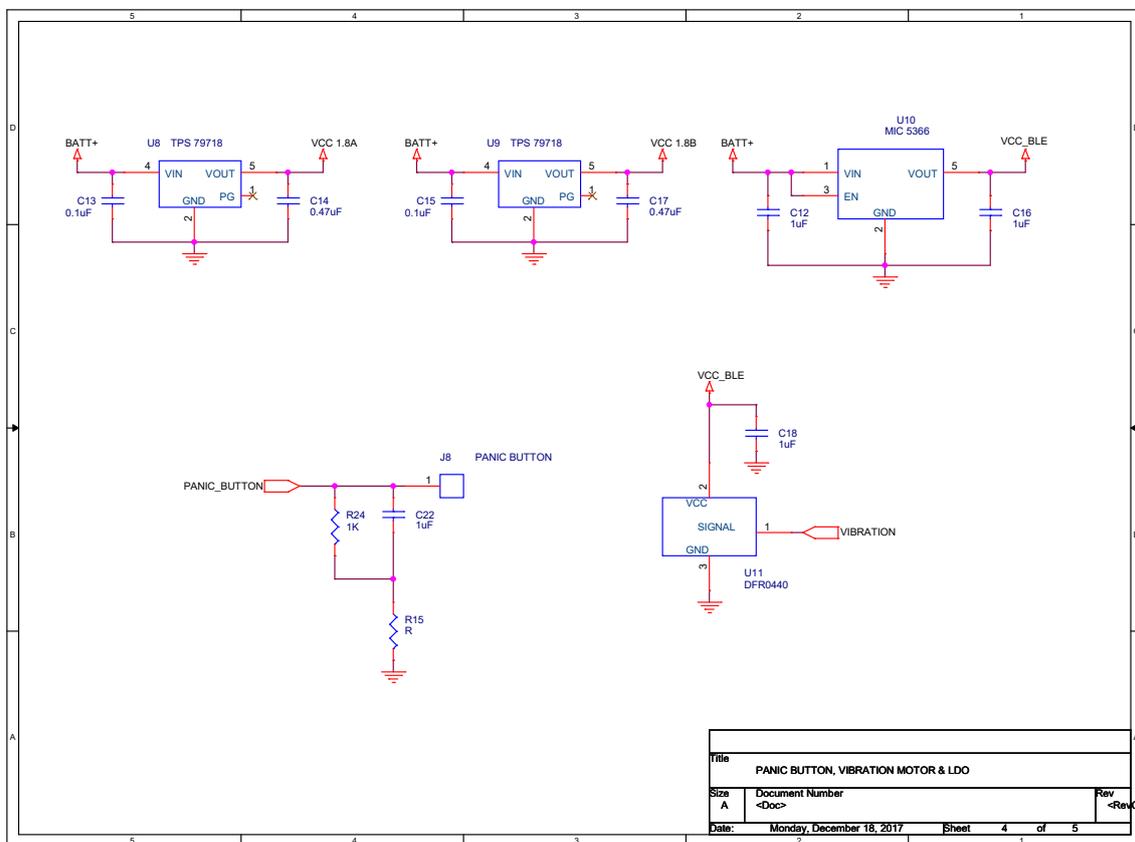
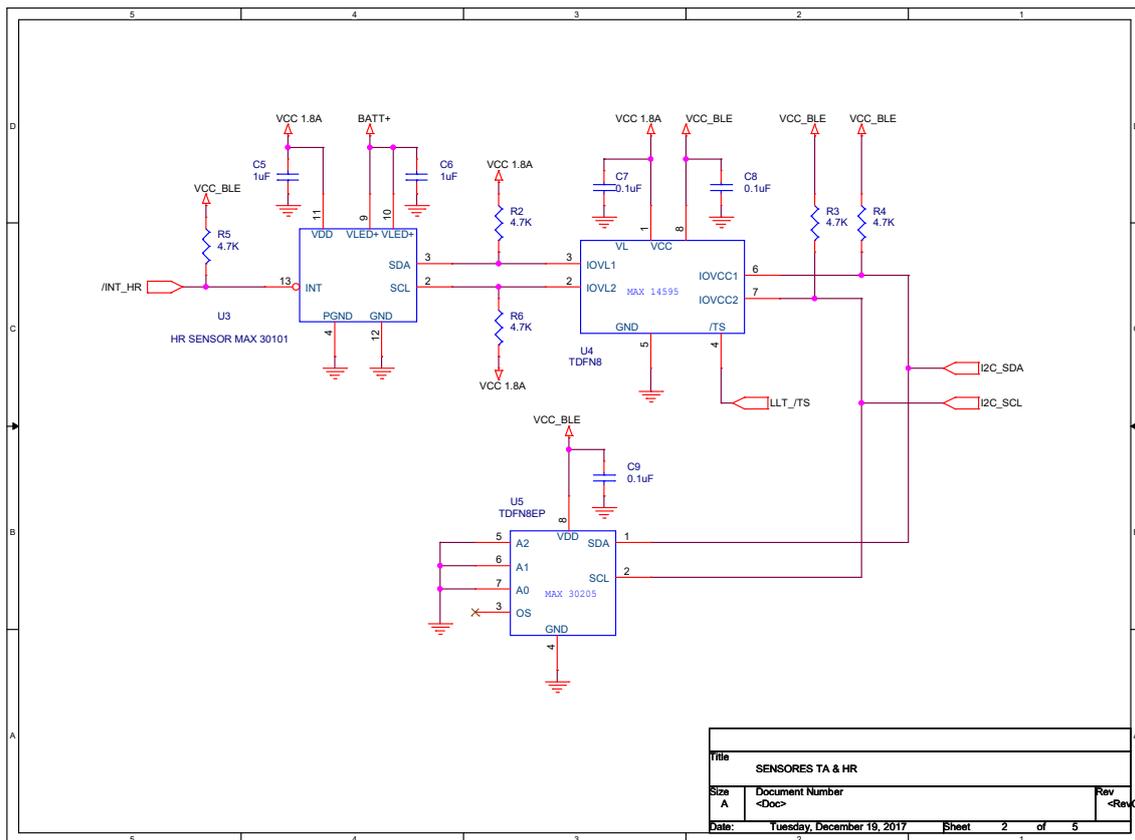
- En caso de recoger autoinformes discretos y dimensionales de los voluntarios, ambas metodologías de etiquetado podrían fusionarse mediante una combinación lineal o no lineal. Esto está en consonancia con el hecho de que tanto las etiquetas discretas como las dimensionales existen, pero tienen propósitos diferentes o caracterizan aspectos distintos de las emociones.
- La investigación y la aplicación de técnicas de detección comprimida para reducir el consumo de energía de la pulsera. Ya se han iniciado los trabajos sobre este tema.
- La investigación e implementación de técnicas de recolección de energía dentro de la pulsera sería interesante para observar el efecto del consumo de energía. El trabajo sobre este tema ya se ha iniciado.
- La investigación y la integración de sistemas de clasificación semi-supervisados destinados a tratar la dinámica de la emoción y/o el aprendizaje débilmente supervisado para el reconocimiento de la emoción de grano fino utilizando señales fisiológicas. Los trabajos sobre este tema ya se han iniciado.
- La integración de la computación neuromórfica en Bindi, como el Akida Neural Processor SoC.
- La implementación de nuevos algoritmos de eliminación de artefactos de movimiento. Por ejemplo, las técnicas de sincronización junto con las redes neuronales de extremo a extremo, que son compatibles con la computación de borde, tienen potencial. Ya se ha empezado a trabajar en este tema.
- La investigación de nuevas técnicas y métodos de personalización del aprendizaje automático aumentaría las posibilidades de despliegue de Bindi. El trabajo sobre este tema ya se ha iniciado.
- El diseño de nuevos factores de forma para llevar puesto, en lugar de una pulsera y un colgante.
- El diseño de un sistema experto que se ejecute en la nube y funcione de forma multivariable. El objetivo de dicho sistema sería corregir o modificar el aprendizaje automático de la computación de borde.

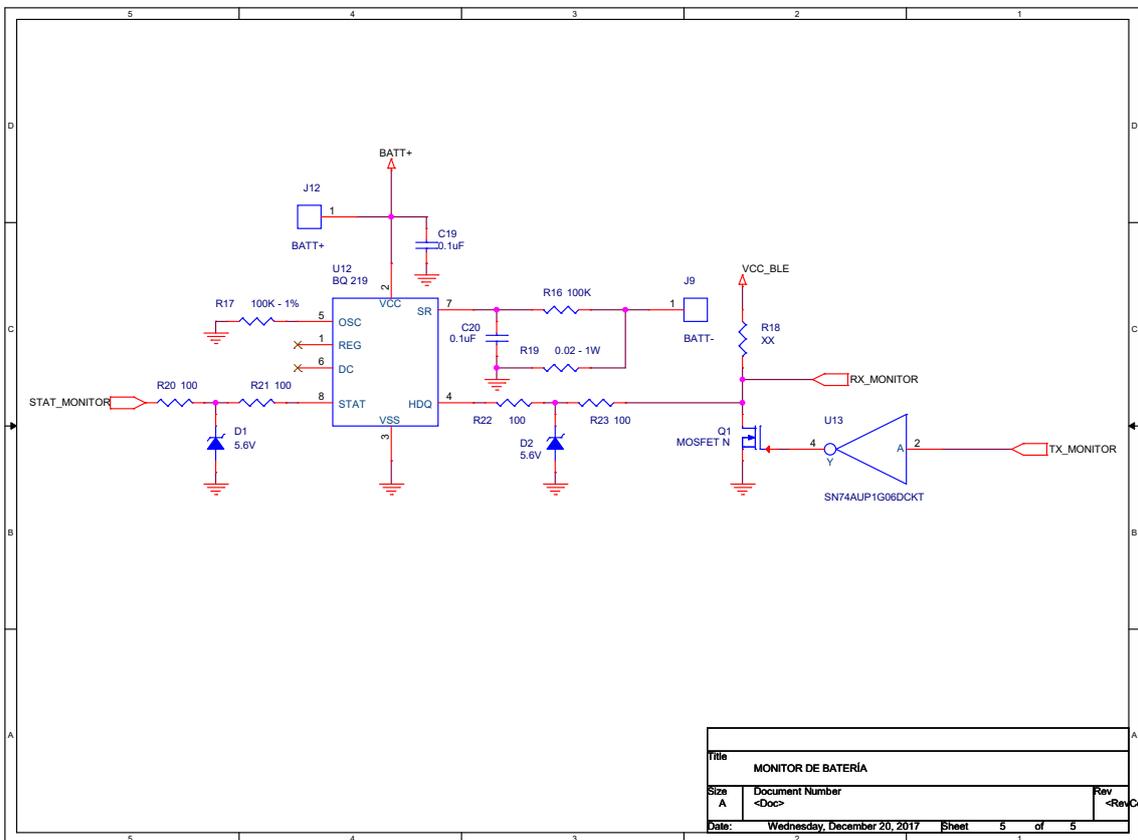
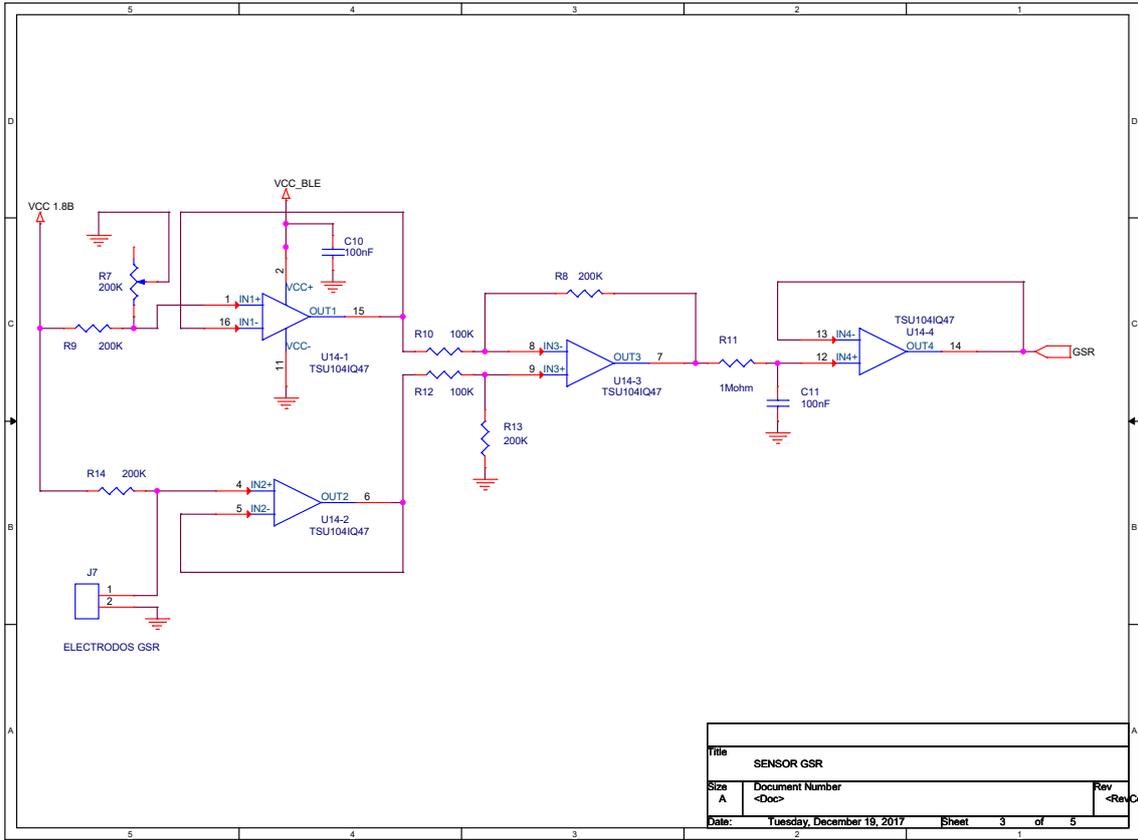
Appendix A

Esquemáticos del brazalete



Appendix A. Esquemáticos del brazalete





Bibliography

- [1] S. G. D. against Gender Violence. (2020) Fact-sheets on gender violence victims killed since 2003. [Online]. Available: violenciagenero.igualdad.gob.es/violenciaEnCifras/victimasmortales/fichamujeres/home.htm
- [2] R. S. Lazarus, “Progress on a cognitive-motivational-relational theory of emotion.” *American Psychologist*, vol. 46, no. 8, pp. 819–834, 1991.
- [3] A. Ellis, *Rational-emotive theory: Albert Ellis.*, ser. Operational theories of personality. Oxford, England: Brunner/Mazel, 1974, pp. x, 421–x, 421.
- [4] I. Bakker, T. van der Voordt, P. Vink, and J. de Boon, “Pleasure, arousal, dominance: Mehrabian and russell revisited,” *Current Psychology*, vol. 33, no. 3, pp. 405–421, Sep 2014.
- [5] J. R. J. Fontaine, K. R. Scherer, E. B. Roesch, and P. C. Ellsworth, “The world of emotions is not two-dimensional,” *Psychological Science*, vol. 18, no. 12, pp. 1050–1057, 2007. [Online]. Available: <http://www.jstor.org/stable/40064702>
- [6] M. M. Bradley and P. J. Lang, “Measuring emotion: The self-assessment manikin and the semantic differential,” *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.
- [7] A. Ferrari, D. Micucci, M. Mobilio, and P. Napoletano, “On the personalization of classification models for human activity recognition,” *IEEE Access*, vol. 8, pp. 32 066–32 079, 2020.
- [8] M. Malik, J. T. Bigger, A. J. Camm, R. E. Kleiger, A. Malliani, A. J. Moss, and P. J. Schwartz, “Heart rate variability: Standards of measurement, physiological interpretation, and clinical use,” *European Heart Journal*, vol. 17, no. 3, pp. 354–381, 03 1996.
- [9] D. of the Spanish Government against Gender Violence, “Dispositivos de control telemático de medidas y penas de alejamiento,” <https://violenciagenero.igualdad.gob.es/informacionUtil/recursos/dispositivosControlTelematico/home.htm>, (Accessed on 04/04/2021).
- [10] *High-Sensitivity Pulse Oximeter and Heart-Rate Sensor for Wearable Health*, Maxim Integrated, 2020, pPG. [Online]. Available: https://www.mouser.es/datasheet/2/744/Seeed_105020003-1217653.pdf
- [11] J. A. Miranda, E. Rituerto-González, M. F. Canabal, A. R. Bárcenas, J. M. Lanza-Gutiérrez, C. Pelaez-Moreno, and C. López-Ongil, “Bindi: Affective internet of things to combat gender-based violence,” *IEEE Internet of Things*, 2022, manuscript submitted for publication.
- [12] J. Lichtenauer and M. Soleymani, “Mahnob-hci-tagging database,” 2011.
- [13] U. Nations, “Declaration on the elimination of violence against women,” 1993.

- [14] E. Commission. (2020) Gender violence definition and forms. [Online]. Available: <https://ec.europa.eu/info/policies/justice-and-fundamental-rights/gender-equality/gender-based-violence/what-gender-based-violence>
- [15] L. Sardinha *et al.*, “Global, regional, and national prevalence estimates of physical or sexual, or both, intimate partner violence against women in 2018,” *The Lancet*, 2 2022. [Online]. Available: [https://doi.org/10.1016/S0140-6736\(21\)02664-7](https://doi.org/10.1016/S0140-6736(21)02664-7)
- [16] H. A. T. DIRECTIVE, “Council directive 75/117/eec of 10 february 1975 on the approximation of the laws of the member states relating to the application of the principle of equal pay for men and women,” *Official Journal L*, vol. 45, no. 19/02, pp. 0019–0020, 1975.
- [17] G. Maganza, “The lisbon treaty: A brief outline,” *Fordham Int’l LJ*, vol. 31, p. 1603, 2007.
- [18] J. Ulla and S. Rosamund, “The istanbul convention: A tool to tackle violence against women and girls,” *European Parliamentary Research Service*, 2020.
- [19] E. Commission. (2020) Actions against women violence. [Online]. Available: <https://ec.europa.eu/justice/saynostopvaw/eu-actions.html>
- [20] M. Segrave and L. Vitis, Eds., *Gender, Technology and Violence*, 1st ed., ser. Routledge Studies in Crime and Society. United Kingdom: Routledge, 2017.
- [21] T. Martínez, “Un recorrido por el sistema institucional en el ámbito de la violencia de género,” *Revista de Estudios Socioeducativos. ReSed*, no. 7, pp. 256–257, 2019.
- [22] J. J. López-Ossorio, J. L. González-Álvarez, and A. Andrés-Pueyo, “Predictive effectiveness of the police risk assessment in intimate partner violence,” *Psychosocial Intervention*, vol. 25, pp. 1 – 7, 04 2016.
- [23] M. de Sanidad. (2020) Atenpro. spanish social services. [Online]. Available: <https://www.mscbs.gob.es/en/ssi/violenciaGenero/Recursos/ATENPRO/home.htm>
- [24] L. Arenas García, “The efficacy of electronic monitoring in gender violence: criminological analysis,” *International e-Journal of Criminal Sciences*, no. 10, 2016.
- [25] R. S. Recio, E. G. Alberola, C. I. F. Guarné *et al.*, “Prevention of violence against women: policies and actions on gender violence,” *Informació Psicològica*, no. 111, pp. 35–50, 2016.
- [26] P. Schmidt, A. Reiss, R. Dürichen, and K. V. Laerhoven, “Wearable-based affect recognition—a review,” *Sensors*, vol. 19, no. 19, 2019. [Online]. Available: <https://www.mdpi.com/1424-8220/19/19/4079>
- [27] C. Morgan, *Introduction to Psychology*. McGraw-Hill, 1961.
- [28] C. Darwin, P. Ekman, and P. Prodger, *The Expression of the Emotions in Man and Animals*. Oxford University Press, 1998.
- [29] A. Colman, *A Dictionary of Psychology*, ser. Oxford Dictionary of Psychology. Oxford University Press, 2009.
- [30] W. B. Cannon, “The james-lange theory of emotions: A critical examination and an alternative theory,” *The American Journal of Psychology*, vol. 39, no. 1/4, pp. 106–124, 1927.
- [31] D. G. MacKay, M. A. Shafto, J. K. Taylor, D. E. Marian, L. Abrams, and J. R. Dyer, “Relations between emotion, memory, and attention: Evidence from taboo stroop, lexical decision, and immediate memory tasks,” *Memory*

- & *Cognition*, vol. 32, pp. 474–488, 2004.
- [32] P. Ekman, “What scientists who study emotion agree about,” *Perspectives on Psychological Science*, vol. 11, no. 1, pp. 31–34, 2016, PMID: 26817724.
- [33] A. Konar and A. Chakraborty, “Introduction to emotion recognition,” in *Emotion Recognition: A Pattern Analysis Approach*, 2015.
- [34] P. Ekman, “An argument for basic emotions,” *Cognition and Emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.
- [35] P. Ekman and W. V. Friesen, “Measuring facial movement,” *Environmental psychology and nonverbal behavior*, vol. 1, no. 1, pp. 56–75, 1976.
- [36] C. E. Izard, *Theories of Emotion and Emotion-Behavior Relationships*. Springer US, 1977, pp. 19–42.
- [37] R. Plutchik, “Emotions: A general psychoevolutionary theory,” *Approaches to emotion*, vol. 1984, pp. 197–219, 1984.
- [38] N. H. Frijda, *The Emotions*. Cambridge University Press, 1986.
- [39] K. Oatley and P. N. Johnson-laird, “Towards a cognitive theory of emotions,” *Cognition and Emotion*, vol. 1, no. 1, pp. 29–50, 1987.
- [40] P. Ekman and D. Cordaro, “What is meant by calling emotions basic,” *Emotion review*, vol. 3, no. 4, pp. 364–370, 2011.
- [41] B. Mesquita, “The legacy of nico h.frijda (1927–2015),” *Cognition and Emotion*, vol. 30, no. 4, pp. 603–608, 2016, PMID: 26943647.
- [42] P. Ekman, “What scientists who study emotion agree about,” *Perspectives on Psychological Science*, vol. 11, no. 1, pp. 31–34, 2016, PMID: 26817724.
- [43] W. Wundt, “Vorselung über die menschen – und tierseele,” *Voss Verlag: Leipzig, Germany*, pp. 145–172, 1863.
- [44] C. A. Clark, “Book reviews : The measurement of meaning by charles e. osgood, george j. suci, and percy h. tannenbaum. urbana, illinois: University of illinois press, 1957. 342 p. \$7.50,” *Educational and Psychological Measurement*, vol. 18, no. 4, pp. 884–886, 1958.
- [45] J. Russell, “A circumplex model of affect,” *Journal of personality and social psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.
- [46] A. Mehrabian, “Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament,” *Current Psychology*, vol. 14, no. 4, pp. 261–292, Dec 1996.
- [47] H. A. Demaree, D. E. Everhart, E. A. Youngstrom, and D. W. Harrison, “Brain lateralization of emotional processing: Historical roots and a future incorporating “dominance”,” *Behavioral and Cognitive Neuroscience Reviews*, vol. 4, no. 1, pp. 3–20, 2005, PMID: 15886400.
- [48] E. Harmon-Jones, C. Harmon-Jones, and E. Summerell, “On the importance of both dimensional and discrete models of emotion,” *Behavioral Sciences*, vol. 7, no. 4, 2017. [Online]. Available: <https://www.mdpi.com/2076-328X/7/4/66>
- [49] T. Thanapattheerakul, K. Mao, J. Amoranto, and J. H. Chan, “Emotion in a century: A review of emotion recognition,” in *Proceedings of the 10th International Conference on Advances in Information Technology*, 2018, pp. 1–8.
- [50] C. Castelfranchi, *Affective Appraisal versus Cognitive Evaluation in Social Emotions and Interactions*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 76–106.
- [51] E. Komulainen, K. Meskanen, J. Lipsanen, J. M. Lahti, P. Jylha, T. Melartin,

- M. Wichers, E. Isometsa, and J. Ekelund, "The effect of personality on daily life emotional processes," *PLOS ONE*, vol. 9, pp. 1–9, 10 2014. [Online]. Available: <https://doi.org/10.1371/journal.pone.0110907>
- [52] M. A. A. Mengual, *Trastorno de estrés postraumático: Daño cerebral secundario a la violencia (mobbing, violencia de género, acoso escolar)*. Ediciones Díaz de Santos, 2007.
- [53] M. Bianchin and A. Angrilli, "Gender differences in emotional responses: A psychophysiological study," *Physiology and Behavior*, vol. 105, no. 4, pp. 925–932, 2012. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0031938411005221>
- [54] L. Lambrecht, B. Kreifelts, and D. Wildgruber, "Gender differences in emotion recognition: Impact of sensory modality and emotional category," *Cognition & emotion*, vol. 28, no. 3, pp. 452–469, 2014.
- [55] X. Chen, H. Yuan, T. Zheng, Y. Chang, and Y. Luo, "Females are more sensitive to opponent's emotional feedback: Evidence from event-related potentials," *Frontiers in Human Neuroscience*, vol. 12, p. 275, 2018. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fnhum.2018.00275>
- [56] S. Dobrišek, R. Gajšek, F. Mihelič, N. Pavešić, and V. Štruc, "Towards efficient multi-modal emotion recognition," *International Journal of Advanced Robotic Systems*, vol. 10, no. 1, p. 53, 2013.
- [57] M. Á. Blanco Ruiz, L. Gutiérrez Martín, J. Á. Miranda Calero, M. F. Canabal Benito, E. Rituerto González, C. Luis Minguenza, J. C. Robredo García, B. Morán González, A. Páez Montoro, A. Ramírez Bárcenas *et al.*, "Uc3m4safety database description," <http://hdl.handle.net/10016/32481>, 2021.
- [58] P. J. Lang, "International affective picture system (iaps): Affective ratings of pictures and instruction manual," *Technical report*, 2005.
- [59] A. Marchewka, Ł. Żurawski, K. Jednoróg, and A. Grabowska, "The nencki affective picture system (naps): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database," *Behavior research methods*, vol. 46, no. 2, pp. 596–610, 2014.
- [60] B. Kurdi, S. Lozano, and M. R. Banaji, "Introducing the open affective standardized image set (oasis)," *Behavior research methods*, vol. 49, no. 2, pp. 457–470, 2017.
- [61] E. S. Dan-Glauser and K. R. Scherer, "The geneva affective picture database (gaped): a new 730-picture database focusing on valence and normative significance," *Behavior research methods*, vol. 43, no. 2, pp. 468–477, 2011.
- [62] M. Wessa, P. Kanske, P. Neumeister, K. Bode, J. Heissler, S. Schönfelder *et al.*, "Emopics: Subjektive und psychophysiologische evaluation neuen bildmaterials für die klinisch-bio-psychologische forschung," *Zeitschrift für Klinische Psychologie und Psychotherapie*, vol. 39, no. Suppl. 1/11, p. 77, 2010.
- [63] A. Haberkamp, J. A. Glombiewski, F. Schmidt, and A. Barke, "The disgust-related-images (dirti) database: Validation of a novel standardized set of disgust pictures," *Behaviour research and therapy*, vol. 89, pp. 86–94, 2017.
- [64] A. C. Samson, S. D. Kreibig, B. Soderstrom, A. A. Wade, and J. J. Gross, "Eliciting positive, negative and mixed emotional states: A film library for affective scientists," *Cognition and emotion*, vol. 30, no. 5, pp. 827–856, 2016.
- [65] A. Di Crosta, P. La Malva, C. Manna, A. Marin, R. Palumbo, M. C. Verroc-

- chio, M. Cortini, N. Mammarella, and A. Di Domenico, “The chieti affective action videos database, a resource for the study of emotions in psychology,” *Scientific data*, vol. 7, no. 1, pp. 1–6, 2020.
- [66] T. L. Gilman, R. Shaheen, K. M. Nylocks, D. Halachoff, J. Chapman, J. J. Flynn, L. M. Matt, and K. G. Coifman, “A film set for the elicitation of emotion in research: A comprehensive catalog derived from four decades of investigation,” *Behavior research methods*, vol. 49, no. 6, pp. 2061–2082, 2017.
- [67] S. Carvalho, J. Leite, S. Galdo-Álvarez, and O. F. Gonçalves, “The emotional movie database (emdb): A self-report and psychophysiological study,” *Applied psychophysiology and biofeedback*, vol. 37, no. 4, pp. 279–294, 2012.
- [68] K. Umla-Runge, H. D. Zimmer, X. Fu, and L. Wang, “An action video clip database rated for familiarity in china and germany,” *Behavior Research Methods*, vol. 44, no. 4, pp. 946–953, 2012.
- [69] T. B. Alakus, M. Gonen, and I. Turkoglu, “Database for an emotion recognition system based on eeg signals and various computer games–gameemo,” *Biomedical Signal Processing and Control*, vol. 60, p. 101951, 2020.
- [70] M. Granato, D. Gadia, D. Maggiorini, and L. A. Ripamonti, “Software and hardware setup for emotion recognition during video game fruition,” in *Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good*, ser. Goodtechs ’18. Association for Computing Machinery, 2018, p. 19–24.
- [71] M. Granato, “Emotions recognition in video game players using physiological information,” Ph.D. dissertation, Università Degli Studi Di Milano, 2019.
- [72] C. Kirschbaum, K.-M. Pirke, and D. H. Hellhammer, “The ‘trier social stress test’—a tool for investigating psychobiological stress responses in a laboratory setting,” *Neuropsychobiology*, vol. 28, no. 1-2, pp. 76–81, 1993.
- [73] L. A. Mitchell, R. A. MacDonald, and E. E. Brodie, “Temperature and the cold pressor test,” *The Journal of Pain*, vol. 5, no. 4, pp. 233–237, 2004.
- [74] F. Scarpina and S. Tagini, “The stroop color and word test,” *Frontiers in psychology*, vol. 8, p. 557, 2017.
- [75] A. L. Shilton, R. Laycock, and S. G. Crewther, “The maastricht acute stress test (mast): Physiological and subjective responses in anticipation, and post-stress,” *Frontiers in psychology*, vol. 8, p. 567, 2017.
- [76] J. Wijsman, B. Grundlehner, H. Liu, J. Penders, and H. Hermens, “Wearable physiological sensors reflect mental stress state in office-like situations,” in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, 2013, pp. 600–605.
- [77] M. Granato, D. Gadia, D. Maggiorini, and L. A. Ripamonti, “Software and hardware setup for emotion recognition during video game fruition,” in *Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good*, 2018, pp. 19–24.
- [78] N. S. Suhaimi, C. T. B. Yuan, J. Teo, and J. Mountstephens, “Modeling the affective space of 360 virtual reality videos based on arousal and valence for wearable eeg-based vr emotion classification,” in *2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA)*. IEEE, 2018, pp. 167–172.
- [79] J. Marín-Morales, J. L. Higuera-Trujillo, A. Greco, J. Guixeres, C. Llinares, E. P. Scilingo, M. Alcañiz, and G. Valenza, “Affective computing in virtual re-

- ality: emotion recognition from brain and heartbeat dynamics using wearable sensors,” *Scientific reports*, vol. 8, no. 1, pp. 1–15, 2018.
- [80] J. Marín-Morales, C. Llinares, J. Guixeres, and M. Alcañiz, “Emotion recognition in immersive virtual reality: From statistics to affective computing,” *Sensors*, vol. 20, no. 18, p. 5163, 2020.
- [81] L. Constantine and H. Hajj, “A survey of ground-truth in emotion data annotation,” in *2012 IEEE International Conference on Pervasive Computing and Communications Workshops*, 2012, pp. 697–702.
- [82] M. Blanco-Ruiz, C. Sainz-de Baranda, L. Gutiérrez-Martín, E. Romero-Perales, and C. López-Ongil, “Emotion elicitation under audiovisual stimuli reception: Should artificial intelligence consider the gender perspective?” *International Journal of Environmental Research and Public Health*, vol. 17, no. 22, 2020. [Online]. Available: <https://www.mdpi.com/1660-4601/17/22/8534>
- [83] J. L. Andreassi, *Psychophysiology: Human behavior and physiological response*. Psychology Press, 2010.
- [84] O. P. Keifer Jr, R. C. Hurt, K. J. Ressler, and P. J. Marvar, “The physiology of fear: reconceptualizing the role of the central amygdala in fear learning,” *Physiology*, vol. 30, no. 5, pp. 389–401, 2015.
- [85] S. D. Kreibig, “Autonomic nervous system activity in emotion: A review,” *Biological psychology*, vol. 84, no. 3, pp. 394–421, 2010.
- [86] L. F. Barrett, “Are emotions natural kinds?” *Perspectives on Psychological Science*, vol. 1, no. 1, pp. 28–58, 2006.
- [87] I. B. Mauss and M. D. Robinson, “Measures of emotion: A review,” *Cognition and Emotion*, vol. 23, no. 2, pp. 209–237, 2009.
- [88] G. Stemmler, “Physiological processes during emotion,” in *The regulation of emotion*. Psychology Press, 2004, pp. 48–85.
- [89] R. W. Levenson, “The autonomic nervous system and emotion,” *Emotion Review*, vol. 6, no. 2, pp. 100–112, 2014.
- [90] A. F. AX, “The physiological differentiation between fear and anger in humans,” *Psychosomatic Medicine*, vol. 15, no. 5, 1953.
- [91] O. Faust and M. G. Bairy, “Nonlinear analysis of physiological signals: a review,” *Journal of Mechanics in Medicine and Biology*, vol. 12, no. 04, p. 1240015, 2012.
- [92] S. Z. Spasić and S. Kesić, “Editorial: Nonlinearity in living systems: Theoretical and practical perspectives on metrics of physiological signal complexity,” *Frontiers in Physiology*, vol. 10, p. 298, 2019. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fphys.2019.00298>
- [93] Q. Yousef, M. Reaz, and M. A. M. Ali, “The analysis of ppg morphology: investigating the effects of aging on arterial compliance,” *Measurement Science Review*, vol. 12, no. 6, p. 266, 2012.
- [94] J. Fine, K. L. Branan, A. J. Rodriguez, T. Boonya-Ananta, J. C. Ramella-Roman, M. J. McShane, G. L. Côté *et al.*, “Sources of inaccuracy in photoplethysmography for continuous cardiovascular monitoring,” *Biosensors*, vol. 11, no. 4, p. 126, 2021.
- [95] G. Tusman, C. M. Acosta, S. Pulletz, S. H. Böhm, A. Scandurra, J. M. Arca, M. Madorno, and F. S. Sipmann, “Photoplethysmographic characterization of vascular tone mediated changes in arterial pressure: an observational study,”

- Journal of clinical monitoring and computing*, vol. 33, no. 5, pp. 815–824, 2019.
- [96] N. Hayashi, N. Someya, T. Maruyama, Y. Hirooka, M. Y. Endo, and Y. Fukuba, “Vascular responses to fear-induced stress in humans,” *Physiology & behavior*, vol. 98, no. 4, pp. 441–446, 2009.
- [97] P. Shi, V. Azorin-Peris, A. S. Echiadis, J. Zheng, Y. Zhu, P.-Y. Cheang, and S. Hu, “Non-contact reflection photoplethysmography towards effective human physiological monitoring,” *Journal of Medical and Biological Engineering*, 2010.
- [98] A. Alzahrani, S. Hu, V. Azorin-Peris, L. Barrett, D. Eslinger, M. Hayes, S. Akbare, J. Achart, and S. Kuoch, “A multi-channel opto-electronic sensor to accurately monitor heart rate against motion artefact during exercise,” *Sensors*, vol. 15, no. 10, pp. 25 681–25 702, 2015.
- [99] V. Rybynok and P. Kyriacou, “Beer-lambert law along non-linear mean light pathways for the rational analysis of photoplethysmography,” in *Journal of Physics: Conference Series*, vol. 238, no. 1. IOP Publishing, 2010, p. 012061.
- [100] T. Shimazaki, S. Hara, H. Okuhata, H. Nakamura, and T. Kawabata, “Cancellation of motion artifact induced by exercise for ppg-based heart rate sensing,” in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2014, pp. 3216–3219.
- [101] H. Nogami, W. Iwasaki, N. Morita, and R. Takigawa, “Relationship between ac/dc ratio and light-blocking structure for reflective photoplethysmographic sensor,” *Sensors and Materials*, vol. 30, no. 12, pp. 3021–3028, 2018.
- [102] V. Hartmann, H. Liu, F. Chen, Q. Qiu, S. Hughes, and D. Zheng, “Quantitative comparison of photoplethysmographic waveform characteristics: Effect of measurement site,” *Frontiers in Physiology*, vol. 10, p. 198, 2019.
- [103] B. Bent, B. A. Goldstein, W. A. Kibbe, and J. P. Dunn, “Investigating sources of inaccuracy in wearable optical heart rate sensors,” *NPJ digital medicine*, vol. 3, no. 1, pp. 1–9, 2020.
- [104] B. Baldaro, M. W. Battacchi, M. Codispoti, G. Tuoizzi, G. Trombini, R. Bolzani, and D. Palomba, “Modifications of electrogastrographic activity during the viewing of brief film sequences,” *Perceptual and motor skills*, vol. 82, no. 3_suppl, pp. 1243–1250, 1996.
- [105] U. Dimberg, “Facial reactions to fear-relevant and fear-irrelevant stimuli,” *Biological psychology*, vol. 23, no. 2, pp. 153–161, 1986.
- [106] B. L. Fredrickson and R. W. Levenson, “Positive emotions speed recovery from the cardiovascular sequelae of negative emotions,” *Cognition & emotion*, vol. 12, no. 2, pp. 191–220, 1998.
- [107] R. Gilissen, M. J. Bakermans-Kranenburg, M. H. van IJzendoorn, and R. van der Veer, “Parent–child relationship, temperament, and physiological reactions to fear-inducing film clips: Further evidence for differential susceptibility,” *Journal of Experimental Child Psychology*, vol. 99, no. 3, pp. 182–195, 2008.
- [108] J. A. Etzel, E. L. Johnsen, J. Dickerson, D. Tranel, and R. Adolphs, “Cardiovascular and respiratory responses during musical mood induction,” *International Journal of psychophysiology*, vol. 61, no. 1, pp. 57–69, 2006.
- [109] Y. Wu, R. Gu, Q. Yang, and Y.-j. Luo, “How do amusement, anger and fear influence heart rate and heart rate variability?” *Frontiers in Neuroscience*,

- vol. 13, p. 1131, 2019.
- [110] E.-H. Jang, S. Byun, M.-S. Park, and J.-H. Sohn, “Predicting individuals’ experienced fear from multimodal physiological responses to a fear-inducing stimulus,” *Advances in cognitive psychology*, vol. 16, no. 4, p. 291, 2020.
- [111] M. J. Christie, “Electrodermal activity in the 1980s: A review,” *Journal of the Royal Society of Medicine*, vol. 74, no. 8, pp. 616–622, 1981, PMID: 7288800. [Online]. Available: <https://doi.org/10.1177/014107688107400812>
- [112] W. Boucsein, *Electrodermal Activity*, ser. The Springer series in behavioral psychophysiology and medicine. Springer US, 2012. [Online]. Available: <https://books.google.es/books?id=6N6rnOEZEEoC>
- [113] M. Schmelz, R. Schmidt, A. Bickel, H. Torebjork, and H. Handwerker, “Innervation territories of single sympathetic c fibers in human skin,” *Journal of neurophysiology*, vol. 79, no. 4, pp. 1653–1660, 1998.
- [114] P. Ellaway, A. Kuppuswamy, A. Nicotra, and C. Mathias, “Sweat production and the sympathetic skin response: Improving the clinical assessment of autonomic function,” *Autonomic Neuroscience*, vol. 155, no. 1, pp. 109 – 114, 2010.
- [115] R. Edelberg, “Electrodermal mechanisms: A critique of the two-effector hypothesis and a proposed replacement,” in *Progress in electrodermal research*. Springer, 1993, pp. 7–29.
- [116] H. F. Posada-Quintero and K. H. Chon, “Innovations in electrodermal activity data collection and signal processing: A systematic review,” *Sensors*, vol. 20, no. 2, 2020. [Online]. Available: <https://www.mdpi.com/1424-8220/20/2/479>
- [117] W. Boucsein, *Methods of Electrodermal Recording*. Boston, MA: Springer US, 2012, pp. 87–258.
- [118] S. Grimnes, A. Jabbari, Ø. G. Martinsen, and C. Tronstad, “Electrodermal activity by dc potential and ac conductance measured simultaneously at the same skin site,” *Skin Research and Technology*, vol. 17, no. 1, pp. 26–34, 2011. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0846.2010.00459.x>
- [119] W. Boucsein, D. Fowles, S. Grimnes, G. Ben-Shakhar, W. Roth, M. Dawson, and D. Filion, “Publication recommendations for electrodermal measurements,” *Psychophysiology*, vol. 49, pp. 1017–34, 08 2012.
- [120] R. Edelberg, T. Greiner, and N. R. Burch, “Some membrane properties of the effector in the galvanic skin response,” *Journal of Applied Physiology*, vol. 15, no. 4, pp. 691–696, 1960, PMID: 13819259. [Online]. Available: <https://doi.org/10.1152/jappl.1960.15.4.691>
- [121] M.-Z. Poh, N. C. Swenson, and R. W. Picard, “A wearable sensor for unobtrusive, long-term assessment of electrodermal activity,” *IEEE transactions on Biomedical engineering*, vol. 57, no. 5, pp. 1243–1252, 2010.
- [122] J. Guerreiro, “A biosignal embedded system for physiological computing,” Ph.D. dissertation, Instituto Superior de Engenharia de Lisboa, 2013.
- [123] G. C. Pope and R. J. Halter, “Design and implementation of an ultra-low resource electrodermal activity sensor for wearable applications,” *Sensors*, vol. 19, no. 11, p. 2450, 2019.
- [124] R. Edelberg, “Electrical properties of the skin,” *Methods in psychophysiology*, 1967.
- [125] J. J. Braithwaite, D. G. Watson, R. Jones, and M. Rowe, “A guide for

- analysing electrodermal activity (eda) and skin conductance responses (scrs) for psychological experiments,” *Psychophysiology*, vol. 49, no. 1, pp. 1017–1034, 2013.
- [126] M. Benedek and C. Kaernbach, “Decomposition of skin conductance data by means of nonnegative deconvolution,” *psychophysiology*, vol. 47, no. 4, pp. 647–658, 2010.
- [127] 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.
- [128] F. Hernando-Gallego, D. Luengo, and A. Artés-Rodríguez, “Feature extraction of galvanic skin responses by nonnegative sparse deconvolution,” *IEEE journal of biomedical and health informatics*, vol. 22, no. 5, pp. 1385–1394, 2017.
- [129] D. M. Alexander, C. Trengove, P. Johnston, T. Cooper, J. August, and E. Gordon, “Separating individual skin conductance responses in a short interstimulus-interval paradigm,” *Journal of neuroscience methods*, vol. 146, no. 1, pp. 116–123, 2005.
- [130] C. A. Frantzidis, E. Konstantinidis, C. Pappas, and P. D. Bamidis, “An automated system for processing electrodermal activity.” *Studies in health technology and informatics*, vol. 150, pp. 787–787, 2009.
- [131] J. Choi, B. Ahmed, and R. Gutierrez-Osuna, “Development and evaluation of an ambulatory stress monitor based on wearable sensors,” *IEEE transactions on information technology in biomedicine*, vol. 16, no. 2, pp. 279–286, 2011.
- [132] M. P. Tarvainen, P. O. Ranta-Aho, and P. A. Karjalainen, “An advanced detrending method with application to hrv analysis,” *IEEE transactions on biomedical engineering*, vol. 49, no. 2, pp. 172–175, 2002.
- [133] A. Dzedzickis, A. Kaklauskas, and V. Bucinskas, “Human emotion recognition: Review of sensors and methods,” *Sensors*, vol. 20, no. 3, p. 592, 2020.
- [134] L. B. et. al., “Allostasis and allostatic load: Woman abuse and chronic illness - learning network - western university,” <http://www.vawlearningnetwork.ca/our-work/briefs/brief-13.html>, (Accessed on 08/20/2021).
- [135] V. Kosonogov, L. De Zorzi, J. Honoré, E. S. Martínez-Velázquez, J.-L. Nandrino, J. M. Martinez-Selva, and H. Sequeira, “Facial thermal variations: A new marker of emotional arousal,” *PLOS ONE*, vol. 12, no. 9, pp. 1–15, 09 2017. [Online]. Available: <https://doi.org/10.1371/journal.pone.0183592>
- [136] C. Goulart, C. Valadão, D. Delisle-Rodriguez, D. Tavares, E. Caldeira, and T. Bastos-Filho, “Emotional state analysis through infrared thermal imaging,” in *XXVI Brazilian Congress on Biomedical Engineering*, R. Costa-Felix, J. C. Machado, and A. V. Alvarenga, Eds. Singapore: Springer Singapore, 2019, pp. 199–203.
- [137] A. Kurz, “Physiology of thermoregulation,” *Best Practice and Research Clinical Anaesthesiology*, vol. 22, no. 4, pp. 627–644, 2008, thermoregulation in Anesthesia and Intensive Care Medicine. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1521689608000554>
- [138] S. Koelstra, C. Mühl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, “Deap: A database for emotion analysis ;using physiological signals,” *IEEE Transactions on Affective Computing*, vol. 3, pp. 18–31, 2012.
- [139] B. A. MacRae, S. Annaheim, C. M. Spengler, and R. M. Rossi, “Skin temper-

- ature measurement using contact thermometry: a systematic review of setup variables and their effects on measured values," *Frontiers in physiology*, vol. 9, p. 29, 2018.
- [140] G. Regalia, F. Onorati, M. Lai, C. Caborni, and R. W. Picard, "Multimodal wrist-worn devices for seizure detection and advancing research: focus on the empatica wristbands," *Epilepsy research*, vol. 153, pp. 79–82, 2019.
- [141] P. Ekman, R. W. Levenson, and W. V. Friesen, "Autonomic nervous system activity distinguishes among emotions," *science*, vol. 221, no. 4616, pp. 1208–1210, 1983.
- [142] R. W. Levenson, P. Ekman, and W. V. Friesen, "Voluntary facial action generates emotion-specific autonomic nervous system activity," *Psychophysiology*, vol. 27, no. 4, pp. 363–384, 1990.
- [143] C. Collet, E. Vernet-Maury, G. Delhomme, and A. Dittmar, "Autonomic nervous system response patterns specificity to basic emotions," *Journal of the autonomic nervous system*, vol. 62, no. 1-2, pp. 45–57, 1997.
- [144] J. Moltó, P. Segarra, R. López, À. Esteller, A. Fonfría, M. C. Pastor, and R. Poy, "Adaptación eapañola del " international affective picture system"(iaps). tercera parte." *Anales de Psicología/Annals of Psychology*, vol. 29, no. 3, pp. 965–984, 2013.
- [145] M. Khomami Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe, "Decaf: Meg-based multimodal database for decoding affective physiological responses", " *IEEE Transactions on Affective Computing*, vol. PP, p. 1, 01 2015.
- [146] S. Vadrevu and M. S. Manikandan, "Real-time ppg signal quality assessment system for improving battery life and false alarms," *IEEE transactions on circuits and systems II: express briefs*, vol. 66, no. 11, pp. 1910–1914, 2019.
- [147] D. Biswas, N. Simões-Capela, C. Van Hoof, and N. Van Helleputte, "Heart rate estimation from wrist-worn photoplethysmography: A review," *IEEE Sensors Journal*, vol. 19, no. 16, pp. 6560–6570, 2019.
- [148] M. Soleymani, F. Villaro-Dixon, T. Pun, and G. Chanel, "Toolbox for emotional feature extraction from physiological signals (teap)," *Frontiers in ICT*, vol. 4, p. 1, 2017.
- [149] P. van Gent, H. Farah, N. van Nes, and B. van Arem, "Heartpy: A novel heart rate algorithm for the analysis of noisy signals," *Transportation research part F: traffic psychology and behaviour*, vol. 66, pp. 368–378, 2019.
- [150] S. A. H. Aqajari, E. K. Naeini, M. A. Mehrabadi, S. Labbaf, N. Dutt, and A. M. Rahmani, "pyeda: An open-source python toolkit for pre-processing and feature extraction of electrodermal activity," *Procedia Computer Science*, vol. 184, pp. 99–106, 2021.
- [151] R. Parasuraman and Y. Jiang, "Individual differences in cognition, affect, and performance: behavioral, neuroimaging, and molecular genetic approaches," *NeuroImage*, vol. 59, no. 1, pp. 70–82, Jan 2012, 21569853[pmid].
- [152] P. Verduyn, E. Delvaux, H. Coillie, F. Tuerlinckx, and I. Mechelen, "Predicting the duration of emotional experience: Two experience sampling studies," *Emotion*, vol. 9, pp. 83–91, 03 2009.
- [153] M. M. Hassan, M. G. R. Alam, M. Z. Uddin, S. Huda, A. Almogren, and G. Fortino, "Human emotion recognition using deep belief network architecture," *Information Fusion*, vol. 51, pp. 10 – 18, 2019.

-
- [154] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu, and X. Yang, “A review of emotion recognition using physiological signals,” *Sensors*, vol. 18, p. 2074, 06 2018.
- [155] J. Zhang, Z. Yin, P. Chen, and S. Nichele, “Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review,” *Information Fusion*, vol. 59, pp. 103–126, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1566253519302532>
- [156] H. P. Martinez, Y. Bengio, and G. N. Yannakakis, “Learning deep physiological models of affect,” *IEEE Computational Intelligence Magazine*, vol. 8, no. 2, pp. 20–33, 2013.
- [157] P. Tzirakis, G. Trigeorgis, M. A. Nicolaou, B. W. Schuller, and S. Zafeiriou, “End-to-end multimodal emotion recognition using deep neural networks,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 8, pp. 1301–1309, 2017.
- [158] A. Bakhshi and S. Chalup, “Multimodal emotion recognition based on speech and physiological signals using deep neural networks,” in *Pattern Recognition. ICPR International Workshops and Challenges*, A. Del Bimbo, R. Cucchiara, S. Sclaroff, G. M. Farinella, T. Mei, M. Bertini, H. J. Escalante, and R. Vezani, Eds. Cham: Springer International Publishing, 2021, pp. 289–300.
- [159] J. Miranda, M. Canabal, L. Gutiérrez-Martín, J. M. Lanza-Gutiérrez, and C. López-Ongil, “A design space exploration for heart rate variability in a wearable smart device,” in *2020 XXXV Conference on Design of Circuits and Integrated Systems (DCIS)*, 2020, pp. 1–6.
- [160] J. Rubin, R. Abreu, S. Ahern, H. Eldardiry, and D. G. Bobrow, “Time, frequency & complexity analysis for recognizing panic states from physiologic time-series,” in *Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare*, ser. PervasiveHealth ’16. Brussels, BEL: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2016, p. 81–88.
- [161] G. Valenza, A. Lanata, and E. P. Scilingo, “The role of nonlinear dynamics in affective valence and arousal recognition,” *IEEE Transactions on Affective Computing*, vol. 3, no. 2, pp. 237–249, 2012.
- [162] E. Kanjo, E. M. Younis, and C. S. Ang, “Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection,” *Information Fusion*, vol. 49, pp. 46–56, 2019.
- [163] L. Santamaria-Granados, M. Munoz-Organero, G. Ramirez-Gonzalez, E. Abdulhay, and N. Arunkumar, “Using deep convolutional neural network for emotion detection on a physiological signals dataset (amigos),” *IEEE Access*, vol. 7, pp. 57–67, 2018.
- [164] F. Li, L. Yang, H. Shi, and C. Liu, “Differences in photoplethysmography morphological features and feature time series between two opposite emotions: Happiness and sadness,” *Artery Research*, vol. 18, pp. 7–13, 2017.
- [165] S. Kotsiantis, “Feature selection for machine learning classification problems: a recent overview,” *Artificial Intelligence Review*, vol. 42, no. 1, pp. 157–176, 2011.
- [166] J. Cai, J. Luo, S. Wang, and S. Yang, “Feature selection in machine learning: A new perspective,” *Neurocomputing*, vol. 300, pp. 70–79, 2018.
- [167] G. Valenza, A. Lanata, and E. P. Scilingo, “The role of nonlinear dynamics

- in affective valence and arousal recognition,” *IEEE transactions on affective computing*, vol. 3, no. 2, pp. 237–249, 2011.
- [168] S. Vijayakumar, R. Flynn, and N. Murray, “A comparative study of machine learning techniques for emotion recognition from peripheral physiological signals,” in *2020 31st Irish Signals and Systems Conference (ISSC)*. IEEE, 2020, pp. 1–6.
- [169] L. Van Der Maaten, E. Postma, J. Van den Herik *et al.*, “Dimensionality reduction: a comparative,” *J Mach Learn Res*, vol. 10, no. 66-71, p. 13, 2009.
- [170] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, “Algorithms for hyperparameter optimization,” in *Proceedings of the 24th International Conference on Neural Information Processing Systems*, ser. NIPS’11. Red Hook, NY, USA: Curran Associates Inc., 2011, p. 2546–2554.
- [171] T. Yu and H. Zhu, “Hyper-parameter optimization: A review of algorithms and applications,” *arXiv preprint arXiv:2003.05689*, 2020.
- [172] J. Bergstra and Y. Bengio, “Random search for hyper-parameter optimization.” *Journal of machine learning research*, vol. 13, no. 2, 2012.
- [173] R. Turner, D. Eriksson, M. McCourt, J. Kiili, E. Laaksonen, Z. Xu, and I. Guyon, “Bayesian optimization is superior to random search for machine learning hyperparameter tuning: Analysis of the black-box optimization challenge 2020,” *arXiv preprint arXiv:2104.10201*, 2021.
- [174] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, “A review of affective computing: From unimodal analysis to multimodal fusion,” *Information Fusion*, vol. 37, pp. 98–125, 2017.
- [175] P. C. Sen, M. Hajra, and M. Ghosh, “Supervised classification algorithms in machine learning: A survey and review,” in *Emerging technology in modelling and graphics*. Springer, 2020, pp. 99–111.
- [176] W. Wang and D. Sun, “The improved adaboost algorithms for imbalanced data classification,” *Information Sciences*, vol. 563, pp. 358–374, 2021.
- [177] R. W. Picard, E. Vyzas, and J. Healey, “Toward machine emotional intelligence: analysis of affective physiological state,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175–1191, 2001.
- [178] R. Subramanian, J. Wache, M. K. Abadi, R. L. Vieriu, S. Winkler, and N. Sebe, “Ascertain: Emotion and personality recognition using commercial sensors,” *IEEE Transactions on Affective Computing*, vol. 9, no. 2, pp. 147–160, 2018.
- [179] 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*, ser. ICMI ’18. New York, NY, USA: Association for Computing Machinery, 2018, p. 400–408.
- [180] J. A. Miranda, M. F. Canabal, J. M. Lanza-Gutiérrez, M. Portela-García, C. López-Ongil, and T. R. Alcaide, “Meaningful data treatment from multiple physiological sensors in a cyber-physical system,” in *DCIS 2017: XXXII Conference on Design of Circuits and Integrated Systems*, 2017, pp. 100–104.
- [181] J. A. Miranda, M. F. Canabal, J. M. Lanza-Gutiérrez, M. P. García, and C. López-Ongil, “Toward fear detection using affect recognition,” in *2019 XXXIV Conference on Design of Circuits and Integrated Systems (DCIS)*, 2019, pp. 1–4.

-
- [182] T. Christy, L. I. Kuncheva, and K. W. Williams, "Selection of physiological input modalities for emotion recognition," *UK: Bangor University*, 2012.
- [183] J. A. Miranda Calero, R. Marino, J. M. Lanza-Gutierrez, T. Riesgo, M. Garcia-Valderas, and C. Lopez-Ongil, "Embedded emotion recognition within cyber-physical systems using physiological signals," in *2018 Conference on Design of Circuits and Integrated Systems (DCIS)*, 2018, pp. 1–6.
- [184] O. Bălan, G. Moise, A. Moldoveanu, M. Leordeanu, and F. Moldoveanu, "Fear level classification based on emotional dimensions and machine learning techniques," *Sensors*, vol. 19, no. 7, 2019.
- [185] L. Petrescu, C. Petrescu, A. Oprea, O. Mitruț, G. Moise, A. Moldoveanu, and F. Moldoveanu, "Machine learning methods for fear classification based on physiological features," *Sensors*, vol. 21, no. 13, p. 4519, 2021.
- [186] J. A. Miranda, M. F. Canabal, L. Gutiérrez-Martín, J. M. Lanza-Gutierrez, M. Portela-García, and C. López-Ongil, "Fear recognition for women using a reduced set of physiological signals," *Sensors*, vol. 21, no. 5, p. 1587, 2021.
- [187] S. Mittal, "A survey of techniques for approximate computing," *ACM Computing Surveys (CSUR)*, vol. 48, no. 4, p. 62, 2016.
- [188] M. T. Valderas, J. Bolea, P. Laguna, R. Bailón, M. Vallverdú *et al.*, "Mutual information between heart rate variability and respiration for emotion characterization," *Physiological measurement*, vol. 40, no. 8, p. 084001, 2019.
- [189] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Physical review letters*, vol. 89, no. 6, p. 068102, 2002.
- [190] J. Kim and E. Andre, "Emotion-specific dichotomous classification and feature-level fusion of multichannel biosignals for automatic emotion recognition," in *2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2008, pp. 114–119.
- [191] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.
- [192] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [193] C. Godin, F. Prost-Boucle, A. Campagne, S. Charbonnier, S. Bonnet, and A. Vidal, "Selection of the most relevant physiological features for classifying emotion," *Emotion*, vol. 40, p. 20, 2015.
- [194] A. Albraikan, D. P. Tobón, and A. El Saddik, "Toward user-independent emotion recognition using physiological signals," *IEEE Sensors Journal*, vol. 19, no. 19, pp. 8402–8412, 2018.
- [195] J. Rottenberg, R. Ray, and J. Gross, "Emotion elicitation using films in: Coan ja, allen jjb, editors. the handbook of emotion elicitation and assessment," 2007.
- [196] W.-H. Lin, D. Wu, C. Li, H. Zhang, and Y.-T. Zhang, "Comparison of heart rate variability from ppg with that from ecg," in *The international conference on health informatics*. Springer, 2014, pp. 213–215.
- [197] G. Lu, F. Yang, J. Taylor, and J. Stein, "A comparison of photoplethysmography and ecg recording to analyse heart rate variability in healthy subjects,"

- Journal of medical engineering & technology*, vol. 33, no. 8, pp. 634–641, 2009.
- [198] M. Bolanos, H. Nazeran, and E. Haltiwanger, “Comparison of heart rate variability signal features derived from electrocardiography and photoplethysmography in healthy individuals,” in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, 2006, pp. 4289–4294.
- [199] M. Soleymani, J. Davis, and T. Pun, “A collaborative personalized affective video retrieval system,” in *2009 3rd international conference on affective computing and intelligent interaction and workshops*. IEEE, 2009, pp. 1–2.
- [200] M. E. Dawson, A. M. Schell, and D. L. Filion, *The Electrodermal System*, 4th ed., ser. Cambridge Handbooks in Psychology. Cambridge University Press, 2016, p. 217–243.
- [201] V. Shusterman, K. P. Anderson, and O. Barnea, “Spontaneous skin temperature oscillations in normal human subjects,” *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 273, no. 3, pp. R1173–R1181, 1997.
- [202] J. Pan and W. J. Tompkins, “A real-time qrs detection algorithm,” *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, 1985.
- [203] C.-K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger, “Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series,” *Chaos: an interdisciplinary journal of nonlinear science*, vol. 5, no. 1, pp. 82–87, 1995.
- [204] C. M. van-den Bleek and J. C. Schouten, “Deterministic chaos: a new tool in fluidized bed design and operation,” *The Chemical Engineering Journal and the Biochemical Engineering Journal*, vol. 53, no. 1, pp. 75 – 87, 1993.
- [205] M. M. Carl Rhodes, “The false nearest neighbors algorithm: An overview,” *Computers & Chemical Engineering*, vol. 21, pp. S1149 – S1154, 1997, supplement to Computers and Chemical Engineering.
- [206] S. Schinkel, O. Dimigen, and N. Marwan, “Selection of recurrence threshold for signal detection,” *The European Physical Journal Special Topics*, vol. 164, pp. 15–53, 10 2008.
- [207] S. J. Raudys and A. K. Jain, “Small sample size effects in statistical pattern recognition: Recommendations for practitioners,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 3, p. 252–264, mar 1991. [Online]. Available: <https://doi.org/10.1109/34.75512>
- [208] F. Nasoz, C. L. Lisetti, K. Alvarez, and N. Finkelstein, “Emotion recognition from physiological signals for user modeling of affect,” in *Proceedings of the 3rd Workshop on Affective and Attitude User Modelling (Pittsburgh, PA, USA)*, 2003.
- [209] G. Chanel, C. Rebetz, M. Bétrancourt, and T. Pun, “Emotion assessment from physiological signals for adaptation of game difficulty,” *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 41, no. 6, pp. 1052–1063, 2011.
- [210] P. Rathod, K. George, and N. Shinde, “Bio-signal based emotion detection device,” in *2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 2016, pp. 105–108.
- [211] B. Zhao, Z. Wang, Z. Yu, and B. Guo, “Emotionsense: Emotion recognition based on wearable wristband,” in *2018 IEEE SmartWorld, Ubiquitous Intel-*

- ligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 2018, pp. 346–355.
- [212] J. A. Miranda, A. P. Montoro, C. López-Ongil, and J. Andreu-Pérez, “Ft2f-sqa: Few-shot type-2 fuzzy-based subject-invariant ppg quality assessment for extreme edge physiological monitoring,” *IEEE TIM*, 2022, manuscript submitted for publication.
- [213] GSMA, “Connected women – the mobile gender gap report 2021,” <https://www.gsma.com/r/wp-content/uploads/2021/07/The-Mobile-Gender-Gap-Report-2021.pdf>, June 2021, (Accessed on 02/15/2022).
- [214] K. Eisenhut, E. Sauerborn, C. García-Moreno, and V. Wild, “Mobile applications addressing violence against women: a systematic review,” *BMJ global health*, vol. 5, no. 4, p. e001954, 2020.
- [215] A. J. Yugueros García, “Violencia de género, seguridad de las víctimas desde la perspectiva psicosocial,” *REVISTA DE GÉNERO E IGUALDAD*, 2021.
- [216] N. Karusala and N. Kumar, “Women’s safety in public spaces: Examining the efficacy of panic buttons in new delhi,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’17. New York, NY, USA: Association for Computing Machinery, 2017, p. 3340–3351.
- [217] M. N. Islam, N. T. Promi, J. M. Shaila, M. A. Toma, M. A. Pushpo, F. B. Alam, S. N. Khaledur, T. T. Anannya, and M. F. Rabbi, “Safeband: A wearable device for the safety of women in bangladesh,” in *Proceedings of the 16th International Conference on Advances in Mobile Computing and Multimedia*, 2018, pp. 76–83.
- [218] Á. González-Prieto, A. Brú, J. C. Nuño, and J. L. González-Álvarez, “Machine learning for risk assessment in gender-based crime,” *arXiv preprint arXiv:2106.11847*, 2021.
- [219] J. A. Miranda, M. F. Canabal, M. Portela García, and C. Lopez-Ongil, “Embedded emotion recognition: Autonomous multimodal affective internet of things,” in *Proceedings of the cyber-physical systems workshop*, vol. 2208, 2018, pp. 22–29.
- [220] E. Rituerto-González, J. A. Miranda, M. F. Canabal, J. M. Lanza-Gutiérrez, C. Peláez-Moreno, and C. López-Ongil, “A hybrid data fusion architecture for bindi: A wearable solution to combat gender-based violence,” in *Multimedia Communications, Services and Security*, A. Dziech, W. Mees, and A. Czyżewski, Eds. Cham: Springer International Publishing, 2020, pp. 223–237.
- [221] M. Felipe Canabal, “iGlove: Plataforma para el desarrollo de investigación sobre la detección de emociones,” in *Master Thesis*. University Carlos III de Madrid, 2019.
- [222] M. T. Quazi, “Human emotion recognition using smart sensors: a thesis submitted in fulfilment of the requirements for the degree of master of engineering in electronics and communication engineering, school of engineering and advanced technology, massey university, palmerston north, new zealand, february 2012,” Ph.D. dissertation, Massey University, 2012.
- [223] A. Ramirez Barcenas, “Configuración de monitor de batería y análisis de con-

- sumo en nodo inalámbrico para prevención de violencia sexual,” in *Bachelor Thesis*. University Carlos III de Madrid, 2018.
- [224] *nRF52832 Product Specification v1.8*, Nordic Semiconductors, 2021, rev. 1.8. [Online]. Available: https://infocenter.nordicsemi.com/pdf/nRF52832_PS_v1.8.pdf
- [225] *Grove - Vibration Motor User Manual*, Seed Studio, 2015, buzzer. [Online]. Available: https://www.mouser.es/datasheet/2/744/Seed_105020003-1217653.pdf
- [226] *Advance Battery Monitor IC*, Texas Instruments, 2003, sLUS465E. [Online]. Available: <https://www.ti.com/lit/ds/symlink/bq2019.pdf>
- [227] *Miniature Single Cell, Fully Integrated Li-Ion, Li-Polymer Charge Management Controller*, MicroChip, 2005, dS21984A. [Online]. Available: <https://cdn-shop.adafruit.com/datasheets/MCP73831.pdf>
- [228] M. van Dooren, J. H. Janssen *et al.*, “Emotional sweating across the body: Comparing 16 different skin conductance measurement locations,” *Physiology & behavior*, vol. 106, no. 2, pp. 298–304, 2012.
- [229] D. Castaneda, A. Esparza, M. Ghamari, C. Soltanpur, and H. Nazeran, “A review on wearable photoplethysmography sensors and their potential future applications in health care,” *International journal of biosensors & bioelectronics*, vol. 4, no. 4, p. 195, 2018.
- [230] G. de Cuerva Camacho, “Eliminación de artefactos de movimiento en señales fotoplethysmográficas para sistemas portables orientados a la detección de emociones,” in *Master Thesis*. University Carlos III de Madrid, 2019.
- [231] S. Beres and L. Hejjel, “The minimal sampling frequency of the photoplethysmogram for accurate pulse rate variability parameters in healthy volunteers,” *Biomedical Signal Processing and Control*, vol. 68, p. 102589, 2021.
- [232] A. Paez Montoro, “Optimización de la medida óptica de pulso cardíaco para su integración en el sistema de detección de emociones Bindi,” in *Bachelor Thesis*. University Carlos III de Madrid, 2019.
- [233] A. Baba and M. Burke, “Measurement of the electrical properties of ungelled ecg electrodes,” *International Journal of Biology and Biomedical Engineering*, vol. 2, 11 2007.
- [234] M. F. Canabal, J. A. Miranda, J. M. Lanza-Gutiérrez, A. I. Pérez Garcilópez, and C. López-Ongil, “Electrodermal activity smart sensor integration in a wearable affective computing system,” in *2020 XXXV Conference on Design of Circuits and Integrated Systems (DCIS)*, 2020, pp. 1–6.
- [235] M. F. Canabal, J. A. Miranda, A. P. Montoro, I. P. Garcilópez, S. P. Álvarez, E. G. Ares, and C. López-Ongil, “Design and validation of an efficient and adjustable gsr sensor for emotion monitoring,” *IEEE Sensors*, 2022, manuscript in progress.
- [236] *Human Body Temperature Sensor*, Maxim Integrated, 2016, sKT. [Online]. Available: <https://datasheets.maximintegrated.com/en/ds/MAX30205.pdf>
- [237] *±0.1°C Accurate, I2C Digital Temperature Sensor*, Maxim Integrated, 2020, sKT. [Online]. Available: <https://datasheets.maximintegrated.com/en/ds/MAX30208.pdf>
- [238] *MAX30208 Evaluation System*, Maxim Integrated, 2019, sKT. [Online]. Available: <https://datasheets.maximintegrated.com/en/ds/MAX30208EVSYSPDF.pdf>
- [239] *S132 Softdevice Specification*, Nordic Semiconductors, 2019, s132. [Online].

- Available: https://infocenter.nordicsemi.com/pdf/S132_SDS_v7.1.pdf
- [240] A. Ramírez-Bárceñas, M. Portela-García, M. García-Valderas, and C. López-Ongil, “System dependability in edge computing wearable devices,” in *2020 XXXV Conference on Design of Circuits and Integrated Systems (DCIS)*, 2020, pp. 1–6.
 - [241] C. Orphanidou, *Signal quality assessment in physiological monitoring: state of the art and practical considerations*. Springer, 2017.
 - [242] M. Elgendi, “Optimal signal quality index for photoplethysmogram signals,” *Bioengineering*, vol. 3, no. 4, 2016.
 - [243] R. Krishnan *et al.*, “Two-stage approach for detection and reduction of motion artifacts in photoplethysmographic data,” *IEEE Trans. on biomedical engineering*, vol. 57, no. 8, pp. 1867–1876, 2010.
 - [244] E. K. Naeini *et al.*, “A real-time ppg quality assessment approach for healthcare internet-of-things,” *Procedia Computer Science*, vol. 151, pp. 551–558, 2019.
 - [245] S. Vadrevu *et al.*, “Real-time ppg signal quality assessment system for improving battery life and false alarms,” *IEEE TCAS2*, vol. 66, no. 11, pp. 1910–1914, 2019.
 - [246] G. Narendra Kumar Reddy *et al.*, “On-device integrated ppg quality assessment and sensor disconnection/saturation detection system for iot health monitoring,” *IEEE TIM*, vol. 69, no. 9, pp. 6351–6361, 2020.
 - [247] S. Alam *et al.*, “On-board signal quality assessment guided compression of photoplethysmogram for personal health monitoring,” *IEEE TIM*, vol. 70, pp. 1–9, 2021.
 - [248] Z. Zhao *et al.*, “Sqi quality evaluation mechanism of single-lead ecg signal based on simple heuristic fusion and fuzzy comprehensive evaluation,” *Frontiers in Physiology*, vol. 9, p. 727, 2018.
 - [249] J. M. Mendel, “Uncertain rule-based fuzzy systems,” *Introduction and new directions*, p. 684, 2017.
 - [250] A. Mueen *et al.*, “The fastest similarity search algorithm for time series subsequences under euclidean distance,” 2017.
 - [251] T. Nakamura *et al.*, “Merlin: Parameter-free discovery of arbitrary length anomalies in massive time series archives,” in *2020 IEEE International Conference on Data Mining (ICDM)*, 2020, pp. 1190–1195.
 - [252] Y. Zhu *et al.*, “Matrix profile xi: Scrimp++: Time series motif discovery at interactive speeds,” in *2018 IEEE International Conference on Data Mining (ICDM)*, 2018, pp. 837–846.
 - [253] S. Lloyd, “Least squares quantization in pcm,” *IEEE Trans. on information theory*, vol. 28, no. 2, pp. 129–137, 1982.
 - [254] F. Liu *et al.*, “Encoding words into interval type-2 fuzzy sets using an interval approach,” *IEEE Trans. on Fuzzy Systems*, vol. 16, no. 6, pp. 1503–1521, 2008.
 - [255] J. Andreu-Perez *et al.*, “Explainable artificial intelligence based analysis for interpreting infant fnirs data in developmental cognitive neuroscience,” *Communications biology*, vol. 4, no. 1, pp. 1–13, 2021.
 - [256] M. Antonelli *et al.*, “Multiobjective evolutionary optimization of type-2 fuzzy rule-based systems for financial data classification,” *IEEE Trans. on Fuzzy Systems*, vol. 25, no. 2, pp. 249–264, 2017.
 - [257] H. Hagrais *et al.*, “An incremental adaptive life long learning approach for type-

- 2 fuzzy embedded agents in ambient intelligent environments,” *IEEE Trans. on Fuzzy Systems*, vol. 15, no. 1, pp. 41–55, 2007.
- [258] H. T. Nguyen *et al.*, “Computing degrees of subsethood and similarity for interval-valued fuzzy sets: Fast algorithms,” in *9th International Conference on Intelligent Technologies*, 2008.
- [259] W. Karlen *et al.*, “Capnabase: Signal database and tools to collect, share and annotate respiratory signals,” in *2010 Annual Meeting of the Society for Technology in Anesthesia*, 2010, p. 27.
- [260] W. Karlen, J. M. Ansermino, and G. Dumont, “Adaptive pulse segmentation and artifact detection in photoplethysmography for mobile applications,” in *2012 Annual Conference of the IEEE EMBS*, 2012, pp. 3131–3134.
- [261] W. Karlen, “Csl pulse oximetry artifact labels,” 2021.
- [262] F. Li, L. Yang, H. Shi, and C. Liu, “Differences in photoplethysmography morphological features and feature time series between two opposite emotions: Happiness and sadness,” *Artery Research*, vol. 18, pp. 7–13, 2017. [Online]. Available: <https://doi.org/10.1016/j.artres.2017.02.003>
- [263] Q. Yousef, M. B. I. Reaz, and M. Ali, “The analysis of ppg morphology: Investigating the effects of aging on arterial compliance,” *Measurement Science Review*, vol. 12, pp. 266–271, 12 2012.
- [264] G. McVeigh, C. Bratteli, D. Morgan, C. Alinder, S. Glasser, S. Finkelstein, and J. Cohn, “Age-related abnormalities in arterial compliance identified by pressure pulse contour analysis: Aging and arterial compliance,” *Hypertension*, vol. 33, no. 6, pp. 1392–1398, 1999, cited By 286.
- [265] F. Foroozan, M. Mohan, and J. S. Wu, “Robust beat-to-beat detection algorithm for pulse rate variability analysis from wrist photoplethysmography signals,” in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 2136–2140.
- [266] M. Soundararajan, S. Arunagiri, and S. Alagala, “An adaptive delineator for photoplethysmography waveforms,” *Biomedical Engineering / Biomedizinische Technik*, vol. 61, pp. 645 – 655, 2016.
- [267] H. S. Shin, C. Lee, and M. Lee, “Adaptive threshold method for the peak detection of photoplethysmographic waveform,” *Computers in Biology and Medicine*, vol. 39, no. 12, pp. 1145–1152, 2009. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0010482509001826>
- [268] D. Morelli, A. Rossi, M. Cairo, and D. A. Clifton, “Analysis of the impact of interpolation methods of missing rr-intervals caused by motion artifacts on hrv features estimations,” *Sensors*, vol. 19, no. 14, p. 3163, 2019.
- [269] Z. Zhang *et al.*, “Troika: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise,” *IEEE Trans. on biomedical engineering*, vol. 62, no. 2, pp. 522–531, 2014.
- [270] D. Biswas *et al.*, “Heart rate estimation from wrist-worn photoplethysmography: A review,” *IEEE Sensors Journal*, vol. 19, no. 16, pp. 6560–6570, 2019.
- [271] U. Satija *et al.*, “A review of signal processing techniques for electrocardiogram signal quality assessment,” *IEEE Reviews in Biomedical Engineering*, vol. 11, pp. 36–52, 2018.
- [272] M. Á. Blanco Ruiz, L. Gutiérrez Martín, J. Á. Miranda Calero, M. F. Canabal Benito, E. Romero Perales, C. Sainz de Baranda Andujar, R. San Segundo Manuel, D. Larrabeiti López, C. Peláez-Moreno, and C. López Ongil.

- (2021) UC3M4Safety Database - List of Audiovisual Stimuli Annotations. [Online]. Available: <https://doi.org/10.21950/CXAAHR>
- [273] M. Á. Blanco Ruiz *et al.* (2021) UC3M4Safety Database - List of Audiovisual Stimuli (Video). [Online]. Available: <https://doi.org/10.21950/LUO1IZ>
- [274] C. S. de Baranda Andújar, M. B. Ruiz, J. Á. M. Calero, L. G. Martín, M. F. C. Benito, R. San Segundo, and C. L. Ongil, “Perspectiva de género y social en las stem: La construcción de sistemas inteligentes para detección de emociones,” *Sociología y tecnociencia: Revista digital de sociología del sistema tecnocientífico*, vol. 11, no. 1, pp. 83–115, 2021.
- [275] L. Gutiérrez Martín, “Gender Perspective in Emotion Elicitation under Audiovisual Stimuli Reception: Fear and Panic Emotions are gender related?” in *Master Thesis*. University Carlos III de Madrid, 2020.
- [276] L. Velasco Gonzalez, “Diseño e implementación de un sistema de recuperación fisiológica para experimentos de reconocimiento de emociones,” in *Bachelor Thesis*. University Carlos III de Madrid, 2021.
- [277] T. Kostoulas, G. Chanel, M. Muszynski, P. Lombardo, and T. Pun, “Dynamic time warping of multimodal signals for detecting highlights in movies,” in *Proceedings of the 1st Workshop on Modeling INTERPERSONAL Synchrony And influence*, 2015, pp. 35–40.
- [278] S. Gashi, E. Di Lascio, and S. Santini, “Using students’ physiological synchrony to quantify the classroom emotional climate,” in *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, 2018, pp. 698–701.
- [279] K. R. Papakannu, “Examining user engagement via facial expressions in augmented reality with dynamic time warping,” Ph.D. dissertation, Arizona State University, 2021.
- [280] R. Satti, N.-U.-H. Abid, M. Bottaro, M. De Rui, M. Garrido, M. R. Raoufy, S. Montagnese, and A. R. Mani, “The application of the extended poincaré plot in the analysis of physiological variabilities,” *Frontiers in Physiology*, vol. 10, 2019. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fphys.2019.00116>
- [281] N. Bu, “Poincaré analysis based on short-term heart rate variability data for stress evaluation,” in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 1–4.
- [282] P. Shi, S. Hu, and H. Yu, “Recovery of heart rate variability after treadmill exercise analyzed by lagged poincaré plot and spectral characteristics,” *Medical & biological engineering & computing*, vol. 56, no. 2, pp. 221–231, 2018.
- [283] F. Shaffer and J. P. Ginsberg, “An overview of heart rate variability metrics and norms,” *Frontiers in public health*, p. 258, 2017.
- [284] J. K. Kim and J. M. Ahn, “New marker for vascular health based on the poincare plot analysis using acceleration plethysmogram,” *International Journal of Applied Engineering Research*, vol. 13, no. 21, pp. 15 417–15 423, 2018.
- [285] M. Toichi, T. Sugiura, T. Murai, and A. Sengoku, “A new method of assessing cardiac autonomic function and its comparison with spectral analysis and coefficient of variation of r–r interval,” *Journal of the autonomic nervous system*, vol. 62, no. 1-2, pp. 79–84, 1997.
- [286] J. Jeppesen, S. Beniczky, P. Johansen, P. Sidenius, and A. Fuglsang-

- Frederiksen, “Using lorenz plot and cardiac sympathetic index of heart rate variability for detecting seizures for patients with epilepsy,” in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2014, pp. 4563–4566.
- [287] A. Aranzana Sánchez, “Implementación de Técnicas de Extracción de Características para el Reconocimiento de Emociones Usando Sensores Fotopletismográficos,” in *Bachelor Thesis*. University Carlos III de Madrid, 2020.
- [288] W. Allay Bakhtaoui, “Implementación de Técnicas de Extracción de Características para el Reconocimiento de Emociones Usando Sensores para la Conductividad de la Piel,” in *Bachelor Thesis*. University Carlos III de Madrid, 2020.
- [289] F. Adrián Hernández Gant, “Diseño de modelos de aprendizaje para detección de miedo en Bindi,” in *Master Thesis*. University Carlos III de Madrid, 2021.
- [290] H. S. Shin, C. Lee, and M. Lee, “Adaptive threshold method for the peak detection of photoplethysmographic waveform,” *Computers in biology and medicine*, vol. 39, no. 12, pp. 1145–1152, 2009.
- [291] S. Kuntamalla and L. R. G. Reddy, “An efficient and automatic systolic peak detection algorithm for photoplethysmographic signals,” *International Journal of Computer Applications*, vol. 97, no. 19, 2014.
- [292] J. T. VanderPlas, “Understanding the lomb–scargle periodogram,” *The Astrophysical Journal Supplement Series*, vol. 236, no. 1, p. 16, 2018.
- [293] O. De Wel, M. Lavanga, A. C. Dorado, K. Jansen, A. Dereymaeker, G. Naulaers, and S. Van Huffel, “Complexity analysis of neonatal eeg using multiscale entropy: applications in brain maturation and sleep stage classification,” *Entropy*, vol. 19, no. 10, p. 516, 2017.
- [294] S. Arunachalam, S. Kapa, S. Mulpuru, P. Friedman, and E. Tolkacheva, “Improved multiscale entropy technique with nearest-neighbor moving-average kernel for nonlinear and nonstationary short-time biomedical signal analysis,” *Journal of healthcare engineering*, vol. 2018, 2018.
- [295] C. Guanghui and Z. Xiaoping, “Multi-modal emotion recognition by fusing correlation features of speech-visual,” *IEEE Signal Processing Letters*, vol. 28, pp. 533–537, 2021.
- [296] H. Pérez-Espinosa, R. Zatarain-Cabada, and M. L. Barrón-Estrada, “Emotion recognition: from speech and facial expressions,” in *Biosignal Processing and Classification Using Computational Learning and Intelligence*. Elsevier, 2022, pp. 307–326.
- [297] Y. Huang, J. Yang, S. Liu, and J. Pan, “Combining facial expressions and electroencephalography to enhance emotion recognition,” *Future Internet*, vol. 11, no. 5, p. 105, 2019.
- [298] A. Muaremi, B. Arnrich, and G. Tröster, “Towards measuring stress with smartphones and wearable devices during workday and sleep,” *Bio-NanoScience*, vol. 3, no. 2, pp. 172–183, 2013.
- [299] A. Exler, A. Schankin, C. Klebsattel, and M. Beigl, “A wearable system for mood assessment considering smartphone features and data from mobile eegs,” in *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: Adjunct*, 2016, pp. 1153–1161.
- [300] J. Kim and E. André, “Emotion recognition using physiological and speech signal in short-term observation,” in *International Tutorial and Research Work-*

- shop on Perception and Interactive Technologies for Speech-Based Systems*. Springer, 2006, pp. 53–64.
- [301] Y. Huang, J. Yang, S. Liu, and J. Pan, “Combining facial expressions and electroencephalography to enhance emotion recognition,” *Future Internet*, vol. 11, no. 5, p. 105, 2019.
- [302] Y. Cimtay, E. Ekmekcioglu, and S. Caglar-Ozhan, “Cross-subject multimodal emotion recognition based on hybrid fusion,” *IEEE Access*, vol. 8, pp. 168 865–168 878, 2020.

