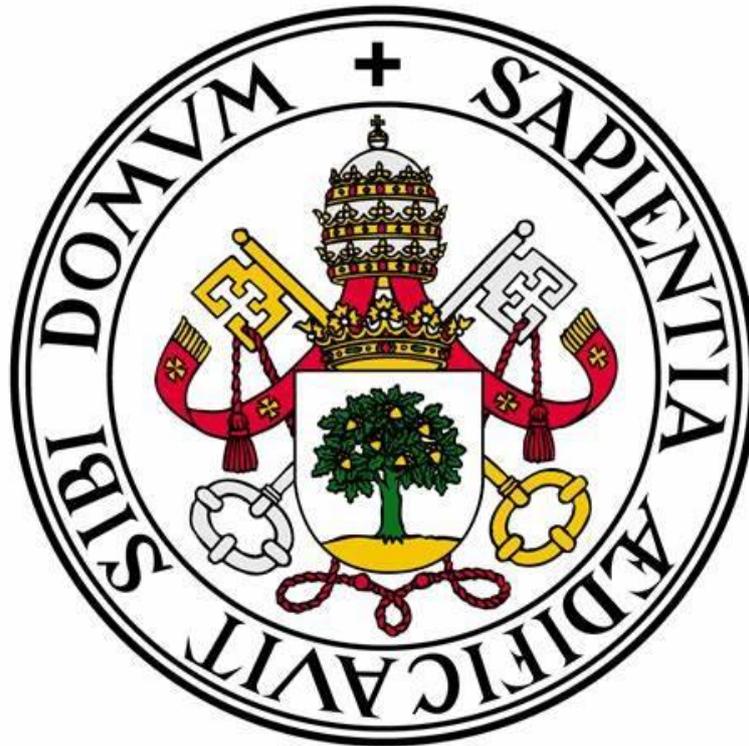


UNIVERSIDAD DE VALLADOLID

FACULTAD DE MEDICINA

ESCUELA DE INGENIERÍAS INDUSTRIALES



TRABAJO DE FIN DE GRADO DE INGENIERÍA BIOMÉDICA

2024

Estimación del estado emocional a través de datos fisiológicos con técnicas de machine learning

Autor: Diego García Muñoz

Tutor: Javier Pérez Turiel

Tutora: Ana Cisnal de la Rica

*A mi familia por apoyarme siempre,
con sus preguntas incrédulas a cerca
del grado
A mis amigos que estuvieron desde
el principio
Y a las infinitas horas de biblioteca.*

RESUMEN

En este trabajo se explora la posibilidad de predecir el estado de ánimo de una persona que está realizando una tarea específica. El método para lograrlo es el desarrollo de un modelo de inteligencia artificial que, utilizando la información proveniente de los órganos periféricos (corazón, músculos, piel, etc.) sea capaz de predecir el estado emocional del sujeto.

El aprendizaje automático, es una disciplina de la inteligencia artificial, que se basa en la extracción de características, que resuman las señales fisiológicas, para entrenar modelos capaces de predecir distintas emociones con un alto grado de precisión. Así pues, el objetivo principal que se persigue es la creación y adaptación de un modelo predictivo completo, desde la selección de la base de datos hasta los resultados del conjunto de validación, que ayude a comprender los mecanismos subyacentes a la detección automática de emociones, a partir de datos electrofisiológicos, como las señales de electrocardiograma (ECG), respiración (RESP), sudoración de la piel (EDA), etc. pues ofrecen una ventana hacia la respuesta emocional del cuerpo.

ABSTRACT

This work explores the possibility of predicting how a person feels while performing a specific task. The method to achieve this is the development of an artificial intelligent model, which uses information from peripheral organs (heart, muscles, skin, etc.) to predict the subject's emotional state.

Machine learning, a discipline of artificial intelligence, is based on extracting features, which that summarize physiological signals to train models capable of predicting different emotions with high degree of accuracy. Therefore, the main objective is to create, adapt and explain a complete predictive model, from the database selection to the validation test, which helps understand the mechanisms behind the artificial intelligence for emotion detection, using electrophysiological data, such as ECG, RESP or EDA signals, as these offer insight into the body's emotional response.

ÍNDICE GENERAL

RESUMEN	ii
ABSTRACT	ii
ÍNDICE GENERAL	iii
ÍNDICE DE FIGURAS	v
ÍNDICE DE TABLAS	vi
1 INTRODUCCIÓN Y CONOCIMIENTOS PREVIOS	1
1.1 INTRODUCCIÓN	1
1.2 CONOCIMIENTOS PREVIOS	2
1.2.1 Estados emocionales	2
1.2.1.1 Clasificación de los Estados Emocionales	2
1.2.2 Señales fisiológicas	6
1.2.2.1 Señal del Corazón	7
1.2.2.2 Sudoración de la piel	8
1.2.2.3 Respiración	9
1.2.2.4 Temperatura de la piel	10
1.2.2.5 Señal del cerebro	10
1.2.2.6 Señales faciales	11
1.2.2.7 Señal de los músculos	12
1.2.3 Bases de datos	12
1.2.3.1 DEAP dataset	13
1.2.3.2 MAHNOB-HCI	14
1.2.3.3 AMIGOSdataset	15
1.2.4 Algoritmos predictivos	16
1.2.4.1 Support Vector Machine (SVM)	16
1.2.4.2 K-Nearest Neighbors (KNN)	17
1.2.4.3 Linear Discriminant Analysis (LDA)	17
1.2.4.4 Random Forest (RF)	18
1.3 DESCRIPCIÓN DEL PROBLEMA	19
2 HIPÓTESIS Y OBJETIVOS	20
2.1 HIPÓTESIS	20
2.2 OBJETIVOS	20
2.2.1 Objetivo principal	20
2.2.2 Objetivos específicos	20
3 MATERIALES Y MÉTODOS	21
3.1 BASE DE DATOS	21
3.2 SEÑALES FISIOLÓGICAS	21
3.2.1 Señal de ECG	21
3.2.2 Señal de EDA	23
3.2.3 Señal de RESP	24
3.2.4 Señal de SKT	25
3.3 EXTRACCIÓN DE CARACTERÍSTICAS (<i>FEATURE EXTRACTION</i>)	26

3.3.1	Características de la señal de ECG	27
3.3.1.1	Características estadísticas	27
3.3.1.2	Características Temporales	28
3.3.1.3	Características Espectrales	29
3.3.1.4	Características no Lineales	31
3.3.2	Características de la señal de EDA	33
3.3.2.1	Características de linealidad	34
3.3.2.2	Características Estadísticas	34
3.3.2.3	Características temporales	35
3.3.2.4	Características espectrales	35
3.3.3	Características de la señal de RESP	38
3.3.3.1	Características estadísticas	38
3.3.3.2	Características temporales	39
3.3.3.3	Características espectrales	40
3.3.4	Características de la señal de SKT	43
3.3.4.1	Características estadísticas	43
3.3.4.2	Características espectrales	43
3.3.5	Extracción de características	44
3.4	SELECCIÓN DE CARACTERÍSTICAS (<i>FEATURE SELECTION</i>)	45
3.4.1	Procesamiento previo a la selección de características	46
3.4.2	Selección de características	46
3.5	TRANSFORMACIÓN DE LOS DATOS (<i>DATA TRANSFORM</i>)	48
3.6	MODELOS PREDICTIVOS	48
3.6.1	SVM	49
3.6.2	KNN	50
3.6.3	PLS (Partial Least Squares)	50
3.6.4	RF	50
3.7	ENTRENAMIENTO Y VALIDACIÓN	51
4	RESULTADOS	53
5	DISCUSIÓN	56
5.1	ANÁLISIS DE RESULTADOS	56
5.1.1	SVR	56
5.1.2	PLS	57
5.1.3	KNN	57
5.1.4	RF	58
5.2	COMPARATIVA ENTRE ESTE Y OTROS TRABAJOS	58
5.3	LÍNEAS FUTURAS	60
6	CONCLUSIÓN	62
	GLOSARIO DE ABREVIATURAS Y ACRÓNIMOS	63
	REFERENCIAS	¡Error! Marcador no definido.

ÍNDICE DE FIGURAS

1	INTRODUCCIÓN Y CONOCIMIENTOS PREVIOS	1
	Figura 1. Modelo proyectado de Plutchik	5
	Figura 2. Modelo de Plutchik con forma de cono invertido [8]	5
	Figura 3. Disposición de las emociones en los ejes de coordenadas.	6
	Figura 4. Señal de BVP	8
	Figura 5. Señal de ECG	8
	Figura 6. Señal de EDA	9
	Figura 7. Bandas de frecuencia delta, theta, alpha y beta (de arriba abajo) [23].	11
	Figura 8. Ejemplos de KNN	17
3	MATERIALES Y MÉTODOS	21
	Figura 9. Señal de ECG original, primeras 2000 muestras (arriba) y filtrada (abajo)	22
	Figura 10. Salida de la función de	23
	Figura 11. Salida de la función de	24
	Figura 12. Salida de la función de	25
	Figura 14. Señal de HRV	30
	Figura 15. FFT de la señal de HRV	30
	Figura 16. FFT de la señal de EDR	36
	Figura 17. FFT de la señal de Resp	40
	Figura 18. FFT de la señal de SKT	44
4	RESULTADOS	53
	Figura 19. Boxplot de MSE	54
	Figura 20. Boxplot del MAE.	54

ÍNDICE DE TABLAS

1 INTRODUCCIÓN Y CONOCIMIENTOS PREVIOS	1
Tabla 1. Características que permiten distinguir las emociones	3
3 MATERIALES Y MÉTODOS	21
Tabla 2. Disposición de las señales fisiológicas en la base de datos	21
Tabla 3. Características extraídas de la señal de ECG con su abreviatura asignada y la señal de la que se extrae.	33
Tabla 4. Características extraídas de la señal de EDA con su abreviatura asignada y la señal de la que se extrae.	38
Tabla 5. Características extraídas de la señal de RESP con su abreviatura asignada y la señal de la que se extrae.	42
Tabla 6. Características extraídas de la señal de SKT con su abreviatura asignada y la señal de la que se extrae.	44
Tabla 7. Mejores características por señal para valence, según su valor de MI.	47
Tabla 8. Mejores características por señal para arousal según su valor de MI.	47
4 RESULTADOS	53
Tabla 9. Mejores características según su valor de MI para valence.	53
Tabla 10. Mejores características según su valor de MI para arousal.	53
Tabla 11. Resultados de entrenamiento.	53
Tabla 12. Métricas de rendimiento de los diferentes modelos	54
5 DISCUSIÓN	56
Tabla 13. Resultados de entrenamiento y validación para SVR.	56
Tabla 14. Resultados de entrenamiento y validación para PLS.	57
Tabla 15. Resultados de entrenamiento y validación para KNN.	57
Tabla 16. Resultados de entrenamiento y validación para RF.	58
Tabla 17. Otros trabajos con DEAP dataset.	59
Tabla 18. Resultados de accuracy de otros trabajos y de este.	60

1 INTRODUCCIÓN Y CONOCIMIENTOS PREVIOS

1.1 INTRODUCCIÓN

En 1997 Rosalind Picard, ingeniera informática del *Massachusetts Institute of Technology* (MIT), publica el libro “*Affective Computing*” [1], en el que se expone, por primera vez, la capacidad de los ordenadores para poder predecir las emociones humanas. A partir de este hito, se abre una nueva rama en el campo de la inteligencia artificial, conocida como *Computación Afectiva*.

La inteligencia artificial (IA) es, según la RAE, “una disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico” [2]. Es decir, estos algoritmos se valen de la emulación de estos comportamientos para solucionar problemas complejos, aprendiendo de los datos y obteniendo resultados cada vez más precisos. La IA engloba tanto el aprendizaje automático (ML, *machine learning*), como el aprendizaje profundo (DL, *deep learning*), aunque este último es en realidad un subcampo dentro de ML. Ambos funcionan de una manera similar; extraen una serie de características de un conjunto de datos y llevan a cabo una fase de entrenamiento para poder realizar predicciones acertadas. La diferencia fundamental entre ambos reside en la automatización del primer paso, en el caso del DL, frente a la necesidad de la intervención humana en ML. Otro punto relevante de la IA es la necesidad o no de etiquetar los datos, esto es, si es necesario asociarlos explícitamente con el grupo al que pertenecen. En este caso, se distingue entre aprendizaje supervisado, con datos etiquetados y no supervisado. Las aplicaciones de la IA son bastante amplias; sistemas de reconocimiento de voz, predicción meteorológica, sistemas de conducción automática, diagnóstico de enfermedades y detección del estado emocional [3].

El estado emocional representa un mecanismo de adaptación del ser humano ante un estímulo importante. Las emociones como la alegría, la tristeza, la ira o el miedo son reacciones naturales que surgen como respuesta ante aquellas situaciones concretas en las que se perciben experiencias sensoriales que son capaces de provocar en las personas una respuesta neurofisiológica consciente.

Por su parte, las señales electrofisiológicas resultan de la medición, generalmente mediante electrodos, de la actividad eléctrica de los órganos periféricos, aquellos no localizados en el sistema nervioso central. La variación del estado emocional tiene la capacidad de influir sobre el funcionamiento de estos órganos, por ejemplo; el miedo puede disminuir la temperatura superficial de la piel o la euforia aumentar la frecuencia cardíaca. Estas señales deberán seguir un proceso de digitalización, remuestreo y procesamiento para poder ser convenientemente empleadas en la detección del estado emocional.

La IA, los estados emocionales y las señales electrofisiológicas constituyen los 3 pilares fundamentales sobre los que se sustenta el trabajo desarrollado. En el contexto de la ingeniería de rehabilitación, dónde se ha demostrado que combinar la terapia con algún tipo de entretenimiento activo mejora los resultados obtenidos, resulta práctico tener una herramienta para valorar objetivamente como se siente el paciente respecto a dicha terapia. De esta forma, un equipo que, mediante un sistema de monitoreo cómodo y no invasivo, pueda recopilar datos del sujeto, para que después un modelo predictivo pueda aproximar cuál es su estado emocional,

es un buen método para controlar y ajustar la dificultad del proceso terapéutico. Existen múltiples formas de desarrollar este modelo predictivo, a parte de las señales fisiológicas, por ejemplo; utilizar la señal de electroencefalograma (EEG) o grabar a los pacientes en vídeo y extraer características de este. La ventaja que ofrecen los datos electrofisiológicos es la sencillez. Es decir, resulta más fácil procesar y extraer características de señales como el electrocardiograma (ECG), la temperatura e impedancia de la piel o la respiración que del EEG o de una grabación en directo. También, resulta más cómodo para el paciente la colocación de unos pocos electrodos en el cuerpo que todo el sistema de registro que acompaña al EEG. Por estos motivos, entrenar un algoritmo de ML con señales fisiológicas para detectar el estado emocional de pacientes es un trabajo que merece la pena llevar a cabo.

1.2 CONOCIMIENTOS PREVIOS

1.2.1 Estados emocionales

En [1] se definen los estados emocionales como una dinámica interna del organismo cuando el individuo experimenta una emoción, produciéndose una serie de cambios no solo mentales sino también fisiológicos. El hecho de que se manifiesten distintos estados emocionales depende de varios factores y actividades.

Los estados emocionales se originan como consecuencia del procesamiento de diferentes estímulos y la posterior respuesta del cerebro ante estos. No todo el órgano se ve implicado en este proceso, sino que hay zonas concretas, que forman circuitos bien definidos, que se encargan de estas tareas. Por ejemplo, el circuito de Papez, descrito en 1937 por el neuroanatomista James Papez, está constituido por diferentes estructuras, entre las que destacan el hipotálamo, el hipocampo o el cíngulo. Actualmente, este modelo se considera incompleto pues se han descubierto nuevas zonas implicadas en procesos emocionales, por ejemplo, la amígdala que es la encargada de analizar el significado de los estímulos [4].

Como se ha mencionado anteriormente, los estados emocionales no se desarrollan exclusivamente en el cerebro, sino que también tienen implicaciones en otros órganos, como el corazón o la piel, sobre los que actúan gracias al sistema nervioso autónomo (SNA). Esta parte del sistema nervioso se encarga de regular las funciones involuntarias del organismo y se divide en dos: SNA simpático (SNAs), relacionado con los estados de alerta, y SNA parasimpático (SNAp), relacionado con estados de reposo [5].

1.2.1.1 Clasificación de los Estados Emocionales

El conjunto de las emociones humanas es un campo muy diverso y subjetivo, por lo que es necesario estandarizarlo. Con este fin, algunos autores han desarrollado diferentes modelos que tratan de simplificar el problema, ya sea mediante la clasificación del estado emocional en unas emociones concretas (modelos discretos o categóricos) o mediante la representación de este en un plano de coordenadas (modelos continuos o dimensionales). A continuación, se hablará sobre los principales modelos que se han ido desarrollando con el paso de los años.

- Modelos discretos o categóricos:

Estos modelos plantean un problema de clasificación, es decir, el objetivo es determinar, de entre un conjunto de emociones, cuál de ellas es a la que corresponde el estado emocional del

paciente bajo estudio. Este enfoque ha sido trabajado por dos psicólogos estadounidenses principalmente, Paul Ekman y Robert Plutchick.

A) Modelo de Ekman

Paul Ekman propuso que las emociones se agrupaban en familias de estados emocionales. Cada familia tiene un “tema”, que es el conjunto de características comunes a cada emoción que permite agruparlas, y unas “variaciones”, que son los factores específicos del individuo que diferencian dichas emociones entre individuos. Además, en [6] describe que existen una serie de nueve características que permiten, no solo diferenciar dichas características entre sí, sino también distinguirlas de otros estados afectivos (Tabla 1).

	Distinctive States	Biological Contribution
1. Distinctive universal signal	X	X
2. Presence in other primates		X
3. Distinctive physiology	X	X
4. Distinctive universals in antecedent events	X	X
5. Coherence among emotional response		X
6. Quick onset		X
7. Brief duration		X
8. Automatic appraisal		X
9. Unbidden occurrence		X

Tabla 1. Características que permiten distinguir las emociones de otros estados emocionales y entre sí.

El modelo de Ekman se compone de 6 emociones que el autor clasifica en dos grupos; positivas, felicidad, y negativas, tristeza, enfado, miedo y disgusto, por otra parte, la sorpresa se considera como una emoción neutra. A continuación, se detallan un poco más [7].

- **Felicidad:** (*happiness*), se describe como una emoción positiva, pues idealmente la mayoría de las personas desean experimentarla. Está íntimamente relacionada con el concepto de placer, que se describe como una sensación física de bienestar, y se puede obtener mediante la realización de un acto satisfactorio o mediante el pensamiento anticipado de la sucesión futura de dicho acontecimiento. Se considera el contrario del aburrimiento e incompatible, generalmente, con el dolor.
- **Sorpresa:** (*surprise*), esta es la más breve de todas las emociones, pues dura poco más que el instante en el que se produce el acontecimiento que causa la sorpresa. Este acontecimiento suele ser algo inesperado, es decir, que ocurre de manera puntual y que no tiene forma de ser anticipado. También puede ser erróneamente esperado, lo que quiere decir que, esperando que ocurra un determinado evento, ocurre otro distinto.
- **Miedo:** (*fear*), las personas tienen miedo a ser dañadas, tanto física como emocionalmente. La primera situación deriva de algún tipo de lesión física, cuya intensidad puede variar, mientras que la segunda puede originarse por un daño psicológico (autoestima, seguridad o confianza) o por una pérdida (familiares, amigos o relaciones amorosas). En cualquier caso, las personas tratarán de evitar la causa de ese sufrimiento.
- **Disgusto:** (*disgust*), esta es una emoción de aversión, las personas muestran rechazo por diferentes estímulos, ya sean sabores, olores, texturas o comportamientos, entre otros. Este rechazo se puede mostrar también por cosas que se vean, a pesar de que, no se sientan. Además, es una sensación que varía dependiendo de la cultura, es decir, no a todas las personas les produce rechazo las mismas cosas.

- Enfado: (*anger*), este es el más peligroso de los estados emocionales pues la gente enfadada tiende a perder el control y atacar a otras personas. El autocontrol en estas situaciones es algo que se enseña desde niños y que posteriormente sirve para etiquetar a los individuos, es decir, a la gente se la cataloga en función de su capacidad para controlar su ira. Por lo general, la frustración desencadena el enfado y este se ve agravado en función de la interpretación que el sujeto realiza de su entorno.
- Tristeza: (*sadness*), en la tristeza se produce un sufrimiento silencioso, por lo general esta emoción no es expuesta públicamente y se sufre en privado. No es de duración breve, sino que se manifiesta durante varios minutos o incluso días y puede estar provocada por varios estímulos, pero fundamentalmente pérdidas, tanto de personas, de oportunidades o de salud, debido a una enfermedad o accidente.

B) Modelo de Plutchik

El trabajo de Robert Plutchik [8] se focaliza en el desarrollo de un sistema avanzado y generalizable para clasificar emociones. Este se basa en las teorías de diferentes etólogos, que describen una serie de comportamientos que los animales llevan a cabo en la naturaleza para poder sobrevivir y reproducirse. La lista de comportamientos a la cual el autor hace referencia fue propuesta por Scott y es modificada por el autor para poder generalizar, como se ha mencionado antes, lo máximo posible el modelo.

Los nuevos comportamientos propuestos se escriben en lo que Plutchik llama lenguaje funcional y son los siguientes:

- Incorporación: (*incorporation*), proceso de alimentarse o asimilar estímulos beneficiosos para el organismo.
- Rechazo: (*rejection*), proceso de expulsar del organismo cualquier elemento dañino para este.
- Protección: (*protection*), los individuos tratan de evitar que se produzca un daño, por lo que hacen todo lo posible por alejarse de aquello que se lo pueda causar.
- Destrucción: (*destruction*), intento de eliminar las barreras que impiden alcanzar algo beneficioso.
- Reproducción: (*reproduction*), proceso de acercarse, relacionarse e intercambiar material genético con el fin de perpetuar la especie.
- Reintegración: (*reintegration*), una vez se ha perdido algo que se poseía y era beneficioso se intenta recuperar.
- Orientación: (*orientation*), adaptación a un entorno nuevo y desconocido.
- Exploración: (*exploration*), capacidad para mapear ese nuevo ambiente.

Estos comportamientos son traducidos en un lenguaje subjetivo el cual se compone de los estados emocionales propiamente dichos. Así pues, se describen 8 emociones, las cuales son muy parecidas a las mencionadas anteriormente por Ekman: alegría, confianza, miedo, sorpresa, tristeza, disgusto, enfado y anticipación. Como se puede ver, las 6 de Ekman prevalecen y a mayores se añaden 2 más (confianza y anticipación).

La diferencia fundamental de este modelo, en comparación con el anterior, no es que se añadan 2 emociones, sino que cada una de ellas tiene asociados unos grados de intensidad (*arousal*), de manera que no solo se describe el estado emocional del individuo, sino que además se puede cuantificar. Las emociones se agrupan en una forma concreta, de cono invertido, en cuya base se disponen las emociones más intensas, suavizándose conforme se acercan al vértice, como se puede ver en las Figuras 1 y 2. Además, no se agrupan de una forma arbitraria, sino que lo hacen por similitud, es decir, las emociones más parecidas entre sí, y que por tanto más fácil podrían confundirse, se colocan contiguamente. Esto permite añadir nuevas emociones que resulten de la combinación de dos adyacentes.

Modelo de Russel

En este modelo se propone que las emociones están determinadas por dos ejes que se cortan, como los ejes cartesianos. Por un lado, el eje horizontal, denominado este-oeste, hace referencia al grado de placer que genera dicha emoción, es decir, cuanto más disfrutable sea más a la derecha se situará, en caso contrario se situará a la izquierda, en otros artículos este eje se le denomina “*valence*”. El eje vertical, norte-sur, hace referencia al grado de excitación del sujeto, de manera que cuanto más arriba este se sentirá más estimulado y cuanto más abajo más dormido. A este eje se le llama “*arousal*”.

En este momento se definen 4 variables, una por cada extremo de los ejes: placer (*pleasure*), miseria (*misery*), en el eje este-oeste y somnolencia (*sleepiness*), despertar (*arousal*). Además, se definen otras 4 variables que ayudan a clasificar los cuadrantes, en sentido antihorario, empezando por el cuadrante superior derecho: emoción (*excitement*), sufrimiento (*distress*), depresión (*depression*) y satisfacción (*contentment*). Estas variables se describen gráficamente en la Figura 3. Se aprecia que se forma un círculo de manera que, al igual que ocurría con el modelo de Plutchik las emociones más diferentes se oponen en el mismo.

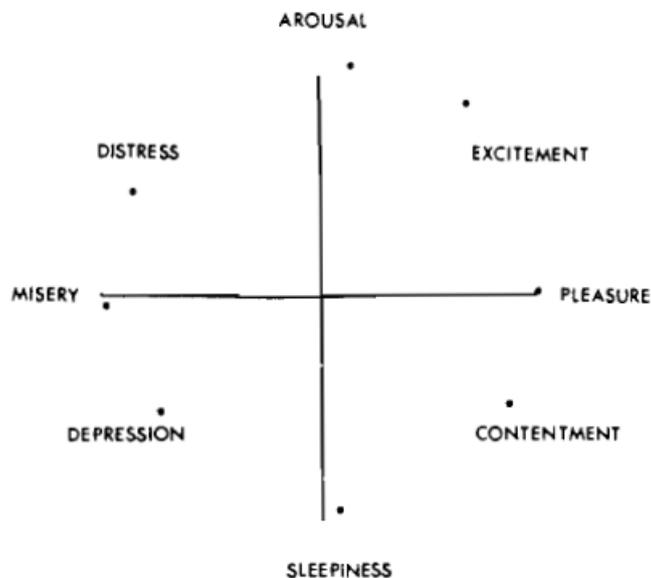


Figura 3. Disposición de las emociones en los ejes de coordenadas.

1.2.2 Señales fisiológicas

Como se ha explicado anteriormente, el estado emocional repercute sobre los órganos a través del sistema nervioso autónomo. La actividad de dichos órganos puede ser medida, obteniendo una serie de señales concretas, cuyo análisis nos permitirá extraer unas características (*features*). Estas, después de pasar por un proceso de selección, serán los *inputs* para entrenar el modelo predictivo, esto se explicará en profundidad más adelante. El empleo de señales fisiológicas de este tipo facilita el estudio del estado emocional, en comparación con otro tipo de registros más complejo como el EEG. Por ejemplo, la variación de la frecuencia cardiaca (HRV, por sus siglas en inglés) es un parámetro sensible en este aspecto, que aumenta en situaciones de estrés debido a la estimulación del nodo sinusal, por la acción del SNA, y disminuye en situaciones de relajación, mediadas por el SNAp. A continuación, se van a analizar las señales más empleadas en estudios anteriores, valorando tanto la naturaleza de las señales, como las técnicas de preprocesado empleadas y las características extraídas.

1.2.2.1 Señal del Corazón

El corazón es un órgano situado en el centro de la cavidad torácica, desplazado a la izquierda, ocupando una región conocida como mediastino. Su función principal es bombear la sangre por todo el organismo. Para ello, se divide en 2 regiones, izquierda y derecha, las cuales deberán impulsar la sangre por diferentes zonas. Cada una de estas partes se divide en 2 cavidades, conocidas como aurículas, superiores, y ventrículos, inferiores, y su función es recibir e impulsar la sangre, respectivamente. La aurícula derecha recibirá sangre desoxigenada que proviene de todo el organismo, a través de las venas cavas, de aquí pasará al ventrículo derecho que bombeará la sangre a los pulmones para que se oxigene, a través de la arteria pulmonar. Por su parte, la aurícula izquierda recibirá esa sangre ya oxigenada, por las venas pulmonares, y se la enviará al ventrículo izquierdo, que la impulsará por la arteria aorta al resto del organismo.

Para que estos procesos se puedan llevar a cabo es necesario que el músculo cardíaco, el miocardio, se contraiga y se relaje, por lo que deberá existir un sistema de conducción capaz de inducir dichos movimientos mediante la despolarización de las células cardíacas. Durante esa despolarización se generarán corrientes eléctricas fácilmente medibles mediante una serie de electrodos, constituyendo la señal de ECG. La adquisición de esta señal puede ser simple, únicamente utilizando tres derivaciones (electrodos) o más compleja, empleando un número superior de hasta 12 derivaciones.

La despolarización del corazón no es la única señal empleada, pues también es posible utilizar la transmisión de la luz a través de la piel para cuantificar el volumen de sangre que eyecta el corazón por cada pulso. Este dispositivo se denomina fotopleletismógrafo (PPG) y está formado por una fuente de luz, un fotodetector y una unidad de procesamiento. Emite luz en el tejido subyacente, y la sangre de los vasos cercanos absorbe parte de esa luz mientras que el resto se dispersa de vuelta al sensor. La información es analizada y procesada, para poder obtener la señal del volumen sanguíneo por pulso (BVP, *Blood Volume Pressure*) [10]. En los trabajos consultados [11]-[12], se emplea un sistema de PPG denominado Empática E4, que permite registrar la señal, sino que también permite preprocesarlas y extraer sus características típicas.

Tanto la señal de ECG, Figura 5, como de BVP, Figura 4, por si solas tienen poco valor a la hora de desarrollar el modelo predictivo, por lo que será necesario extraer algunas de sus componentes principales. Destacan fundamentalmente dos: la frecuencia cardíaca (HR) y la variabilidad de la frecuencia cardíaca (HRV).

- Frecuencia cardíaca: es una característica muy típica y fácil de extraer, tanto del ECG como del BVP. Representa el número de latidos del corazón por unidad de tiempo, generalmente 1 minuto. De forma normal, oscila entre 60 y 80 latidos por minuto (lpm), pero puede haber situaciones en que aumente, por ejemplo, ante estrés o excitación. En estados de máxima relajación lo normal es que disminuya.
- Variabilidad de la frecuencia cardíaca: esta es una variable bastante más compleja que aporta información más amplia. Se basa en la distancia temporal que hay entre latidos, para ello es necesario detectar los picos R, en caso del ECG o los valores máximos de la BVP. Para ello se puede utilizar diversas técnicas, como algoritmos concretos, en el caso de [13], se usó el de Pan-Tompkins, en [14] se usó el recomendado por el fabricante del dispositivo de registro (Empática E4) y en [15] se empleó un *teager energy operator*. Otra técnica, más lenta, es detectarlos de forma manual, por observación, y contar el tiempo de separación. Este método es más lento y además menos preciso, pues es fácil cometer errores en la apreciación de los picos R o valores máximos de BVP. De esta característica se puede obtener un gran número de nuevas características, tanto en los dominios temporal y frecuencial o también características no lineales.

En [13] y [15] se menciona la extracción de unas bandas de interés derivadas de la densidad espectral de potencia (PSD) de la HRV, la PSD se obtiene generalmente mediante la transformada corta de Fourier, basada en el método del periodograma de Welch. De estas se puede sacar información interesante relacionada con la actividad del SNA y son:

- Very Low Frequency (VLF): <0.04 Hz
- Low Frequency (LF): 0.04-0.15 Hz
- High Frequency (HF): 0.12-0.4Hz

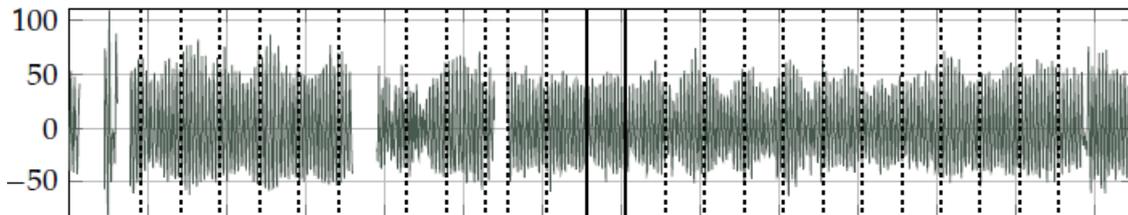


Figura 4. Señal de BVP, en la que cada máximo se corresponde con los picos que se deben contar, el eje de abscisas está en segundos [3].

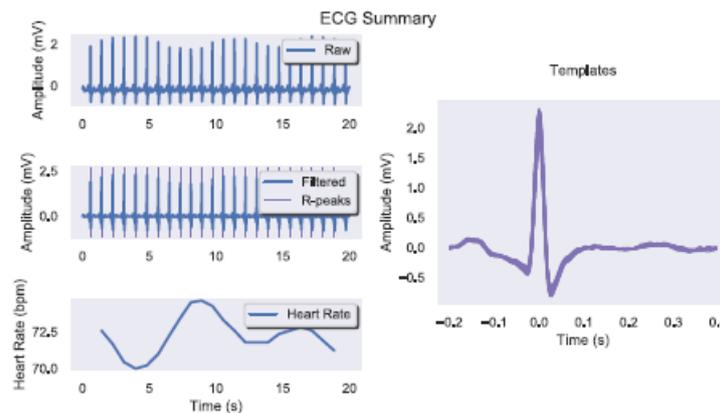


Figura 5. Señal de ECG, sin procesar (columna izquierda, primera imagen), señal de ECG procesada (columna izquierda, segunda imagen), señal de HR (columna izquierda, tercera imagen) y señal típica de ECG de un solo pulso (columna derecha) [6].

1.2.2.2 Sudoración de la piel

La piel humana está formada, entre otras cosas, por las glándulas sudoríparas, que son pequeñas estructuras capaces de secretar sudor cuando son estimuladas por acción del SNAs. Por lo general, la sudoración se produce como respuesta a un aumento de la temperatura corporal, ya sea por un ambiente caluroso o por la realización de algún tipo de actividad física. Pero, también existe la posibilidad de que se desencadene debido a estímulos externos que alteren el estado emocional de dicha persona.

Para medir el grado de sudoración se utiliza un sistema formado por 2 electrodos, que aplican corriente continua constantemente, de forma inocua para el paciente. El propósito de esto es determinar la conductancia de la piel, de manera que cuando el individuo suda esta aumenta. A la señal obtenida se la ha denominado convencionalmente respuesta galvánica de la piel (GSR, *Galvanic Skin Response*) [16].

Actualmente, el nombre de la señal está en discusión, pues hay autores [17] que sugieren que el término GSR es impreciso, pues no trata toda la amplitud del fenómeno electrodermal, además de que tampoco se describirían ciertos aspectos que influyen en la conductancia de la piel como la actividad eléctrica muscular y nerviosa. Por tanto, estos autores proponen el concepto de

actividad electrodermal (EDA, *Electrodermal Activity*) para referirse a la señal y por tanto será el término que yo utilice.

En la señal de EDA se distingue una parte que se conoce como nivel electrodermal (EDL, *Electrodermal level*) que es la parte de baja frecuencia del registro y que se asocia al proceso de regulación de la temperatura corporal, para este estudio resulta de poco interés pues no permite analizar la actividad del SN. Por otro lado, encontramos la respuesta electrodermal (EDR, *electrodermal response*) que es la que se surge como contestación a un estímulo y la que permite estudiar la función del sistema nervioso [18].

En [15] y [18] se propone realizar un remuestreo a 20 muestras con la finalidad de separar EDL de EDR. Una vez separadas ambas componentes, se obtiene una señal con picos que desciende eventualmente de 0. Para detectar eventos de EDR se buscan 2 ceros seguidos y se calcula la amplitud de ese evento hallando el máximo de ese pico. Por lo general, se descartan aquellos picos cuya amplitud no supere el 10% de la amplitud máxima registrada. El resultado gráfico se puede ver en la Figura 6.

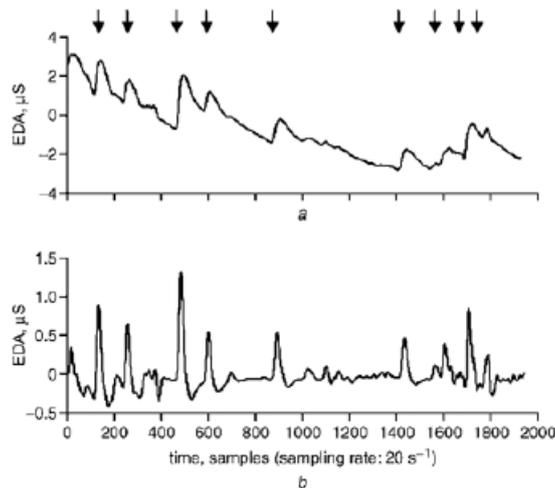


Figura 6. Señal de EDA típica, sin remuestreo (a) y con los eventos de EDR marcados con flechas y señal remuestreada para ver la señal de EDR [15]

1.2.2.3 Respiración

El ser humano realiza un proceso de intercambio gaseoso conocido como respiración. Este se lleva a cabo mediante la contracción y relajación de una serie de músculos ubicados en el tórax, los músculos respiratorios, destacando principalmente el diafragma. Estos se contraen en la inspiración, expandiendo la caja torácica, facilitando la entrada de aire a los pulmones al generar una diferencia de presión entre su interior y el exterior. Posteriormente, los músculos se relajan expulsando el aire inspirado, en un proceso que se conoce como espiración.

El indicador más característico de la respiración es la frecuencia respiratoria, que se define como el número de ciclos respiratorios, inspiración y espiración, que se producen a lo largo de 1 minuto. Esta frecuencia oscila, en el adulto y de manera normal, entre 12 y 20 respiraciones por minuto (rpm) si bien puede estar condicionada por diferentes factores como la actividad física, la temperatura corporal o el estado emocional.

Con relación a las emociones la frecuencia de la respiración aumenta cuando el sujeto se encuentra excitado, es decir, su estado de ánimo se ubica en la mitad superior del eje de coordenadas de Russel explicado anteriormente. En este sentido, sentimientos de terror o emoción provocan una disminución del volumen de aire inhalado [19], [20].

Para poder analizar la relación entre la respiración y las emociones computacionalmente, se pueden obtener diversos parámetros como la media del número de ciclos respiratorios por minuto, es decir, la frecuencia respiratoria, el tiempo de duración de dichos ciclos, la existencia de apneas (momentos en los que la persona no respira) y otras estadísticas más complejas del dominio de la frecuencia o no lineales [18].

1.2.2.4 Temperatura de la piel

La piel es el órgano más grande del organismo y el más expuesto al exterior, por lo que su temperatura está condicionada por la temperatura ambiente. Esta no es la única causa que influye en la temperatura de la piel (SKT, *Skin Temperature*), pues el estado emocional también tiene su importancia. En este caso el SNA tiene capacidad para influir en el calibre de los vasos sanguíneos periféricos. Esto quiere decir, que en función de cómo se encuentre el individuo el grosor de estos aumentará o disminuirá.

Los cambios de diámetro se denominan vasoconstricción o vaso dilatación. En el primer caso disminuye, provocando una bajada de la temperatura local debida al menor aporte sanguíneo. El caso contrario aumenta el grosor, provocando que llegue un mayor flujo sanguíneo y aumente la temperatura local.

En [21] se estudiaron los cambios de temperatura de la piel en manos y cara después de someter a un grupo de 32 mujeres a ver una serie de videos con temáticas diferentes. Se observó que, ante películas de terror, situaciones estresantes, se producía una disminución de la temperatura de los dedos, mientras que el resto del organismo permanecía igual. Por su parte las películas con contenido positivo generan situaciones de relajación y producen el calentamiento de las manos y la cara.

Estos cambios tienen correlación con el sistema nervioso simpático y parasimpático, pues el primero actúa en situaciones de estrés y miedo produciendo una cierta vasoconstricción periférica, mientras que el parasimpático induce la vasodilatación en situaciones de relajación.

1.2.2.5 Señal del cerebro

Como se ha mencionado anteriormente, los estados emocionales se desarrollan en el cerebro, por lo que es de esperar que la señal registrada en este, la electroencefalografía, sea de gran relevancia a la hora de estudiarlos. Al igual que la señal de ECG, se realiza un registro de la actividad eléctrica de las células, en este caso se mide la despolarización de las neuronas, responsables de la transmisión del impulso nervioso que, en última instancia, es el responsable de que se induzcan los cambios fisiológicos pertinentes en el resto de los órganos. Para el registro se emplean una serie de electrodos que se disponen de forma concreta sobre una malla que se empleará, a modo de casco, como soporte estructural.

El resultado de la adquisición es un conjunto de señales de alta frecuencia, una por cada canal, que pueden ser procesadas. El tratamiento de las señales es variado, pero se sintetiza en la aplicación de filtros, para eliminar el ruido relevante o para focalizar la señal, y en la extracción de las bandas de interés típicas del EEG. Estas bandas de frecuencia se obtienen mediante el empleo de las transformadas tiempo-frecuencia para poder visualizarlas con más claridad, otra forma de obtenerlas es mediante el análisis de componentes independientes (ICA) que permite dividir la señal original en varias señales nuevas. [13], [18], [22].

Las bandas de frecuencia mencionadas son 5, cuatro de ellas visibles en la Figura 7 y cada una contiene información relevante y distinta [23]:

- **Banda delta (δ):** característica del sueño profundo, son ondas lentas (0.5-4 Hz) de gran amplitud.
- **Banda theta (θ):** estas son ondas más rápidas (4-8 Hz), no muy típicas en humanas, pero que se asocian a estados emocionales o cognitivos. Tienen una morfología concreta, como de dientes de sierra.
- **Banda alpha (α):** características de los estados de vigilia en reposo, es decir, desaparecen cuando el sujeto realiza algún tipo de esfuerzo visual o cognitivo. Se dan entre 8 y 13 Hz.
- **Banda beta (β):** entre 13 y 30 Hz se asocian a estados de alerta y concentración
- **Banda gamma (γ):** la más rápida de las ondas (> 30 Hz) se asocia a procesos de asimilación de la información y al inicio de los movimientos.

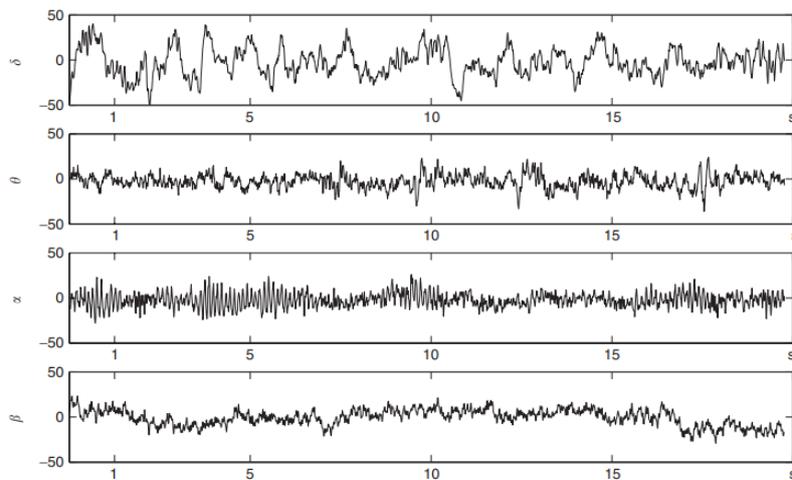


Figura 7. Bandas de frecuencia delta, theta, alpha y beta (de arriba abajo) [23].

1.2.2.6 Señales faciales

Una forma típica de exteriorizar los sentimientos es a través de los gestos y las expresiones faciales. Estos son fácilmente reconocibles, amplias sonrisas indican que el sujeto está feliz y excitado, mientras que ceños fruncidos y descensos de las comisuras de los labios indican disgusto o enfado. De esta forma, en estudios como [24] se graba la cara de los individuos para posteriormente, poder extraer información relevante de las mismas.

La metodología consiste en disponer una serie de puntos de referencia, 49 en el estudio anteriormente mencionado, en zonas concretas de la cara, como nariz, ojos, cejas o boca para después analizar estas zonas para extraer características determinadas mediante algoritmos complejos.

Dentro de las señales faciales destaca una, que es la señal de *Eye Gazing*. Esta señal se centra en el análisis de la mirada, fundamentalmente, se mide la dilatación de la pupila y se cuenta el número de parpadeos por minuto [25].

En el caso de la pupila (efecto Hippus) se produce una conexión directa con el sistema nervioso autónomo, de forma que en situaciones de estrés o miedo actúa el sistema nervioso simpático, provocando la dilatación de la pupila, mientras que en estados de relajación el parasimpático la contrae. En el caso del parpadeo la situación es parecida, pues situaciones de estrés producen un aumento de la frecuencia mientras que estados de relajación la disminuyen.

Otro punto que se tuvo en cuenta en el estudio fue el movimiento de los sujetos respecto de la cámara, es decir, si se acercaban o se alejaban de ella. Obviamente, se acercan cuando ven cosas agradables, mientras que se alejan en caso contrario.

1.2.2.7 Señal de los músculos

A parte del corazón, existen muchos otros músculos, que en combinación con huesos, tendones y cartílagos forman el aparato locomotor. Estos músculos, llamados músculos esqueléticos, se contraen y se relajan para permitir el movimiento de las articulaciones. Es decir, al acortarse el músculo, el tendón que lo une con el hueso se tensa y tira de este para moverlo. Este movimiento se produce gracias a una estructura localizada en el interior de las células o fibras musculares denominada sarcómero. Esta estructura se compone de 2 proteínas principales, la actina y la miosina, que se intercalan y se unen para poder desplazarse sobre sí mismas permitiendo el acortamiento de las fibras. Para que se realice este proceso es necesaria la intervención del Sistema Nervioso Periférico (SNP), que induce la despolarización de las fibras musculares a través de neuronas motoras [26].

Al igual que en el caso del corazón, la despolarización de las células musculares se puede medir y registrar, obteniendo una señal denominada electromiografía (EMG). A cada despolarización se le denomina Potencial de Acción de la Unidad Motora (MUAP). Las fibras musculares se reclutan conforme aumenta la exigencia del movimiento, no es igual la fuerza necesaria para levantar una pesa de 2 kg que una de 10 kg, por lo que el número de fibras reclutadas en el segundo caso será mayor. Al aumentar las demandas musculares del movimiento se aprecia una variación considerable de la morfología de la señal, pues se ensancha. La forma más común de registrar el EMG es mediante electrodos de superficie, pero en ocasiones la profundidad de los músculos requiere la inserción de una aguja-electrodo [27].

Los principales músculos que se registran, en este caso son; el trapecio y el músculo cigomático. El primero ubicado entre el cuello y la parte alta de la espalda, permite registrar movimientos que realiza el sujeto con la cabeza, generalmente cuando una situación le causa rechazo, lo asusta o lo atrae. Por su parte, el músculo cigomático se encuentra en la cara, controla la elevación del labio superior y, por tanto, regula expresiones como la alegría, el desdén o la sorpresa, mediante el control de microexpresiones de la boca. Aunque ambos músculos no están regulados por el SNA, existe una correlación entre su activación y el estado emocional.

1.2.3 Bases de datos

En el desarrollo de un modelo de clasificación o de regresión es imprescindible implementar una etapa de entrenamiento; para que este sea capaz de clasificar o cuantificar, en función del modelo emocional seleccionado, correctamente nuevos datos que no formen parte de dicho entrenamiento, pues a fin de cuentas es el objetivo de desarrollar el algoritmo. Para ello se precisa de una fuente de datos, es decir, un repositorio obtenido a partir de la realización de diferentes procedimientos sobre sujetos voluntarios, con el fin de conseguir información relevante sobre una determinada materia. En este caso, existen varios conjuntos de datos que someten a los individuos a diferentes estímulos que provocan en ellos una serie de emociones, que serán representadas por adjetivos concretos, en el caso de los modelos emocionales categóricos, o como valores numéricos. Simultáneamente, se registran una serie de señales fisiológicas, descritas en el apartado anterior, y que variarán en función del estudio seleccionado.

1.2.3.1 DEAP dataset

La primera de las bases de datos consultadas fue [28], desarrollada por investigadores de Reino Unido, Suiza y Países Bajos. Para realizarla se seleccionaron 32 participantes los cuales vieron una serie de vídeos musicales, cada uno seleccionado con intención de generar un estado emocional en el sujeto, y a partir de ellos se registraron una serie de señales fisiológicas: EEG, ECG, GSR, EMG, EOG, respiración, SKT e incluso se registró en vídeo las caras de los participantes.

El proceso de selección de los vídeos se compuso de varias etapas. En primer lugar, se utilizó una web, llamada *Last.fm*, que lo que hace es hacer un seguimiento de los gustos musicales de los usuarios y, además, asigna etiquetas a las canciones, en base a los sentimientos que provoca. De cada etiqueta, seleccionada se eligen las 10 canciones más representativas, hasta un total de 1084 canciones. El segundo paso, es hacer una selección manual de 15 canciones por cada cuadrante del modelo bidimensional de emociones elegido, similar al modelo de Russel, expuesto en el apartado 1. A continuación, se seleccionaron 60 vídeos manualmente, de nuevo 15 por cuadrante, que, en combinación con otros 60 vídeos, conseguidos a partir de los videoclips de las canciones seleccionadas, constituyeron el conjunto inicial de estímulos visuales. El cuarto paso fue reducir el número de vídeos seleccionados y la duración de estos. Para esta tarea se siguieron 2 procedimientos. Por un lado, se implementó un algoritmo que detectaba las partes de los vídeos capaces de sacar mejor las emociones [29], acertando la duración de estos y, por otra parte, se proyectaron de manera aleatoria algunos vídeos a algunos de los voluntarios para que fueran puntuados del 1 al 9, en función de que sentimientos les provocaban. Como resultado se seleccionaron 40 vídeos, finalmente.

Para la obtención del *dataset* se seleccionaron 32 participantes, mitad hombres y mitad mujeres, de entre 19 y 37 años. Antes de comenzar con la experimentación, al individuo se le explicó el protocolo de la prueba, se le colocaron los electrodos de registro e hizo una pequeña prueba para familiarizarse con el equipo. El experimento comenzó con la obtención, durante 2 minutos, de la línea base de las señales, mediante la visualización de una pantalla con una cruz en el medio. Posteriormente, se realizaron una serie de 40 pruebas, una por vídeo, que se dividían en 4 partes. En primer lugar, durante 2 segundos se mostraba el número de prueba, después, se volvía a obtener durante 5 segundos la línea de base. En tercer lugar, se mostraba el vídeo de 1 minuto correspondiente a la prueba y, por último, el sujeto debía realizar un cuestionario basado en 4 puntos: *arousal*, *valence*, grado de familiaridad con la canción, si la había escuchado o no, y su opinión respecto al vídeo y la canción, es decir, si le gustaba o no, independientemente de cómo se hubiese sentido.

Para la construcción de la base de datos se emplearon 3 tipos de señales: electroencefalografía, análisis de contenido multimedia y señales fisiológicas periféricas. Estas últimas, son de interés y son, por tanto, en las que se centrará el análisis. En primer lugar, tenemos la señal de GSR, que como se explicó en el apartado anterior es la medida de la respuesta galvánica de la piel, fue registrada mediante la colocación de 2 electrodos en las falanges terminales de los dedos índice y corazón. Para la señal de ECG se utilizó un fotopletimógrafo colocado en el dedo pulgar, que registra el volumen sanguíneo en esa zona, y de ahí se puede derivar la señal de electrocardiograma. Se midió la amplitud de la respiración y EMG; especialmente el trapecio, para registrar movimientos de la cabeza, y el músculo cigomático, que permite valorar si el paciente se ríe. Por último, se midió la señal de electrooculografía (EOG), que registra el estrés del sujeto.

La toma de datos se realizó con una frecuencia de muestreo de 512 Hz y, posteriormente, se realizó un submuestreo a 128 Hz, para reducir la cantidad de datos. Se realizó un preprocesado,

consistente en la eliminación de las derivas de baja frecuencia de la señal de GSR y ECG y se extrajeron hasta 103 características de todas las señales.

En cuanto a los resultados, cada una de las modalidades de datos extraídas desempeñaron un buen papel en uno de los parámetros analizados. El EEG mostró buenos resultados de *arousal*, las señales periféricas en *valence* y el análisis del contenido multimedia para el grado de apreciación de los vídeos.

1.2.3.2 MAHNOB-HCI

Otra de las bases de datos consultadas fue [30], en esta se emplean a 27 sujetos, 11 hombres y 16 mujeres, a los cuales se les proyectó un total de 20 vídeos. El objetivo de estos, como en el caso anterior, fue provocar una serie de emociones y obtener, a partir de estas, una serie de registros de EEG, señales periféricas, grabaciones del movimiento del cuerpo, la cara y los ojos y archivos de audio.

Los vídeos utilizados para provocar las emociones fueron extractos de 21 películas comerciales que fueron elegidos en un proceso de selección, que partía de un conjunto inicial de 155 video clips. Estos videos fueron vistos por un grupo de 50 personas que valoraban como se habían sentido al respecto, generando una serie de anotaciones, a partir de un sistema de autoevaluación basado en “SAM Manikins”. Los 14 vídeos que más etiquetas acumularon fueron seleccionados, junto con otros 6. De estos últimos 3 eran vídeos populares y el resto pronósticos meteorológicos de YouTube que fueron utilizados como vídeos neutrales.

A parte de las grabaciones de los movimientos de la cara, el cuerpo y los ojos, se registraron señales fisiológicas periféricas como el ECG y GSR, cuya deriva de base fue eliminada mediante el filtrado paso-alto de la señal. Todas las señales fueron muestreadas a 1.024 Hz y posteriormente, submuestreadas a 256Hz. La señal de electrocardiograma fue registrada con un sistema de 3 electrodos, ubicados en las esquinas superiores del pecho y la parte alta del abdomen. Para la GSR se utilizan 2 electrodos que se colocan en las falanges distales del dedo corazón. Un sensor se colocó en el dedo meñique para registrar la temperatura de la piel y, por último, para medir la respiración se empleó un cinturón en torno al tórax.

En cuanto al diseño experimental, se empleó un sistema de grabación que registró todas las señales e imágenes deseadas. Se llevaron a cabo dos fases experimentales, por un lado, los participantes debían contestar un cuestionario de 5 preguntas de respuesta múltiple al finalizar la visualización de cada vídeo. Por otra parte, en la segunda fase, únicamente debían responder cuestiones de sí o no.

El procedimiento partió de informar a los sujetos sobre que iba a consistir el experimento y, posteriormente, realizaron un preentrenamiento para familiarizarse con el interfaz. La fase experimental consistió en la visualización del video neutral, el pronóstico meteorológico mencionado anteriormente, a continuación, los individuos procedían a la visualización de los vídeos y, para finalizar, realizaban la autoevaluación de la primera fase. Esta consistía en 5 preguntas en las que la persona, para empezar, debía etiquetar emocionalmente el vídeo. El resto de las preguntas consistían en valorar del 1 al 9 diferentes parámetros de los vídeos como *arousal*, *valence*, *dominance* y la predictibilidad. En total la prueba duraba unos 50 minutos, más otros 30 de preparación.

El siguiente paso es la extracción de características de las señales extraídas para realizar un proceso de clasificación. Para las señales fisiológicas se seleccionaron un total de 102. La clasificación se realizó mediante un modelo de Support Vector Machine (SVM), que se verá más adelante. Además, se planteó; por un lado, un sistema de clasificación basado en arousal y

otro en *valence*. Los mejores resultados se obtuvieron al combinar las señales de EEG y el movimiento ocular. Por su parte, las señales fisiológicas tuvieron el peor desempeño.

1.2.3.3 AMIGOSdataset

En este *dataset* [31] participaron 40 personas, 13 mujeres y 27 hombres de entre 21 y 40 años, los cuales fueron sometidas a un doble proceso experimental, por un lado, visualizaron 16 vídeos cortos emocionales y en el segundo caso vieron 4 vídeos de mayor duración. En este último caso con la peculiaridad de que algunos participantes los vieron en solitario y otros formando grupos.

Para la selección de vídeos participaron 72 voluntarios cuya misión era visualizar 36 vídeos y determinar cuál era el cuadrante del eje de coordenadas *valence-arousal* para cada vídeo. Tras este proceso, se seleccionaron 12 vídeos que fueron cumplimentados con otros 4, uno por cuadrante, hasta un total de 16. Para la segunda parte, se seleccionaron 8 fragmentos de películas, de unos 20 minutos, y que tuvieran un fuerte contenido afectivo, es decir, que la combinación entre música, color y efectos especiales fuera llamativa. De este conjunto inicial se seleccionaron 4 vídeos.

Las señales periféricas tomadas en este caso son 2: ECG y GSR. Para la primera se empleó un sistema de registro de 3 electrodos, colocados; uno en el tobillo y los otros 2 en las caras internas de los codos, cuando están plegados. El registro se realizó a 256Hz. En el caso de la señal de impedancia de la piel (GSR) se colocaron 2 electrodos, uno en la segunda falange del dedo índice de la mano izquierda y el otro en la misma posición, pero en el dedo corazón. El muestreo se realizó a 128 Hz. A mayores se registraron; la señal de EEG y grabaciones de audio y vídeo.

El proceso experimental consistió en 2 partes, ambas realizadas en un ambiente de laboratorio con condiciones ambientales controladas. En primer lugar, se visualizaron los vídeos cortos. Para esta parte, los participantes dieron el consentimiento para realizar la experimentación, fueron informados acerca del procedimiento y recibieron un pequeño entrenamiento para familiarizarse con los dispositivos. A continuación, los experimentadores les colocaron todos los utensilios de registro y realizaron una pequeña autoevaluación, con objeto de saber que sentían antes de comenzar la prueba. Una vez finalizada la etapa de preparación, se les mostró a los participantes, una pantalla con una cruz, para obtener la línea de base y tras 5 segundos comenzó el vídeo. Una vez visto, los individuos realizan un test de autoevaluación donde juzgan como se han sentidos en términos de: *arousal*, *valence*, *dominance*, si les ha gustado el vídeo o no y también tuvieron que elegir entre un conjunto de emociones básicas cual era la que más definía su estado emocional. Para la parte de los vídeos largos, el procedimiento fue bastante similar. Las únicas diferencias fueron que en este caso 17 personas vieron las grabaciones en solitario y los 20 restantes, hubo 3 que no pudieron completar esta fase, lo vieron en 4 grupos de 5. Al término del segundo vídeo realizaron un descanso de 15 minutos. Los grupos tenían permitido hablar, pero no fueron informados de este detalle con el fin de que las conversaciones fueran naturales y espontáneas.

Una vez se registraron las señales, se pasó a la extracción de características. Para el ECG se calculó la frecuencia cardíaca y la variabilidad de esta y se calcularon otras 77 características. En la señal de GSR, por su parte, se calculó el valor de la conductancia de la piel y se obtuvieron otras 31. Estas características se utilizaron para entrenar un modelo predictivo *arousal-valence* con un algoritmo de Naive-Bayes y otro de SVM. Los resultados fueron mejores para las características de EEG, seguidos por los de ECG y posteriormente los de GSR. En conjunto las señales fisiológicas rindieron peor que las de EEG. Los mejores resultados se obtuvieron mezclando las señales fisiológicas y las grabaciones en vídeo de los sujetos. Por su

parte, la extracción de características de las señales originadas a partir de vídeos cortos dio mejores resultados que las de los vídeos largos. Tampoco resultó muy eficaz la combinación entre vídeos cortos y largos.

1.2.4 Algoritmos predictivos

Todo lo anterior representa las características que describen el trabajo a realizar. Es decir, queremos determinar el estado emocional de un sujeto usando algunas de las señales fisiológicas descritas anteriormente. El siguiente paso, es construir el medio para conectar ambos elementos y poder determinar con precisión cuales son las emociones de la persona. Para ello desarrollaremos un algoritmo predictivo, que en base al modelo empleado (discreto o continuo) puede ser de clasificación o de regresión.

Los modelos de clasificación son aquellos cuya salida puede ser una de las opciones posibles. Este es el algoritmo utilizado en los modelos continuos, pues el estado emocional solo puede ser clasificado como uno de todos los que se proponen. Por su parte la salida de la regresión es un conjunto de valores comprendidos dentro de un rango concreto. En cierta medida, la regresión se puede ver como un problema de clasificación, pero con muchas más clases [32]. En este caso los modelos continuos emplean una doble e incluso una triple regresión, para representar en los ejes de coordenadas del modelo de Russel el estado emocional del sujeto.

En el contexto de ML, existen una gran variedad de algoritmos que deberán ser entrenados, usando para ello las señales fisiológicas, debidamente etiquetadas, es decir, cada una de las señales de entrenamiento va asociada a un estado emocional previamente definido. Posteriormente, se deberá realizar un proceso de validación, con un set de datos independiente, para comprobar que el modelo es generalizable y que no hay ningún tipo de sobreajuste. Vamos a analizar ahora los algoritmos de ML más empleados en diferentes estudios.

1.2.4.1 Support Vector Machine (SVM)

Este es algoritmo más empleado en la bibliografía consultada. Se emplea fundamentalmente en problemas de clasificación [11], [14], [15], pero también tiene un buen desempeño en la regresión [13], [33].

El objetivo del SVM es representar las características de cada sujeto en un espacio de coordenadas y encontrar un método para separar dichas características en grupos concretos, es decir, para clasificar cada sujeto en su estado emocional correspondiente. La mayoría de los estudios no usan modelos de 2 o 3 emociones, que serían fácilmente separables mediante una curva o un plano, sino que se desarrollan modelos más complejos que requieren de funciones matemáticas más complejas como los hiperplanos. Este no es seleccionado arbitrariamente, sino que se realiza un exhaustivo proceso de selección, en el que se emplean algoritmos de optimización con la finalidad de hallar el hiperplano que garantice la máxima separación interclase, garantizando el menor error posible en la clasificación [34].

En cuanto a la precisión obtenida, varía dependiendo del modelo emocional elegido. En los casos en los que se utilizó *valence/arousal* [14] y [13] los resultados fueron del 75% y 85% en el caso de la *arousal* y del 71.2% y 80% para la *valence*. Por su parte, los modelos categóricos varían en función del número de emociones seleccionadas: en [14] y en [15] se eligieron modelos de 4 emociones con un 75% y un 61.8% respectivamente, en [15] también se seleccionó un modelo de 3 emociones, obteniendo un 78.4% y en [33] se empleó un modelo de 6 emociones, el modelo de Ekman completo, obteniendo la clasificación en dichas emociones a

través de un sistema *arousal-valence* y representando sobre el mapa de coordenadas los estados emocionales, superando el 80%.

1.2.4.2 K-Nearest Neighbors (KNN)

Este es también un algoritmo bastante utilizado, debido a que es bastante intuitivo y ofrece resultados bastante buenos. En algunos estudios como en [18] se plantea como el mejor de los utilizados con un rendimiento superior al 90% en cuanto a la precisión. En otro artículo [35] la precisión obtenida ronda el 80%, pero en este caso se realiza la selección de características con dos algoritmos diferentes, PCA (*Principal Component Analysis*) y JMI (*Joint Mutual Information*), el desempeño en ambos es similar, mejorando algo con JMI, 79.61% y 81.63% respectivamente. En [36] se emplean 2 algoritmos, este y LDA (*Linear Discriminant Analysis*), del que se hablará más adelante, en este caso se obtuvieron mejores resultados con KNN.

El algoritmo puede ser tanto de clasificación como de regresión. En el primer caso se debe seleccionar un valor para la k , es decir, se define el tamaño del vecindario. A continuación, el algoritmo procede de tal manera que selecciona las k muestras más cercanas a la muestra objetivo (x), la que se quiere clasificar, y cuenta las distintas clases con las que están etiquetados esos k valores. Tras ese conteo la clase mayoritaria es a la que se asigna la muestra x . La elección de la k es un factor determinante, pues una k pequeña mejora la precisión, al etiquetar correctamente muestras de una clase que caen entre varias muestras de otra clase, pero hace que el modelo sea susceptible al ruido y poco generalizable. Por su parte, una k grande mejora el coste computacional, pero reduce la precisión al no tener en cuenta esas muestras dispersas, estas características del modelo son visibles en la Figura 8 [32].

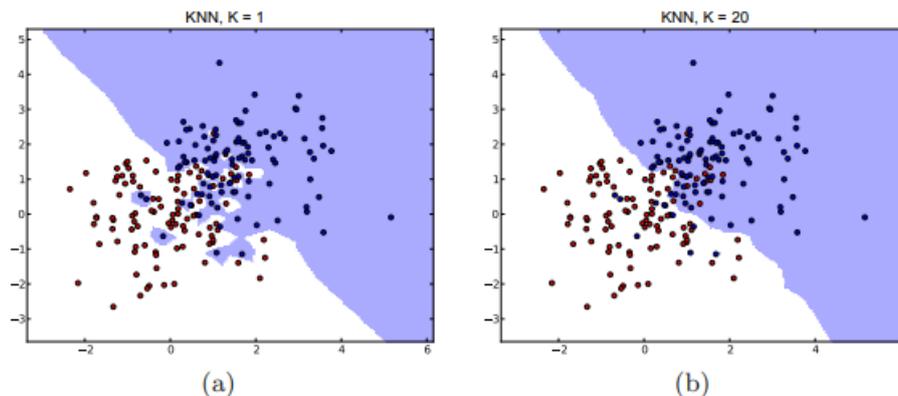


Figura 8. Ejemplos de KNN para una $k = 1$ y $k = 20$ [32].

Por su parte, la regresión es bastante similar a la clasificación, se seleccionan los k vecinos más cercanos y se evalúa el valor de cada uno de ellos. La principal diferencia es que en este caso la muestra x se clasifica como el valor promedio de todos esos vecinos seleccionados. El valor de la k también es determinante, en este caso valores pequeños son propensos al sobreajuste del modelo [32].

1.2.4.3 Linear Discriminant Analysis (LDA)

En [36] se propone utilizar este tipo de algoritmo para compararlo con KNN. La técnica de LDA es similar a la PCA, pero enfocada a la clasificación en lugar de para la selección de características. Esta consiste en el cálculo de una matriz de transformación, que nos permita proyectar las características originales en un nuevo espacio de características, con la finalidad de maximizar la separación interclase.

Este algoritmo se puede desarrollar a partir de 2 enfoques: transformación dependiente de clases y transformación no dependiente de clases. La diferencia principal entre ambas radica

fundamentalmente, en que en el primer caso se tiene en cuenta la separación interclase y en el segundo no, únicamente se valora la separación intraclase. En este último, puede pasar que si las clases no están bien diferenciadas el algoritmo falle a la hora de clasificar las muestras futuras.

El método LDA presenta la ventaja particular de manejar adecuadamente secuencias de frecuencia variable, es decir, aquellas cuyas clases están formadas por distinto número de muestras. En el contexto de la computación afectiva, en el que los estados afectivos pueden variar mucho entre personas resulta interesante valorar este enfoque para realizar el modelo predictivo [37].

1.2.4.4 Random Forest (RF)

En [33] se utiliza, además del ya mencionado SVM, la técnica de RF para llevar a cabo una regresión, *arousal-valence*. A partir de aquí, se dibuja un eje de coordenadas, sobre el que se representan los estados emocionales, estando el estado más relajado en el centro y el resto en los demás cuadrantes. Los resultados obtenidos, superan a los del SVM alcanzando el 86% de precisión.

Un modelo de RF está formado por un conjunto de árboles de decisión. Estos árboles emulan la estructura de uno real, es decir, poseen una raíz (*root*), una serie de ramas (*branches*) que se unen en nodos y una serie de hojas que constituyen las diferentes clases, en un problema de clasificación, o los diferentes valores, en un problema de regresión. El funcionamiento es sencillo, se introduce el conjunto de datos de entrenamiento en el nodo raíz, el primero de todos. En este punto el algoritmo selecciona una característica con la que dividir el conjunto inicial en 2, en función de un umbral concreto. Cada grupo formado constituirá una rama del árbol y cada una de ellas llegará hasta el siguiente nodo. Este puede, al igual que el primero, usar otra característica para volver a dividir el grupo o puede ser un nodo final, si alcanza un criterio de parada.

Un nodo final supone que todas las muestras que se encuentren en él serán asignadas a una clase concreta o cuantificadas como un valor en el caso de la regresión. El problema que se asocia a los árboles de decisión es que son propensos a sobreajuste (*overfitting*), lo que impediría al modelo ser generalizable, esto es, que no sería aplicable a un conjunto de datos distinto al conjunto de entrenamiento. Esto supone un problema considerable, por lo que se utilizan métodos de *ensemble learning* para solucionarlo. Estos métodos son más complejos computacionalmente, pero ofrecen una mejora significativa respecto de los árboles de decisión comunes.

Dentro de estos métodos de *ensemble learning*, encontramos el algoritmo de *RF*. Este no es más que una agrupación de árboles de decisión. La entrada a cada árbol es un subconjunto de datos. Es decir, mediante una técnica denominada *Bagging (Bootstrap Aggregation)* se seleccionan, aleatoriamente, un número reducido de muestras del conjunto de entrenamiento. La selección de estas muestras se hace con remplazo, lo que significa que cada una que se coge es devuelta al conjunto inicial, no se elimina. Esto supone que una misma muestra puede formar parte de todos los subconjuntos de entrenamiento y estar en todos los árboles simultáneamente. La ventaja de este método es asegurar la variabilidad entre los subconjuntos de entrenamiento, reduciendo sesgos.

En RF se tienen en cuenta las salidas de todos los árboles, es decir, los nodos finales u hojas. Posteriormente, se realiza una votación, es decir, se busca la clase más repetida entre estas hojas o se promedia el valor de estas. El resultado de la votación es el que se asigna como la salida del algoritmo. Por tanto, este método no solo ofrece mejores resultados que otros, sino que además previene el sobreajuste del modelo y los sesgos [38].

1.3 DESCRIPCIÓN DEL PROBLEMA

El objetivo es desarrollar un algoritmo de ML para predecir el estado emocional de un sujeto, utilizando señales fisiológicas. El procedimiento para hacerlo comenzará por elegir una de las bases de datos como conjunto de entrenamiento, DEAP dataset en este caso. Es importante que sea adecuada, es decir, que tenga las señales correctamente registradas, digitalizadas y preprocesadas. Las señales que se van a utilizar son: ECG, EDA, RESP y SKT. El segundo paso, consiste en analizar los datos y, si es preciso, realizar el procesamiento adecuado, para mejorar su calidad. A continuación, se extraen las características del conjunto de datos inicial. Estas se van a dividir en 4 grupos: estadísticas, temporales, espectrales y no lineales. Para cada señal seleccionada se extraerán las características pertinentes de cada grupo, que pueden coincidir entre sí. Si el número extraído es demasiado elevado se debe reducir la dimensionalidad; pasar de un número inicial de características a otro más pequeño, para ello se selecciona un método de elección de características. Además, tras este paso es importante escalar los datos, lo que significa que hay que transformarlos para que sean más manejables para el algoritmo predictivo. Para la predicción, se van a desarrollar 4 algoritmos de regresión (SVM, KNN, PLS y RF) para poder comparar entre ellos cuales rinden mejor. Al elegir la regresión en lugar de la clasificación se deberá tener en cuenta que el problema es multivariado, lo que significa que habrá que ejecutar el proceso 2 veces, una para la *arousal* y otra para la *valence*. Por último, se utilizará otro conjunto de datos para validar los modelos, es decir, comprobar si tienen buena capacidad de generalización y se adaptan a datos con los que no han sido entrenados.

2 HIPÓTESIS Y OBJETIVOS

2.1 HIPÓTESIS

Una alteración en el estado de ánimo, motivada por un estímulo puntual, tiene consecuencias directas sobre el SNA y los diferentes órganos que regula. Estas se traducen como variaciones en el funcionamiento de los órganos, que pueden ser capturadas mediante dispositivos electrónicos. Utilizando algoritmos de *machine learning*, es posible desarrollar un modelo predictivo que, a partir del análisis de señales electrofisiológicas periféricas monitorizadas bajo estímulos constantes, detecte con alto grado de precisión el estado emocional de una persona.

2.2 OBJETIVOS

2.2.1 Objetivo principal

El objetivo principal es el desarrollo, evaluación y comparación de un conjunto de modelos predictivos de *machine learning*, para saber cuál de ellos presenta un mejor rendimiento en este tipo de problema de regresión multivariada.

2.2.2 Objetivos específicos

A parte del objetivo principal se definen los objetivos específicos, que permitirán evaluar el trabajo de forma secuencial.

1. **Análisis y procesamiento de los datos:** los datos deben ser adecuados para utilizarlos en todas las fases del proceso, por lo que hay que analizar cómo son esos datos y si es necesario realizar algún tipo de operación. Estas pueden ser: eliminación de ruido, cancelación de artefactos, corrección de derivas de la línea base, etc.
2. **Extracción de características:** para entrenar los modelos hacen falta una serie de características, por lo que se debe implementar una etapa en la que conseguir las. En esta etapa se utilizarán métodos matemáticos con los que tratar de representar las señales originales de una forma más sencilla.
3. **Selección de características y escalado de datos:** la información extraída es, por lo general, excesiva, redundante y sobredimensionada. Para solucionar estos problemas hay que, en primer lugar, elegir un número reducido de características que aporten la mayor cantidad de información diferente posible y, en segundo lugar, enmarcar el tamaño de los datos en un rango común, para que sea más sencillo el entrenamiento de los algoritmos.
4. **Elaboración de los modelos predictivos:** programar, entrenar y ajustar los diferentes modelos de ML, para conseguir que sean lo más precisos y generalizables posible.
5. **Validación de los modelos:** una vez están entrenados los modelos toca ver si son lo suficientemente generalizables, es decir, si para otros datos no vistos en la fase de entrenamiento el modelo también es válido y no se ha producido sobreajuste.

3 MATERIALES Y MÉTODOS

Anteriormente, se han explicado los principales elementos necesarios para el desarrollo del modelo y, ahora, es el momento de seleccionar cuales de estos son los más adecuados.

3.1 BASE DE DATOS

Lo primero es buscar los datos de entrenamiento, para ello analizamos las bases de datos anteriormente mencionadas en busca de la que ofrezca la información más completa y variada. En este sentido, nos vamos a quedar con DEAP dataset, pues ofrece una amplia variedad de señales preprocesadas; descargables tanto para Python como para Matlab. En este caso vamos a emplear la versión 3.11.7 de Python. Para acceder a los datos es necesario firmar un consentimiento para que, posteriormente se permita el acceso a todos los datos. Estos se componen de las señales originales, los datos preprocesados y los registros de vídeo. Para el modelo, se descargan los datos preprocesados para Python.

La carpeta descargada consta de 32 archivos, uno por individuo, que debido al formato precisa del uso de la librería *pickle* para poder cargarlo en *spyder*. Cada uno de los archivos de los sujetos están compuestos por una matriz tridimensional de 40x40x8064. En primer lugar, están representados los diferentes vídeos que tienen que ver los pacientes, el segundo número equivale a las señales extraídas (tabla 2) y el último al número de muestras extraídas de cada una de ellas, después de que fuesen submuestreadas. Las señales extraídas fueron las siguientes:

FILAS	SEÑALES
1-32	Canales de EEG
33-34	EOG
35	EMG
36	ECG
37	EDA
38	RESP
39	BVP
40	SKT

Tabla 2. Disposición de las señales fisiológicas en la base de datos.

Las señales fueron submuestreadas a 128 Hz, por lo que teniendo en cuenta el número de muestras, el tiempo de adquisición de todas ellas fue de 64 s. De todas estas las señales con las que se va a trabajar son: ECG, EDA, RESP y SKT. A continuación, se va a describir el trabajo previo a la extracción de características de cada señal.

3.2 SEÑALES FISIOLÓGICAS

3.2.1 Señal de ECG

Esta señal se corresponde con la fila 36 del conjunto de datos, pero en Python, como la primera fila de una matriz es la 0 y no la 1, el indicador para seleccionarla será el 35. Al visualizar las señales, empleando para ello el comando '*matplotlib.pyplot.plot*' de la librería *matplotlib* (a partir de este punto, el término *matplotlib.pyplot* se resumirá como *plt*), se aprecia que la línea de

base del ECG no se alinea en torno al 0, por lo que es necesario remover esta deriva. Para ello, utilizamos el algoritmo descrito en [39]. En este se aplica un método morfológico, en el que se realiza un suavizado de las características de corta duración (complejo QRS) y de larga duración (ondas P y T). A partir de aquí, se puede calcular la morfología de la línea de base y extraerla sin distorsionar la señal. También es necesario realizar un filtrado paso-bajo con frecuencia de corte de 45 Hz debido a que algunas señales de ECG son un poco ruidosas. Para ello se utiliza un filtro de tipo Butterworth de orden 5. El resultado se ve en la Figura 9.

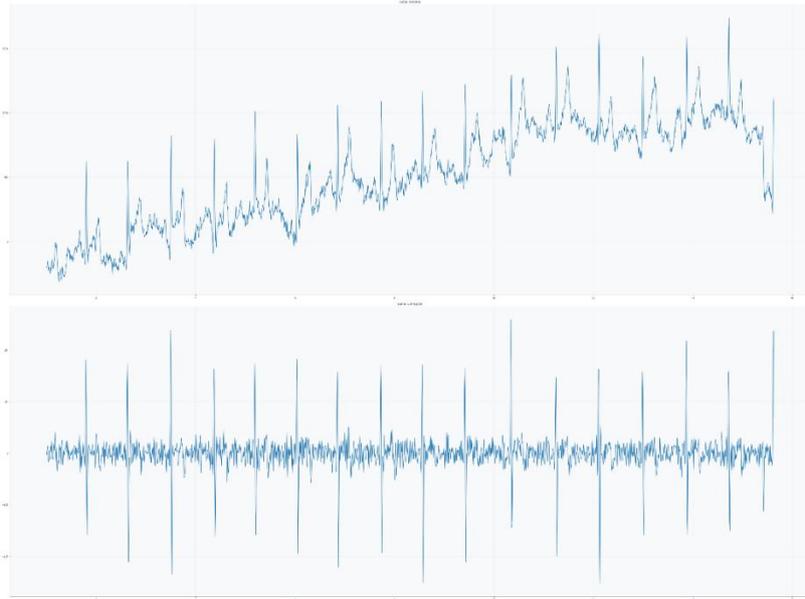


Figura 9. Señal de ECG original, primeras 2000 muestras (arriba) y filtrada (abajo)

Al procesar la señal se ve que esta pierde calidad, es decir las estructuras típicas del ECG, como las ondas P y T parecen desaparecer, pero para este trabajo no son necesarias, pues el propósito no es realizar un estudio electrofisiológico, sino analizar la correlación entre el SNA y el corazón. Para esto es necesario que los picos R sean claramente visibles, que es el caso, pues la distancia entre ellos indica como varía la frecuencia cardíaca, de esto se hablará más adelante.

A continuación, se utiliza la librería *BioSPPy* [40] para extraer información relevante del ECG. Esta librería está diseñada para trabajar con señales fisiológicas y será utilizada también en las señales de EDA y RESP. A la función de la librería que vamos a utilizar hay que pasarle la señal de ECG y la frecuencia de muestreo. La salida de la función es: la señal preprocesada, la detección de los picos R de la señal, la representación de la frecuencia cardíaca y los *templates* que consisten en la superposición de los diferentes ciclos cardíacos, estos dos últimos elementos no son utilizados. A parte de estos elementos, que se devuelven en forma de variables, se proporciona una representación gráfica, que se puede ver en la Figura 10. A pesar de que la función realizase el procesamiento de la señal, se optó por remover la deriva de base previamente, debido a que en el algoritmo fallaba en la detección de algunos picos R, confundiendo algunas partes de las ondas P y T con estos.

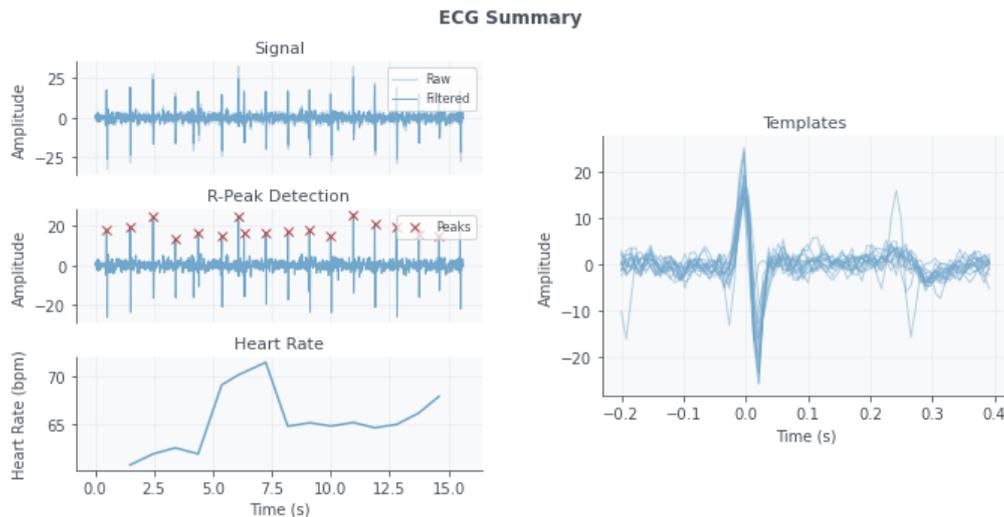


Figura 10. Salida de la función de *BioSPPy* para ECG. señal original de ECG (1000 muestras) y filtrada (arriba izda.), picos R (medio izda.), frecuencia cardiaca (abajo izda.) y templates (dcha.).

La distancia entre picos R sucesivos (intervalos RR) equivale al tiempo que dura un latido cardíaco. Cuanto más corto sea el valor, más rápido late el corazón y viceversa, lo que dependerá de la actividad del SNA. De esta forma, tenemos en esta medida una forma fiable y sencilla de relacionar la señal cardíaca con el sistema nervioso.

La función de *BioSPPy* devuelve los índices en los que se encuentra cada pico, es decir, cuál de los 8064 valores que forman cada señal es un pico R. El modelo que se utiliza no se especifica, pero el ejemplo más claro de este procedimiento es el algoritmo de Pan-Tompkins [41]. Una vez se tienen los índices se deben calcular los intervalos RR; para ello se define una función, compuesta por un bucle que en cada iteración (n) le resta el coeficiente del pico n al del pico n + 1. Hay que tener en cuenta que, de esta forma, se necesitaría un último pico R (N + 1, siendo N el número de picos y el último de ellos), que no existe. Por tanto, el bucle no puede realizar un número de iteraciones igual al número de picos, sino que deberá ser una unidad menor. Otra consideración es que, los resultados obtenidos no se dan en tiempo, sino que se dan en muestras. Teniendo en cuenta que la frecuencia de muestreo son 128Hz y no 1000 cada muestra no equivale a 1ms, sino a 7.8125ms (1), por lo que habrá que tener en cuenta este factor a la hora de realizar los cálculos.

$$sampling\ interval = \frac{1}{f} * 1000 \quad (1)$$

Sobre esos valores se vuelven a calcular las distancias, utilizando una función muy similar a la anterior, con el fin de comparar como varía la diferencia de 2 picos R consecutivos.

3.2.2 Señal de EDA

En este caso, se vuelve a utilizar *BioSPPy* para extraer características. Al igual que con el ECG a la función hay que introducirle la señal de EDA y la frecuencia de muestreo. Como salidas en este caso se obtienen:

- Eventos detectados: cuando se produce una variación en la impedancia de la piel, se registra en los electrodos y se produce un cambio en la tendencia de la señal. Estos cambios se denominan *onsets* (comienzos) y son los que se registran. Además, se

detecta también el máximo valor que alcanza la señal en el intervalo comprendido entre 2 *onsets* consecutivos.

- **Amplitudes:** representan los cambios en la conductividad eléctrica de la piel, generalmente asociados con respuestas emocionales o estados de excitación del SNA. Un aumento en la amplitud indica una mayor respuesta conductiva, que puede estar vinculada a emociones como el estrés, la ansiedad, o la sorpresa [17].
- **Descomposición de la señal:** como se menciona en el apartado en el que se describe la señal, esta se puede descomponer en una fase de baja frecuencia (EDL) y otra de alta frecuencia (EDR), la función devuelve ambas componentes separadas.
- **Representación:** al igual que en el caso del ECG se puede habilitar, mediante el comando *'True'* que la función devuelva una representación gráfica de los elementos anteriores, como se puede ver en la Figura 11.

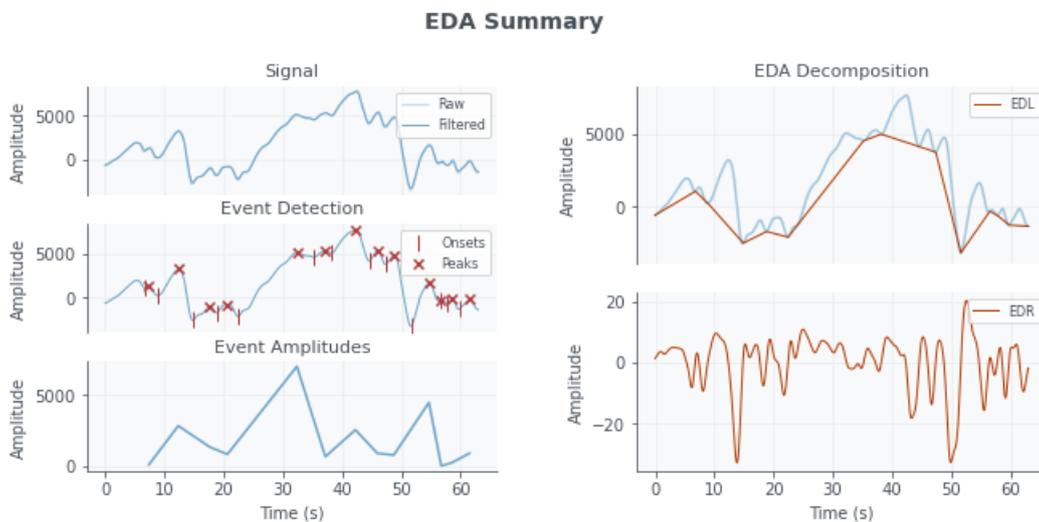


Figura 11. Salida de la función de *BioSPPy* de la señal de EDA. Señal original y filtrada (arriba, izda.), onsets y picos (medio, izda.), amplitudes (abajo, izda.), señal descompuesta (abajo, dcha.).

Uno de los problemas que aparecieron a la hora de procesar esta señal fue que en algunas señales la función de *BioSPPy* no era capaz de detectar los *onsets*, generando un error en Python. Este error se debe a que la función de alguna de las características extraídas solicita la variable que contenía estos onsets, que estaba vacía. Para solucionarlo, se sustituyeron las características de esa muestra por las de la anterior. Bien es cierto que esta medida hace que la variabilidad de los datos disminuya, pero dado el tamaño del conjunto de entrenamiento esto no va a tener gran repercusión y además es bastante fácil de implementar.

3.2.3 Señal de RESP

Las señales de RESP también presentan una deriva en la línea base, por lo que se elimina antes de utilizar la función correspondiente de *BioSPPy*. La salida esta señal ofrece la señal filtrada, con buena calidad, los coeficientes de la señal donde se hace 0 y la función de la frecuencia respiratoria. También se muestra una representación gráfica, mostrada en la Figura 12.

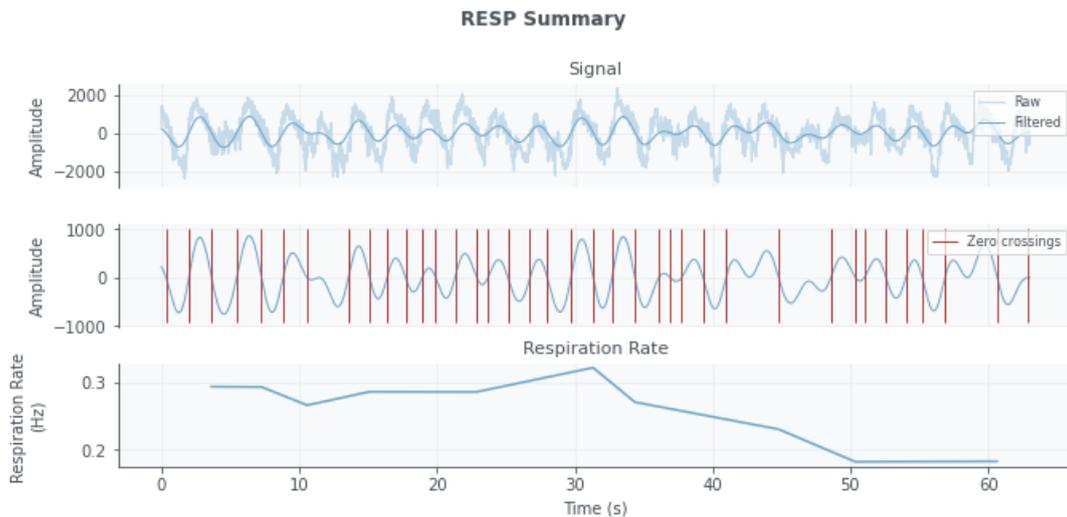


Figura 12. Salida de la función de *BioSPPy* de la señal de RESP. Señal original y filtrada (arriba), detección de ceros (barras rojas) en la señal filtrada (medio) y la frecuencia respiratoria (abajo).

Un ciclo respiratorio está formado por una fase de inspiración (entrada del aire) y otra de espiración (salida del aire). El algoritmo de la función para la señal de RESP de la biblioteca *BioSPPy* no devuelve que coeficientes corresponden a cada una de estas fases, sino, como se ha señalado antes, en que coeficientes empiezan o terminan. Tampoco indican si en esas muestras empieza una parte del ciclo u otra. Así pues, es necesaria una función que permita separar y clasificar los ciclos adecuadamente.

La función parte de crear una matriz de ceros, de longitud la mitad del vector de ceros que devuelve *BioSPPy* (a esta longitud la llamaremos $N/2$). A continuación, se describe un bucle 'for' que compare el valor máximo y el mínimo, en valor absoluto, de medio ciclo, entendiendo por medio ciclo la parte de la señal comprendida entre 2 ceros. Si el máximo es mayor que el valor absoluto del mínimo, la fase es de inspiración, en caso contrario será espiración. La parte final consiste en crear una matriz de 2 columnas y $N/2$ filas, dónde cada elemento de la primera columna representa el inicio de una espiración y el segundo el inicio de la inspiración. Así pues, tenemos los semiciclos correctamente clasificados en espiratorios o inspiratorios, pero ahora queda determinar la longitud de dichos semiciclos.

Para el paso descrito anteriormente, se desarrolla otra función que lo que va a hacer es buscar las posiciones definidas en la función anterior, como inspiración o espiración, y las va a restar apropiadamente para devolver la longitud de cada fase respiratoria.

3.2.4 Señal de SKT

Para esta señal no hace falta ningún tipo de procesado, debido a que no se aprecia ninguna fuente de ruido destacable, ni tampoco existe una señal de *BioSPPy* con la que trabajar por lo que se extraen directamente las características.

3.3 EXTRACCIÓN DE CARACTERÍSTICAS (*FEATURE EXTRACTION*)

Una vez las señales han sido revisadas, procesadas y se ha obtenido información relevante se procede a extraer las características de cada una de ellas. El objetivo de este apartado es reducir la cantidad de información de la que disponemos, resumiéndola en parámetros matemáticos que aporten la mayor cantidad de información posible. Es decir, de cada vídeo se han seleccionado 4 señales con 8064 muestras cada una, lo que hace un total de 32.256 parámetros, y lo que queremos es reducir este número de parámetros a una cantidad mucho menor.

Para llevar a cabo este proceso recurrimos a diversas técnicas que podemos agrupar:

- Características estadísticas: se fundamentan en el cálculo de valores matemáticos que permiten hacer inferencia sobre poblaciones a partir de muestras concretas. Estas características son calculadas de manera común a todas las señales, salvo algunas excepciones.
- Características temporales: estas son individualizadas, es decir, cada señal tiene las suyas propias. Consisten en describir tendencias o patrones, visibles a simple vista y que varían o se repiten con el tiempo.
- Características espectrales: están basadas en el cálculo de la densidad espectral de potencia de una función (PSD, *Power Spectral Density*). Las señales, convencionalmente, se obtienen y se observan en el dominio temporal (Figuras 9, 10, 11, 12), lo que significa que, vemos como la señal captada cambia conforme pasa el tiempo. Sin embargo, existen diversas técnicas matemáticas, como la transformada de Fourier o la transformada de Welch, que permiten dar un giro a la interpretación de estas señales, transformándolas al dominio de la frecuencia. En este contexto, la PSD nos permite medir y representar cómo se distribuye la potencia de la señal en diferentes frecuencias, proporcionando una visión detallada de la influencia de cada frecuencia en la composición total de la señal [42].

Para el cálculo de la transformada se emplea la expresión de síntesis de la transformada rápida de Fourier (FFT, *Fast Fourier Transform*) (1). En este caso, al tratarse de señales discretas se emplea un sumatorio que recorre la señal a lo largo de todas sus muestras.

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-jk\frac{2\pi}{N}n}, \quad k = 0, 1, 2, \dots, N - 1 \quad (2)$$

Fórmula 2. Ecuación de síntesis de la transformada rápida de Fourier (FFT).

$X(k)$ es la función en el dominio de la frecuencia y para su análisis se calcula la potencia por de la señal por bandas de frecuencia, que variarán en función del tipo de señal.

- Características no lineales: Similares a las temporales, pero en este caso se pretende localizar tendencias y patrones que no se pueden ver a simple vista, es decir, se buscan correlaciones internas del conjunto de datos. Al igual que los estadísticos algunas de ellas son comunes a todas las señales, pero no siempre se pueden implementar [43].

3.3.1 Características de la señal de ECG

Como se ha mencionado anteriormente, de cada señal se pueden extraer señales temporales, estadísticas, espectrales y no lineales. En el caso del ECG las características a extraer han sido las mencionadas en [13] y [18], que se describirán más adelante. Cabe destacar que de cada señal existen varias componentes que se pueden extraer y analizar, en este caso nos centraremos en la señal de HRV.

3.3.1.1 Características estadísticas

La señal de ECG en sí no es de nuestro interés, por lo que vamos a calcular las características sobre las distancias calculadas en el apartado 1.1, de esta sección. En este aspecto, se calculó:

- **Media:** indica el valor promedio de la frecuencia cardíaca. Cuanto más bajo sea el valor significa que los latidos son más cortos y, por tanto, el corazón late más deprisa. La media se calcula con la función de *numpy* (como en el caso de *matplotlib*, *numpy* será definido como `np`) '*np.mean*', que recibe como argumento de entrada las distancias.
- **Desviación estándar:** en este caso se mide la variabilidad de la frecuencia cardíaca (HRV), que sirve para analizar la salud del corazón. La HRV será analizada más en profundidad más adelante, ya que es una función propia que aporta información relevante. Se calcula mediante la función '*np.std*' que recibe como argumento de entrada la distancia.
- **Mediana:** aporta información adicional a la media, evitando el sesgo de los valores extremos. La función para calcularla es '*np.median*' que recibe por entrada la misma que en los casos anteriores.
- **Moda:** debido a que lo que aporta es información sobre cuál es el valor más repetido se pueden detectar patrones. En este caso, es necesaria la librería *scipy.stats* para calcularla, a través de la función '*stats.mode*'. Como argumento de entrada serán necesarias las distancias, pero además hay que tener en cuenta que la función te devuelve el valor de la moda y el número de veces que se repite, por lo que habrá que indicar con un 0 que el valor de interés es el primero.
- **Rango intercuartílico (IQR):** es la diferencia entre el primer y tercer cuartil, 25% de valores más altos y más bajos, respectivamente. Con esto se consigue captar parte de la variabilidad de los datos. Para el cálculo se recurre a la función de *numpy* '*np.percentile*', a la que hay que indicarle el valor del percentil que se quiere calcular (25 y 75 en este caso) y posteriormente hay que restar las salidas de la función.
- **Asimetría (*skewness*):** indica la simetría de la distribución, de manera que valores positivos indica que la señal está sesgada hacia la derecha, lo que significa que predominan las secuencias cortas. Para calcularla, se emplea la función de *scipy.stats* '*stats.skew*' que utiliza las distancias como entrada.
- **Curtois (*kurtosis*):** es una medida de como de apuntada está la señal, de manera que valores más altos indican que los extremos tienen mayor relevancia, por lo que se podrían detectar más valores anómalos que en una distribución más normal. Para calcularla se emplea una función muy similar a la anterior denominada: '*stats.kurtosis*'.
- **Varianza:** se trata de la raíz cuadrada de la desviación estándar por lo que la información que aporta es bastante parecida y se calcula con la función '*np.var*'.
- **RMS (*Root Mean Square*):** La RMS es una medida de la magnitud promedio de los intervalos RR, elevando cada intervalo al cuadrado antes de promediar y luego sacando la raíz cuadrada. Este valor proporciona una medida de la magnitud promedio que da más peso a las variaciones más grandes, lo que puede resaltar episodios de mayor

irregularidad en los intervalos RR. Para calcularlo introducimos la fórmula manualmente (3).

$$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad (3)$$

- **RMSSD (*Root Mean Square of Successive Differences*):** es la raíz cuadrada de la media de las diferencias al cuadrado entre intervalos RR sucesivos. Es una medida estándar de la variabilidad de la frecuencia cardíaca que refleja la actividad del sistema nervioso autónomo, especialmente el componente parasimpático. Un valor alto indica una mayor variabilidad de la frecuencia cardíaca, lo que generalmente es saludable. Al igual que en el caso anterior, la fórmula se introduce manualmente (4).

$$RMSSD = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (x_{i+1} - x_i)^2} \quad (4)$$

- **Índice Triangular:** el índice triangular es el número total de intervalos RR dividido por el valor modal del histograma de los intervalos RR. Este índice es una medida de la distribución de los intervalos RR. Un índice más alto sugiere una mayor dispersión en los intervalos RR, lo que puede ser interpretado como un indicador de una mayor variabilidad y, potencialmente, de un sistema cardiovascular más adaptable.

3.3.1.2 Características Temporales

Las características temporales son aquellas que aportan información en el dominio temporal. Esto es, se analizan las señales y se buscan patrones, detalles o secuencias que aporten alguna información relevante. Por lo general, estas características se obtienen realizando operaciones sencillas, como sumas acumuladas, mediciones o comparaciones periódicas. Las características extraídas en este caso fueron:

- **NN20:** para obtener este parámetro se deben medir, antes de nada, las diferencias entre dos intervalos RR consecutivos. En el apartado 3.2 de esta sección se habla de cómo calcular el tiempo que transcurre entre 2 picos R consecutivos, es decir, como calcular los intervalos RR. Ahora, vamos a coger esos tiempos o distancias y vamos a restar aquellos que sean adyacentes, como en el caso anterior. Para realizar este procedimiento utilizaremos un código bastante similar. El resultado será un vector cuyos elementos expresan la variación de la distancia de un pico R respecto del siguiente. Esto es, de qué manera varía la frecuencia cardíaca entre 2 ciclos cardíacos.
El objetivo de este proceso es contar cuantas veces 2 intervalos RR distan en más de 20ms. Esta medida aporta información acerca de cómo un estímulo puede provocar cambios en la frecuencia cardíaca. De esta manera, si el número es elevado, significa que la variabilidad “intra-latido” es alta. Dicho con otras palabras, el vídeo que se está visualizando en el momento genera cambios rápidos en el estado emocional del sujeto, es decir, que se suceden diferentes estímulos.
- **NN50:** esta característica es igual que la anterior, solo que, en este caso, en vez de buscar el número de veces de que las diferencias entre distancias son superiores a 20ms buscamos aquellas superiores a 50ms. Esta distancia, al ser mayor, sugiere cambios más bruscos en el estado emocional, lo que se puede explicar si el vídeo contiene estímulos chocantes como sustos.

- **pNN20:** en este caso calculamos que proporción de diferencias de los intervalos RR mayores a 20ms hay, respecto del total. Para obtener el valor dividimos NN20 entre la longitud del vector de diferencias. El resultado aporta la misma información que el NN20, pero en este caso ofrece una visión cuantitativa de dicho valor. Es decir, permite observar si realmente el valor es elevado o no.
- **pNN50:** se calcula de la misma forma que la característica anterior y aporta la misma información, pero referida al parámetro NN50.
- **pNN50/pNN20:** esta característica mide la relación entre las proporciones anteriores. El valor obtenido puede estar comprendido entre 0 y 1. Si se acerca a 1 quiere decir que hay muchos eventos superiores a 50ms. La importancia de este parámetro reside en la comparación directa de las diferencias temporales de los picos R. Hay que tener en cuenta que cada evento superior a 50ms es también superior a 20ms, por lo que gracias a esta característica se comprueba cual es el peso real del NN50, permitiendo descifrar información acerca de si el vídeo contiene estímulos chocantes o no, que será así en el caso de tener un valor cercano a 1.
- **Valores mínimos y máximos:** para finalizar con el apartado de las características temporales que emplean las distancias de los picos R, calculamos cuales son los valores máximos y mínimos, tanto de las distancias temporales como de sus variaciones. Para ello se utilizan las funciones nativas de *Python* ‘*min*’ y ‘*max*’. Esta información ayuda a dimensionar la magnitud de los datos que estamos tratando. De manera que, valores muy extremos significan la variabilidad total es muy elevada y que, por tanto, los estímulos recibidos fueron muy intensos.
- **TINN (Triangular Interpolation of the NN Interval Histogram):** es una métrica utilizada para evaluar la variabilidad de los intervalos RR en una señal de ECG. Se calcula a partir del histograma de los intervalos RR. Primero, se genera un histograma, que representa la frecuencia de los diferentes intervalos RR, cuantas veces se repite un valor concreto. Luego, se identifica el bin modal, es decir, el intervalo RR más frecuente. A partir de este punto, se realiza una interpolación lineal a ambos lados del pico modal para encontrar los puntos de intersección con la base del histograma. El TINN se obtiene midiendo la distancia entre estos dos puntos de la base, reflejando así la amplitud del triángulo formado. Un TINN más amplio sugiere una mayor variabilidad cardíaca, lo cual es generalmente un indicador de buena capacidad de respuesta adaptativa del corazón. Por el contrario, un TINN reducido puede reflejar un aumento del estrés o una disminución en la regulación autónoma, asociada a emociones negativas como la ansiedad o la depresión. Así, el TINN proporciona una medida indirecta pero valiosa del impacto de las emociones en la función cardíaca.

3.3.1.3 Características Espectrales

Como se había comentado con anterioridad, el análisis espectral de una señal suponía un cambio en la interpretación de esta, pues permitía verla desde otra perspectiva. En el caso del ECG necesitamos calcular una señal previa, la señal de variabilidad de la frecuencia cardíaca (HRV, *Heart Rate Variability*). Hasta ahora, se había trabajado con los intervalos RR y con las diferencias entre estos directamente. En este momento, se obtiene una nueva señal que permite ver de qué forma varía la frecuencia cardíaca con el paso del tiempo, influenciada por el estado emocional. Las variaciones las pueden inducir otros factores como la respiración, lo que podría guardar cierta relación con el estado emocional, pero nos centraremos en analizar los que son provocados directamente por las emociones [44]. Para obtener la HRV se interpolan los intervalos RR, mediante la función ‘*interp1d*’ de la librería *SciPy* [45]. El resultado se puede ver en la Figura 14.

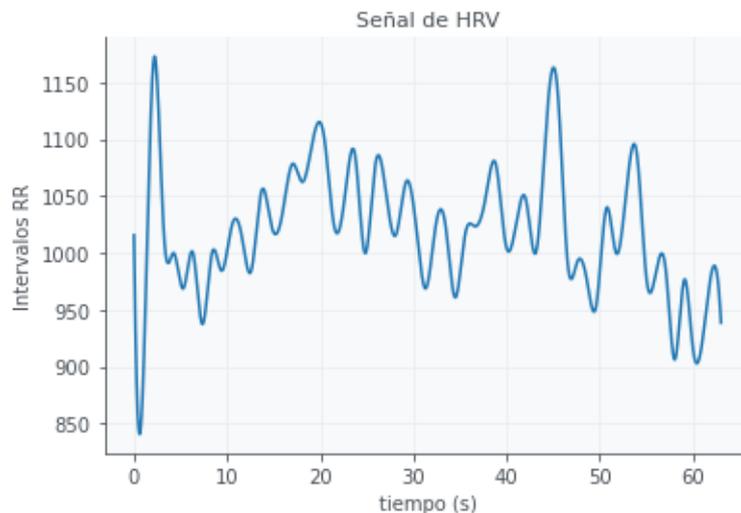


Figura 14. Señal de HRV.

Una vez se ha obtenido la señal se utiliza la FFT (2) para obtener su transformada en el dominio de la frecuencia, también se probó la transformada Welch, pero con peores resultados. Para ello, se emplea la función `'np.fft.fft'` y `'np.fft.fftfreq'`. La primera, devuelve los valores de la señal transformada y la segunda, devuelve las frecuencias a las que se dan esos valores, estas constituirán el eje de abscisas. Un problema que aparece es que las funciones anteriormente mencionadas, devuelven la señal duplicada, lo que significa que proporcionan un mismo valor para 2 frecuencias idénticas, pero cambiadas de signo. La solución a este problema es sencilla, pues basta con excluir aquellos valores negativos, que solo aportan información redundante, mediante el comando `'abs'`, que elimina, tanto los valores negativos, como los no reales de la entrada de la función. Posteriormente, se representa la señal para ver su forma (Figura 15).

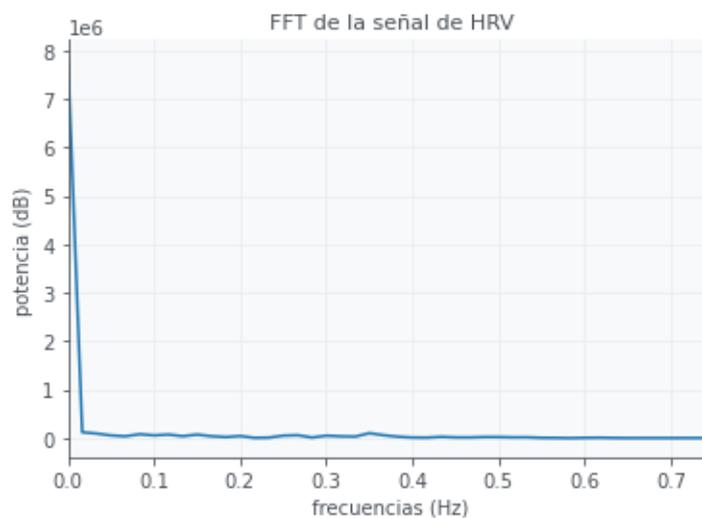


Figura 15. FFT de la señal de HRV.

En la representación obtenida se puede ver que en la FFT de esta señal predominan las bajas frecuencias, pues se ve un gran pico en torno al 0. Por lo demás, es una señal bastante lisa, con algunos pequeños picos por debajo de los 0.4 Hz y después se aplana hasta los 64 Hz, por lo que conviene limitar el eje x para poder visualizarla mejor. En este caso, se limita entre 0 y 0.75 Hz.

Ahora toca extraer información relevante de la FFT, por lo que se calcula la potencia de una serie de bandas de interés. La potencia se puede definir como el área bajo la curva de la señal

comprendida entre dos frecuencias concretas, o sea, las bandas de interés. Estas están definidas en [18], [33], [46] y se dividen en 3:

- **Muy baja frecuencia (VLF, *Very Low Frequency*):** está comprendida entre los 0 y los 0.04 Hz, coincidiendo con el pico máximo de la señal. Esta va a ser la banda que acumule una mayor potencia, porque se asocia a los mecanismos de regulación, teniendo en cuenta que la HRV es la regulación de la frecuencia cardíaca en respuesta a un estímulo.
- **Baja frecuencia (LF, *Low Frequency*):** entre los 0.04 y los 0.15 Hz, se asocia a la influencia del SN simpático.
- **Alta frecuencia (HF, *High Frequency*):** esta banda está comprendida entre los 0.15 y 0.4 Hz, coincidiendo con ese valor que habíamos resaltado como el final de las ondulaciones de la FFT, es decir, dónde acaba la información relevante. Esta banda se asocia a la activación del SN parasimpático

De las bandas de frecuencia sacamos una serie de características, algunas de ellas rápidas como la potencia, que se calcula como la suma acumulada de los elementos de la FFT comprendidos entre las frecuencias máxima y mínima de la banda, el máximo de la banda y la proporción de dicha banda respecto del total de la señal.

Otras características que se pueden extraer son las relaciones entre las potencias. En concreto se calcula:

- **Ratio LF/HF:** que permite comparar las influencias del SN simpático y parasimpático entre sí.
- **Potencia conjunta LF y HF:** permite diferenciar entre la actividad del SN y la propia activación de modulación del corazón que se representaba a través de la VLF.
- **Potencia total de las bandas de interés:** se suma VLF, LF y HF lo que permite valorar que cantidad de la FFT es relevante.

Por último, se normaliza la LF y HF, dividiendo cada una de ellas entre la suma conjunta.

3.3.1.4 Características no Lineales

Estas son métricas que capturan patrones complejos e impredecibles que no pueden ser descritos por métodos lineales. Estas características permiten entender mejor las dinámicas de los sistemas fisiológicos. Es decir, hacemos un análisis temporal, similar al del punto 2.1.2, pero realizando cálculos complejos que ayuden a descifrar información oculta a simple vista. En este contexto se extrajeron las siguientes características:

- **Plot de Poincaré SD1:** representa la desviación estándar de los puntos perpendiculares a la línea de identidad en un diagrama de Poincaré, esta es una herramienta gráfica utilizada para analizar la variabilidad de señales fisiológicas, especialmente la de HRV. Esto significa que el diagrama de Poincaré es un tipo de representación gráfica que muestra la relación entre valores consecutivos de una señal, por lo general, intervalos RR. Refleja la variabilidad a corto plazo en la señal, lo que puede estar relacionado con respuestas emocionales rápidas [18].
- **Plot de Poincaré SD2:** en este caso se evalúa la desviación estándar de los puntos a lo largo de la línea de identidad del diagrama de Poincaré. Refleja la variabilidad a largo plazo, que puede estar asociada con tendencias más persistentes en el estado emocional. Ambas métricas permiten distinguir entre las emociones pasajeras, evocadas por los estímulos de los vídeos, y aquellas que no varían, o por lo menos no lo hacen tan deprisa, que se podrían asociar con el estado emocional de la persona a largo plazo, el que no está sujeto a un estímulo puntual [18].

- **Sample entropy (SampEn):** mide la complejidad de una señal, evaluando la probabilidad de que patrones similares de datos permanezcan similares en la siguiente evolución temporal. El valor de esta característica permite clasificar algunos tipos de emociones. De esta forma, una baja entropía indica una señal más predecible, mientras que una alta entropía sugiere un sistema más caótico. Esto es relevante en la detección de emociones, ya que estados emocionales como la ansiedad pueden estar asociados con un mayor desorden [18].
- **Approximate Entropy (ApEn):** similar a la SampEn, mide la regularidad y la impredecibilidad de fluctuaciones en una señal, pero es más sensible a la longitud de los datos y puede sobreestimar la complejidad. Así pues, la ApEn se utiliza para evaluar la irregularidad de la señal; una mayor ApEn puede indicar una mayor variabilidad emocional o estado [1], en el que se expone, por.
- **Detrended Fluctuation Analysis (DFA):** la DFA es un método para cuantificar las correlaciones a largo plazo en señales fisiológicas. Evalúa cómo las fluctuaciones de la señal cambian con diferentes escalas de tiempo después de eliminar tendencias. Los coeficientes alfa (α_1 para correlaciones a corto plazo y α_2 para largo plazo) indican si la señal es más caótica o persistente, lo cual es útil para diferenciar entre estados emocionales como calma y ansiedad [43].
- **DFA Alpha 1 (α_1):** Representa la escala de correlación a corto plazo (menor a 11 latidos en señales de variabilidad del ritmo cardíaco).
- **DFA Alpha 2 (α_2):** Representa la escala de correlación a largo plazo (mayor a 11 latidos). Valores de α_1 y α_2 indican la presencia de autocorrelaciones en diferentes escalas temporales, lo que permite clasificar estados emocionales basados en la dinámica de la señal.

A continuación, se presenta la Tabla 3 donde se hace un resumen de las diferentes características que se han extraído. En la tabla se incluye una abreviatura para estas y la señal o componente de las que se extraen.

NOMBRE	ABREVIATURA	SEÑAL
Media	Mean_ECG	Intervalos RR
Desviación estándar	Std_ECG	Intervalos RR
Mediana	Med_ECG	Intervalos RR
Moda	Mod_ECG	Intervalos RR
Rango intercuartílico	Iqr_ECG	Intervalos RR
Asimetría	Skew_ECG	Intervalos RR
Curtosis	Kurt_ECG	Intervalos RR
Varianza	Var_ECG	Intervalos RR
RMS	RMS_ECG	Intervalos RR
RMSSD	RMSSD_ECG	Intervalos RR
Índice triangular	Triang_ECG	Intervalos RR
NN20	NN20	Diferencia entre intervalos RR consecutivos
NN50	NN50	Diferencia entre intervalos RR consecutivos
pNN20	pNN20	Diferencia entre intervalos RR consecutivos
pNN50	pNN50	Diferencia entre intervalos RR consecutivos
pNN50/pNN20	50/20	Diferencia entre intervalos RR consecutivos
Máximo intervalo RR	Max_RR	Diferencia entre intervalos RR consecutivos
Mínimo intervalo RR	Min RR	Diferencia entre intervalos RR consecutivos
TINN	Tinn	Intervalos RR
Potencia	Power_ECG	FFT de HRV (VLF, LF, HF)
Máximo de banda	Max_ban_ECG	FFT de HRV (VLF, LF, HF)
Mínimo de banda	Min_ban_ECG	FFT de HRV (VLF, LF, HF)
Proporción de banda	Prop_ECG	FFT de HRV (VLF, LF, HF)
Ratio LF/HF	LF/HF	FFT de HRV
Potencia conjunta LF+HF	LF+HF	FFT de HRV
Potencia total bandas	Tot_power	FFT de HRV
Plot de Poincaré SD1	SD1	Intervalos RR
Plot de Poincaré SD2	SD2	Intervalos RR
Sample entropy	SampEn	Intervalos RR
Approximate entropy	ApEn	Intervalos RR
Detrended fluctuation analysis	DFA	HRV
DFA α 1	DFA1	HRV
DFA α 2	DFA2	HRV

Tabla 3. Características extraídas de la señal de ECG con su abreviatura asignada y la señal de la que se extrae.

3.3.2 Características de la señal de EDA

Al igual que con la señal de ECG, en este caso vamos a extraer características temporales, espectrales y estadísticas, aparte de estudiar la linealidad de la componente de SCL. Además, habrá que tener en cuenta que esta señal se puede dividir en varias partes, como se mostró en el apartado 1. De esta circunstancia debemos tener en cuenta que, si extrajáramos las mismas características de todos los componentes de estas señales tendríamos un elevado número de características, lo que elevaría el coste computacional y la complejidad del problema, por lo que deberemos elegir qué características extraeremos.

Las características que se repitan respecto del ECG no van a ser explicadas a fondo, simplemente se citará la información aportada. Hay que destacar que la información desarrollada a continuación fue consultada principalmente en [18] y [33].

3.3.2.1 Características de linealidad

Para comenzar es necesario calcular la componente basal de la señal de EDA, la SCL (*Skin Conductance Level*). Esta aporta información sobre el nivel de sudoración que no es fruto de un estímulo concreto y como consecuencia es de variación lenta. Lo que vamos a conseguir con esto es analizar la morfología de la señal de EDA, pudiendo distinguir lo que corresponde a los estímulos y lo que no [17].

Para calcular la SCL utilizamos un filtro pasabajo ‘*Butterworth*’ de primer orden y frecuencia de corte 0,05 Hz, como se ha comentado la señal de SCL es de variación lenta por lo que nos quedamos solo con las bajas frecuencias. Posteriormente, calculamos los siguientes parámetros: R^2 , MSE (*Mean Square Error*), MAE (*Mean Absolute Error*) y RMSE (*Root Mean Square Error*), estas métricas serán explicadas en el apartado de resultados, pues son las que se utilizan para evaluar el rendimiento del modelo. Estas 4 características se basan en comprobar como de parecidos son las señales de EDA y SCL, en el caso de R^2 y en el caso de los otros lo que se hace es ver como de distintos son, es decir, en el primer caso cuanto mayor sea el valor más parecidas serán y en el segundo caso al revés. Lógicamente, cuanto más parecidas sean quiere decir que los estímulos provocan una reacción menor en los sujetos [47].

3.3.2.2 Características Estadísticas

Como se ha comentado anteriormente, para el cálculo de los estadísticos no vamos a tener solamente en cuenta la señal de EDA filtrada, sino que vamos a utilizar también la señal de las amplitudes y de EDR. En este último caso vamos a sacar alguna característica diferente.

- **Media:** indica el nivel electrodermal general, a lo largo del tiempo. Permite estimar como es la señal en un solo valor.
- **Mediana:** es similar a la media, pero sin incluir los valores extremos, por lo que se presenta como una medida más robusta. Especialmente representativa de las amplitudes, aunque se calcula en las otras 2 señales.
- **Desviación estándar:** al representar el grado de variación de la señal, resulta especialmente útil en vídeos en los que el estímulo principal busca generar excitación en el sujeto.
- **Moda:** al igual que la mediana es bastante utilizada en la señal de amplitudes.
- **Curtosis:** las componentes de amplitudes y EDR presentan regiones bastante apuntadas, por lo que la curtosis es una buena forma de ver como son esas regiones y que influencia tienen sobre la señal.
- **Asimetría:** una simetría positiva sugiere respuestas bajas con picos ocasionales altos (eventos de estrés), mientras que una simetría negativa indica una actividad sostenida alta con algunas respuestas bajas. Valores cercanos a cero reflejan una distribución equilibrada de las respuestas electrodérmicas.
- **Varianza:** indica fluctuaciones relevantes en la actividad, es decir, en qué forma se aleja la señal de su valor promedio.
- **RMS:** indica la intensidad promedio de la señal.

Por último, se calculan los valores máximos y mínimos de las señales, que al igual que en la señal de ECG y sus componentes permiten ver las dimensiones de las señales observando el rango de valores que las componen.

A parte de estas características, para la EDR se extraen 2 a mayores [30]. La primera de ellas consiste en calcular la media de los valores negativos de la señal. Para ello localizamos aquellos valores menores que 0 con *'np.where'*, a continuación se suman y por último se dividen entre el número total de estos. La otra característica consiste en calcular la proporción de valores negativos que hay respecto del total, es decir, dividir la longitud del vector de valores negativos entre el número de muestras totales.

3.3.2.3 Características temporales

A diferencia de las estadísticas, todas las características temporales fueron extraídas sobre la señal filtrada de EDA y se basaron fundamentalmente en el cálculo de tiempos.

- **Tiempo hasta que se alcanza el pico máximo de la señal:** detectamos el pico máximo de la señal con el comando *'max'* y después localizamos su posición dentro de la señal con *'np.where'*. Con esto conseguimos detectar en qué momento se produce el estímulo más potente de la señal y su magnitud. A mayores tendremos que multiplicar por el factor de muestreo (1) que habíamos calculado, pues el resultado del bucle es la posición dentro de la señal y no el tiempo. Con esto, lo que se puede ver es la reacción del sujeto, comparando la respuesta con el resto. En vídeos más relajados este valor será menor y se encontrará más aleatoriamente repartido dentro de la señal, que en aquellos vídeos donde los estímulos están preparados a conciencia, es decir, sean muy intensos y en momentos concretos.
- **Tiempo que tarda la señal en decaer un 50% del pico máximo (*Recovery time 50*):** medimos el tiempo que pasa hasta que el valor de la señal es igual o inferior a la mitad del pico máximo. Una vez detectado el máximo calculamos el 50% de este, multiplicando por 0,5, y lo buscamos con un bucle que recorre la señal y cuando localiza un valor igual o inferior a este sale del bucle, quedando registrado en una variable el último valor, que es el que deseamos. Al igual que en el caso anterior deberemos multiplicar el resultado por el factor de muestreo. Esta característica informa a cerca de como el sujeto se recompone del estímulo principal. Si el tiempo es largo, significa que el vídeo tiene mayor actividad emocional que impide que el valor caiga. Por el contrario, valores más pequeños suponen que el vídeo tiene pocos estímulos, seguramente concentrados al final. Otra característica es el *recovery time 60* que es lo mismo que este, pero con una caída hasta 60%.
- **Área bajo la curva de la señal, a partir del inicio del primer evento:** el primer evento coincide con el primer elemento del vector de *onsets*, por lo que se calcula el área bajo la curva de la señal a partir de este punto. Dependiendo del valor también podemos obtener información acerca de los estímulos. Valores altos suponen que estos son bastante sostenidos en el tiempo, la señal no se limita a un único evento sino a varios.
- **Otras características:** también se extrajeron otras características que aportan información sobre la morfología de la señal como son: el número de eventos, la suma de los tiempos que se tarda en alcanzar el máximo de un *onset* y la diferencia entre el valor de la señal al principio del evento y cuando alcanza el pico máximo del *onset*. Este último valor informa sobre como de larga es la subida, como de intenso es el estímulo.

3.3.2.4 Características espectrales

Para calcular la señal de EDR en el dominio de la frecuencia utilizamos el mismo método que en el caso del ECG, es decir la FFT (Figura 16). Transformamos la EDR y no la EDA, porque la primera tiene el componente basal de la señal, el que no se corresponde con los estímulos, que aportaría información irrelevante.

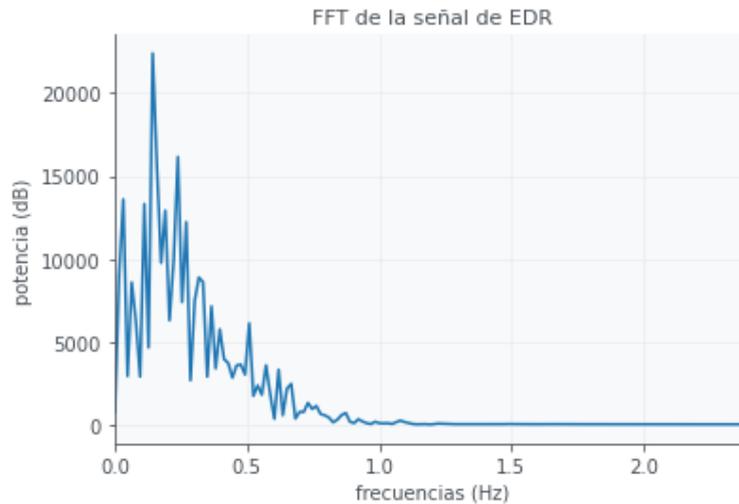


Figura 16. FFT de la señal de EDR.

En este caso también hay 3 bandas de frecuencia principales.

- **Respuesta muy lenta de la conductancia de la piel (SCVSR, *skin conductance very slow response*):** entre 0 y 0.08 Hz. Se asocia cambios en la impedancia lentos, lo que significa que, es un reducto de la respuesta basal de la EDR, una parte que no se pudo separar de la EDL.
- **Respuesta lenta de la conductancia de la piel (SCSR, *skin conductance slow response*):** entre 0 y 0.2 Hz. Sigue siendo una frecuencia bastante baja, pero en este caso la información que se aporta a mayores es la capacidad de adaptación del sujeto a un estímulo concreto. Como se ve en la Figura 16 esta parte de la señal tiene mayor potencia que la anterior.
- **Banda principal:** entre 0 y 2.4 Hz. No solo contiene la información anterior, sino que además contiene la frecuencia de los cambios rápidos de la señal producidos por los estímulos.

Además de la potencia, potencia relativa, potencia relativa, etc. que se calculó, de la misma manera para el ECG, ahora se calculan otras nuevas características espectrales más complejas. Estas son:

- **Entropía espectral (*spectral entropy*):** esta es una medida del grado de desorden de una señal o, en este caso, una distribución de potencia. Esto significa que se va a cuantificar de qué manera se distribuye la propia señal (su potencia) a lo largo del rango de frecuencias. De manera que, a menor entropía mayor orden. Esto es debido a que la potencia se acumula en un rango de frecuencias pequeño y localizado. Por el contrario, si el valor fuese alto, supondría que la potencia se dispersa por un conjunto más amplio de frecuencias. La información que se puede extraer de esta característica es la capacidad de respuesta del SNA, es decir, como se adapta a un estímulo concreto. Para calcularla se emplea la siguiente fórmula.

$$H = - \sum_{i=1}^N P(f_i) \log(P(f_i)) \quad (5)$$

donde P representa la distribución de potencia, la FFT de la señal de EDR y la N es el rango de frecuencias.

- **Distribución de potencia (deviation):** es una métrica de evaluación de la desviación estándar de la señal. Por lo general, informa de a qué tipo de estímulos ha sido sometido el paciente, es decir, un valor elevado supone que estos son rápidos y repetitivos, mientras que valores más bajos representan una respuesta más uniforme o específica.
- **Asimetría:** la información aportada es bastante similar a la expuesta en el apartado 2.2.2.
- **Curtosis:** también es similar a la explicada en el apartado 2.2.2, pero en este caso hay que tener en cuenta a mayores que al haberse eliminado la información procedente de la respuesta basal, la curtosis va a indicar de qué manera el sujeto reacciona ante los estímulos.
- **Frecuencia dominante:** revela la frecuencia en la que ocurre la mayor actividad en la señal EDR. Una frecuencia dominante más alta puede asociarse con respuestas rápidas y agudas a estímulos, mientras que una más baja puede estar relacionada con respuestas más lentas o sostenidas, proporcionando información sobre el tipo de activación fisiológica.
- **Magnitud media de la señal:** indicador de la fuerza o intensidad general de la actividad electrodérmica en el dominio de la frecuencia. Un RMS alto puede sugerir que hay alta reactividad del sistema nervioso simpático, reflejando potencialmente altos niveles de excitación o estrés.
- **Ancho de banda de potencia (bandwidth):** Indica la gama de frecuencias donde se concentra la mayor parte de la potencia de la señal. Un ancho de banda más amplio sugiere una respuesta fisiológica más compleja o variada, mientras que uno más estrecho puede reflejar una respuesta más focalizada a ciertos estímulos.

A continuación, se presenta la Tabla 4 donde se hace un resumen de las diferentes características que se han extraído para la señal de EDA. En la tabla se incluye una abreviatura para estas y la señal o componente de las que se extraen.

NOMBRE	ABREVIATURA	SEÑAL
Media	Mean_EDA	EDA, EDR, Amplitudes
Desviación estándar	Std_EDA	EDA, EDR, Amplitudes
Mediana	Med_EDA	EDA, EDR, Amplitudes
Moda	Mod_EDA	EDA, EDR, Amplitudes
Rango intercuartílico	Iqr_EDA	EDA, EDR, Amplitudes
Asimetría	Skew_EDA	EDA, EDR, Amplitudes
Curtosis	Kurt_EDA	EDA, EDR, Amplitudes
Varianza	Var_EDA	EDA, EDR, Amplitudes
RMS	RMS_EDA	EDA, EDR, Amplitudes
RMSSD	RMSSD_EDA	EDA, EDR, Amplitudes
Media de valores negativos	-Mean	EDR
Proporción valores negativos	-Prop	EDR
Tiempo pico máximo	Tpmax	EDA
Tiempo señal <50% pico máximo	Tpmax_50	EDA
Integral después del primer evento	Int	EDA
Características morfológicas	morfo	EDA
Potencia	Power_EDA	FFT de EDA (SCVSR, SCSR)
Máximo de banda	Max_ban_EDA	FFT de EDA (SCVSR, SCSR)
Mínimo de banda	Min_ban_EDA	FFT de EDA (SCVSR, SCSR)
Proporción de banda	Prop_EDA	FFT de EDA (SCVSR, SCSR)
Entropía espectral	SpectEntro	FFT de EDA (main)
Distribución de potencia	PowerDist	FFT de EDA (main)
Asimetría	Skew_FFT_EDA	FFT de EDA (main)
Curtosis	Kurt_FFT_EDA	FFT de EDA (main)
Frecuencia dominante	MainFreq	FFT de EDA (main)
Magnitud media de la señal	MeanMag	FFT de EDA (main)
Ancho de potencia	Bandwidth	FFT de EDA (main)

Tabla 4. Características extraídas de la señal de EDA con su abreviatura asignada y la señal de la que se extrae.

3.3.3 Características de la señal de RESP

Al igual que en los casos anteriores las características extraídas serán temporales, estadísticas y espectrales. Las características son extraídas tanto de la señal principal, como de su frecuencia y de los ciclos respiratorios, enteros y separados en fases.

3.3.3.1 Características estadísticas

El número de estas características va a ser bastante elevado, porque como se ha mencionado en el párrafo anterior se utilizan 5 variables distintas por lo que habrá hasta 60 características.

- **Media:** refleja el nivel promedio de la respiración. Un cambio en la media puede indicar una variación en el ritmo respiratorio, asociado a diferentes estados emocionales.
- **Desviación estándar:** mide la variabilidad de la respiración. Una alta desviación estándar sugiere variaciones significativas en la amplitud respiratoria, lo que podría asociarse con emociones intensas o cambios abruptos en el estado emocional.

- **Mediana:** proporciona información sobre la tendencia central de la señal de RESP, menos sensible a los picos extremos que la media. Es útil para analizar respiraciones consistentes frente a respiraciones más esporádicas o irregulares.
- **Moda:** indica la frecuencia de la amplitud respiratoria más común. Puede ser útil para identificar patrones repetitivos en la respiración, como respiraciones rápidas o lentas en estados emocionales específicos.
- **Rango intercuartílico (IQR):** mide la dispersión central de los datos, proporcionando una idea de la variabilidad en la respiración sin ser afectado por valores atípicos. Una mayor variabilidad podría indicar estrés o excitación.
- **Asimetría:** proporciona información sobre la simetría de la señal respiratoria. Una asimetría positiva o negativa puede estar relacionada con un predominio de respiraciones profundas o superficiales, respectivamente, indicativas de estados emocionales como relajación o ansiedad.
- **Curtosis:** indica la como de apuntada está la distribución de la señal respiratoria. Una alta curtosis sugiere la presencia de picos o respiraciones anómalas que podrían asociarse con respuestas emocionales extremas o de alerta.
- **Varianza:** similar a la desviación estándar, mide la dispersión de la señal. Una alta varianza en la respiración puede sugerir cambios rápidos en el estado emocional.
- **RMSE:** proporciona una medida de la magnitud de la señal respiratoria. Es útil para evaluar la intensidad global de la respiración y sugiere cambios en el esfuerzo respiratorio relacionados con diferentes estados emocionales.
- **RMS:** similar al RMSE, puede utilizarse para evaluar la energía total de la señal respiratoria, que podría correlacionarse con niveles de excitación o calma.
- **Valor mínimo y máximo:** indican el nivel más bajo y más alto de la señal respiratoria. El primero de ellos puede revelar momentos de respiración superficial o apneas, asociadas a estados de relajación profunda o estrés intenso. Por el contrario, el pico máximo, se asocia a respiraciones profundas o forzadas, características del estrés o el miedo.

3.3.3.2 Características temporales

En el caso de esta señal las características temporales están especialmente relacionadas con los ciclos y sus partes, estableciéndose comparación entre ellos y midiendo algunas partes concretas.

- **Proporción de las fases:** las fases de los ciclos respiratorios no son iguales, no duran lo mismo. De esta manera, se pretende comprobar que influencia tiene cada semiciclo. Para ello, se calcula la proporción de la fase inspiratoria y espiratoria respecto al ciclo al que pertenecen, esto es, la división de la longitud de cada semiciclo entre la longitud total del ciclo. Posteriormente, se calcula la media de todas las proporciones, separando las inspiratorias de las espiratorias. Por otro lado, se suman todos los tiempos de las inspiraciones y se dividen entre el tiempo total del registro. Se hace lo mismo con las espiraciones. Los resultados, aunque similares, no coinciden. Como se distribuyen estas fases respiratorias aporta bastante información acerca de cómo un estímulo puede influir en un sujeto. De manera que, una fase inspiratoria muy larga puede suponer estados de relajación y, por el contrario, una fase inspiratoria corta supone miedo o ansiedad. Este parámetro es un buen indicador de la *arousal* de un estado emocional, porque representa si una persona se encuentra excitada o no ante un estímulo [48].
- **Amplitud de las fases:** es la distancia que hay entre el inicio de la fase respiratoria, en 0, y el punto mínimo de la espiratoria y el máximo de la espiratoria. Esto en combinación es, la profundidad del ciclo respiratorio. A mayor valor quiere decir que

las respiraciones son más profundas, hay mayor grado de relajación que si fueran más ligeras [48].

- **Otras características:** aparte de las anteriores, calculamos algunas otras características menos relevantes, como pueden ser: el número de ciclos y los valores máximos y mínimos de las proporciones calculadas anteriormente. Estas no aportan información muy específica, pero ponen en contexto la magnitud de las señales.

3.3.3.3 Características espectrales

Al igual que en los casos anteriores, utilizamos la FFT para transformar la señal de RESP al dominio frecuencial, el resultado se puede ver en la Figura 17.

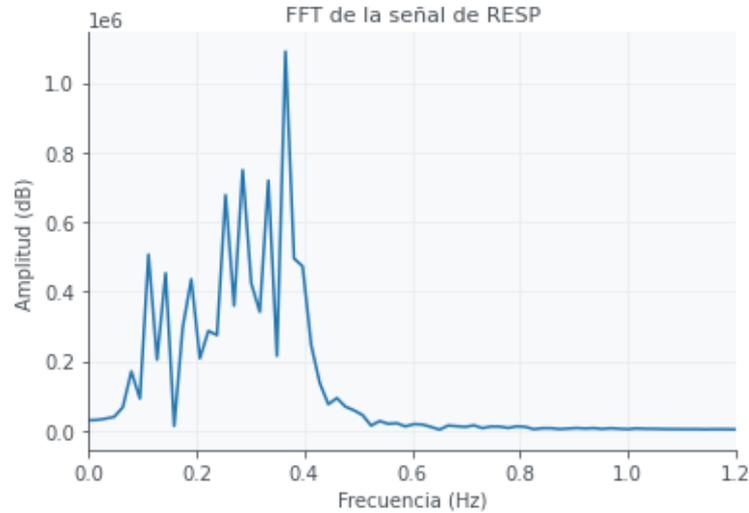


Figura 17. FFT de la señal de Resp

Las bandas de potencia que se calculan en este punto son 2:

- **Baja frecuencia (LF, Low Frequency):** comprendida entre 0 Hz y 0.25 Hz. En la señal de RESP están relacionadas principalmente con el ritmo respiratorio más lento y las variaciones más amplias de la respiración. Esta banda de frecuencia puede reflejar respiraciones profundas y lentas, típicas de estados de relajación, meditación o sueño. Los cambios en la potencia de esta banda pueden indicar transiciones entre estados emocionales de calma y estados de estrés moderado
- **Alta frecuencia (HF, High Frequency):** comprendida entre 0.25Hz y 5Hz. En la señal de RESP reflejan respiraciones más rápidas y variables. Esta banda puede estar asociada con respuestas emocionales de alta excitación, como ansiedad, estrés, miedo, o estados de alerta, donde la respiración tiende a ser rápida y superficial. A medida que aumenta la frecuencia respiratoria, se incrementa la potencia en esta banda, indicando cambios rápidos en el ritmo respiratorio. En resumen, una potencia alta en esta banda es sinónimo de la activación del SN simpático.

A mayores, se calcula una tercera banda. Esta coincide con la última calculada en la señal de EDA pues está comprendida entre 0 y 2.4 Hz. Esta aporta información más general de la señal pues abarca partes de las frecuencias anteriores, lo que resulta útil para poder obtener información cruzada de ambas bandas, viendo cómo se relacionan entre sí.

También se extrajeron otras características más complejas. Estas fueron las mismas que en la señal de EDA, por lo que no se va a explicar en detalle que son estas características. Estas fueron las siguientes:

- **Entropía espectral:** una alta entropía indica una señal más impredecible y caótica, lo que puede estar relacionado con emociones de alta excitación, estrés o ansiedad. Por el contrario, una baja entropía sugiere una señal más ordenada y predecible, típica de estados emocionales más tranquilos o relajados.
- **Desviación:** una desviación alta indica que la potencia de la señal está distribuida de manera dispersa entre diferentes frecuencias, lo cual puede asociarse con respiraciones irregulares y respuestas emocionales variables. Una desviación baja sugiere una distribución más concentrada de la potencia, común en patrones respiratorios más constantes, vinculados a emociones estables y prolongadas en el tiempo.
- **Asimetría:** una asimetría positiva o negativa puede indicar un predominio de frecuencias más bajas o más altas en la señal respiratoria. Por ejemplo, una asimetría positiva podría sugerir un predominio de respiraciones rápidas y cortas, típicas de estados de ansiedad o miedo.
- **Curtosis:** una alta curtosis indica la presencia de picos espectrales pronunciados, lo que puede señalar respiraciones anómalas o episodios de hiperventilación, típicos de estados emocionales extremos como el pánico. Por su parte, una curtosis baja puede estar asociada con una distribución más plana y menos concentrada, indicativa de estados emocionales más calmados.
- **Frecuencia máxima:** frecuencias más altas están asociadas con respiraciones rápidas y suelen correlacionarse con estados de alta excitación o estrés, mientras que frecuencias más bajas indican respiraciones lentas y profundas, típicas de la relajación o el sueño.
- **Valor cuadrático medio:** el RMS del espectro normalizado proporciona una medida de la magnitud general de la señal respiratoria en el dominio frecuencial. Un valor alto de RMS puede indicar una mayor energía total de la respiración, lo que puede estar asociado con un mayor esfuerzo respiratorio, característico de emociones intensas o estados de alerta.
- **Ancho de banda:** un ancho de banda más amplio puede reflejar una respiración más variada y, por tanto, una mayor variabilidad emocional. Por otro lado, un ancho de banda más estrecho puede estar asociado con un patrón respiratorio más regular, reflejando un estado emocional estable y controlado.

A continuación, se presenta la Tabla 5 donde se hace un resumen de las diferentes características que se han extraído. En la tabla se incluye una abreviatura para estas y la señal o componente de las que se extraen.

Media	Mean_RESP	RESP, inspiración, espiración, frecuencia respiratoria, ciclo respiratorio
Desviación estándar	Std_RESP	RESP, inspiración, espiración, frecuencia respiratoria, ciclo respiratorio
Mediana	Med_RESP	RESP, inspiración, espiración, frecuencia respiratoria, ciclo respiratorio
Moda	Mod_RESP	RESP, inspiración, espiración, frecuencia respiratoria, ciclo respiratorio
Rango intercuartílico	Iqr_RESP	RESP, inspiración, espiración, frecuencia respiratoria, ciclo respiratorio
Asimetría	Skew_RESP	RESP, inspiración, espiración, frecuencia respiratoria, ciclo respiratorio
Curtosis	Kurt_RESP	RESP, inspiración, espiración, frecuencia respiratoria, ciclo respiratorio
Varianza	Var_RESP	RESP, inspiración, espiración, frecuencia respiratoria, ciclo respiratorio
RMS	RMS_RESP	RESP, inspiración, espiración, frecuencia respiratoria, ciclo respiratorio
RMSE	RMSE_RESP	RESP, inspiración, espiración, frecuencia respiratoria, ciclo respiratorio
Proporciones de las fases	FaseProp	Ciclo respiratorio, espiración, inspiración
Amplitud de las fases	FaseAmpl	Ciclo respiratorio, espiración, inspiración
Otras morfológicas	Morfo	Ciclo respiratorio, espiración, inspiración
Potencia	Power_RESP	FFT de RESP (LF, HF)
Máximo de banda	Max_ban_RESP	FFT de RESP (LF, HF)
Mínimo de banda	Min_ban_RESP	FFT de RESP (LF, HF)
Proporción de banda	Prop_RESP	FFT de RESP (LF, HF)
Entropía espectral	SpectEntro_RESP	FFT de RESP (LF, HF)
Distribución de potencia	PowerDist_RESP	FFT de RESP (LF, HF)
Asimetría	Skew_FFT_RESP	FFT de RESP (LF, HF)
Curtosis	Kurt_FFT_RESP	FFT de RESP (LF, HF)
Frecuencia dominante	MainFreq_RESP	FFT de RESP (LF, HF)
Magnitud media de la señal	MeanMag_RESP	FFT de RESP (LF, HF)
Ancho de potencia	Bandwidth_RESP	FFT de RESP (LF, HF)
RMS	RMS_FFT_RESP	FFT de RESP (LF, HF)
Frecuencia máxima	Max_freq_RESP	FFT de RESP (LF, HF)

Tabla 5. Características extraídas de la señal de RESP con su abreviatura asignada y la señal de la que se extrae.

3.3.4 Características de la señal de SKT

Esta es la última de las señales y menos compleja de las 4, por lo que el número de características extraídas es pequeño. Solamente se extraen las mismas características estadísticas que las otras señales y se calcula la potencia de las bandas frecuenciales de interés.

3.3.4.1 Características estadísticas

- **Media:** cambios en la media pueden indicar variaciones en la regulación térmica asociadas a diferentes estados emocionales, como el estrés (donde la temperatura puede bajar debido a la vasoconstricción) o la relajación (donde la temperatura puede aumentar).
- **Desviación estándar:** una alta desviación estándar sugiere fluctuaciones significativas en la temperatura, que podrían estar asociadas con respuestas emocionales intensas o cambios rápidos en el estado emocional.
- **Mediana:** es útil para analizar la temperatura basal de la piel en condiciones de estabilidad emocional o para identificar cambios sutiles en la temperatura durante diferentes estados emocionales.
- **Moda:** una moda que coincide con la temperatura promedio puede indicar un estado térmico estable, mientras que diferencias significativas entre la moda y la media pueden sugerir episodios de cambios repentinos en la temperatura relacionados con emociones transitorias.
- **Rango intercuartílico (IQR):** un IQR amplio puede indicar una alta variabilidad en la temperatura, potencialmente vinculada a fluctuaciones emocionales, mientras que un IQR estrecho sugiere una temperatura más constante, característica de un estado emocional estable.
- **Asimetría:** una asimetría positiva o negativa puede reflejar un predominio de temperaturas más altas o más bajas, respectivamente, lo que podría estar relacionado con estados emocionales específicos como excitación (mayor temperatura) o ansiedad (menor temperatura).
- **Curtosis:** Una alta curtosis sugiere la presencia de valores extremos (picos), que pueden estar asociados con eventos emocionales extremos o respuestas emocionales repentinas.
- **Varianza:** una alta varianza puede indicar fluctuaciones rápidas en la temperatura de la piel, lo que podría estar vinculado con cambios rápidos en el estado emocional.
- **RMSE:** un alto valor de RMS puede reflejar cambios significativos en la temperatura debido a respuestas emocionales intensas o prolongadas.
- **RMS:** proporciona una evaluación de la energía total de la señal de temperatura de la piel, que puede correlacionarse con variaciones sostenidas en la temperatura relacionadas con diferentes estados emocionales.

3.3.4.2 Características espectrales

Se emplea la FFT para obtener la transformada de Fourier de la señal de RESP (Figura 18). La representación de la señal en el dominio de la frecuencia es el siguiente:

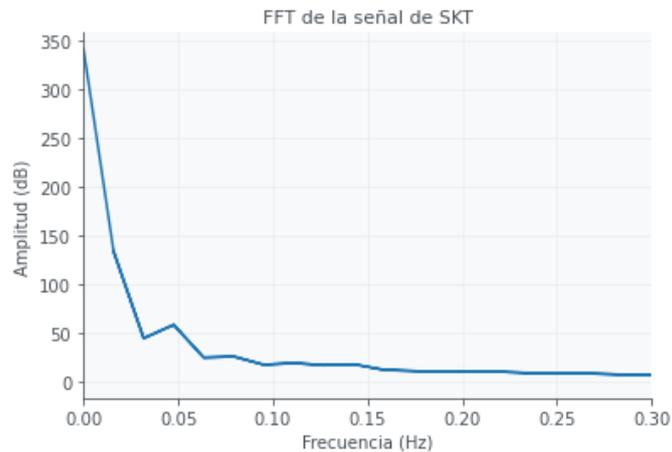


Figura 18. FFT de la señal de SKT

Las bandas de frecuencia analizadas en este caso fueron la alta y baja frecuencia comprendidas entre los 0 y 0.1 Hz y 0.1 y 0.2 Hz, respectivamente. La información que ambas aportan está relacionada con como de rápidas son las fluctuaciones de la señal, de modo que una potencia baja muy elevada sugiere que estas son más lentas, la señal es más lineal y los estímulos menos bruscos. Por el contrario, potencias altas elevadas sugieren que la temperatura corporal fluctúa rápidamente y, por consiguiente, el sujeto se encuentra ante una situación de estrés, donde la temperatura se precipita.

A continuación, se presenta la Tabla 6 donde se hace un resumen de las diferentes características que se han extraído. En la tabla se incluye una abreviatura para estas y la señal o componente de las que se extraen.

NOMBRE	ABREVIATURA	SEÑAL
Media	Mean_RESP	SKT
Desviación estándar	Std_RESP	SKT
Mediana	Med_SKT	SKT
Moda	Mod_SKT	SKT
Rango intercuartílico	Iqr_SKT	SKT
Asimetría	Skew_SKT	SKT
Curtosis	Kurt_SKT	SKT
Varianza	Var_SKT	SKT
RMS	RMS_SKT	SKT
RMSE	RMSE_SKT	SKT
Potencia	Power_SKT	FFT de SKT (LF, HF)
Máximo de banda	Max_ban_SKT	FFT de SKT (LF, HF)
Mínimo de banda	Min_ban_SKT	FFT de SKT (LF, HF)
Proporción de banda	Prop_SKT	FFT de SKT (LF, HF)

Tabla 6. Características extraídas de la señal de SKT con su abreviatura asignada y la señal de la que se extrae.

3.3.5 Extracción de características

Para extraer todas las características anteriores se diseñaron 4 *scripts*, uno por cada señal. Cada uno de estos se compone de una serie de funciones que calculan las características descritas en los apartados anteriores. Generalmente, para ahorrar espacio y generar menos variables, las funciones agrupan aquellas características pertenecientes a categorías comunes. Es decir, las estadísticas se calculan en una sola función, las temporales en otra, las espectrales en otra, etc.

Las funciones descritas en el párrafo anterior eson invocadas en otro *script* cuya función es crear la matriz de características. Para ello, en el código se crearon 4 funciones que agrupan el cálculo de las características de cada señal, de manera que al final quedan 4 variables. Teniendo en cuenta que hay 40 vídeos por cada uno de los 32 individuos, hay que crear 2 bucles '*for*' para automatizar el proceso. Debido a la alta cantidad de datos el proceso es bastante largo y como resultado se obtiene una variable formada por 32 matrices, una por sujeto. Cada una de estas matrices tiene 227 filas y 40 columnas, representando así las características y etiquetas y el número de vídeos respectivamente.

En total se extrajeron 225 características, contando con las extraídas de todas las componentes de cada señal, estas se repartieron de la siguiente manera: 44 de la señal de HRV, componente de la variación de la frecuencia cardíaca extraída de la señal de ECG; 72 de la señal de EDA, entre la propia señal filtrada, la señal de EDR y las amplitudes; 89 de la señal de RESP, contando con características de la señal filtrada, los ciclos y fases respiratorias y la frecuencia respiratoria y, por último, 20 características extraídas de la señal de SKT.

3.4 SELECCIÓN DE CARACTERÍSTICAS (*FEATURE SELECTION*)

Cómo se describe al principio del apartado 2 de esta sección (pág. 23) el objetivo de la extracción de características era reducir el número de elementos que componían la base de datos original. Ciertamente, pasar de las 32.256 muestras que componían las 4 señales con las que se ha trabajado a 225 características supone una reducción del 99.3% de los elementos originales, pero todavía se puede disminuir más.

Para conseguirlo, vamos a implementar una etapa de selección de características, lo que significa que, vamos a elegir cuales son aquellas características más valiosas. Ya sea porque aportan una información única o porque dicha información está muy sintetizada. Con esto, además, se consigue reducir la carga computacional del modelo predictivo y mejorar su rendimiento [49].

Existen varias formas de hacer la elección de características. En [49] se describen algunas de ellas, dependiendo del tipo de problema de predicción frente al que nos encontremos. En este caso, se proponen alternativas tanto para los problemas de clasificación como para los de regresión. Este último es el de interés, pues vamos a crear un modelo en el que tratemos de emular el modelo de Russel, descrito en la sección anterior. Así pues, el texto nos sugiere dos opciones para realizar la elección.

La primera de estas opciones es la correlación (*correlation*). Mide la fuerza y dirección de la relación lineal entre dos variables. En otras palabras, evalúa si los cambios en una variable están asociados con cambios en otra, y en qué dirección (positiva o negativa). Para medir esta relación, se utiliza el coeficiente de correlación de Pearson, que toma valores entre -1 y 1. Un valor de 1 indica una correlación positiva perfecta (ambas variables aumentan o disminuyen juntas de manera lineal), un valor de -1 indica una correlación negativa perfecta (una variable aumenta mientras la otra disminuye linealmente), y un valor de 0 indica que no hay correlación lineal entre las variables. Es importante destacar que la correlación no implica causalidad; simplemente indica que las dos variables tienden a variar juntas [49].

La otra opción es la información mutua (MI, *mutual information*) que mide cuánta información sobre una variable puede obtenerse conociendo otra variable, o en otras palabras, como se reduce la incertidumbre de una variable dada una segunda variable. Es una medida simétrica de dependencia que puede capturar tanto relaciones lineales como no lineales. En este caso, los

resultados obtenidos oscilan entre 0 e infinito, pero por lo general, al menos en este trabajo, el valor máximo no sobrepasa un valor de 0'15. Las características que obtengan valor más elevados serán consideradas como las más relevantes [49].

Entre estas dos se escoge la segunda, la MI, porque tiene en consideración las relaciones no lineales entre las características y el vector de *targets*. Este hecho permite que un mayor número de características obtengan una puntuación superior a cero. Posteriormente, entre el conjunto de características que no sean descartadas por su bajo resultado se seleccionarán las más relevantes.

3.4.1 Procesamiento previo a la selección de características

Antes de introducir la matriz de características en el algoritmo que calculará la MI es necesario realizar algunas modificaciones.

En primer lugar, se eliminan las matrices de características de los sujetos 2, 16 y 31, debido a que los registros de ECG son bastante malos y la función de *BioSPPy* consideraba como picos R puntos que no lo eran. Para reducir la carga computacional, en la fase anterior, la extracción de características, el bucle que recorre los 32 sujetos de la base de datos salta estos 3. No hay un motivo concreto que explique por qué las matrices de estos individuos, completamente formadas por ceros, sean eliminadas en este paso y no en el anterior. Si bien es cierto, que al guardar en la memoria del ordenador la matriz de características, una vez construida, hubiese ocupado menos espacio si se hubiera prescindido de 3 sujetos, pero tampoco supone un problema excesivo.

El siguiente paso es imputar los valores perdidos durante el proceso de extracción. Estos valores son aquellas características para las que, por la naturaleza de la señal, la función que las calculaba ha fallado y en lugar de dar un valor numérico concreto aparece un 'NaN' (*Not a Number*) o un 'Inf' (valor infinito). Para eliminarlos, como no son muchos, en lugar de revisar todas las funciones, reparar los posibles errores y volver a construir la matriz de características, proceso en el que se perdería mucho tiempo, se imputan esos valores por la media de las filas. Dicho de otra manera, como todos los elementos de una fila son la misma característica, pero para diferentes vídeos, el cálculo de su media, sin contar con NaN ni Inf, es un resultado aproximado al que podría haber sido el real.

El último paso, es reorganizar la matriz para que el algoritmo de selección pueda trabajar con ella. El objetivo es pasar de una variable de 3 dimensiones a otra de 2, básicamente, eliminar las matrices de cada sujeto y tener una única matriz formada por todas ellas. Se probó con la función '*reshape*' que sirve para reorganizar las variables, pero no se consiguió el resultado esperado, que debía ser las matrices de los individuos concatenadas. Para solucionarlo se creó un bucle que las fuera colocando una al lado de la otra hasta formar esa variable de 225 filas y 1160 columnas. Para las etiquetas también se tuvo que realizar este proceso, pero como es una variable de 2 dimensión si sirvió con '*reshape*'.

3.4.2 Selección de características

Para la selección de características recurrimos al algoritmo de la biblioteca *Scikit-Learn* dedicado a la MI '*mutual_info_regression*', que necesita la matriz de características y el vector de etiquetas como entradas, ambos en las dimensiones adecuadas. Antes de continuar, hay que separar los vectores de etiquetas, por un lado, la *arousal* y por el otro la *valence*, pues el grado de información que aporta una misma característica en ambos casos no es el mismo. La salida de esta función es un vector formado por los coeficientes de MI obtenidos. Cabe destacar que la

matriz de características reestructurada fue transpuesta previamente. A continuación, en las Tablas 7 y 8 se muestran las características más relevantes para cada señal, en función de su valor de MI.

CARACTERÍSTICA	VALOR DE MI
ECG	
Tinn	0,0875
DFA 2	0,0802
Max_RR	0,0736
EDA	
Bandwidth	0,1102
Prop_EDA	0,1082
MeanMag	0,0937
RESP	
FaseProp	0,1243
Morfo	0,1098
Max_ban_RESP (LF)	0,0972
SKT	
Ptop_SKT (HF)	0,105
Std_SKT	0,0863
Iqr_SKT	0,0856

Tabla 7. Mejores características por señal para valence, según su valor de MI.

CARACTERÍSTICA	VALOR DE MI
ECG	
DAFA1	0,1251
DFA	0,1156
DFA2	0,0844
EDA	
Var_EDA	0,1333
Kurt_EDA	0,1332
Med_EDA	0,0897
RESP	
Mod_RESP	0,1927
Med_RESP	0,1659
Std_RESP	0,1622
SKT	
Std_SKT	0,0683
Med_SKT	0,062
Prop_SKT (HF)	0,0614

Tabla 8. Mejores características por señal para arousal según su valor de MI.

Entre ambas tablas podemos ver diferencias entre las características con los valores de MI, solo coinciden DFA1 y Std_SKT, por lo que podemos asumir que éstas son características robustas para la señal a la que representan. Si bien el valor de MI no muy elevado, en el caso de *arousal*. Por su parte, los valores de MI para *arousal* son superiores a los de *valence*. Esto se explica porque la primera es más fácil de detectar, pues tiene mayor impacto sobre los órganos periféricos, especialmente en la frecuencia cardíaca o respiratoria.

3.5 TRANSFORMACIÓN DE LOS DATOS (*DATA TRANSFORM*)

A menudo, es frecuente realizar un escalado de los datos de entrenamiento para que las características tengan una escala común. Entre otros beneficios, la escala común de las características permite mejorar el rendimiento y eficacia del modelo, ofreciendo mejores resultados en menos tiempo. Por otro lado, un problema de no escalar los datos reside en que un elevado tamaño de los pesos, que son los parámetros que relacionan las salidas con las entradas, puede conducir al modelo a un comportamiento inestable (demasiado sensible a pequeños cambios en la entrada) y a una pobre generalización (incapacidad para etiquetar correctamente datos que no sean los de entrenamiento) [49].

Existen dos métodos de escalado: la estandarización y la normalización. El primero de ellos consiste en restar la media y dividir entre la desviación estándar (6) y el segundo consiste en hacer que los datos oscilen en un rango entre 0 y 1, para ello se resta el valor mínimo y se divide entre el valor máximo menos el mínimo (7).

$$X' = \frac{X - \text{media}(X)}{\text{desviación estándar}(X)} \quad (6)$$

$$X' = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (7)$$

Dónde X' es el conjunto de valores escalados y X los valores sin escalar.

En [49] se dice que para los inputs de algoritmos predictivos como SVM o KNN es necesario realizar la estandarización de los datos de entrenamiento, por lo que este será el proceso que se lleve a cabo. Para hacerlo se crea un *script* en el que, en primer lugar, se va a preguntar por el percentil deseado para la selección de características. Lo que significa que, en función del valor introducido, se escogerá un número de características u otras. En este caso se escogerán aquellas superiores al tanto por ciento introducido, es decir, si el percentil introducido es 75 se cogerán aquellas características cuyo valor de MI sea superior al 75%. Esto se hace para no tener que estandarizar todas las características sino solo las de interés. Posteriormente se crea un bucle 'for' que realizará el proceso, característica por característica. Este paso hay que repetirlo 2 veces pues hay que hacerlo tanto para las características seleccionadas para *arousal* y para la *valence*. Finalmente, la función de estandarización devuelve las matrices estandarizadas.

3.6 MODELOS PREDICTIVOS

Este es el último paso, previo a los resultados, la elaboración de los distintos modelos predictivos de los que se habló en la Parte 1.2.4. Estos son los algoritmos con los que se realiza el proceso de regresión. Para ello, se van a diseñar 4 modelos predictivos, que se describirán a continuación. Estos modelos pasarán por una fase de entrenamiento, con los datos y las características descritos en apartados anteriores. Del entrenamiento se obtienen unas variables que se deberán guardar y posteriormente se testarán con un nuevo conjunto de datos, para

valorar el rendimiento, la generalización y la calidad de los modelos. El motivo de utilizar 4 modelos diferentes es, analizar como diferentes algoritmos pueden adaptarse a un problema complejo, como es la clasificación del estado emocional y poder descubrir cuál de ellos ofrece los mejores resultados.

3.6.1 SVM

Este es el primero de los algoritmos utilizados. En [49] se menciona que para algoritmos de este tipo es recomendable la utilización de datos estandarizados, por lo que en un primer momento los utilizaremos para entrenar, aunque posteriormente se pueden hacer modificaciones para ajustar los modelos.

El primer paso, es cargar las matrices con las características elegidas, tanto para *arousal* como para *valence* y en las dimensiones apropiadas, tanto estandarizadas como sin estandarizar, por si hiciera falta cambiar el conjunto de entrenamiento. Además, junto con la matriz de características se cargan también los vectores de targets. Todo esto se va a hacer desde el *script* en el que se realizó la estandarización, pues estaba formado por una función cuyas salidas eran estas. A continuación, es necesario trasponer las matrices y los vectores, pues las funciones con las que se realiza la regresión precisan que estas variables tengan las características ordenadas por columnas y no por filas, que es el resultado de reestructurar la matriz de características para pasar de una de 3 dimensiones a otra de 2.

Una vez se cargan los datos; proceso que se repite exactamente igual en el resto de los algoritmos, se procede a desarrollar el algoritmo. En este caso, se recurre a la biblioteca *Scikit-Learn* donde ya existen funciones para desarrollar un modelo de SVM en *Python*. Para empezar, se utiliza la función '*SVR*', que es una versión adaptada de SVM para regresión. Esta función necesita una cadena de caracteres como entrada. Esta será el nombre o la abreviatura de un *kernel*. Un *kernel* es una función que se encarga de determinar cómo se transforman los datos en un espacio de características, los tipos se desarrollarán en el apartado de resultados.

Lo siguiente es de definir un diccionario en el que contenga una serie de hiperparámetros que se irán probando durante el entrenamiento del modelo. Esto significa, que se buscarán cuáles son los valores más adecuados. Los hiperparámetros que se definen son:

- **C:** controla la penalización de los errores en el margen, estos son puntos de datos que se encuentran dentro del margen definido por las líneas de soporte o que se encuentran en el lado incorrecto del margen. Valores más altos de C significan un margen más estrecho y menos errores permitidos.
- **E (épsilon):** Controla la sensibilidad de la pérdida de regresión; es el margen en el que no se penalizan los errores durante la regresión.
- **Γ (gamma):** determina la influencia de un solo ejemplo de entrenamiento. Tiene 2 opciones: 'scale' es el valor predeterminado y se calcula como $1/(n_features * X.var())$. La opción 'auto' establece $\gamma = 1/n_features$. Dónde $n_features$ es el número de características y $X.var()$ es la varianza de los datos de entrenamiento.

A continuación, se va a ajustar el modelo buscando los mejores valores para los hiperparámetros. Para ello, se utiliza la función '*GridSearchCV*' que necesita como argumentos: tanto la variable creada con la función '*SVR*' y el *kernel*, como el diccionario creado anteriormente. Este modelo se va a entrenar utilizando los datos de entrenamiento, es decir, la matriz de características seleccionadas y los vectores de etiquetas, a través del comando '*fit*'. Este proceso se debe realizar tanto para *arousal* como para *valence*.

3.6.2 KNN

Otro de los modelos desarrollados fue KNN, que en [49] también se sugiere que se estandaricen las matrices de características antes de entrenar el modelo. Para cargar los datos se lleva a cabo el mismo procedimiento descrito anteriormente. Y para crear el modelo, hay que seguir una serie de pasos ligeramente diferentes, pues para esta ocasión utilizamos simplemente la función ‘*KNeighborsRegressor*’ de *Scikit-Learn*, que necesita como parámetros de entrada el número de vecinos y una métrica con la que calcular la distancia entre dos puntos en un espacio n -dimensional. Con todo esto, se consigue que el modelo de KNN tenga las características deseadas. El último procedimiento, es entrenarlo, de la misma manera que para SVM utilizando el comando ‘*fit*’, la matriz de características y el vector de etiquetas.

3.6.3 PLS (Partial Least Squares)

En la parte 1.2.4 se habla acerca de LDA como otro posible modelo predictivo a ejecutar, el problema es que este no es adecuado para realizar regresión multivariada porque su objetivo principal es la clasificación y no la predicción continua. LDA se centra en encontrar fronteras de decisión lineales para distinguir entre categorías discretas y no en modelar relaciones continuas entre variables dependientes y predictoras. En contraste, PLS es una técnica diseñada específicamente para problemas de regresión multivariada, ya que su funcionamiento consiste en encontrar componentes latentes que expliquen la máxima covarianza entre las variables predictoras (X) y las variables de respuesta (Y). Es decir, busca direcciones en el espacio de características que no solo capturan la mayor variación en las variables de entrada, sino que también están altamente correlacionadas con las variables de salida, permitiendo manejar datos con alta multicolinealidad y obtener predicciones precisas para múltiples variables dependientes. Por lo tanto, PLS es más adecuado para la regresión multivariada donde el objetivo es predecir valores continuos de múltiples variables de respuesta a partir de un conjunto de variables predictoras.

Para crear el modelo se utiliza la función ‘*PLSRegression*’ de *Scikit-Learn* que necesita una variable de entrada llamada *n_components*. Esta variable determina cuántas direcciones latentes se extraen de los datos para capturar la mayor cantidad de varianza en las variables predictoras (X) y, al mismo tiempo, maximizar la correlación con las variables de respuesta (Y). Posteriormente quedaría entrenar el modelo como en los casos anteriores.

3.6.4 RF

Este es el último de los algoritmos que se van a desarrollar, en cuanto al escalado de los datos se dice que no es necesario llevar a cabo ningún procedimiento, aunque es recomendable. El procedimiento es el mismo que en los casos anteriores, se utiliza la función ‘*RandomForestRegressor*’ que precisa de 2 entradas: *n_estimators*, que especifica el número de árboles de decisión que se construirán, y *random_state*, que controla la aleatoriedad de las muestras utilizadas en cada árbol.

3.7 ENTRENAMIENTO Y VALIDACIÓN

Una vez están diseñados todos los modelos predictivos y se han seleccionado las características más relevantes, empieza la fase de entrenamiento. El objetivo de este apartado es hacer que los algoritmos predictivos capturen la variabilidad en los datos y sean capaces de realizar predicciones. Para conseguirlo se divide el conjunto de datos inicial en 2. El primer subconjunto, el de entrenamiento, es el más grande y estará formado por el 70% de los vídeos. El segundo subconjunto, el de test, formado por el 30% de los vídeos y servirá para validar el entrenamiento de los modelos.

Para el entrenamiento, se ejecutan los scripts donde se definieron los algoritmos, citados en el apartado anterior. En el programa se utiliza el modelo entrenado y el comando `‘.predict’` para predecir las etiquetas de los datos. En un primer momento, esto se realiza con los datos de entrenamiento, para poder ajustar los diferentes hiperparámetros que se explicaron en el apartado anterior, así como para poder elegir el número de características más adecuado. La forma de determinar cuáles son los mejores valores es entrenar varias veces y comprobar mediante una serie de métricas, de qué forma mejoran o empeoran los resultados de la predicción. Las métricas empleadas en este punto fueron [50]:

- **Error cuadrático medio (MSE, mean squared error):** es un estadístico simple, que calcula las diferencias entre los valores predichos y los valores objetivo y, posteriormente los promedia. Hay que tener en cuenta que la resta se eleva al cuadrado (8), por lo que el error se sobredimensiona y puede conducir a sobreajustes por tratar de reducirlo, sin tener en consideración la naturaleza de los datos de entrenamiento. Al tratarse de una discrepancia entre valores cuanto mayor sea este número, menos capacidad de predicción tiene el modelo.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (8)$$

En la fórmula, y_i es el valor objetivo e \hat{y}_i es el resultado de la predicción.

- **Error absoluto medio (MAE, mean absolute error):** este es una métrica similar a la anterior, salvo que, en esta ocasión no se elevan al cuadrado las diferencias, sino que se calcula su valor absoluto (9). Lo que se consigue con esto es penalizar errores demasiado grandes, haciendo que todos tengan la misma influencia sobre el valor medio.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (9)$$

- **R al cuadrado (R^2):** esta medida permite comparar como de buena es la predicción comparándola con una base constante. Es decir, se va a calcular el MSE, de una forma diferente, entre las predicciones y el modelo objetivo más sencillo posible, que es la media de todas las variables objetivo (10). El valor obtenido oscila entre $-\infty$ y 1, siendo el segundo caso una situación en la que el error es cero y podríamos decir que ideal.

$$R^2 = 1 - \frac{MSE(modelo)}{MSE(baseline)} \quad (10)$$

El MSE (baseline) se calcula de la siguiente manera:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y}_i)^2$$

Siendo \bar{y}_i la media de todos los valores objetivos.

Cuando las métricas se consideran lo suficiente buenas, es decir, los resultados para el conjunto de entrenamiento son lo más elevados posible, se graban los modelos entrenados y pasa a la fase de validación. En este punto se debe tener en cuenta que los valores tienen que ser adecuados, pero no deben conducir al sobreajuste del modelo. Los valores del modelo para los datos de entrenamiento se recogen en la Tabla 11.

Para valorar el rendimiento y la generalización de los modelos se va a utilizar el siguiente método: se guardan los algoritmos entrenados en formato *pickle*, mediante el comando '*pickle.dump*', para posteriormente cargarlos con '*pickle.load*'. Con estos cargados se va a hacer la predicción de las etiquetas, en esta ocasión con el conjunto de validación. Tras la división, explicada en el apartado 3.7, se obtienen 348 vídeos para test y 812 para entrenamiento. Con todo esto, las métricas explicadas en el punto anterior y una cantidad de características elegidas de 54, los resultados obtenidos se muestran en la Tabla 12.

A simple vista, se puede ver que salvo el RF para *arousal* los modelos ofrecen unos resultados bastante pobres. El significado, la causa y las posibles soluciones se analizarán y discutirán en el siguiente apartado.

4 RESULTADOS

A continuación, se van a presentar todos los resultados obtenidos a lo largo del trabajo. Tanto los de entrenamiento y validación como los del proceso de *feature selection*.

En primer lugar, las Tablas 9 y 10 muestra cuales son las 10 características más relevantes, ordenadas de mayor a menor, en función de su valor de MI y tanto para *valence* como para *arousal*.

POSICIÓN	CARACTERÍSTICA	VALOR DE MI	SEÑAL
1	FaseProp	0,1243	RESP
2	Bandwidth	0,1102	EDA
3	Morfo	0,1098	RESP
4	Prop_EDA	0,1082	EDA
5	Max_ban_RESP (LF)	0,0972	RESP
6	MeanMag	0,0937	EDA
7	PowerDist	0,0937	EDA
8	Mean_EDA (Amplitudes)	0,0895	EDA
9	Kurt_EDA (Amplitudes)	0,0811	EDA
10	DFA2	0,0802	ECG

Tabla 9. Mejores características según su valor de MI para *valence*.

POSICIÓN	CARACTERÍSTICA	VALOR DE MI	SEÑAL
1	Mod_RESP	0,1927	RESP
2	Med_RESP	0,1659	RESP
3	Std_RESP	0,1622	RESP
4	Var_EDA	0,1333	EDA
5	Kurt_EDA	0,1332	EDA
6	DFA1	0,1251	ECG
7	Med_EDA	0,1249	EDA
8	DFA	0,1156	ECG
9	DFA2	0,0844	ECG
10	Mean_EDA	0,0828	EDA

Tabla 10. Mejores características según su valor de MI para *arousal*.

De todas las características extraídas solamente se repite DFA2, por lo que podemos afirmar que esta es bastante robusta y aporta una información bastante valiosa para el problema de regresión. Por lo demás, destaca que el valor de MI para *arousal* es superior, el motivo por el que esto ocurre se ha explicado en apartados anteriores y se tratará en mayor profundidad más adelante.

Como se ha explicado en apartados anteriores, el entrenamiento es la fase previa a la validación por lo que en la Tabla 11 se presentan los resultados del entrenamiento, tomando como número de características aquellas cuyo valor de MI sea superior al 75%.

	AROUSAL			VALENCE		
	MSE	MAE	R ²	MSE	MAE	R ²
SVR	3,661	1,583	0,122	4,457	1,775	0,035
PLS	3,821	1,622	0,084	4,331	1,756	0,062
KNN	3,736	1,594	0,104	4,316	1,754	0,065
RF	0,54	0,604	0,87	3,283	1,523	0,29

Tabla 11. Resultados de entrenamiento.

Por último, una vez los modelos están entrenados, se realiza el proceso de validación o *test*, cuyos resultados se presentan en la Tabla 12.

	AROUSAL			VALENCE		
	MSE±SD	MAE±SD	R ² ±SD	MSE±SD	MAE±SD	R ² ±SD
SVR	3,508±0,56	1,539±0,13	0,035±0,05	4,69±0,1	1,816±0,22	-0,011±1,14
PLS	3,637±0,78	1,569±0,15	0±0,04	6,098±0,78	1,984±0,17	-0,314±0,14
KNN	3,617±0,54	1,546±0,13	0,055±0,04	4,817±1,09	1,837±0,22	-0,038±0,08
RF	0,687±0,73	0,712±0,16	0,81±0,02	4,046±1,06	1,705±0,21	0,124±0,07

Tabla 12. Métricas de rendimiento de los diferentes modelos, tras la validación.

Por último, se va a representar en las Figuras 19 y 20 unos boxplots para visualizar de forma gráfica los resultados de MSE y MAE obtenidos, tanto para *valence* como para *arousal*.

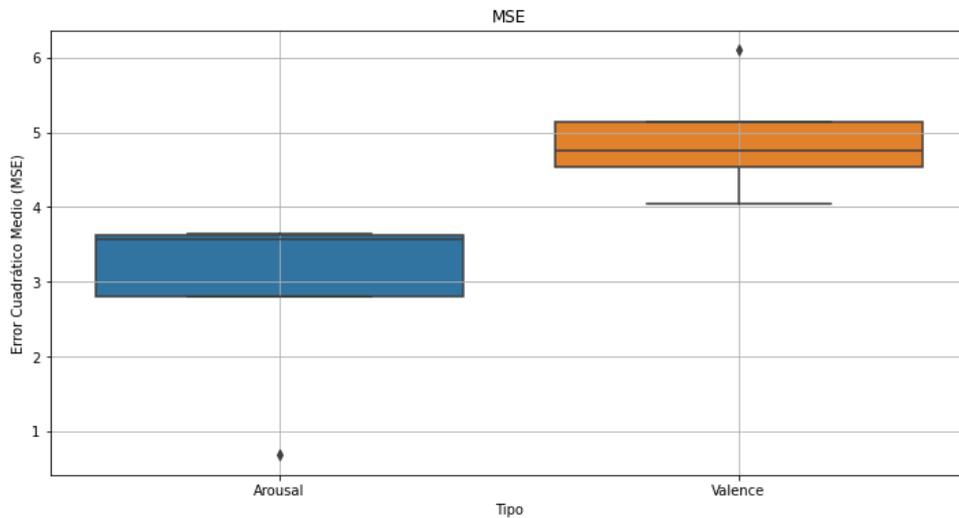


Figura 19. Boxplot de MSE.

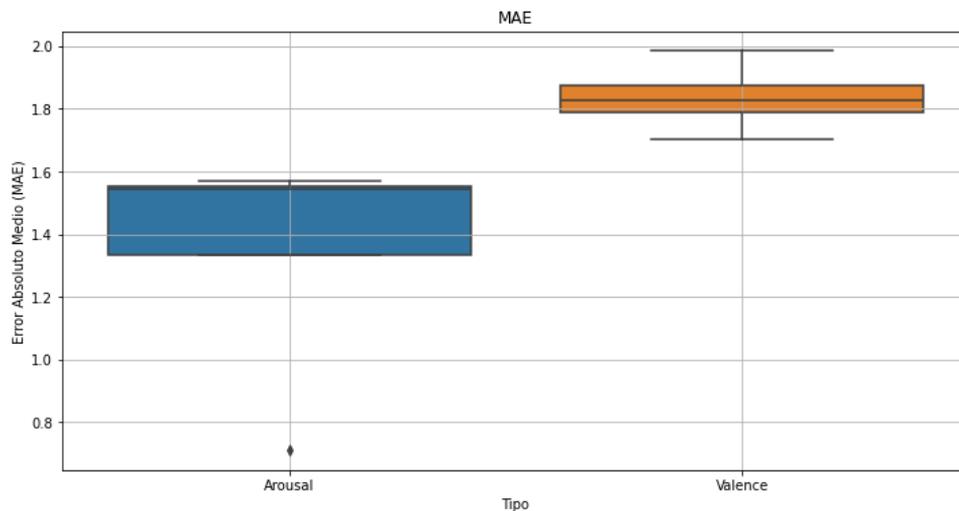


Figura 20. Boxplot del MAE.

En el primer gráfico (Figura 19), el MSE revela que *arousal* presenta una menor variabilidad en el error, con valores que oscilan entre aproximadamente 2,5 y 4, además de un valor atípico que indica un caso con un error particularmente bajo, asociado al valor de MSE de RF. En contraste, *Valence* tiene un rango de MSE más alto, entre 4,5 y 6, con un *outlier* que representa un error

más elevado, el de PLS. Esto sugiere que el modelo de *arousal* es más preciso y estable, mientras que el de *valence* muestra mayor error promedio y dispersión en sus resultados.

El segundo gráfico (Figura 20), que muestra el MAE, respalda estas conclusiones, evidenciando que *arousal* mantiene un rango de error más bajo, de aproximadamente 1,3 a 1,6, y presenta un *outlier* con un error absoluto significativamente menor a 1. Por otro lado, *valence* tiene un rango de error mayor, entre 1,8 y 2,0, sin valores atípicos destacados. En general, ambos *boxplots* indican que el modelo de *arousal* es más efectivo, con menores errores en ambas métricas, mientras que el modelo para *valence* no solo tiene un error promedio más alto, sino que también presenta mayor variabilidad, sugiriendo predicciones menos precisas y consistentes para esta dimensión.

5 DISCUSIÓN

5.1 ANÁLISIS DE RESULTADOS

Como se resaltó en el apartado anterior los resultados obtenidos, en general, no son buenos. Antes de hacer un análisis pormenorizado de cada uno de los modelos predictivos, cabe destacar que el hecho de que *arousal* sea la variable del modelo emocional elegido con un mejor rendimiento. Para demostrarlo, se observa la media de cada métrica, que ofrece mejores resultados en el primer caso, de forma que se puede sentenciar que, de manera general la *arousal* es más sencilla de predecir. Esto se debe a que es más fácil medir el grado de excitación de un individuo frente a una situación concreta. Esto es comprobable, pues, por ejemplo, ante estímulos que provocan sobresaltos en los sujetos las fases respiratorias se acortan, el ritmo cardíaco se acelera y aumenta la sudoración. Estas respuestas permiten saber si el individuo está más o menos estimulado, pero no se puede determinar si este está disfrutando de la experiencia o no. Se podría decir que *valence* es algo intrínseco del cerebro y que, por tanto, no tiene mucha influencia sobre los órganos periféricos; es algo difícil de captar con señales fisiológicas periféricas.

5.1.1 SVR

Los resultados de este modelo no son especialmente positivos. Este hecho reside en la naturaleza no lineal del grado de excitación. Es decir, a pesar de que la *arousal* es fácilmente distinguible a “simple vista”, computacionalmente presenta una serie de variaciones abruptas y espontáneas, debidas a la naturaleza imprevisible del estímulo subyacente que complica el proceso para captarla en esencia. Por su parte, *valence* es una variable que tiene poca expresión sobre los órganos periféricos. A SVR le cuesta clasificar más *arousal* que *valence* porque el algoritmo busca un hiperplano que separe las muestras y este proceso es bastante más sencillo si la relación entre estas es lineal.

Otro aspecto importante de los modelos predictivos es su capacidad de generalización, como en alguna ocasión se ha explicado, esto significa si son capaces, o no, de hacer buenas predicciones con conjuntos de datos nunca vistos en la fase de entrenamiento. En este caso las predicciones no van a ser buenas, por lo que se comprobará si los resultados no empeoran significativamente. Para valorar la capacidad de generalización, se calcula el rendimiento de los modelos utilizando el conjunto de datos de entrenamiento para predecir.

	AROUSAL			VALENCE		
	MSE	MAE	R ²	MSE	MAE	R ²
Entrenamiento	3,661	1,583	0,122	4,457	1,775	0,035
Validación	3,508	1,539	0,035	4,69	1,816	-0,011
Diferencias	0,153	0,044	0,087	0,388	0,041	0,046

Tabla 13. Resultados de entrenamiento y validación para SVR.

La discrepancia entre los resultados es bastante pequeña, aun así, no se puede decir que el modelo sea generalizable, pues el rendimiento debería ser bueno para los datos de validación. El detalle que llama especialmente la atención es que los valores de MAE y MSE son superiores para el conjunto de entrenamiento que de validación, es más probable que el error cometido entre el valor de las etiquetas objetivo y las resultantes de la predicción, de alguna de las muestras sea

mayor y, por consiguiente, el valor de MSE aumente. El hecho de que MAE sea superior en el conjunto de entrenamiento se explica de la misma manera que MSE, solo que, en esta ocasión, como se toma el valor absoluto de las diferencias entre los resultados de ambos conjuntos, la variación es menor. Por su parte R^2 es superior para el conjunto de entrenamiento, porque los valores que se utilizan para hacer la predicción ya son conocidos por el algoritmo, por lo que tiene mayor rendimiento. Para *valence* cambia el hecho de que MAE y MSE son superiores en el conjunto de validación, esto puede ser porque a el tratamiento que hace SVR de los datos de entrenamiento es distinto para esta segunda variable. Las razones de porqué ocurre esto no las tengo del todo claras.

5.1.2 PLS

Este modelo ofrece los peores resultados, tanto para *arousal* y para *valence*. La explicación a este fenómeno reside en que PLS es un modelo de reducción de dimensionalidad que busca relaciones lineales entre los conjuntos de datos. Si tenemos en cuenta que la influencia del SNA sobre los órganos periféricos no viene descrita de forma que se pueda representar sobre una proyección lineal, es lógico pensar que sea el peor de los modelos diseñados.

En cuanto a la generalización procedemos de la misma manera que antes, mostrando la Tabla 14.

	AROUSAL			VALENCE		
	MSE	MAE	R^2	MSE	MAE	R^2
Entrenamiento	3,821	1,622	0,084	4,331	1,756	0,062
Validación	3,637	1,569	0	6,098	1,984	-0,314
Diferencias	0,184	0,053	0,084	1,767	0,228	0,376

Tabla 14. Resultados de entrenamiento y validación para PLS.

En este caso, para *arousal* se ve una situación parecida a la anterior, en la que los valores de MAE, MSE y R^2 son superiores en los datos de entrenamiento, las razones de este hecho pueden ser similares a las anteriores. Por su parte, los datos de *valence* son peores. Destaca el valor de MSE. Esto supone que la capacidad de generalización de PLS es mala.

5.1.3 KNN

La explicación para que los resultados de KNN no sean del todo positivos puede ser de naturaleza diversa. Por un lado, la maldición de la dimensionalidad (*curse of dimensionality*) provoca que para un conjunto de datos de entrenamiento elevado las distancias entre las muestras sean menos significativa, lo que hace que el rendimiento del modelo sea peor. Una solución a este problema sería reducir el número de características seleccionadas, pero como el objetivo es determinar cuál es el mejor modelo las condiciones iniciales deben ser iguales, por lo que este factor no se podría modificar sin comprometer la calidad del resto de modelos. Otra opción es, al igual que en los casos anteriores, la mala respuesta del modelo ante problemas de regresión con datos no lineales.

En la Tabla 15 vemos el estudio sobre la generalización del modelo.

	AROUSAL			VALENCE		
	MSE	MAE	R^2	MSE	MAE	R^2
Entrenamiento	3,736	1,594	0,104	4,316	1,754	0,065
Validación	3,617	1,546	0,055	4,817	1,837	-0,038
Diferencias	0,421	0,048	0,049	0,501	0,083	0,103

Tabla 15. Resultados de entrenamiento y validación para KNN.

La generalización de KNN sigue la línea de los anteriores, pues los datos son mejores para *arousal* que para *valence*. A diferencia de PLS las diferencias entre las métricas son menores, por lo que parece que se ajusta mejor a los datos de entrada, pero tampoco se puede decir que sea un modelo más generalizable.

5.1.4 RF

Los resultados para RF son los mejores, para ambas variables. Destaca el caso de la *arousal* para la que el modelo logra un rendimiento bastante bueno, con errores escasos y un alto valor de R^2 . Para explicar estos datos es necesario entender el funcionamiento de RF, que como se explica en el apartado 1.2.2.4 consiste en una serie de árboles de decisión que en cada rama dividen en subconjuntos de datos más pequeños los datos originales. La ventaja que presenta este modelo frente a los anteriores es que en ese proceso de división se trata de minimizar el error mediante umbralización. Este criterio no tiene por qué ser de carácter lineal, sino que puede modelar relaciones bastante más complejas. Teniendo en cuenta que la *arousal* viene caracterizada por relaciones no lineales entre el SNA y los órganos periféricos, RF viene a ser un buen predictor. Por su parte *valence*, como se trata de la variable que presenta menor correlación fisiológica entre el SNA y los órganos periféricos es entendible que RF ofrezca peor rendimiento. A pesar de no ser los mejores resultados, los de RF siguen siendo los mejores para predecir *valence*

	AROUSAL			VALENCE		
	MSE	MAE	R^2	MSE	MAE	R^2
Entrenamiento	0,54	0,604	0,87	3,283	1,523	0,29
Validación	0,687	0,712	0,81	4,046	1,705	0,124
Diferencias	0,147	0,108	0,06	0,763	0,182	0,166

Tabla 16. Resultados de entrenamiento y validación para RF.

En cuanto a la capacidad de generalización del modelo, como se puede ver, los datos son positivos. Por un lado, los datos de *arousal* se parecen bastante, las diferencias entre ellos no parecen muy significativas. Por su parte, los valores para *valence* sí que presentan más discrepancia, especialmente MSE. Esto puede significar que el conjunto de validación contiene datos para los que RF comete un error grande a la hora de realizar la regresión, lo que hace que el valor de MSE aumente.

5.2 COMPARATIVA ENTRE ESTE Y OTROS TRABAJOS

En este apartado se pretende informar a cerca de otras investigaciones que se han realizado, utilizando *DEAP dataset* como conjunto de datos, predicciones sobre el estado emocional. La tarea de comparar los resultados entre el trabajo presentado en este documento y los seleccionados es complicada, pues se emplean diferentes señales biomédicas, diferentes modelos predictivos, diferentes métodos de entrenamiento e incluso diferentes métricas de rendimiento.

Se han seleccionado 4 artículos en los que se trabaja con el *dataset*, cada uno con características diferentes. Se presentan en la Tabla 17.

Trabajos	Tipo de IA	Señales usadas	Algoritmos	Métricas
[51]	ML	GSR, BVP, EMG y RESP	MIL (NB, SVM y RF)	Accuracy y F1 score
[52]	ML	RESP, PPG y SKT	DLF (RF, SVM y LR)	Accuracy
[53]	ML	EEG	SVM, LR, RF, KNN, LDA	Accuracy, Precision, Recall y F1 score
[54]	DL	EEG, ECG, GSR, SKT, RESP	CNN	Accuracy

Tabla 17. Otros trabajos con DEAP dataset.

Cada uno de estos trabajos aborda el mismo problema, pero introduciendo diferentes elementos que les aportan un extra en su rendimiento. El primero de ellos, [51], utiliza una técnica denominada *multiple instance learning* (MIL). Esta técnica parte de la agrupación en distintas bolsas (*bags*) las instancias del conjunto de entrenamiento. Estas bolsas son etiquetadas como positivas, si alguna de las instancias es positiva, o como negativa, si ninguna de ellas es positiva. En el caso de la computación afectiva, se etiquetará una bolsa como positiva si presenta un valor de *arousal* o *valence* alto. Posteriormente, se utilizaría un modelo predictivo para clasificar cada una de las bolsas, en este caso, se empleó SVM, combinado con diferentes modalidades de MIL y tamaños de las bolsas. La *accuracy* y el *F1 score* se calcularon tanto para los modelos predictivos convencionales: RF, SVM y NB (*Naive-Bayes*), sin usar MIL y para las diferentes modalidades de MIL con SVM. A simple vista, el empleo de MIL con 3 instancias por bolsa mejora los resultados de los modelos convencionales, pero no de una manera significativa.

El siguiente de los artículos, [52], propone utilizar *decision level fusion* (DLF). Esta técnica consiste en combinar los resultados de diferentes tipos de clasificadores, como: SVM, RF y LR (*logistic regression*) para realizar la predicción final. Para el entrenamiento se utilizan las señales por separado y combinadas, mejorando los resultados en este último caso. En general, la *accuracy* para *valence* como para *arousal* oscila entre el 67% y el 73%.

El tercer artículo, [53], plantea un trabajo similar a este, pues utiliza una serie de algoritmos de predicción para detectar el estado emocional. La diferencia fundamental es que se emplean señales de EEG. Los resultados son bastante amplios, pues se ofrecen los de cada clasificador y, además, se separan las señales en tramos de 15 segundos. Todos los valores ofrecidos superan el 50%, tanto para *arousal* como para *valence*.

Por último, [54], es el que propone la metodología más diferente respecto pues se aleja del aprendizaje automático (ML) y se adentra en el campo del aprendizaje profundo (DL). Se diseñan 2 *convolutional neural network* (CNN), la primera de 2 dimensiones, para procesar todos los canales del EEG y otra de una dimensión para las señales fisiológicas. Los resultados son comparados con un algoritmo de SVM, siendo mejores para la CNN, tanto 2D, que es el mejor método de regresión, y para la de 1D.

Para comparar el resultado de este trabajo con los anteriores, se calcula la *accuracy* del modelo. Hay que tener en cuenta que los valores de salida de la regresión son continuos, lo que significa que, pueden ser cualquier número comprendido entre 1 y 9. De esta forma, se realiza una discretización de los datos. Esto se consigue restando las etiquetas predichas y las objetivo. Todas aquellas diferencias cuyo valor absoluto sea inferior a un determinado umbral son seleccionadas. Por último, para calcular la *accuracy* se divide el número seleccionado

anteriormente entre el número de etiquetas totales (348 para el conjunto de test) y se multiplica por 100. Los resultados de *accuracy* obtenidos se presentan en la Tabla 18.

	AROUSAL	VALENCE
Modelos desarrollados en el TFG		
SVR	10,63	11,21
PLS	7,76	7,47
KNN	9,2	6,9
RF	27,34	21,92
Modelos desarrollados en [50]		
SVM	59,1	59,5
RF	60,1	59,8
mil-SVM	60,65	60,2
mi – SVM	59,7	62,15
MI-SVM	58,95	60,65
EMDD-SVM	57,4	57,65
Modelos desarrollados en [51]		
DLF	73,08	72,18
Modelos desarrollados en [52]		
SVM	64,4	61,22
RF	61,86	60,81
KNN	58,75	59,4
Modelos desarrollados en [53]		
SVM	52,72	53,31
1D-CNN	59,67	58,08
2D-CNN	61,5	58

Tabla 18. Resultados de *accuracy* de otros trabajos y de este.

Para resumir, algunos de los datos aquí presentados son el resultado de calcular medias de los obtenidos en los trabajos originales. Se observa que los resultados son mejores en todos los trabajos, comparados con este. No se aprecian diferencias muy significativas entre cual es más fácil de predecir, si *arousal* o *valence*, aunque parece que algo mejor la *arousal*. Por otro lado, destaca la técnica de DLF como la que mejores resultados aporta.

5.3 LÍNEAS FUTURAS

Del apartado anterior, se pueden extraer numerosas ideas para tratar de mejorar el rendimiento del modelo. Por un lado, se puede trabajar en optimizar el código, analizando en profundidad las características extraídas, descartando aquellas que supongan un alto coste computacional y no aporten información especialmente relevante o diseñando manualmente los modelos predictivos, para adaptarlos de forma más específica al problema de la computación afectiva, en lugar de utilizar aquellos predefinidos en librerías concretas, de carácter más general y que se adaptan peor a los datos de entrenamiento.

A parte de mejorar el código, hay diversas técnicas que se pueden implementar para mejorar los resultados. En los trabajos del apartado 4.2 se propone la separación de las señales fisiológicas en intervalos temporales, con resultados variables. También se plantea utilizar diferentes tamaños de conjuntos de entrenamiento y test, en lugar de plantear una sola cantidad, se podría comparar entre varias. Técnicas más complejas como MIL y DLS, en los artículos encontrados no parece que mejora de forma notable el rendimiento de los modelos, pero se podría estudiar su

implementación para ver si en este caso particular, donde el rendimiento es malo, podrían mejorar los resultados.

Otras posibilidades, que se alejan del propósito del trabajo, son la implementación de información más compleja como la señal de EEG o las grabaciones de vídeo y audio. Para esto es probable que haga falta hacer un proyecto entero, pues la combinación de estas señales con las obtenidas de los órganos periféricos y con algoritmos de ML, parece una idea bastante ambiciosa. En este sentido, también alejado del tema que nos ocupa, estaría el desarrollo de un algoritmo de DL, que gracias a la extracción automática de características ahorraría bastante tiempo y coste computacional, lo que sí que haría viable el uso de toda la información anteriormente citada. Este es el caso de [54] que aúna EEG y señales periféricas.

6 CONCLUSIÓN

El trabajo realizado, que comprende desde el análisis de las bases de datos hasta la evaluación de los modelos diseñados, es extenso, complejo y en ocasiones repetitivo, pues el proceso de extraer las características de cada una de las señales, que como se reflejó en el punto 3.3 en muchas ocasiones es el mismo, o el de entrenar los diferentes modelos supone reutilizar y adaptar código desarrollado en algún momento anterior. Por otro lado, la extracción de características de todas las señales supone un alto coste computacional, acompañado de varias horas de ejecución del código que en ocasiones da errores, perdiéndose buena parte del trabajo. Otro de los detalles que complican este trabajo con otros realizados en el grado es el elevado número de variables, que por cuestiones de similitud son llamadas de forma parecida y que hasta el final generan confusiones.

Los objetivos descritos en el apartado 2.2.1 creo que están parcialmente cumplidos. El peor de ellos, el entrenamiento de los modelos, pues los resultados no son positivos y probablemente en este punto es donde se encuentren los principales fallos. Otra posibilidad es que las características extraídas no sean las más adecuadas en el contexto de este problema y con estos datos. En conclusión, en cuanto al objetivo principal no se puede afirmar cuál de los modelos realizados es mejor, pues a pesar de que se ven claras diferencias en RF, el hecho de que en otros trabajos [51], [52], [53], [54] realizados con la misma base de datos, se consigan resultados mejores hace pensar que ha habido fallos en el proceso y que existe un sesgo que hace que el análisis no sea posible.

En referencia a la ingeniería de rehabilitación es más práctico monitorizar las señales de los órganos periféricos por comodidad. Bien es cierto que algunas como SKT o EDA, al recogerse los datos en los dedos de la mano, presentarían inconvenientes en pacientes con lesiones que afecten al miembro superior, pues lo más probable es que se indujesen artefactos por movimiento. Esto hace necesario buscar otros puntos del organismo dónde tomar los datos de manera accesible y cómoda. Otra forma de predecir el estado emocional es utilizar registros de vídeo y audio que registren la reacción del paciente. Este procedimiento permitiría desarrollar una terapia sin cables que dificulten la consecución de los objetivos.

Por último, respecto a la base de datos me surge la duda de si la mejor forma de etiquetar los datos es la más adecuada. Es decir, autodescribir el estado emocional como un número comprendido entre 1 y 9, en este caso, me parece arbitrario, poco preciso y subjetivo. En primer lugar, considero que en un rango de valores tan amplio a una persona le debe costar elegir entre valores próximos por lo que al final puede haber un componente de aleatoriedad en su decisión. En ese sentido, creo que el desarrollo de un modelo de aprendizaje no supervisado ayudaría a eliminar el sesgo que suponen las etiquetas de los datos.

En conclusión, la computación afectiva es un campo interesante, con múltiples aplicaciones, pero todavía en desarrollo. Poder haber realizado este trabajo, a pesar de que los resultados no son los que me gustaría, me ha ayudado a entender el funcionamiento de la inteligencia artificial. Además, comprobar como la tecnología tiene capacidad para predecir algo tan complejo como son las emociones humanas, pone de manifiesto la relevancia de la ingeniería biomédica y su capacidad para integrar la ingeniería en otros campos de la ciencia como la medicina y la psicología.

GLOSARIO DE ABREVIATURAS Y ACRÓNIMOS

ApEn – Approximate entropy
BVP – Blood volume pressure
DL – Deep learning
DLF – Decision level fusion
DFA – Detrended fluctuation analysis
EDA – Electrodermal activity
EDL – Electrodermal level
EDR – Electrodermal response
ECG – Electrocardiograma
EEG – Electroencefalograma
EMG – Electromiografía
EOG – Electrooculografía
FFT – Fast Fourier transform
GSR – Galvanic skin response
HF – High Frequency
HR – Heart rate
HRV – Heart rate variability
ICA – Independent components analysis
IA – Inteligencia artificial
Inf – Valor infinito
JMI – Joint mutual information
KNN – K-nearest neighbors
LDA – Linear discriminant analysis
LF – Low Frequency
Lpm – Latidos por minuto
LR – *Logistic Regression*
MAE – Mean absolute error
MIL - Multiple instance learning
MIT – Massachusetts Institute of Technology
ML – Machine learning

MI – Mutual information
MSE – Mean square error
MUAP – Motor unit action potential
NaN – Not a number
NB – Naïve-Bayes
PCA – Principal component analysis
PPG – Fotopletismógrafo
PSD – Densidad espectral de potencia
RESP - Respiración
RF – Random forest
RMS – Root mean square
RMSE – Root mean square error
RMSSD – Root mean square of successive differences
Rpm – Respiraciones por minuto
SampEn – Sample entropy
SCL – Skin conductance level
SKT – Skin temperature
SNA – Sistema nervioso autónomo
SNA_p – Sistema nervioso autónomo parasimpático
SNA_s – Sistema nervioso autónomo simpático
SNP – Sistema nervioso periférico
SVM – Support vector machine
VLf – Very Low Frequency

REFERENCIAS

- [1] R. W. Picard, "Affective computing," p. 292, 2000, Accessed: Sep. 16, 2024.
- [2] "inteligencia | Definición | Diccionario de la lengua española | RAE - ASALE." Accessed: Sep. 16, 2024. [Online]. Available: <https://dle.rae.es/inteligencia#2DxmhCT>
- [3] "¿Qué es la Inteligencia Artificial (IA)? | IBM." Accessed: Sep. 16, 2024. [Online]. Available: <https://www.ibm.com/mx-es/topics/artificial-intelligence>
- [4] N. S. Consedine, "Health-Promoting and Health-Damaging Effects of Emotions: The View from Developmental Functionalism," *Handbook of Emotions*, pp. 676–690, 2008, Accessed: Sep. 16, 2024. [Online]. Available: https://books.google.com/books/about/Handbook_of_Emotions.html?hl=es&id=ulQQskejGwUC
- [5] McCorry LK. Physiology of the autonomic nervous system. *Am J Pharm Educ*. 2007 Aug 15;71(4):78. doi: 10.5688/aj710478. PMID: 17786266; PMCID: PMC1959222.
- [6] P. Ekman, "An argument for basic emotions," *Cogn Emot*, vol. 6, no. 3–4, pp. 169–200, May 1992, doi: 10.1080/02699939208411068.
- [7] P. Ekman and W. V Friesen, "Unmasking the face: A guide to recognizing emotions from facial expressions," *Ann Phys (N Y)*, p. 212, 1975, Accessed: Sep. 16, 2024. [Online]. Available: https://books.google.com/books/about/Unmasking_the_Face.html?id=TukNoJDgMTUC
- [8] R. Plutchik, "A psychoevolutionary theory of emotions," *Social Science Information*, vol. 21, no. 4–5, pp. 529–553, Jul. 1982, doi: <https://doi.org/10.1016/B978-0-12-558701-3.50007-7>.
- [9] "Rueda-Emociones-2D.png (1000×1500)." Accessed: Sep. 16, 2024. [Online]. Available: <https://ayuda-psicologica-en-linea.com/wp-content/uploads/2017/09/Rueda-Emociones-2D.png>
- [10] A. A. R. Kamal, J. B. Harness, G. Irving, and A. J. Mearns, "Skin photoplethysmography — a review," *Comput Methods Programs Biomed*, vol. 28, no. 4, pp. 257–269, Apr. 1989, doi: 10.1016/0169-2607(89)90159-4.
- [11] P. Romaniszyn-Kania *et al.*, "Affective State during Physiotherapy and Its Analysis Using Machine Learning Methods," *Sensors 2021, Vol. 21*, vol. 21, no. 14, p. 4853, Jul. 2021, doi: 10.3390/S21144853.
- [12] A. Albraikan, D. P. Tobón, and A. El Saddik, "Toward User-Independent Emotion Recognition Using Physiological Signals," *IEEE Sens J*, vol. 19, no. 19, 2019, doi: 10.1109/JSEN.2018.2867221.

- [13] J. Marín-Morales *et al.*, “Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors OPEN”, doi: 10.1038/s41598-018-32063-4.
- [14] Marín-Morales J, Higuera-Trujillo JL, Greco A, Guixeres J, Llinares C, Scilingo EP, Alcañiz M, Valenza G. Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors. *Sci Rep.* 2018 Sep 12;8(1):13657. doi: 10.1038/s41598-018-32063-4. PMID: 30209261; PMCID: PMC6135750..
- [15] K. H. Kim, S. W. Bang, and S. R. Kim, “Emotion recognition system using short-term monitoring of physiological signals,” *Med Biol Eng Comput*, vol. 42, no. 3, pp. 419–427, May 2004, doi: 10.1007/BF02344719.
- [16] Sharma, M., Kacker, S., & Sharma, M. (2016). A Brief Introduction and Review on Galvanic Skin Response. *International Journal of Medical Research Professionals*, 2(6). <https://doi.org/10.21276/ijmrp.2016.2.6.003>.
- [17] “Electrodermal Activity - Wolfram Boucsein - Google Libros.” Accessed: Sep. 17, 2024. [Online]. Available: https://books.google.es/books?hl=es&lr=&id=6N6rnOEZEEoC&oi=fnd&pg=PR3&dq=electrodermal+activity&ots=B4ixkyOPIG&sig=nwnNLvPj4hPmC0baQmYMwkHHbgg&redir_esc=y#v=onepage&q=electrodermal%20activity&f=false
- [18] P. J. Bota, C. Wang, A. L. N. Fred, and H. Placido Da Silva, “A Review, Current Challenges, and Future Possibilities on Emotion Recognition Using Machine Learning and Physiological Signals,” *IEEE Access*, vol. 7, pp. 140990–141020, 2019, doi: 10.1109/ACCESS.2019.2944001.
- [19] F. A. Boiten, “The effects of emotional behaviour on components of the respiratory cycle,” *Biol Psychol*, vol. 49, no. 1–2, pp. 29–51, Sep. 1998, doi: 10.1016/S0301-0511(98)00025-8.
- [20] I. Homma and Y. Masaoka, “Breathing rhythms and emotions,” *Exp Physiol*, vol. 93, no. 9, pp. 1011–1021, 2008, doi: 10.1113/EXPPHYSIOL.2008.042424.
- [21] S. E. Rimm-Kaufman and J. Kagan, “The psychological significance of changes in skin temperature,” *Motiv Emot*, vol. 20, no. 1, pp. 63–78, 1996, doi: 10.1007/BF02251007.
- [22] Z. Khalili and M. H. Moradi, “Emotion detection using brain and peripheral signals,” *2008 Cairo International Biomedical Engineering Conference, CIBEC 2008*, 2008, doi: 10.1109/CIBEC.2008.4786096.
- [23] K. Blinowska and P. Durka, “ELECTROENCEPHALOGRAPHY (EEG)”.
- [24] M. Soleymani, S. Asghari-Esfeden, Y. Fu, and M. Pantic, “Analysis of EEG Signals and Facial Expressions for Continuous Emotion Detection,” *IEEE Trans Affect Comput*, vol. 7, no. 1, pp. 17–28, Jan. 2016, doi: 10.1109/TAFFC.2015.2436926.
- [25] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, “A multimodal database for affect recognition and implicit tagging,” *IEEE Trans Affect Comput*, vol. 3, no. 1, pp. 42–55, Jan. 2012, doi: 10.1109/T-AFFC.2011.25.

- [26] Brooks SV. Current topics for teaching skeletal muscle physiology. *Adv Physiol Educ.* 2003 Dec;27(1-4):171-82. doi: 10.1152/advan.2003.27.4.171. PMID: 14627615.
- [27] R. F. M. Kleissen, J. H. Buurke, J. Harlaar, and G. Zilvold, "Electromyography in the biomechanical analysis of human movement and its clinical application," *Gait Posture*, vol. 8, no. 2, pp. 143–158, Oct. 1998, doi: 10.1016/S0966-6362(98)00025-3.
- [28] S. Koelstra *et al.*, "DEAP: A Database for Emotion Analysis ;Using Physiological Signals," in *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18-31, Jan.-March 2012, doi: 10.1109/T-AFFC.2011.15
- [29] M. Soleymani, J. J. M. Kierkels, G. Chanel, and T. Pun, "A Bayesian framework for video affective representation," *Proceedings - 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009, 2009*, doi: 10.1109/ACII.2009.5349563.
- [30] M. Soleymani, J. Lichtenauer, T. Pun and M. Pantic, "A Multimodal Database for Affect Recognition and Implicit Tagging," in *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42-55, Jan.-March 2012, doi: 10.1109/T-AFFC.2011.25.
- [31] J. Abdon Miranda-Correa, S. Member, M. Khomami Abadi, N. Sebe, S. Member, and I. Patras, "AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups," *IEEE Trans Affect Comput*, no. 1, Accessed: Sep. 17, 2024. [Online]. Available: <http://www.eecs.qmul.ac.uk/mmv/datasets/amigos/>
- [32] O. Kramer, "K-Nearest Neighbors," pp. 13–23, 2013, doi: 10.1007/978-3-642-38652-7_2.
- [33] K. Sun, J. Zhu, and J. Liang, "Emotion Recognition of Physical Activities for Health Monitoring Using Machine Learning," *APCCAS 2022 - 2022 IEEE Asia Pacific Conference on Circuits and Systems*, pp. 400–403, 2022, doi: 10.1109/APCCAS55924.2022.10090279.
- [34] W. S. Noble, "What is a support vector machine?," *Nature Biotechnology* 2006 24:12, vol. 24, no. 12, pp. 1565–1567, Dec. 2006, doi: 10.1038/nbt1206-1565.
- [35] P. Romaniszyn-Kania *et al.*, "Affective State during Physiotherapy and Its Analysis Using Machine Learning Methods," *Sensors* 2021, Vol. 21, Page 4853, vol. 21, no. 14, p. 4853, Jul. 2021, doi: 10.3390/S21144853.
- [36] Z. Khalili and M. H. Moradi, "Emotion detection using brain and peripheral signals," *2008 Cairo International Biomedical Engineering Conference, CIBEC 2008, 2008*, doi: 10.1109/CIBEC.2008.4786096.
- [37] Alaa Tharwat, Tarek Gaber, Abdelhameed Ibrahim, and Aboul Ella Hassanien. 2017. Linear discriminant analysis: A detailed tutorial. *AI Commun.* 30, 2 (2017), 169–190. <https://doi.org/10.3233/AIC-170729>
- [38] J. Ali, R. Khan, N. Ahmad, and I. Maqsood, "Random Forests and Decision Trees," 2012, Accessed: Sep. 17, 2024. [Online]. Available: www.IJCSI.org

- [39] M. H. Sedaaghi, M. Khosravy, M. H. Sedaaghi, and M. Khosravi, "Morphological ECG Signal Pre-processing with More Efficient Baseline Drift Removal," 2003, Accessed: Sep. 17, 2024. [Online]. Available: <https://www.researchgate.net/publication/230851489>
- [40] "Welcome to BioSPPy — BioSPPy 0.6.1 documentation." Accessed: Sep. 17, 2024. [Online]. Available: <https://biosppy.readthedocs.io/en/stable/>
- [41] "J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," in *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230-236, March 1985, doi: 10.1109/TBME.1985.325532
- [42] Dr. D. Sundararajan, "Power Spectral Density," *Digit Signal Process*, pp. 157–194, 2024, doi: 10.1007/978-3-031-56740-7_5.
- [43] G. Valenza, A. Lanata, and E. P. Scilingo, "The role of nonlinear dynamics in affective valence and arousal recognition," *IEEE Trans Affect Comput*, vol. 3, no. 2, pp. 237–249, 2012, doi: 10.1109/T-AFFC.2011.30.
- [44] A. M. Bianchi, L. Mainardi, E. Petrucci, M. G. Signorini, M. Mainardi, and S. Cerutti, "Time-Variant Power Spectrum Analysis for the Detection of Transient Episodes in HRV Signal," *IEEE Trans Biomed Eng*, vol. 40, no. 2, pp. 136–144, 1993, doi: 10.1109/10.212067.
- [45] R. J. Ellis, B. Zhu, J. Koenig, J. F. Thayer, and Y. Wang, "A careful look at ECG sampling frequency and R-peak interpolation on short-term measures of heart rate variability," *Physiol Meas*, vol. 36, no. 9, p. 1827, Aug. 2015, doi: 10.1088/0967-3334/36/9/1827.
- [46] Z. German-Sallo, "Wavelet Transform based HRV Analysis," *Procedia Technology*, vol. 12, pp. 105–111, Jan. 2014, doi: 10.1016/J.PROTCY.2013.12.462.
- [47] F. Pouromran, S. Radhakrishnan, and S. Kamarthi, "Exploration of physiological sensors, features, and machine learning models for pain intensity estimation," *PLoS One*, vol. 16, no. 7, p. e0254108, Jul. 2021, doi: 10.1371/JOURNAL.PONE.0254108.
- [48] F. A. Boiten, "The effects of emotional behaviour on components of the respiratory cycle," *Biol Psychol*, vol. 49, no. 1–2, pp. 29–51, Sep. 1998, doi: 10.1016/S0301-0511(98)00025-8.
- [49] "Data Preparation for Machine Learning." Accessed: Sep. 25, 2024. [Online]. Available: <https://machinelearningmastery.com/data-preparation-for-machine-learning/>
- [50] "Métricas de Regresión para ML - sitiobigdata.com." Accessed: Sep. 20, 2024. [Online]. Available: <https://sitiobigdata.com/2018/08/27/machine-learning-metricas-regresion-mse/>
- [51] "Mauri M, Magagnin V, Cipresso P, Mainardi L, Brown EN, Cerutti S, Villamira M, Barbieri R. Psychophysiological signals associated with affective states. *Annu Int Conf IEEE Eng Med Biol Soc*. 2010;2010:3563-6. doi: 10.1109/IEMBS.2010.5627465. PMID: 21096828; PMCID: PMC3059749.

- [52] D. Ayata, Y. Yaslan, · Mustafa, and E. Kamasak, “Emotion Recognition from Multimodal Physiological Signals for Emotion Aware Healthcare Systems,” *J Med Biol Eng*, vol. 40, pp. 149–157, 2020, doi: 10.1007/s40846-019-00505-7.
- [53] Doma, V., Pirouz, M. A comparative analysis of machine learning methods for emotion recognition using EEG and peripheral physiological signals. *J Big Data* **7**, 18 (2020). <https://doi.org/10.1186/s40537-020-00289-7>.
- [54] Tongshuai Song, Guanming Lu, and Jingjie Yan. 2020. Emotion Recognition Based on Physiological Signals Using Convolution Neural Networks. In Proceedings of the 2020 12th International Conference on Machine Learning and Computing (ICMLC '20). Association for Computing Machinery, New York, NY, USA, 161–165. <https://doi.org/10.1145/3383972.3384003>