



Universidad
Zaragoza

Trabajo Fin de Grado

Reconocimiento y Clasificación de Emociones
Empleando Redes Neuronales

Emotion Recognition and Classification Using
Neural Networks

Autor

Ilmira Iembergenova

Director

Dr. David Asiain Ansorena

Escuela Universitaria Politécnica La Almunia

Septiembre 2023



**Escuela Universitaria
Politécnica** - La Almunia
Centro adscrito
Universidad Zaragoza

**ESCUELA UNIVERSITARIA POLITÉCNICA
DE LA ALMUNIA DE DOÑA GODINA (ZARAGOZA)**

MEMORIA

Reconocimiento y clasificación de Emociones
Empleando Redes Neuronales

Emotion Recognition and Classification Using
Neural Networks

424.22.28

Autor: Ilmira Iembergenova

Director: Dr. David Asiain Ansorena

Fecha: 09 2023

RESUMEN

El presente trabajo se enfoca en el reconocimiento de emociones humanas a partir del análisis de señales fisiológicas mediante el uso de redes neuronales. Esto se lleva a cabo a través de un estudio exhaustivo de diversos tipos de entradas y modelos de redes neuronales destinados a la identificación de las seis emociones primordiales.

El método se fundamenta en el ensayo con un grupo de personas 18-25 años. Estos individuos se exponen a diecisiete vídeos de distintos géneros cinematográficos con el propósito de evocar emociones. Este estudio se lleva a cabo durante el período de una semana, en este mismo se hace uso de una variedad de sensores para medir la actividad del corazón (ECG), la conductancia de la piel (EDA) y la obtención de imágenes.

Durante la fase de recopilación se realizan evaluaciones de emociones e intensidad para cada uno de los sujetos y una vez obtenidos los datos integrales, se aplican metodologías de depuración para extraer características relevantes de cada tipo de señal. Estos datos se almacenan en archivos de formatos de tipo CSV, que posteriormente se emplean como capas de características y de etiquetas en la red neuronal.

Por otro lado, para el procesamiento de los datos, se selecciona una red de Memoria a Corto y Largo Plazo Bidireccional (BLSTM), cuya configuración se ajusta en dos modalidades: de forma manual, a través de estudios y conocimientos técnicos, y de manera automática, utilizando GridSearchCV, con el propósito de determinar los hiperparámetros óptimos para el modelo.

Además, se lleva a cabo un estudio para analizar la influencia que ejercen las etiquetas usadas en la red neuronal, evaluando el desempeño tanto con un enfoque multi-etiqueta como con etiquetas únicas. Para medir el rendimiento del modelo, se efectúan cálculos de pérdida y precisión.

Es importante incidir que las entradas a la red neuronal se dividen en dos conjuntos esenciales: el conjunto de entrenamiento y el conjunto de validación. Estos conjuntos se emplean, respectivamente, para el proceso de entrenamiento y para validar el rendimiento de la red. Con esto mismo, permitiendo estudiar la influencia que diversos factores, como la cantidad y tipo de datos de entrada, ejercen sobre el comportamiento y el rendimiento de la red neuronal bidireccional.

PALABRAS CLAVE

Reconocimiento de emociones, redes neuronales, reconocimiento emocional con imágenes, actividad electrodérmica, variabilidad de frecuencia cardíaca, E4 Empatica, MsWH

ABSTRACT

This paper focuses on the recognition of human emotions from the analysis of physiological signals using neural networks. This is done through an exhaustive study of various types of inputs and neural network models aimed at identifying the six primary emotions.

The method used to obtain the data is based on testing a group of 18-25 years old. These individuals are exposed to seventeen videos of different film genres for the purpose of evoking emotions. This study is conducted over a period of one week, using a variety of sensors to measure heart activity (ECG), skin conductance (EDA) and imaging.

During the collection phase, emotion and intensity assessments are performed for each of the subjects, and once the comprehensive data is obtained, debugging methodologies are applied to extract relevant features from each type of signal. These data are stored in CSV file formats, which are then used as feature and label layers in the neural network.

On the other hand, for data processing, a Bidirectional Short and Long Term Memory (BLSTM) network is selected, whose configuration is adjusted in two ways: manually, through studies and technical knowledge, and automatically, using GridSearchCV, in order to determine the optimal hyperparameters for the model.

In addition, a study is carried out to analyse the influence exerted by the labels used in the neural network, evaluating the performance both with a multi-label approach and with single labels. To measure the performance of the model, loss and accuracy calculations are performed.

It is important to note that the inputs to the neural network are divided into two essential sets: the training set and the validation set. These sets are used, respectively, for the training process and to validate the performance of the network. This allows us to study the influence that various factors, such as the amount and type of input data, exert on the behaviour and performance of the bidirectional neural network.

KEY WORDS

Emotion recognition, neural networks, emotional recognition with images, electrodermal activity, heart rate variability, E4 Empatica, MsWH

ÍNDICE DE CONTENIDO BREVE

1. INTRODUCCIÓN	1
2. ESTADO DE ARTE	5
3. DESARROLLO	51
4. RESULTADOS	65
5. CONCLUSIONES	85
6. ABREVIATURAS	87
7. BIBLIOGRAFÍA	91

ÍNDICE DE CONTENIDO

1. INTRODUCCIÓN	1
1.1. MOTIVACIÓN	1
1.2. OBJETO	2
1.3. METODOLOGÍA	3
2. ESTADO DE ARTE	5
2.1. CONCEPTOS BÁSICOS	6
2.1.1. <i>Introducción al reconocimiento y clasificación de emociones</i>	6
2.1.2. <i>Fundamentos teóricos de las señales</i>	8
2.1.2.1. Actividad electrodérmica - EDA	8
2.1.2.2. Actividad cardíaca – ECG, HRV, IBI, HR	11
2.1.2.3. Vídeos e imágenes	13
2.1.3. <i>Métodos de reconocimiento y clasificación de emociones</i>	15
2.1.3.1. Sistemas de inteligencia artificial	15
2.1.3.1.1. Sistemas de reconocimiento de emociones con inteligencia artificial	16
2.1.3.2. Obtención y selección de datos	19
2.1.3.3. Procesamiento y extracción de características de las señales	19
2.1.3.4. Clasificación de emociones	29
2.1.3.5. Redes neuronales	29
2.1.3.5.1. El concepto de red neuronal artificial	29
2.1.3.5.2. Neurona biológica y neurona artificial	30

INDICES

2.1.3.5.3. Red neuronal aplicada en Python	35
2.2. ESTUDIOS RELACIONADOS	40
2.2.1. Reconocimiento y clasificación de imágenes	40
2.2.1.1. Reconocimiento de emociones usando Deep Learning [29]	41
2.2.1.2. Metodologías de reconocimiento facial en modelos bidimensionales [30]	42
2.2.1.3. Sistema de reconocimiento de expresiones faciales que emplea rasgos faciales robustos de videos profundos y Deep learning [31]	44
2.2.2. Señales EDA	45
2.2.2.1. Aprendizaje automático en el diseño de un detector de estrés a partir de señales biomédicas [32]	45
2.2.2.2. Reconocimiento de emociones a través de las señales fisiológicas del observador [33]	46
2.2.2.3. Reconocimiento de emociones utilizando señales de actividad electrodérmica y red neuronal convolucional profunda multiescalar [34]	47
2.2.3. Electrocardiogramas- ECG	47
2.2.3.1. Utilidad de la forma de onda ECG completa para la clasificación de estrés [37]	48
2.2.3.2. Algoritmos de aprendizaje automático basados en ECG para la clasificación de los latidos del corazón [38]	49
3. DESARROLLO	51
3.1. RECOPIACIÓN DE DATOS	51
3.1.1. Entorno	51
3.1.2. Sistema de sensores	52
3.1.2.1. Colocación de los sensores	53
3.1.2.2. Ejecución del ensayo	55
3.1.2.2.1. Descripción de la prueba	55
3.2. SELECCIÓN DE DATOS	57
3.3. INSTALACIÓN DE SOFTWARES	57
3.4. PROCESAMIENTO DE DATOS	57
4. RESULTADOS	65
4.1. HIPÓTESIS DE RED NEURONAL CON DOBLE ETIQUETA	73
4.2. HIPÓTESIS DE RED NEURONAL CON UNA ÚNICA ETIQUETA	75
4.2.1. Emociones según MTCNN	75
4.2.1.1. Optimización manual MTCNN	75
4.2.1.2. Optimización automatizada MTCNN	77
4.2.2. Emociones según VideoID	79
4.2.2.1. Optimización manual VideoID	80
4.2.2.2. Optimización automatizada VideoID	81
5. CONCLUSIONES	85



	INDICES
5.1. TRABAJOS FUTUROS	86
6. ABREVIATURAS	87
7. BIBLIOGRAFÍA	91

ÍNDICE DE ILUSTRACIONES

Ilustración 1: EDA con respuestas de conductancia cutánea detectadas (SCR - puntos negros) [5]	8
Ilustración 2: ECG de un corazón en ritmo sinusal normal, creado por Anthony Atkielski	23
Ilustración 3: Partes de un esquema de una neurona biológica	30
Ilustración 4: Propagación hacia delante	32
Ilustración 5: Propagación hacia atrás	33
Ilustración 6: Esquema una Red Neuronal Convolutiva	36
Ilustración 7: Red de Memoria a Corto y Largo Plazo Bidireccional	38
Ilustración 8: Ensayo en un entorno controlado en la EUPLA	51
Ilustración 9: MsWH [39].....	52
Ilustración 10: Muse 2 [40]	53
Ilustración 11: Empatica E4 [41]	53
Ilustración 12: Puntos de colocación de Muse 2 y OV2640	54
Ilustración 13: Puntos de ubicación de sensores sobre la mano y muñeca izquierdas.....	54
Ilustración 14: Sam grado de activación [46]	56
Ilustración 15: Diagrama de flujo de datos.....	58
Ilustración 16: Facial Emotion Recognition (MTCNN) (1)	59
Ilustración 17: Resample and filter (2)	60
Ilustración 18: Feature extraction (3).....	61
Ilustración 19: Resample and filter (4)	62
Ilustración 20: Feature extraction (5).....	62
Ilustración 21: Register and readjustment of scale (6)	63
Ilustración 22: Artificial Neural Network (7)	64
Ilustración 23: Frecuencia de picos en EDA de 3 sujetos.....	65

Ilustración 24: Frecuencia de picos en HR de 3 sujetos.....	66
Ilustración 25: Sujeto 1: frecuencia de picos en EDA y HR.....	67
Ilustración 26: Sujeto 2: frecuencia de picos en EDA y HR.....	68
Ilustración 27: Sujeto 3: frecuencia de picos en EDA y HR.....	70
Ilustración 28: Emoción: Enfado - cambios bruscos en señal EDA.....	70
Ilustración 29: Emoción: Sorpresa - cambios bruscos y medios en señal EDA...	71
Ilustración 30: Emoción: Alegría - cambios inestables en señal EDA.....	71
Ilustración 31: Emoción: Miedo - cambios bruscos, variedad de picos en señal EDA	71
Ilustración 32: Amplitudes de EDA en función de las emociones de 3 sujetos ...	72
Ilustración 33: Etiqueta doble. Entrenamiento y validación de red: pérdidas. ...	74
Ilustración 34: Etiqueta doble. Entrenamiento y validación de red: precisión. ...	74
Ilustración 35: Etiqueta única: MTCNN. Entrenamiento y validación de red: pérdidas. Optimización manual.....	76
Ilustración 36: Etiqueta única: MTCNN. Entrenamiento y validación de red: precisión. Optimización manual.	77
Ilustración 37: Etiqueta única: MTCNN. Entrenamiento y validación de red: pérdidas. Optimización GridSearchCV.....	78
Ilustración 38: Etiqueta única: MTCNN. Entrenamiento y validación de red: precisión. Optimización GridSearchCV.	79
Ilustración 39: Etiqueta única: VideoID. Entrenamiento y validación de red: pérdidas. Optimización manual.....	80
Ilustración 40: Etiqueta única: VideoID. Entrenamiento y validación de red: precisión. Optimización manual.	81
Ilustración 41: Etiqueta única: VideoID. Entrenamiento y validación de red: pérdidas. Optimización GridSearchCV.....	82
Ilustración 42: Etiqueta única: VideoID. Entrenamiento y validación de red: precisión. Optimización GridSearchCV.	83

ÍNDICE DE TABLAS

Tabla 1: Sujeto 1, características de HR y EDA	66
Tabla 2: Sujeto 2, características de HR y EDA	67
Tabla 3: Sujeto 3, características de HR y EDA	69
Tabla 4: Etiqueta doble. Entrenamiento y testeo de red, pérdidas y precisiones.	73
Tabla 5: Etiqueta única: MTCNN. Entrenamiento y testeo de red, pérdidas y precisiones. Optimización manual.	76
Tabla 6: Etiqueta única: MTCNN. Entrenamiento y testeo de red, pérdidas y precisiones. Optimización GridSearchCV.	78
Tabla 7: Etiqueta única: VideoID. Entrenamiento y testeo de red, pérdidas y precisiones. Optimización manual.	80
Tabla 8: Etiqueta única: VideoID. Entrenamiento y testeo de red, pérdidas y precisiones. Optimización GridSearchCV.	82

1. INTRODUCCIÓN

El crecimiento de volúmenes de información, así como la expansión de la gama de problemas de toma de decisiones técnicamente complejas, requieren la sistematización de los métodos existentes y el desarrollo de nuevos métodos y algoritmos que aporten soluciones óptimas.

Este trabajo de fin de grado explora las diferentes redes neuronales para abordar la comprensión y clasificación de las emociones humanas, destacando la importancia de la tecnología y la ciencia de datos en esta área en constante evolución.

Es importante destacar los avances en el reconocimiento de las emociones humanas combinando la técnica de la doble etiqueta, demostrando una disminución en las pérdidas y mejorando la precisión, asimismo proporcionando un material preparado para investigaciones futuras.

El trabajo describe en la sección 2. el estado de arte, el cual a su vez se subdivide en dos grandes bloques, el 2.1 de conceptos básicos y 2.2 de estudios relacionados, donde se explican todos los estudios que se tienen en cuenta para el desarrollo del propio.

A continuación, el apartado 3. consta de todo el proceso de desarrollo, desde la captación de datos hasta el procesamiento de los dichos. El análisis de estos datos se encuentra en el apartado 4. de resultados y la conclusión en el apartado 5, junto con las propuestas y trabajos futuros.

1.1. MOTIVACIÓN

La inteligencia artificial es uno de los temas que más en auge se encuentra hoy en día, sobre todo el relacionado con las emociones humanas. El tema ha sido propuesto por el profesor y director de este Trabajo de Fin de Grado, Dr. David Asiain Ansorena, el cual indicó buscar un tipo de inteligencia artificial que sea capaz de establecer relación entre los diferentes tipos de señales físicas y biomédicas para determinar emociones. La justificación de su realización viene dada por la búsqueda de soluciones para un tema que engloba dos áreas de estudio particulares y poco vistas en la EUPLA, tratamiento de señales biomédicas y aplicación de redes neuronales para su solución.

Cabe destacar que los ensayos se desarrollan en la EUPLA por el Grupo de Investigación en Interfaces Avanzadas AFFECTIVE LAB del Instituto de Investigación en Ingeniería de Aragón, del cual son miembro Dr. David Asiain Ansorena, Mg. Gustavo Astudillo y Dr. Jesús Ponce de León Vázquez.

1.2. OBJETO

El objetivo principal del trabajo es analizar los gestos de un rostro humano, las señales electrodérmicas y la frecuencia cardíaca del corazón para identificar emociones. Para ello, se busca seleccionar modelos de redes neuronales adecuados e implementar algoritmos que permitan el reconocimiento preciso a partir de estas señales fisiológicas.

En este sentido, se pretende encontrar una red neuronal óptima en términos de su estructura interna y el manejo eficiente de los flujos de información entre las neuronas. Estos modelos seleccionados serán utilizados para resolver problemas prácticos relacionados con el reconocimiento de emociones.

Por ello, los objetivos específicos de este trabajo son:

- 1) Analizar los métodos existentes para el reconocimiento y clasificación de emociones mediante el reconocimiento facial a través de imágenes, electrocardiogramas (ECG o EKG), actividad electrodérmica (EDA), y seleccionar los enfoques más adecuados para cada uno de ellos.
- 2) Realizar un estudio exhaustivo de los diferentes tipos de redes neuronales que sean capaces de manejar conjuntos de datos fisiológicos, así como comprender sus métodos de funcionamiento.
- 3) Determinar las características más relevantes de cada tipo de señal. Extraer la amplitud, los picos y la variabilidad de la señal EDA para comprender los cambios en la conductancia de la piel asociados a las emociones. En el caso de las imágenes, es relevante analizar los patrones de expresiones faciales y los rasgos que indican diferentes estados emocionales. En cuanto a la frecuencia cardíaca, se debe analizar su ritmo y variabilidad para identificar posibles correlaciones con las emociones experimentadas.

Y los relacionados con los Objetivos de Desarrollo Sostenible (ODS):

- ODS 3.4, responsable de la Salud y Bienestar, para mejorar la salud mental y emocional al aplicar técnicas de inteligencia artificial y reconocimiento de emociones para comprender y abordar las emociones humanas.
- ODS 9.1, responsable de la Innovación y la Infraestructura, promoviendo la innovación y el desarrollo de infraestructuras tecnológicas resilientes como las redes neuronales y tecnologías avanzadas en la resolución práctica de problemas en el reconocimiento de emociones.

1.3. METODOLOGÍA

Este apartado se centra en el procesamiento de las señales fisiológicas y el análisis de estas, extrayendo las características más relevantes en función del tipo de señal analizada. Posteriormente, las señales junto con sus características se introducen en un modelo de inteligencia artificial que se entrena y se testea a base de datos de entrada. Para poder desarrollarlo es necesario tener una base teórica y perspectivas para resolver el problema de reconocimiento automático de emociones y empleo de diferentes señales para dicha tarea. Junto con este se mencionan las tecnologías y softwares que son empleados para la resolución de cada uno de los problemas.

Dentro de la revisión teórica de los métodos y conceptos más relevantes es importante mencionar que para el desarrollo de este proyecto se toman las 6 emociones principales que propuso Paul Ekman más el estado neutral. [1][2]

El desarrollo práctico implica el procesamiento previo de datos que son posteriormente empleados para el entrenamiento de redes neuronales y, consecuentemente, presentación de resultados y las pruebas del sistema desarrollado.



2. ESTADO DE ARTE

El reconocimiento y clasificación de emociones juega un papel crucial en numerosas áreas de investigación, desde la psicología hasta la inteligencia artificial. En los últimos años, se han llevado a cabo investigaciones significativas que se centran en el uso de señales de actividad electrodérmica (EDA), frecuencia cardíaca (HR) e imágenes para comprender y capturar las experiencias emocionales humanas. En este estado de arte, se presenta una revisión exhaustiva de los avances más relevantes en este campo, explorando los diferentes enfoques y técnicas utilizados en la literatura científica, entre los cuales destacan las redes neuronales y aprendizaje automático.

En este contexto, el análisis de imágenes se ha explorado ampliamente en el reconocimiento de respuestas emocionales. Se ha evidenciado que las imágenes son una fuente de información vital para capturar expresiones y características visuales relacionadas con las emociones. Acuerdo a esto, se han utilizado técnicas de visión especializada por ordenador y redes neuronales para desarrollar modelos de clasificación de emociones basados en imágenes.

En paralelo, el estudio de las señales EDA ha sido objeto de gran atención en relación con las emociones. Las investigaciones fueron centradas en emplear técnicas de procesamiento de señales para extraer características relevantes y clasificar las emociones experimentadas por los individuos a partir de EDA. Aunque estos enfoques han demostrado resultados prometedores, enfrentan desafíos en la variabilidad interindividual y adaptación a diferentes contextos.

Por otro lado, un aspecto fundamental en el estudio de reconocimiento y clasificación de respuestas emocionales es la frecuencia cardíaca, esta es estudiada como medida del número de latidos que el corazón realiza en un tiempo determinado, comúnmente expresado en cantidad de latidos por minuto (LPM). La frecuencia cardíaca se ve influenciada y asociada a las emociones y el estrés, puesto que esta ayuda en la comprensión del funcionamiento de los sistemas cardiovascular y nervioso autónomo en situaciones que involucren actividad emocional. En situaciones de excitación o estrés, se ha observado que la frecuencia cardíaca tiende a aumentar, reflejando así la influencia directa de las emociones en el sistema cardiovascular.

2.1. CONCEPTOS BÁSICOS

2.1.1. Introducción al reconocimiento y clasificación de emociones

La emoción es un proceso mental que refleja la actitud evaluativa subjetiva de una persona hacia diversas situaciones y objetos.

La principal diferencia entre emociones y sentimientos es que las primeras son una experiencia subjetiva temporal de situaciones y eventos que son importantes para él, y las segundas son una actitud emocional estable de un individuo hacia cualquier objeto.

A continuación, se presentan algunos aspectos clave o acciones principales que resaltan la importancia de las emociones:

- Comunicación emocional y toma de decisiones. Es gracias a las emociones que se entiende en qué estado mental y físico se encuentra una persona y, dependiendo de esto, se elige una u otra línea de comportamiento para la comunicación. Este último aspecto influye en las decisiones morales y éticas.
- Regulación o motivación. Las emociones inducen al individuo a la acción, guían y determinan el comportamiento posterior y pueden bloquear el pensamiento.
- Señalización. La comunicación no verbal, como las expresiones faciales, gestos, tono de voz, lenguaje corporal y señales biomédicas, durante las emociones señalan a los individuos que se encuentren dentro del ambiente del sujeto los deseos y necesidades de este.
- También es importante destacar el efecto protector de las emociones puesto que este induce una reacción instantánea para salvar a una persona en situaciones de peligro o la puede paralizar para realizar una acción.

Por lo que el reconocimiento y clasificación de emociones es un área de investigación amplia, la cual sigue creciendo. Hoy en día este campo de estudio se realiza de manera interdisciplinaria, en la cual combina la psicología, la inteligencia artificial y la tecnología.

Estas disciplinas se complementan entre sí, ya que el estudio de las emociones en la psicología brinda una base teórica, mientras que la inteligencia artificial y la tecnología proporcionan herramientas y técnicas para su detección y análisis.

En este contexto, EDA, HR e imágenes tienen una gran relevancia y presentan una gran variedad de aplicaciones en diversos campos de estudio. En primer lugar, en el ámbito de la salud mental, estas técnicas permiten identificar y monitorear el estado emocional de las personas, lo que puede ser útil en el diagnóstico y tratamiento de trastornos como la depresión, la ansiedad y el estrés. Además, en la investigación psicológica, el reconocimiento de emociones proporciona una herramienta objetiva para estudiar las respuestas emocionales y comprender mejor los procesos cognitivos y afectivos.

En el campo de interacción humano-máquina (human-to-machine), el reconocimiento de emociones permite crear sistemas más intuitivos y adaptativos. Por ejemplo, las aplicaciones de asistencia virtuales que permiten a los usuarios brindar respuestas personalizadas y mostrando cierto grado de empatía acorde a las emociones que haya identificado la aplicación. Asimismo, en el ámbito del entretenimiento, el reconocimiento de emociones puede mejorar la experiencia del usuario al adaptar los contenidos y la interacción en función de las emociones detectadas, como en videojuegos. Un ejemplo es el "Silent Hill shattered memories", desarrollado por Climax Studios y publicado por Konami Digital Entertainment [3], este juego de miedo se centra en la exploración de la mente del jugador, esto es, se enfoca en la psicología del usuario en el cual en función de la acción, guiada por la emoción, que se realice en tiempo real por el jugador las respuestas son correspondientemente adaptadas al entorno que cree el propio usuario con sus acciones, y, por tanto, en este caso, los monstruos que son creados durante la partida varían en función del individuo.

2.1.2. Fundamentos teóricos de las señales

El objetivo es establecer una base sólida de conocimientos teóricos que respalde la implementación, evaluación de características, aplicaciones en el reconocimiento y análisis de emociones, al mismo tiempo que se exploran los algoritmos y modelos de redes neuronales. De esta manera, se busca contribuir al avance en la comprensión y aplicación de estas señales en el ámbito emocional, impulsando la investigación y el desarrollo de herramientas eficaces para la detección de emociones.

2.1.2.1. Actividad electrodérmica - EDA

La actividad electrodérmica (EDA) es una propiedad del cuerpo humano que provoca un cambio permanente en las características eléctricas de la piel. Históricamente, la EDA también se conoce como conductancia de la piel, respuesta galvánica de la piel (GSR), respuesta electrodérmica (EDR), reflejo psicogalvánico (PGR), conductancia de la piel (SCR), respuesta simpática de la piel (SSR) y nivel de conductancia de la piel (SCL).[4] La conductancia (SCR) y el nivel de la conductancia de la piel (SCL) se pueden visualizar en la Ilustración 1.

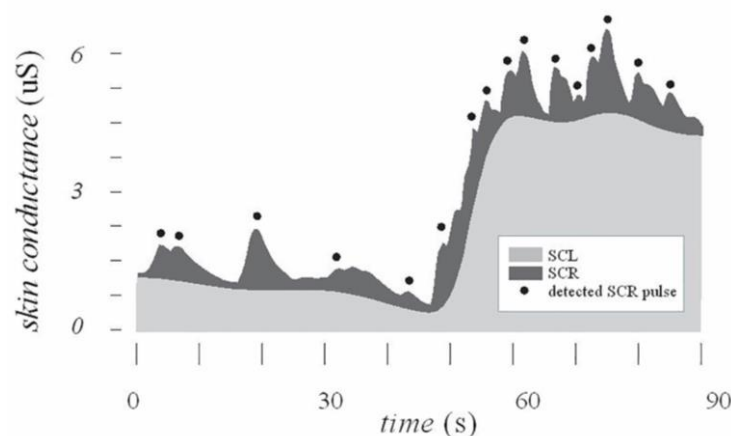


Ilustración 1: EDA con respuestas de conductancia cutánea detectadas (SCR - puntos negros) [5]

Una larga historia de investigación sobre las propiedades eléctricas activas y pasivas de la piel por parte de varias disciplinas ha dado como resultado una gran cantidad de nombres que ahora están estandarizados y denominados como actividad electrodérmica (EDA).

La teoría establece que la resistencia de la piel depende de la condición de las glándulas sudoríparas en la piel.

La sudoración está controlada por el sistema nervioso simpático [6], y la conducción cutánea es un indicador de excitación psicológica o fisiológica. Si la rama simpática del sistema nervioso autónomo está fuertemente excitada, también aumenta la actividad de las glándulas sudoríparas, lo que, a su vez, aumenta la conductividad de la piel. Por lo tanto, la conductancia de la piel puede ser una medida de la respuesta emocional y simpática [7][8].

Se ha descubierto que existe una relación entre la excitación emocional y la actividad simpática, aunque el cambio eléctrico en sí mismo no determina qué emoción se provoca. Estos cambios simpáticos autonómicos alteran el sudor y el flujo sanguíneo, lo que a su vez afecta a GSR y GSP. El número de glándulas sudoríparas varía según el cuerpo humano, sobre todo en la zona de los brazos y las piernas (200-600 glándulas sudoríparas por centímetro cuadrado) [9]. La reacción de la piel y el tejido muscular a los estímulos externos e internos puede provocar un cambio en la conductividad de varios microsiemens. Un dispositivo calibrado correctamente puede registrar y mostrar cambios sutiles.

En el ámbito de estudio de las respuestas emocionales, se han observado los cambios combinados entre la resistencia electrodérmica y el potencial electrodérmico constituyen la actividad electrodérmica. La resistencia galvánica de la piel (GSR) es un término más antiguo que se refiere a la resistencia eléctrica registrada entre dos electrodos cuando una corriente muy baja fluye constantemente entre ellos. Los electrodos generalmente están separados por una pulgada y la resistencia registrada varía según el estado emocional del paciente. El potencial galvánico de la piel (GSP) se refiere al voltaje medido entre dos electrodos sin corriente aplicada externamente y su rango de frecuencia no supera los 10 Hz. Se mide conectando los electrodos a un amplificador de voltaje. Esta tensión también depende del estado emocional de la persona.

Además, la EDA abarca dos tipos de actividades como la actividad simpática tónica que cambia lentamente como la actividad simpática fásica que cambia rápidamente. La actividad tónica se puede expresar en unidades de nivel electrodérmico (SCL), mientras que la actividad fásica se expresa en unidades de respuesta electrodérmica (EDR). [10]

Las técnicas que se emplean para la captación de este tipo de señales son las siguientes:

- GSR: esta técnica mide la conductancia eléctrica de la piel, que cambia en respuesta a la excitación emocional y otros procesos

psicológicos. En este caso, se emplean electrodos de contacto que se colocan sobre la piel. Estos registran los cambios de actividad en las glándulas sudoríparas; estos cambios son causa de las reacciones emocionales y fisiológicas que alteran la conductividad de la piel.

Una técnica similar es la de la potencia electrodermal, que consiste en medir el potencial eléctrico entre dos electrodos de Ag/AgCl de la piel, en general en estos casos se emplea un gel especial para aumentar la conducción y el contacto entre la piel y el electrodo. Esta técnica proporciona información sobre la variabilidad de la propia actividad electrodermal.

- Impedancia cutánea: esta técnica de medición de la actividad electrodérmica consiste en medir la resistencia eléctrica de la piel. Esta técnica registra que una mayor actividad sudorípara disminuye la resistencia y se interpreta como un aumento en los niveles de actividad electrodermal.

Respecto al análisis de la propia señal, los cambios fásicos (SCR) son cambios a corto plazo en EDA que aparecen en respuesta a un estímulo particular. Las SCR también pueden ocurrir espontáneamente sin estímulos externos visibles. Estos tipos de SCR se denominan "SCR no específicos" (NS.SCR). [11] El SCR fásico es útil en el estudio de procesos de atención multifacéticos. [12] Por otro lado, los cambios tónicos (SDL) se basan en parámetros de fase. Las fluctuaciones espontáneas en la SDR no específica se pueden utilizar para evaluar la EDA tónica. Más específicamente, usar la frecuencia de "SDR no específica" como una medida de EDA durante un período de tiempo, como 30 a 60 segundos. Los SCL EDA se considera útil en el estudio de la excitación y el estado de alerta general.

Es importante destacar que los factores externos, como la temperatura y la humedad, afectan las mediciones de EDA, lo que puede generar señales y resultados inconsistentes. [13]

Los factores intrínsecos, como los medicamentos y la hidratación, también pueden alterar las mediciones de EDA para mostrar inconsistencias con el mismo nivel de estímulo. Además, en el sentido clásico, la EDA se trató como si representara un cambio homogéneo en la excitación de todo el cuerpo, pero de hecho, diferentes lugares para medirlo pueden conducir a diferentes respuestas; por ejemplo, las respuestas de las muñecas izquierda y derecha son provocadas por diferentes áreas del cerebro, proporcionando múltiples fuentes de

estimulación; por lo tanto, la EDA medida en diferentes partes del cuerpo varía no solo según la densidad de las glándulas sudoríparas, sino también según las diversas fuentes principales de excitación. Esas señales son las de respuesta del sistema nervioso simpático las cuales involucran la activación fisiológica durante situaciones emocionales o estresantes. [14]

Por último, según los estudios, las reacciones electrodérmicas se retrasan entre 1 y 3 segundos. [15] Lo que muestran la dificultad para determinar la relación entre la EDA y la actividad simpática.

2.1.2.2. Actividad cardíaca – ECG, HRV, IBI, HR

La HR (frecuencia cardíaca) es una medida que examina la cantidad de latidos que realiza el corazón por minuto (LPM). También se le suele conocer como pulso o ritmo cardíaco refiriéndose a la frecuencia cardíaca, pero los conceptos de cada uno de los términos varían, puesto que el pulso se refiere a un latido o expansión de las arterias por un impulso de sangre que bombea el corazón, el ritmo cardíaco es, por otro lado, el patrón con el que se ejecutan los latidos. Es importante destacar que la frecuencia cardíaca no es lo mismo que la variabilidad cardíaca, ya que la variabilidad es una medida empleada para evaluar las fluctuaciones en el intervalo entre latidos continuados. [16][17]

Para determinar la frecuencia cardíaca es necesario tener en cuenta que esta información se extrae de los electroencefalogramas (ECG o EKG) que son pruebas no invasivas que consiste en realizar un registro dinámico de la carga eléctrica que hace funcionar el corazón. Para evaluar las características de esta carga, el registro se realiza a partir de varias secciones del músculo cardíaco. [18] Para ello se utilizan electrodos que se aplican en diferentes partes del tórax, muñecas y tobillos. Los electrodos más empleados son los de sujeción. Estos electrodos se adhieren a la piel y requieren de un gel para poder realizar una mejor conducción de la señal, y los textiles, son fabricados a partir de un material textil conductor que se colocan también sobre la piel y permiten grabar las señales de manera continua sin la necesidad de realizar reajustes.[19][20]

Las metodologías de captura de este tipo de datos más empleadas son las siguientes:

- ECG: este método se explica anteriormente, que consiste en colocar electrodos sobre zonas estratégicas del cuerpo para registrar la actividad eléctrica del corazón. La información que

proporciona sobre los intervalos de tiempo entre latidos (IBI) es detallada y consecuentemente permite extraer la frecuencia cardíaca.

- PPG: esta técnica se emplea principalmente en medir los cambios de volumen en los vasos sanguíneos mediante un sensor óptico. Este sensor se suele colocar sobre la muñeca o los dedos. La HR calcula analizando los cambios de volumen de la sangre, esto es, en el momento en el cual el corazón se relaja, el volumen de sangre en los vasos disminuye y ocurre lo contrario en caso de contracción, el volumen aumenta y la absorción de luz del sensor es mayor.

En los estudios se ha determinado que las emociones intensas, como la excitación, el enfado y el miedo, pueden influir en la frecuencia cardíaca, ya que se activa el sistema nervioso simpático [21], el cual regula las respuestas involuntarias del organismo, incluyendo la aceleración del ritmo cardíaco en situaciones de peligro. Sin embargo, es importante tener en cuenta que la relación entre las emociones y los cambios en la frecuencia cardíaca no es lineal ni exclusiva, ya que, intervienen otros factores.

Con relación a la variabilidad de la frecuencia cardíaca (HRV), diversos estudios indican que está influenciada por diferentes factores, incluyendo la actividad del sistema nervioso parasimpático y simpático. Una HRV alta se ha asociado con una regulación emocional adecuada, toma de decisiones y atención, mientras que una HRV baja puede reflejar lo contrario, aunque también pueden influir otros mecanismos fisiológicos y psicológicos en esta medida.

Es elemental considerar la complejidad de la regulación de la frecuencia cardíaca y la variabilidad de la frecuencia cardíaca, así como la influencia de múltiples factores en estos procesos. La relación entre las emociones y los cambios en la frecuencia cardíaca es dinámica y está sujeta a una interacción compleja entre diversos mecanismos fisiológicos, psicológicos y contextuales.

2.1.2.3. Vídeos e imágenes

Las imágenes o vídeos de personas son representaciones visuales de uno o varios individuos capturados mediante un medio visual. Estas imágenes se componen de píxeles, que son las divisiones mínimas que la forman. Cada píxel representa un punto de color en la imagen y almacena información sobre su posición espacial y el valor del color correspondiente. El tamaño y la resolución de la imagen se determinan según la cantidad de píxeles que la conforman.

Adicionalmente, una imagen puede contener canales de color. En el caso de una imagen en escala de grises, cada píxel tiene un único valor que representa la intensidad de gris en esa ubicación. Por otro lado, en una imagen a color, cada píxel cuenta con tres canales de color: rojo, verde y azul ("RGB", Red Green Blue). Cada canal refleja la intensidad de uno de esos colores primarios, y al combinarlos en diferentes proporciones se obtiene el color final de cada píxel. Las metodologías esenciales para extraer características de las imágenes son las siguientes:

- Metodología de extracción basada en píxeles: analiza las propiedades de los píxeles de una imagen, como intensidad, color y textura, utilizando algoritmos de procesamiento de imágenes para extraer información relevante.
- Metodología de extracción basada en regiones: segmenta la imagen en regiones, o ROI, y extrae características específicas de cada región, como forma, tamaño y simetría, mediante técnicas de segmentación y detección de bordes.
- Metodología de extracción basada en descriptores visuales: usa descriptores visuales predefinidos, como la transformada de características invariantes a escala (SIFT) y el histograma de gradientes orientados (HOG), para capturar información relevante de la imagen, como puntos de interés y características textuales, para su posterior representación y comparación en tareas de reconocimiento y clasificación.
- Metodología de extracción basada en redes neuronales convolucionales (CNN): se emplean arquitecturas de aprendizaje profundo para extraer características automáticamente de las imágenes, aprendiendo representaciones discriminativas a través de capas convolucionales, que pueden ser usadas en clasificadores o sistemas de detección de emociones.

Para realizar el reconocimiento de emociones, es esencial extraer características relevantes que reflejen los rasgos emocionales presentes. Algunas de las características esenciales que se consideran son las expresiones faciales, como las cejas, la mirada, sonrisas, pueden proporcionar información sobre las emociones que experimenta el individuo. Por otro lado, los movimientos oculares, el seguimiento puede revelar información sobre el enfoque y la atención del individuo, esto puede estar relacionado con emociones de interés, sorpresa, miedo, alegría o disgusto. Alternativamente, los patrones de color y las texturas faciales pueden contener información valiosa para la detección de emociones de felicidad, enojo, miedo, entre otras.

En relación con el análisis de imágenes para extraer características para un reconocimiento de emoción puede variar según la metodología y el enfoque empleado. Comúnmente se trabaja a nivel de píxeles donde se aplican algoritmos de procesamiento de imágenes y técnicas de extracción de características para obtener información de la respuesta emocional. En el análisis de reconocimiento de emociones, influyen diversos factores, como el contexto emocional, la variedad de expresiones, ambigüedad emocional, influencia cultural, etc.

En los estudios, se observa que se emplean diferentes metodologías relacionadas con el aprendizaje máquina en el cual destaca la última mencionada de la lista, las redes neuronales convoluciones (CNN), puesto que tienen la capacidad de aprender características relevantes directamente de los datos de entrada y, por tanto, se consideran eficaces en este tipo de tareas. Sin embargo, se encuentran limitaciones cuando se trata de la subjetividad en la etiquetación de datos puesto que la asignación de una emoción a una imagen es subjetiva y esta influenciada por las percepciones individuales de cada uno.

Por otro lado, en contexto al tamaño mínimo de píxeles se establece en los estudios que es suficiente con una resolución de 48x48 píxeles para realizar una clasificación de emociones básicas, [22] aunque está sujeto a limitaciones como la pérdida de detalles, la dificultad de discriminación, la variabilidad de representación e incluso impacto de ruido y distorsión.

2.1.3. Métodos de reconocimiento y clasificación de emociones

Hoy en día existen varios métodos empleados para el reconocimiento de clasificación de emociones, los más comunes son basados en expresiones faciales (imágenes o vídeos), procesamiento de voz, patrones fisiológicos (EDA y ECG), basados en textos y sensores corporales. A continuación, se presenta una clasificación de sistemas de inteligencia artificiales basados en diferentes factores.

2.1.3.1. Sistemas de inteligencia artificial

Los sistemas de inteligencia artificial poseen unos rasgos muy característicos:

Por un lado, los sistemas de IA permiten una comunicación natural entre el usuario y el sistema, permitiendo solicitudes y diálogos arbitrarios, gracias a habilidades de comunicación desarrolladas.

Por otro lado, estos tienen la capacidad de aprender automáticamente a partir de la experiencia y los datos acumulados, entrenando redes neuronales con grandes conjuntos de datos procesados.

Por último, es importante destacar que la IA es capaz de abordar problemas poco formalizados, que carecen de soluciones específicas y dependen de varios factores poco definidos. Se utilizan enfoques no estándar y redes neuronales artificiales para resolver eficazmente este tipo de problemas.

Las definiciones de los modelos de redes neuronales varían según su uso y aplicación, sin una definición única para toda la familia de modelos. La red neuronal original puede considerarse como un modelo basado en el aprendizaje automático, que adquiere propiedades en el proceso de entrenamiento al ajustar los pesos según una regla específica teniendo como objetivo principal de resolver problemas y encontrar patrones en los datos de entrada.

Las redes neuronales también son consideradas como una herramienta para la computación paralela, es decir, que es una red de procesadores, neuronas simples, distribuidos de manera masiva y dispuestas en paralelo, estas redes poseen una propiedad natural de almacenar la información empírica y tornarla disponible para un nuevo uso. El proceso

de funcionamiento de las neuronas es simple, periódicamente reciben y envían señales a otras/siguientes neuronas. [23]

Al construir redes neuronales, se puede utilizar tanto el aprendizaje supervisado, para perceptrones multicapa, como el aprendizaje no supervisado, para redes de Kohonen.

Muy a menudo, las redes neuronales se utilizan para resolver problemas de tipo:

- Aproximación de funciones. Recuperación de dependencias funcionales a partir de datos de entrenamiento.
- La clasificación. Determinación de si la imagen de entrada (objeto) representada por el vector de características pertenece a una de las clases predefinidas.
- La agrupación en clústeres. Agrupación de objetos basada en la proximidad de sus propiedades.
- Pronóstico. Predicción de un determinado valor.
- La optimización. Encontrar una solución que satisfaga el sistema de restricciones y maximice o minimice la función objetivo o final.
- La memoria asociativa. Memoria direccionable por contenido que se utiliza en sistemas de búsqueda ultrarrápidos.
- El control. Cálculo de tal acción de entrada en el sistema, en el que sigue la trayectoria deseada.
- Reconocimiento de voz e incluso traducción automática.
- La implementación de las redes neuronales se puede realizar tanto por medio de software como por hardware, como los neuro chips.

2.1.3.1.1. Sistemas de reconocimiento de emociones con inteligencia artificial

Usando la teoría de reconocimiento y clasificación de emociones básicas, así como su modificación, es posible describir el modelo matemático de una emoción particular.

Cualquier expresión facial es analizada y traducida a un lenguaje comprensible para el sistema inteligente. Los pasos principales e interconectados para crear un modelo matemático son:

1. Identificación y encuadre de los rostros o de las personas.
2. Compresión por el sistema artificial que ante la presencia de un giro antinatural del rostro o de iluminación insuficiente, la persona del encuadre es la misma persona.
3. Identificación de características únicas de la cara que permitan distinguir a una persona de otras personas (factores como la forma de los ojos, la forma de la cara, la posición de la boca, etcétera).

El cerebro humano hace todo esto de forma automática e instantánea. De hecho, los humanos son capaces de reconocer rostros, incluso relacionan rasgos faciales en objetos cotidianos.

Los ordenadores, por el momento, no son capaces de un nivel tan alto de reconocimiento, por lo que es necesario enseñarles cada paso de este proceso por separado. [24]

Existen diferentes puntos de vista sobre la determinación del número de etapas en el desarrollo y diseño de sistemas inteligentes. Depende de muchos factores, en particular, de la naturaleza de las funciones del futuro sistema inteligente, el área de uso, la disponibilidad de herramientas avanzadas, etc.

El proceso de construcción de sistemas de inteligencia artificial se realiza en varias etapas.

1. Identificación de tareas y de sus características. Se han de desarrollar los términos de referencia del sistema que se está diseñando; el círculo de usuarios del sistema suele ser limitado.
2. Identificación de los principales conceptos del área temática, que reflejan el conocimiento del círculo del personal experto. El ingeniero del conocimiento define los medios formales de representación del conocimiento y los procedimientos para obtener soluciones. Se identifican y formulan conceptos que determinan la elección de un esquema característico para representar el conocimiento experto sobre el área temática. La principal fuente de conocimiento sobre el área del problema es un profesional que fue ilustrado en dicha área temática, libros,

descripciones tecnológicas, artículos, documentos científicos, métodos de lluvia de ideas, métodos de llenado automatizado de bases de datos.

Otra fuente importante de conocimiento es Internet (la búsqueda tradicional de la información y el conocimiento necesarios, así como agentes inteligentes (robots de software desarrollados)).

3. Elección del formalismo de representación del conocimiento y determinación del mecanismo de inferencia de decisión. La estructura desarrollada para la representación del conocimiento es la base para la implementación de la siguiente etapa: la construcción directa de la base de conocimiento del sistema.
4. Elección o desarrollo del lenguaje de representación del conocimiento. Después de formular y presentar las reglas en el lenguaje de representación elegido, el ingeniero lo ha de programar y configurar el software.
5. Probar el sistema resolviendo problemas de prueba específicos.

Las etapas de creación de sistemas inteligentes no están estrictamente reguladas, es difícil trazar un límite temporal y/o cuantitativo.

El desarrollo de un sistema de reconocimiento de emociones con inteligencia artificial utilizando una red neuronal requiere un entrenamiento obligatorio de la red.

Para crear la red neuronal para una tarea específica, es necesario seleccionar los valores de los parámetros de las conexiones interneuronales.

Este proceso requiere una gran cantidad de datos para mostrar conjuntos de características faciales y un alto rendimiento de la computadora.

2.1.3.2. Obtención y selección de datos

La mayoría de los estudios obtienen datos de diversas maneras dependiendo de la naturaleza de los recursos y el campo de estudio, pero normalmente se realizan en entornos controlados que se encuentran adaptados específicamente para que el individuo se encuentre sometido a unos estímulos y tareas determinados para que pueda generar respuestas emocionales o psicológicas lo más puras posible. En función de los tipos de datos que se quieran grabar se emplean diferentes dispositivos y sensores.

En varias investigaciones se han desarrollado los ensayos en entornos controlados, donde los sujetos no se encuentren expuestos a factores externos que puedan interferir en la alteración de cualquiera de las señales que se estén registrando.

Por otro lado, también se desarrollan los ensayos en entornos más realistas y cercanos a la vida cotidiana. Aunque, estos ensayos presentan desafíos adicionales debido a la presencia de factores externos no controlados, proporcionan información valiosa sobre el comportamiento y las respuestas de los individuos en situaciones más naturales. Estos ensayos pueden llevarse a cabo en entornos como hogares, lugares de trabajo o espacios públicos, lo que permite una comprensión más completa de cómo las personas interactúan con su entorno y cómo se ven afectadas por diversos estímulos.

2.1.3.3. Procesamiento y extracción de características de las señales

En el procesamiento de señales, se utilizan diversos métodos con el fin de analizar y manipular la información contenida en las señales.

Procesamiento de las imágenes.

Manejo de formato. Esta es una conversión de formato en caso de que sea necesario, puesto que no todos los entornos de desarrollo del estudio soportan archivos con un tamaño que no sea el permitido por el propio entorno. La mayoría de los datos se suelen manejar en .jpg, .jpeg o .png. para facilitar el procesamiento posterior.

Redimensionamiento de la imagen. Para ello se emplean algoritmos como K-NearestNeighbors, interpolaciones bilineales o interpolaciones bicúbicas.

En la mayoría de los estudios se observa que se emplean correcciones de balances de color, por lo que se usan algoritmos de balance de blancos automático u otra solución es realizar la grabación en blanco y negro para evitar su implementación [25].

Procesamiento de las señales EDA.

Remuestreo. En función de la frecuencia a la que haya sido grabada se remuestrea acorde a las marcas de tiempo puesto que es necesario para poder sincronizar la señal con otros eventos que emplean marcas de tiempo como referencia.

Filtrado de las señales. Esto es para eliminar los ruidos que puedan existir, para ello se usan algoritmos de filtros no causales (filtros de respuesta al impulso finita, FIR, o filtros de respuesta al impulso infinita, IIR); la fórmula de un filtro FIR en el dominio n es la siguiente:

$$Y_n = \sum_{k=0}^{N-1} b_k x_{n-k} \quad (2.1)$$

donde,

- $N - 1$: es el orden del filtro
- Y_n : es la salida en el instante de tiempo n
- b_k : es el coeficiente del filtro en el índice b
- x_{n-k} : es la señal de entrada en el instante de tiempo $n - k$

Para expresar como la convolución de la señal de entrada aplicando la transformada Z se obtiene la siguiente fórmula:

$$H[z] = \sum_{k=0}^{N-1} h_k z^{-k} = h_0 + h_1 z^{-1} + \dots + h_{N-1} z^{-(N-1)} \quad (2.2)$$

donde,

- h_n : son los coeficientes

y para la generación de los coeficientes de filtrado de Butterworth, un filtro de paso bajo.

La función de transferencia para este filtro es la siguiente:

$$H[s] = \frac{1}{\prod_{k=1}^n (s - s_k)/w_c} \quad (2.3)$$

donde,

- s_k : son los polos del filtro
- k : son los valores hasta los que llega la s

Detección de anomalías. Antes de normalizar cualquier señal se eliminan las zonas de valores elevados que podrían causar pérdida de datos en el resto del tiempo, para ello se realizan técnicas de eliminación para detectar los componentes de la señal no deseados o se realiza de forma manual.

La normalización de la señal. Esto se usa para tener los datos dentro de un rango determinado por ejemplo entre 0-1. En caso de poseer más de una señal se emplea la correlación, que es una técnica para la detección de señales preconocidas contra un fondo de ruido, denominado como el filtrado óptimo. La correlación cruzada es muy similar a la convolución, pero cada una se calcula de una manera diferente y, por consiguiente, las áreas de aplicación también difieren.

$$H_{x,y}(\tau) = \int_{-\tau}^{\tau} x(t)y(t + \tau)dt \quad (2.4)$$

donde,

- $H_{x,y}$: es la función de correlación cruzada. Mide el grado de semejanza entre las señales.
- x, y : son las señales que se están comparando
- τ : es el parámetro de desplazamiento

Segmentación de la señal. Se establece una ventana de trabajo, siendo una ventana una partición de la señal, para poder trabajar con el conjunto de datos.

Extracción de las características. Las características son los picos y las amplitudes de la señal física, la cual se encuentra estrechamente relacionada con las emociones repentinas.

Detección de las características. Se aplican algoritmos de detección que pueden variar en función del entorno dónde se esté procesando la señal.

Una de las transformadas matemáticas más relevantes para la detección de picos es la de wavelet,

$$W[a, b] = \sum_{k=0}^N x_n \psi\left[\frac{n-b}{a}\right] \quad (2.5)$$

donde,

- $W[a, b]$: representa el coeficiente de la transformada de wavelet en el tiempo b y a escala a
- x_n : es la señal de entrada
- $\psi(t)$: es la función de wavelet primaria, que es una función de onda de corta duración y alta frecuencia

La transformada de wavelet se obtiene al desplazar y modificar la escala de la función wavelet primaria a lo largo de la señal de entrada. De esta manera, se logra capturar información específica que está localizada en distintas escalas y momentos temporales de la señal.

El hecho de que esta función es capaz de proporcionar información detallada sobre los cambios locales en diferentes escalas y tiempos la convierte en una técnica útil que se puede utilizar en muchas aplicaciones, desde el procesamiento de señales, hasta análisis de imágenes y compresión de datos.

Procesamiento de las señales IBI.

Detección de características. Consiste principalmente en la detección de los picos RR, siendo estos los puntos de mayor amplitud [Ilustración 2], los cuales pertenecen al ECG. Los algoritmos que se emplean para su detección son algunos de los que se mencionan anteriormente como el algoritmo de umbral, de la derivada, de la transformada de Wavelet, transformada de Hillbert, etc.

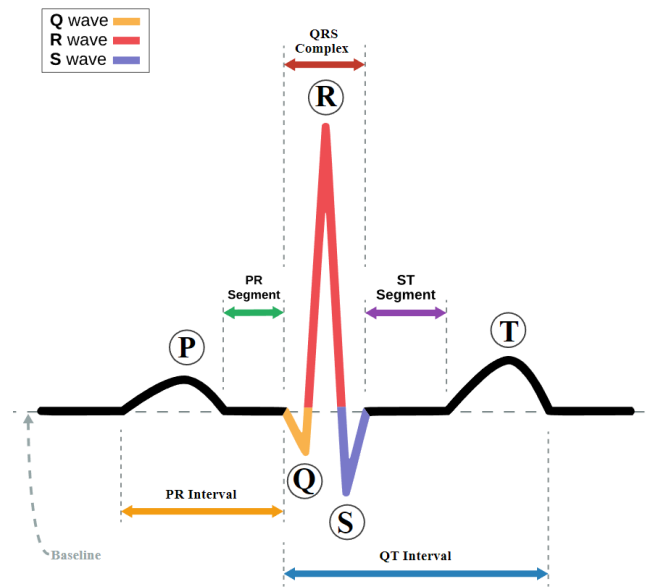


Ilustración 2: ECG de un corazón en ritmo sinusal normal, creado por Anthony Atkielski

Filtrado y eliminación de alteraciones de la señal. Esto se aplica puesto que existen ruidos causados por el movimiento o interferencias eléctricas. Para los filtrados se emplean filtros de banda o filtro adaptativo para eliminar o reducir el ruido.

Para la eliminación de ruidos se aplica la fórmula correspondiente al filtro adaptativo LMS,

$$y[n] = x_n - W_n u_n \quad (2.6)$$

donde,

- $y[n]$: es el valor de la señal en el instante n
- x_n : es la señal de entrada original en ese instante n
- W_n : es el vector de coeficientes adaptativos del filtro
- u_n : es el vector de entrada del filtro

El objetivo principal es ajustar los coeficientes del vector W_n de manera iterativa para poder minimizar el error entre la señal original y la filtrada.

Es importante destacar que existen otros filtros adaptativos como Mínimos Cuadrados Medios Normalizados (NLMS) o Mínimos Cuadrados Recursivos (RLS) que se emplean para las mismas aplicaciones.

Interpolación para completar datos. Esto se aplica en caso de que falten partes de la señal debido a errores de detección de picos. Para ello se usan algoritmos de interpolación, como la interpolación lineal para estimar valores de intervalos que faltan y mantener una continuidad de la secuencia del IBI.

La interpolación lineal sigue la siguiente fórmula,

$$Y_n = y_1 + (y_2 - y_1) * t \quad (2.7)$$

donde,

- Y_n : es el valor estimado de un punto intermedio
- y_1 : es el valor del punto de la señal en una fracción de tiempo t_1
- y_2 : es el valor del punto de la señal en una fracción de tiempo t_2
- t : es la fracción de tiempo entre t_1 y t_2 en la ubicación del punto intermedio

Hay que tener en cuenta que este tipo interpolación asume una relación lineal, como su nombre indica.

Remuestreo de la señal acorde al tiempo. Cada dato es grabado y asociado un tiempo determinado, por lo que se remuestrea, se ajustan los valores a su tiempo o epoch de captura.

Eliminación de ruidos. Para ello se aplica el filtrado de Butterworth, anteriormente mencionado y se aplica un filtro no causal de paso bajo.

Para obtener la frecuencia cardíaca se aplica la siguiente fórmula.

$$HR = \frac{60}{IBI} \quad (2.8)$$

donde,

- HR : es la señal que se obtiene de la frecuencia cardíaca
- IBI : señal de intervalos entre latidos

Segmentación de la señal. Se aplica para poder trabajar de manera más cómoda y realizar un análisis más específico de la frecuencia cardíaca. Para ello, se examina la señal y se realizan cortes al principio y al final de

la señal en función de la cantidad de señal menos restrictiva de los tres tipos de señales que se empleen.

Análisis de HR. Se aplican métricas que proporcionen información como la media de la frecuencia, desviación estándar de la frecuencia o de los intervalos RR.

La media de la frecuencia cardíaca se calcula como:

$$Mean_{HR} = \frac{\Sigma(HR)}{N} \quad (2.9)$$

donde,

- $Mean_{HR}$: es la media de la frecuencia cardíaca
- $\Sigma(HR)$: es el sumatorio de todos los valores de HR
- N : es el número total de valores

La media proporciona la estimación de la HR durante un período de tiempo determinado.

Por otro lado, la desviación estándar de la HR es similar a la variabilidad de la frecuencia cardíaca, pero difiere en que la primera es a corto plazo, a menores duraciones.

Para calcular la desviación estándar (SD HR) se aplica la siguiente fórmula:

$$SD HR = \sqrt{\frac{\Sigma_1^N (HR - Mean_{HR})^2}{N - 1}} \quad (2.10)$$

donde,

- $SD HR$: es la desviación estándar
- $Mean_{HR}$: es la media de la frecuencia cardíaca
- HR : señal de frecuencia cardíaca
- $\Sigma(HR)$: es el sumatorio de todos los valores de la resta de los valores de frecuencia cardíaca menos el valor medio
- N : es el número total de valores de intervalos

Obtención de los picos de la señal. Se emplean diversos algoritmos, como el del umbral, de la derivada, la de transformada de Hilbert, de Wavelet, etc.

Determinación de picos por umbral. Se emplea el umbral basado en el porcentaje, que consiste en establecer que, a partir de un determinado porcentaje de la amplitud de la señal, se considera o no, pico. Por otro lado, también se emplea el umbral basado en la desviación estándar, que usa la propia desviación como señal de referencia.

Cuantificación de variabilidad cardíaca. Se aplican métricas para el cálculo del índice de variabilidad cardíaca o el espectro de frecuencia. Esto se calcula mediante algoritmos como el análisis de dominio del tiempo o de frecuencia; Transformada Rápida de Fourier (FFT) de una secuencia de datos; este último es el más empleado teniendo la siguiente fórmula:

$$Y_k = \sum_{n=0}^{N-1} [y_n * e^{-2\pi*i*k*\frac{n}{N}}] \quad (2.11)$$

donde,

- Y_k : es la k-ésima componente espectral de la señal de entrada
- y_n : es la muestra n de la señal
- N : es el número total de muestras en la secuencia de datos
- i : es la unidad imaginaria
- k : es el índice de frecuencia que va desde 0 hasta N-1

El cálculo de la FFT divide la secuencia de datos en subconjuntos menores y se aplican operaciones recursivas para calcular las componentes espectrales en un orden más eficiente.

Densidad espectral de potencia (PSD) es un método para analizar la distribución de energía en las diferentes frecuencias en una señal. Para obtener la PSD se aplica el algoritmo de estimación espectral, siendo este la transformada de Fourier. En otros estudios se observa que se emplea el método de autoregresión (AR).

Densidad espectral de potencia (PSD)

$$P_{sd} = |FFT(x_n)|^2 \quad (2.12)$$

donde,

- P_{sd} : es la señal de salida, el espectro de potencia en la frecuencia
- $FFT(x_n)$: es la señal de entrada pasada por la Transformada de Fourier

Se toma el valor absoluto de la señal para obtener el espectro de frecuencia y después se eleva al cuadrado para poder obtener la potencia.

Esta es una medida útil para analizar y caracterizar las componentes frecuenciales presentes en el IBI, lo que proporciona información sobre la variabilidad de la frecuencia cardíaca y otros fenómenos relacionados

Obtención de bandas. La banda de baja frecuencia (LF) y de alta frecuencia (HF), esta última se encuentra típicamente en un rango de frecuencia entre 0.15 y 0.4 Hz y se encuentra fuertemente relacionada con las emociones puesto que se encuentra ligada con la actividad del SNP (sistema nervioso parasimpático).

Índice de Alta Frecuencia (HF)

$$H_{f0} = r_a \left(\frac{\text{frec. corte inferior}}{f_r} \right) \quad (2.13)$$

$$H_{f1} = r_b \left(\frac{\text{frec. corte superior}}{f_r} \right) \quad (2.14)$$

donde,

- H_{f0}/H_{f1} : es el índice inicial y final del rango de alta frecuencia
- r_a/r_b : es el redondeo normal y estándar hacia arriba independientemente del signo
- f_r : es la frecuencia de resolución, siendo esta a su vez:

$$f_r = \frac{\text{Tasa de muestreo}}{\text{nº puntos de datos}} \quad (2.15)$$

Extracción de la potencia en la banda de alta frecuencia. Para obtenerla se integra la PSD en la banda de la frecuencia.

Potencia de Alta Frecuencia (HF)

$$P_{hf} = \sum (P_{sd}[H_{f0}, H_{f1+1}]) \quad (2.16)$$

donde,

- P_{hf} : es la potencia de alta frecuencia calculada
- P_{sd} : es la densidad de espectral de potencia
- H_{f0} : es el índice inicial del rango de alta frecuencia
- H_{f1} : es el índice final del rango de alta frecuencia

La banda se puede normalizar en caso deseado y se suele expresar en potencia relativa o potencia absoluta (ms^2).

2.1.3.4. Clasificación de emociones

Para detectar y clasificar las emociones, se evalúan las señales integrándolas en un modelo de inteligencia artificial, los cuales determinan los resultados.

2.1.3.5. Redes neuronales

2.1.3.5.1. El concepto de red neuronal artificial

Una red neuronal artificial se entiende como un modelo matemático, así como su correspondiente implementación en entorno de software y hardware.

La red neuronal artificial (RNA) fue construida sobre el principio de las redes neuronales biológicas; las células nerviosas de un organismo vivo. Este concepto surgió al tratar de modelar los procesos que ocurren en el cerebro humano. Una RNA es un sistema de procesadores simples (neuronas artificiales) conectados que interactúan entre sí. Cada uno de los procesadores de la red se ocupa de las señales que periódicamente llegan o se transmiten a otros procesadores (neuronas). Una gran red es capaz de resolver problemas complejos en tiempos pequeños. Cuanto más se entrena la red más se asemejará el modelo.

Desde un punto de vista matemático, las redes neuronales son una forma de resolver problemas de optimización no lineal. La cibernética utiliza la teoría de las redes neuronales para resolver problemas de control adaptativo, construyendo algoritmos que se emplean principalmente en áreas de la robótica y automática.

En el área de la programación, una red neuronal es una de las formas de resolver el problema del paralelismo eficiente. La programación de las redes neuronales artificiales significa entrenar la red con grandes cantidades de datasets. Es gracias a esto que la red es capaz de identificar las relaciones entre datos de entrada (input) y salida (output), aparte de generalizar, filtrar los resultados y emplear el conocimiento para dividir tareas extremadamente complejas en otras derivadas más simples.

2.1.3.5.2. Neurona biológica y neurona artificial

El cerebro humano y su sistema nervioso está formado por neuronas conectadas por fibras nerviosas, que transmiten impulsos eléctricos entre ellas cuando se produce un traspaso de información, incluida la acción del pensamiento.

La estructura de una neurona biológica [Ilustración 3]:

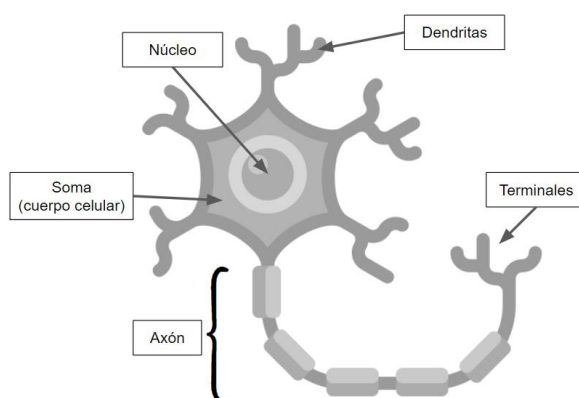


Ilustración 3: Partes de un esquema de una neurona biológica

Dendritas - son las encargadas de recibir los impulsos neuronales.

Axón - se encarga de transmitir el impulso de la misma neurona.

Sinapsis - son las formaciones que afectan la fuerza del impulso, por el contacto del axón y la dendrita. Durante el paso de la sinapsis, la fuerza del impulso cambia un determinado número de veces (el "peso" de la sinapsis). Cuando los impulsos llegan a una neurona a través de varias dendritas, se suman. Y si el impulso total supera el umbral, la neurona entra en estado de excitación, generando su propio impulso y enviándolo a lo largo de todo el axón.

El comportamiento de la neurona puede cambiar en función de las condiciones de contorno, ya que los pesos de las sinapsis tienden a cambiar con el tiempo, adaptándose al entorno y acorde a los inputs y outputs.

2.1.3.5.2.1. Funcionamiento de las redes neuronales artificiales

Respecto al funcionamiento en la práctica, la red neuronal artificial, en contraste de la red neuronal biológica, realiza la transformación de datos, cuyo tipo está determinado por los pesos de las conexiones

interneuronales, el tipo de función de activación de las neuronas, la arquitectura y la configuración de la red.

De nuevo, se incide en que el funcionamiento de las redes se asemeja a la de un cerebro y destaca en dos puntos importantes:

- El conocimiento es adquirido a través de un proceso de aprendizaje de la propia red.
- Para el almacenamiento de lo aprendido, se emplean conexiones interneuronales de diferentes importancias, sean estos, los pesos sinápticos.

Por tanto, se puede definir de manera general como un modelo matemático que imita el funcionamiento de un cerebro dotando al sistema de inteligencia artificial con una estructura que consta de neuronas artificiales conectadas entre sí y con el entorno externo mediante conexiones, cada una de las cuales tiene un coeficiente, pesos, por el cual se multiplica el valor que llega a través de ella.

Las redes neuronales se entrenan con un "maestro". En este caso el maestro es el propio input, el conjunto de datos de entrenamiento que contiene ejemplos con valores verdaderos: etiquetas, clases, indicadores, etc. Las variables de entrada son multiplicadas por sus pesos que a su vez van cambiando acorde a la función matemática que se calcula durante el entrenamiento de la red.

Las funciones, anteriormente, se generaban de manera manual. Cuantas más características y pesos seleccionadas con mayor precisión, más preciso es el resultado. Hoy en día este proceso se ha automatizado.

Las redes neuronales artificiales constan de tres componentes:

- Capa de entrada
- Capas ocultas (blackbox)
- Capa de salida

Y posee dos etapas de entrenamiento:

- Propagación de errores hacia delante.
- Propagación hacia atrás de un error.

Durante la propagación de errores hacia delante, se realiza una predicción de la respuesta. Con la retro propagación, se minimiza el error entre la respuesta real y la predicha.

2.1.3.5.2.2. La propagación de errores hacia delante

A continuación, se observa un diagrama de bloques en el cual se muestran las partes de la red tanto de la biológica, en la parte superior de los bloques, como de la artificial, parte inferior.

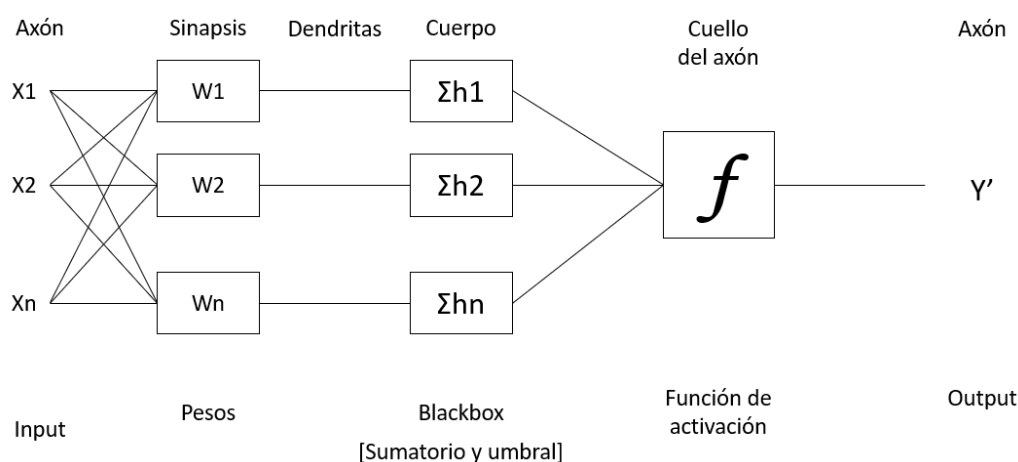


Ilustración 4: Propagación hacia delante

Se establecen pesos iniciales aleatoriamente, siendo estos w_1, w_2 y w_n , estos son las sinapsis de una red neuronal biológica. [Ilustración 4]

A continuación, se multiplican los datos de entrada por los pesos para formar la capa oculta, o lo que se denomina la blackbox. Obteniendo las siguientes ecuaciones:

$$h_1 = (x_1 * w_1) + (x_2 * w_1) + (x_n * w_1) \quad (2.17)$$

$$h_2 = (x_1 * w_2) + (x_2 * w_2) + (x_n * w_2) \quad (2.18)$$

$$h_n = (x_1 * w_n) + (x_2 * w_n) + (x_n * w_n) \quad (2.19)$$

La respuesta de la capa oculta, o impulso de cada una de la neurona, se pasa a través de una función no lineal, denominada como la función de activación, para obtener la respuesta final de la red, o el denominado output:

$$y' = f(h_1, h_2, h_n) \quad (2.20)$$

Por tanto, se obtiene un conjunto de vectores como entradas de las siguientes neuronas.

Es necesario mencionar que una neurona artificial recibe muchas señales en su entrada, cada una de las cuales es simultáneamente la salida de otra neurona. Por esto, se observa que cada entrada se multiplica por sus pesos correspondientes, posteriormente sumándose los productos y determinando el nivel de activación de cada neurona.

2.1.3.5.2.3. La propagación hacia atrás

A continuación, se muestra un diagrama de bloques de la propagación hacia atrás [Ilustración 5]:

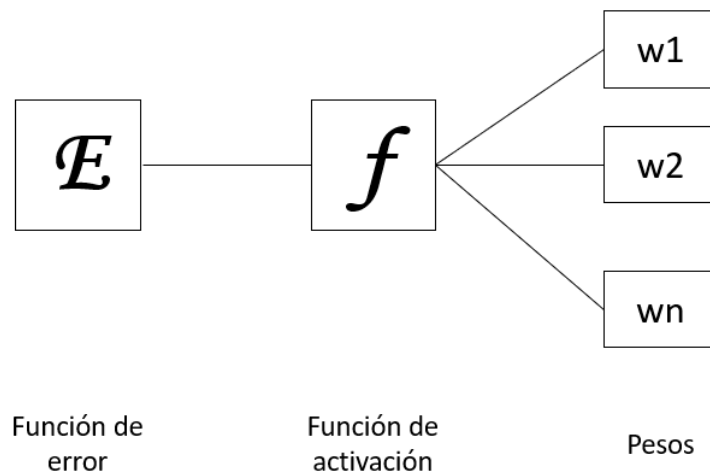


Ilustración 5: Propagación hacia atrás

Siendo:

$$E = \text{función de pérdida } (y, y') \quad (2.21)$$

$$f(E, wx) = \eta * \partial(E) / \partial(wx) \quad (2.22)$$

El error total (ET) se calcula como la diferencia entre el valor esperado de "y", el conjunto de datos de entrenamiento, y el valor recibido "y'", el cuál ha sido calculado en la etapa de propagación de errores hacia adelante, pasando por la función de costo.

La derivada parcial del error se calcula sobre cada uno de los pesos, estos diferenciales parciales reflejan la contribución de cada uno de los pesos al error total, ET.

Posteriormente, los diferenciales se multiplican por un valor denominado como "learning rate" o tasa de aprendizaje "η"

El resultado obtenido, finalmente se resta de los pesos correspondientes para poder obtener los pesos "actualizados".

$$w1 = w1 - (\eta * \partial(E)/\partial(w1)) \quad (2.23)$$

$$w2 = w2 - (\eta * \partial(E)/\partial(w2)) \quad (2.24)$$

$$wn = wn - (\eta * \partial(E)/\partial(wn)) \quad (2.25)$$

Cabe destacar que la retropropagación tiene el mismo resultado que la serie de Taylor. La única diferencia es que en lugar de una serie infinita se pretende optimizar únicamente su primer término.

Las compensaciones son los pesos añadidos a las capas ocultas. Que también se inicializan y se actualizan aleatoriamente, de la misma manera que la capa oculta.

La función principal de la capa oculta es determinar la forma de la función base en los datos, mientras que la función del desplazamiento es desplazar la función encontrada hacia un lado para que coincida parcialmente con la función original.

2.1.3.5.2.4. *Derivadas parciales*

Respecto a las derivadas parciales, se observa que se pueden realizar los cálculos con bastante sencillez y observar que se posee una contribución importante al error para cada uno de los pesos de las neuronas.

2.1.3.5.2.5. *Hiperparámetros*

Se emplea una red neuronal para automatizar la selección de funciones, pero algunos parámetros se han de configurar manualmente.

2.1.3.5.2.6. *Tasa de aprendizaje (learning rate)*

La tasa de aprendizaje es un hiperparámetro de gran importancia, puesto que esta es la que determina la rapidez con la que se va a mejorar los resultados.

Si la tasa es demasiado baja, incluso después de entrenar la red neuronal durante mucho tiempo, estará lejos de obtener resultados óptimos. Por otro lado, si la tasa de aprendizaje es demasiado alta, la red dará respuestas muy rápidamente y el hecho de que funcione el sistema puede ser muy irregular.

2.1.3.5.2.7. *Función de activación*

La función de activación es una de las herramientas más poderosas que afecta el poder atribuido a las redes neuronales. En parte, determina qué neuronas se activarán, es decir, qué información se transmitirá a las capas posteriores.

Sin funciones de activación, las redes profundas pierden gran parte de su capacidad de aprendizaje. La no linealidad de estas funciones es responsable de aumentar el grado de libertad, lo que permite generalizar problemas de alta dimensión a dimensiones más bajas.

Para finalizar, es importante incidir en que el modelo de neuronas artificiales ignora la mayoría de las propiedades de la neurona biológica. Por ejemplo, retrasos de tiempo que afectan la dinámica del sistema. Las señales de entrada generan inmediatamente una señal de salida. Además, la neurona artificial no tiene en cuenta los efectos de la función sincronizadora de la neurona biológica. Sin embargo, cabe señalar la excepcional similitud entre una neurona de un ser vivo y una artificial. Para determinar el lugar de las redes neuronales en el campo de la tecnología de la información, es necesario referirse a la clasificación de los sistemas de inteligencia artificial.

La comparativa entre las neuronas biológicas y las artificiales realizada en este apartado constituye una parte importante del análisis documental de los avances en el desarrollo de redes neuronales que se han llevado a cabo en este milenio.

2.1.3.5.3. Red neuronal aplicada en Python

2.1.3.5.3.1. *Redes Neuronales Convolucionales*

La red neuronal convolucional (CNN) [Ilustración 6] es una clase de redes neuronales profundas que se utilizan a menudo en el análisis de imágenes visuales. Las redes neuronales convolucionales son un tipo de perceptrón multicapa que utiliza operaciones de convolución. Han encontrado aplicaciones en reconocimiento de imágenes y videos, sistemas de recomendación, clasificación de imágenes, NLP (procesamiento del lenguaje natural) y análisis de series temporales.

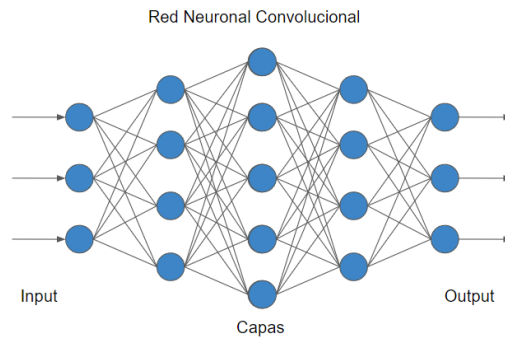


Ilustración 6: Esquema una Red Neuronal Convolutiva

2.1.3.5.3.1.1. Funcionamiento del CNN

La operación de la convolución se puede representar mediante el siguiente algoritmo:

1. Una ventana deslizante, llamada filtro, con tamaño (n, n) se mueve a lo largo de la entidad de entrada. El número de movimientos está determinado por el número especificado de filtros.
2. Cada plantilla resultante tiene la forma (n,n,d) , donde d es la profundidad de la entidad de entrada.
3. Cada patrón se multiplica por su propio núcleo de convolución, lo que da como resultado un mapa de características de salida. El mapa de características de salida resultante tiene la forma (h,w,N) , donde h y w son la longitud y el ancho resultantes del recorte, y N es la cantidad de filtros.

El número de filtros es un hiperparámetro, por lo que se elige de forma independiente. Por lo general, se elige una potencia de dos, y el número de filtros aumenta a medida que aumenta la profundidad de la arquitectura. Y los núcleos de convolución son parámetros que se pueden aprender.

2.1.3.5.3.1.2. Propiedades de las CNN

Las redes neuronales convolucionales han podido ganar su popularidad debido a las jerarquías de representación coincidentes a medida que se aprenden los patrones locales. CNN tiene varias propiedades:

- Las representaciones resultantes son invariantes con respecto a la transferencia. En la imagen, el individuo puede estar en

cualquier lugar disponible de la imagen y la red convolucional no recuerda su posición, CNN solo conoce sus representaciones (ojos, rostro, pómulos, etc.)

- El modelo CNN es espacialmente jerárquico. Las primeras capas estudian las plantillas locales y las siguientes las plantillas obtenidas de las primeras capas.

2.1.3.5.3.2. *Redes Neuronales Recurrentes*

La red neuronal recurrente (RNN) es una arquitectura de redes neuronales profundas utilizada en tareas que involucran datos secuenciales o temporales. A diferencia de las redes neuronales convolucionales, las RNN pueden mantener y propagar información a lo largo del tiempo gracias a sus conexiones recurrentes. Son especialmente útiles en el procesamiento del lenguaje natural, generación de texto, traducción automática y análisis de series temporales.

Las RNN se aplican en diversas áreas, como el análisis de sentimientos, generación de texto y traducción automática en el procesamiento del lenguaje natural. Pueden capturar dependencias temporales, lo que permite entender el contexto a largo plazo y generar secuencias coherentes. Además, de ser eficientes en el análisis de series temporales, es decir, pueden predecir valores futuros basados en patrones pasados, siendo útiles en campos como la predicción de diversos tipos de datos.

Las RNN también tienen arquitecturas más complejas, como las redes neuronales recurrentes bidireccionales (BRNN), que capturan información tanto del pasado como del futuro de una secuencia. Esto es beneficioso en tareas donde el contexto bidireccional es importante, como el análisis de sentimientos en conjuntos de datos secuenciales largos.

Una variante de este tipo de redes es la Long Short-Term Memory (LSTM) y Bidireccional Long Short-Term Memory (BLSTM) [Ilustración 7], esta última destaca en que es capaz de capturar las dependencias temporales de los eventos en ambas direcciones para poder reconocer patrones de una forma más precisa, en otras palabras, se puede decir que posee un contexto temporal más completo lo que permite detectar cambios sutiles en las secuencias de datos. Por otro lado, es importante incidir que estas redes son capaces de predecir datos con más precisión al tener en cuenta la dinámica de las secuencias de las señales de entrada.

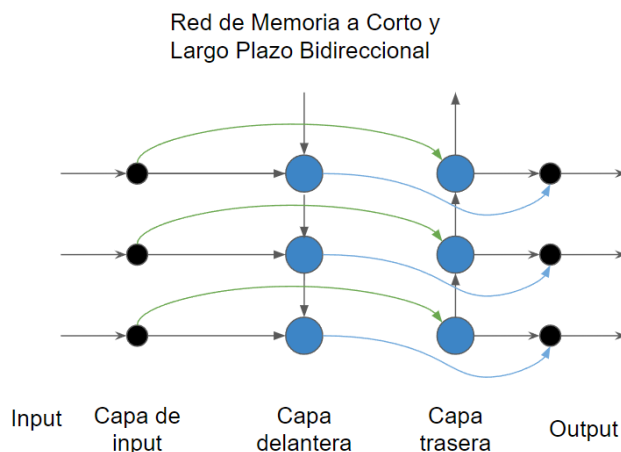


Ilustración 7: Red de Memoria a Corto y Largo Plazo Bidireccional

2.1.3.5.3.2.1. Funcionamiento de la BLSTM

El funcionamiento de la red neuronal recurrente se diferencia de las redes convolucionales en su estructura y funcionamiento. Esta red consta de dos redes de tipo LSTM que se conectan en sentido inverso. Otra de las diferencias de la CNN es que procesa la secuencia de entrada tanto en inversa como en directa, permitiendo así la captura del contexto pasado y futuro de cada una de las posiciones.

El funcionamiento se puede resumir en los siguientes pasos:

1. Propagación hacia adelante (véase apartado "2.1.3.5.2.2. Propagación hacia adelante") y posteriormente la propagación hacia atrás (véase apartado "2.1.3.5.2.3. Propagación hacia atrás").
2. Método de procesamiento. La capacidad de la bidireccional actúa de la siguiente manera: la capa delantera procesa la secuencia en orden directo capturando la información del contexto del pasado y la capa hacia atrás procesa el orden inverso capturando el futuro. En este proceso, se aplica la concatenación de ambas salidas para formar una representación combinada de secuencias.
3. Clasificación. En la salida se obtiene la clasificación o predicción en función de los datos de entrada seleccionados. El rendimiento de la clasificación puede variar en función del ajuste de parámetros como pesos, funciones de pérdida y métricas a utilizar.

2.1.3.5.3.2.2. Propiedades de la BLSTM frente a CNN

La BLSTM, como se indica anteriormente, es una variante de las RNN que emplea memoria de largo plazo para aprender de las secuencias de datos en ambas direcciones. Su ventaja principal frente a las CNN es que posee la capacidad para modelar dependencias a largo plazo en los conjuntos de datos de entrada. Esta red, como se menciona, procesa secuencias y toma en cuenta tanto contextos temporales pasados como futuros, tomando así una forma más compleja y de mayor capacidad de procesamiento en comparación con las CNN.

Por otro lado, las CNN son útiles para aplicarlas en clasificación de imágenes. Esta red posee un filtro convolucional de extracción de características locales de entrada lo que permite realizar una detección de objetos de manera más precisa y eficiente. Es importante incidir en que las CNN se emplean principalmente en conjuntos de datos visuales que permitan identificar patrones a escalas variables permitiendo así una extracción de características discriminativas.

En síntesis, la BLSTM se emplea para un conjunto de datos que tenga en cuenta el factor de secuencia futuro y pasado, aparte de requerir datos secuenciales como se indica en los párrafos anteriores. Desde otra perspectiva, la CNN se encuentra más enfocada hacia datos de tipo visual que permitan identificar características puesto que los datos se encuentran organizados de manera jerárquica como los son las imágenes.

2.2. ESTUDIOS RELACIONADOS

Antes de proceder con el desarrollo y los resultados se analizan las diferentes investigaciones, teorías y estudios más relevantes hasta la fecha relacionados con el presente proyecto.

La idea de los antecedentes es determinar conceptos y términos en los que se podría basar para poder avanzar con la investigación recogiendo información de las diferentes fuentes fiables de información como bien se había mencionado anteriormente, estudios, investigaciones, trabajos de fin de grado, de máster doctorado y teorías.

Los temas para resolver son el reconocimiento y clasificación de imágenes, análisis de señales de tipo EDA y finalmente de los datos de electrocardiogramas; este último es una señal reciente que se emplea a modo de método de detección de la variabilidad de las emociones

2.2.1. Reconocimiento y clasificación de imágenes

Para que haya posibilidad de clasificación de la emoción a través de imágenes es necesaria la detección facial, puesto que, a nivel general, este es el indicador que pone en evidencia las emociones y los estados de ánimo de los individuos. Es importante destacar que el rostro humano puede poseer hasta 16 mil configuraciones distintas empleando diferentes músculos del rostro. Y aunque todas esas configuraciones puedan mostrar alguna combinación de emociones, en este proyecto nos centraremos en el reconocimiento y clasificación de las 6 emociones básicas a través de obtención y procesamiento de imágenes.

Por tanto, a la hora de reconocer los rostros nos centraremos en la boca, algunas arrugas características, pómulos y ojos o mirada.

Se han considerado algunos de los sistemas de software existentes para el reconocimiento automático de emociones:

Google Cloud Vision AI. Este software ayuda con la determinación de emociones de rostros, comprensión mejorada de textos entre otros con el requerimiento de poseer las imágenes, que vayan a ser analizadas, en el Google Cloud. Cabe destacar que el servicio de visión artificial emplea aprendizaje automático para realizar la predicción de emociones con mayor precisión. [26]

Tobii Pro Sticky. Este software ofrece una plataforma de autoservicio; esto es, combina encuestas con un seguimiento de la mirada u ocular de la cámara y el reconocimiento de emociones. De forma general se encuentra dedicada para desarrolladores e investigadores que estén interesados en este tipo de servicio, aunque también se ha aplicado en otras áreas como la industrial. La recogida de datos la realiza la plataforma y es capaz de elaborar respuestas con elevada rapidez, incluso cuando se tratan alcances globales. [27]

FaceReader. Este, al igual que los anteriores softwares, se emplea para obtener datos confiables de expresiones faciales. Se puede combinar con seguimientos oculares y otros datos fisiológicos. Cabe destacar que es capaz de analizar los datos en vivo, vídeos e imágenes por lo que permite un considerable ahorro de tiempo de adaptación y procesamiento repetitiva de datos. Es importante fijarse en que este software en general está dedicado para equipos de seguridad, pero también para investigadores que estén trabajando con emociones, también posee la capacidad de reconocer las seis expresiones universales. [28]

2.2.1.1. Reconocimiento de emociones usando Deep Learning [29]

Emotions Recognition using Deep Learning

En este trabajo los datos son extraídos de kaggle bajo el nombre de Competición de Reconocimiento de Expresiones Faciales, son imágenes de tamaño 48x48 en escala de grises, el 80% de las imágenes se empleó para entrenamiento y el 20% para la validación. Para obtener más datos se aplica el Data Augmentation, que es una técnica de espejo para generar imágenes adicionales a partir de las originales. A continuación, añade el cropping, que lo que hace es realizar un recorte de las imágenes para simular desplazamientos. De nuevo, se generan diferentes recortes de cada una de las imágenes y se emplean en los conjuntos de entrenamiento y prueba. Los algoritmos empleados para el entrenamiento y la prueba son las redes convolucionales (CNN) dentro de la cual se emplean diferentes arquitecturas de red para el reconocimiento de emociones. Estas redes se implementan utilizando el framework Theano. Para mejorar las redes se emplean un algoritmo de optimización, usan el Theano como base para crear una capa intermedia y lanzar experimentos.

El algoritmo se encuentra implementado en Python, y, por consiguiente, emplea bibliotecas bastante conocidas como:

- Numpy. Es una biblioteca escrita en C que permite facilitar y acelerar el procesamiento de los códigos. Principalmente funciona para los vectores y matrices de gran tamaño.
- Pickle. Es una biblioteca que permite realizar implementaciones binarias de protocolos para la serialización o no de la estructura de objetos.
- H5py. Es una interfaz mejorada escrita en lenguaje C que es capaz de procesar formatos de datos binarios HDF5 de gran tamaño.
- Cv5. Es un adaptador proveniente de OpenCV para solventar problemas con la interfaz y la visión artificial.
- Gzip. Es otra biblioteca empleada para comprimir y extraer archivos. Esta biblioteca es bastante práctica pues mejora la velocidad de carga de tiempo de las aplicaciones.
- Timeit. Es un módulo que permite obtener el tiempo de ejecución necesario para un trozo de código determinado.
- Theano. Por último, esta biblioteca permite definir, optimizar y evaluar expresiones matemáticas de dimensiones múltiples empleadas en la biblioteca de NumPy.

2.2.1.2. Metodologías de reconocimiento facial en modelos bidimensionales [30]

En este estudio los datos utilizados se obtuvieron de la Yale Face Database B, que consiste en imágenes de 38 personas con diferentes expresiones faciales. Las imágenes de la database fueron realizadas desde un ángulo frontal en relación con el rostro de las personas. Cada imagen tiene un tamaño de 192x168 píxeles, lo que equivale a un total de 32,256 píxeles. Las imágenes se preprocesaron para asegurar su calidad y estandarización antes de aplicar métodos de reconocimiento facial.

Para el procesamiento se usó la biblioteca OpenCV en Python, que proporciona métodos para detectar características específicas en una fotografía, como el rostro completo y los ojos de una persona. Aunque la detección funcionaba correctamente en la mayoría de las imágenes, en

algunos casos se producían detecciones anormales, como confundir los agujeros nasales con los ojos. Si no se dispusiera de esta base de datos específica, sería necesario realizar una revisión y corrección manual de los datos después de realizar el procesamiento.

Es importante destacar que se ha aplicado Análisis de Componentes Principales (PCA) para reducir la dimensionalidad de los datos, se ha realizado un estudio de la varianza acumulada de las componentes principales. Para el cálculo de las componentes principales se empleó la biblioteca de sklearn de Python, la técnica de descomposición en valores singulares (SVD) para sacar la matriz de covarianza. Para evaluar la relevancia de cada componente principal se realizó la proyección inversa empleando PCA para reconstruir el conjunto de datos original. Se ha observado que utilizar más componentes principales mejora los resultados de la reconstrucción. Respecto a los algoritmos utilizados los principales son de Vecinos más Cercanos (KNN) y Redes Neuronales. Para KNN se realizó una búsqueda de parámetros para encontrar el número óptimo de vecinos y se ha evaluado su rendimiento en términos de precisión. Para las redes neuronales, se ha llevado a cabo una búsqueda de la arquitectura de red que proporciona los mejores resultados concluyendo en la investigada, CNN, puesto que son más efectivas para mayor número de datos y ofrecen mejores resultados frente a la regresión y los clasificadores de KNN y Naïve Bayes.

Las bibliotecas que ha empleado son las similares a las del Trabajo Fin de Máster anterior:

- NumPy. Se encuentra definida anteriormente.
- Matplotlib. Es una biblioteca perteneciente al entorno de Python para la creación de distintos tipos de visualizaciones.
- Keras. Es un paquete específico y perteneciente a las redes neuronales. Es una API escrita en Python que se emplea para el desarrollo de modelos de Deep Learning.
- TensorFlow. Al igual que el anterior es una biblioteca de Python empleada para la computación numérica rápida. Se suele emplear también para crear modelos de tipo de Deep Learning.
- Sklearn. Es una biblioteca creada para el entorno de Python que se suele emplear para aprendizaje de tipo aprendizaje automático.

2.2.1.3. Sistema de reconocimiento de expresiones faciales que emplea rasgos faciales robustos de videos profundos y Deep learning [31]

A facial expression recognition system using robust face features from depth videos and deep learning.

En este estudio se empleó una cámara de profundidad para adquirir videos. Las imágenes que se emplean la mayoría son de sorpresa y disgusto para ilustrar el proceso. En cuanto a la extracción de las características, se menciona que se empleó el Patrón Direccional Local Modificado (MLDP) para cada una de las imágenes. El MLDP funciona de la siguiente manera, se asigna un código binario de seis bits a cada uno de los píxeles en una imagen, que representa las fortalezas de borde relativas del píxel en ocho direcciones. Se emplean también máscaras de Kirsch típicas para calcular las fortalezas de borde cada píxel. Estos códigos MLDP se organizan en orden descendente y se consideran los signos correspondientes. Es importante que para el procesamiento se ha aplicado el método de clasificación supervisado denominado Análisis Discriminante Generalizado (GDA) para mejorar las características. El GDA es una generalización del LDA, Análisis Discriminante Lineal; busca separar las muestras de diferentes expresiones en un espacio lineal o no lineal empleando un kernel. La función de criterio GDA tiene como objetivo maximizar la dispersión entre clases y minimizar la dispersión dentro de las propias clases. Para el reconocimiento se ha empleado Deep Belief Network como modelo discriminativo. Para realizar el entrenamiento se han empleado dos fases, preentrenamiento y ajuste. Para el entrenamiento emplearon Maquinas de Boltzmann Restringidas (RBM) apiladas capa por capa, se emplea también el algoritmo de divergencia contrastiva para entrenar la RBM dentro de la DBN. El algoritmo ayuda a determinar la aproximación del gradiente basado en una cadena de Márkov. Los pasos de reajuste incluyen inicialización de un vector de sesgo y una matriz de pesos para la capa oculta. Se calculan los valores binarios de la capa oculta a partir de la capa visible. Se reconstruyen los valores de la capa visible empleando los de la capa oculta y finalmente se ajusta la capa oculta empleando los valores de estado de la capa visible reconstruida. Los pesos de la capa actual se calculan multiplicando los pesos por las entradas en la capa anterior sumándolas. Para concluir con el reajuste se emplea un algoritmo de retropropagación (backpropagation) como entrenamiento supervisado.

El enfoque propuesto se comparó con los enfoques tradicionales demostrando superioridad al obtener una tasa de éxito de un 96.25% mientras que los otros alcanzaron un máximo de 91.67%

2.2.2. Señales EDA

Las señales electrodérmicas, o EDA, son las señales que se obtienen de la actividad eléctrica biológica, que son cambios o alteraciones de las características eléctricas que se producen sobre la superficie de la piel y las cuales son medidas a través de unos sensores especiales.

Se observa que los estudios e investigaciones en este campo es amplio por lo que la mayoría de los proyectos pertenecen a TFM y doctorados, que analicen las señales EDA y o artículos pertenecientes al tema y que son interesantes a la hora de basar el estudio.

2.2.2.1. Aprendizaje automático en el diseño de un detector de estrés a partir de señales biomédicas [32]

Para el presente trabajo se han empleado los datos proporcionados por Robert Bosch GmbH y University Siegen. Estos datos se encontraban en archivos pkl y estaban organizados según el detector que se ha empleado en el trabajo para poder detectar las distintas señales (pecho o muñeca), el tipo de sensor y el sujeto. Se realizaron entrenamientos de la red neuronal utilizando conjuntos de datos de los detectores ajustando parámetros como el tamaño del lote y el número de épocas.

Para la selección de datos se ha empleado Python3 y el entorno de notebook de Jupyter para el diseño de la Red Neuronal Artificial del detector de estrés. Para la red se emplearon perceptrones o neuronas como unidades básicas de procesamiento. Se establecieron características clave de la red neuronal como la cantidad de neuronas en cada capa, la cantidad de capas, la función de activación (ELU, SELU, RELU), el optimizador y la función de pérdida.

Para que la red pudiese aprender de forma correcta se emplearon algoritmos de entrenamiento supervisado, entre los cuales destacan el Adadelta, Nadam y Rmsprop, y funciones de costo para la evaluación de resultados como error logarítmico cuadrático medio (ELCM), error absoluto medio (MAE), porcentaje de aciertos y Hinge que es una función de pérdida que se suele emplear en problemas de clasificación.

Para la visualización, se generaron varios tipos de gráficas de las salidas reales frente a las predichas, este tipo de gráficas fueron de todos los diferentes sensores empleados en este trabajo.

La calidad de precisión que se ha logrado al evaluar los estados de estrés, no estrés y otros, ronda el 99% para el sensor de la banda y el 97-98% en el sensor de la pulsera. Los resultados son positivos en comparación con la Universidad de Siegen que ronda un 75-76% de éxito.

Las bibliotecas empleadas en este proyecto son las siguientes: Matplotlib, NumPy, Pandas, Keras, Sklearn y sus correspondientes ramas, Pickle, etc. Todas estas bibliotecas se encuentran descritas o mencionadas con anterioridad en el apartado de la reconocimiento y detección de imágenes.

2.2.2.2. Reconocimiento de emociones a través de las señales fisiológicas del observador [33]

Emotion recognition through observer's physiological signals

En este estudio, se recopilaron datos fisiológicos usando la pulsera E4 y el seguimiento de la mirada Eye Tribe. Se emplearon trozos de películas del conjunto de datos AFEW como estímulos para provocar seis emociones básicas (enfado, asco, miedo, felicidad, tristeza, sorpresa) y una clase neutral. Para el estudio se obtuvo el consentimiento de los participantes y se recopilaron señales fisiológicas como EDA, temperatura de la piel (SKT), PPG, intervalos entre latidos y frecuencia cardíaca (HR). Para el procesamiento de las señales se extrajeron las características de las secuencias de EDA, incluyendo variables estadísticas básicas (máximos, mínimos, medias y varianzas), variables estadísticas de la primera diferencia y variables estadísticas de la segunda diferencia. Estas características cubren algunas de las características típicas informadas en la literatura de fisiología de las emociones. Es importante destacar que también se usó la normalización de min-max para reducir las diferencias entre los participantes. Para el entrenamiento se introdujo una red de tres capas y extracción de características como línea base para este conjunto de datos, logrando un rendimiento del 42,08% en el conjunto de datos de un solo participante. Se investigó la precisión de clasificar el conjunto de datos por activación o valencia, logrando un 68,66% y un 72,72% respectivamente. Se encontró que agregar información de activación o valencia ayuda a mejorar la precisión de reconocimiento.

2.2.2.3. Reconocimiento de emociones utilizando señales de actividad electrodérmica y red neuronal convolucional profunda multiescalar [34]

Emotion Recognition Using Electrodermal Activity Signals and Multiscale Deep Convolutional Neural Network

El estudio se basa en la clasificación de las emociones empleando la base de datos de DEAP, que contiene señales fisiológicas de 32 individuos. Las señales fueron registradas mientras los individuos veían 40 estímulos audiovisuales de un minuto de duración cada uno. Las señales incluyen actividad electrodérmica, SKT, pulso de volumen sanguíneo, frecuencia respiratoria, electromiografía (EMG) y electrooculograma. Es importante incidir que también se han recopilado calificaciones subjetivas de los participantes sobre las emociones que se han experimentado durante la visualización de los trozos de vídeos.

Para la extracción de las características, se han seguido cuatro pasos, procesamiento, enfoque de granulado grueso, extracción de características basada en Red Neuronal Convolucional Multiescalar (MSCNN) y finalmente la clasificación.

En el procesamiento, se normalizan las señales EDA y lo mismo se aplica a las calificaciones subjetivas de las dimensiones de valencia. Se aplica un enfoque de granulado grueso para descomponer la señal en diferentes escalas de tiempo. Se aplica la MSCNN junto con el submuestreo (pooling) para extraer y clasificar las características.

Para poder evaluar los resultados se ha empleado la técnica de la validación cruzada "leave-one-subject-out" (LOSO) para evaluar el rendimiento. También se utilizaron métricas como precisión (accuracy), recall, F1 score y área bajo la curva (AUC). Los resultados mostraron que el enfoque hacia las MSCNN superó el CNN en términos de rendimiento, puesto que la MSCNN trabaja de manera más efectiva con señales mayores.

2.2.3. Electrocardiogramas- ECG

Los electrocardiogramas son los resultados de las electrocardiografías, los indicadores de los ECG más informativos para evaluar la señal son la frecuencia cardíaca (FC), formas de onda P, Q, R, S y T, siendo las R las

partes más prominentes de las ondas [Ilustración 2], y la variabilidad de la frecuencia cardíaca (HRV) [35]. Un aumento de la actividad emocional suele llevar a una subida de la frecuencia cardíaca. Sin embargo, debido a que todos los individuos son diferentes, emociones como la ansiedad y el miedo pueden ir acompañados de una disminución de la frecuencia cardíaca ("reacción de congelación"). Uno de los indicadores del trabajo del corazón, analizado en relación con los estados de ansiedad, es la frecuencia cardíaca. [36]

Un fuerte nivel de estrés emocional puede cambiar significativamente la frecuencia cardíaca del individuo. Por lo tanto, ésta sirve como un indicador objetivo confiable del grado de estrés emocional en una persona en dos condiciones: la experiencia emocional se caracteriza por un fuerte estrés y no se acompaña de actividad física.

Por tanto, se procederá a analizar los métodos y estudios que se han aplicado respecto al presente tema.

2.2.3.1. Utilidad de la forma de onda ECG completa para la clasificación de estrés [37]

Utility of the Full ECG Waveform for Stress Classification

En este estudio se usó un conjunto de datos público disponible en PhysioNet, que contenía registros de ECG de 57 participantes con miedo a las arañas. Se utilizaron estampas de tiempo para etiquetar cada muestra como estresada o no estresada.

Cada registro de ECG se filtró y se dividió en muestras de 10 segundos. Se utilizó el algoritmo de detección de picos Pan-Tompkins para identificar los picos RR. Se extrajeron características de dominio temporal relacionadas con los intervalos RR, para ello se empleó el paquete de Python "neurokit2". Es importante destacar que se empleó un algoritmo de selección de características basado en criterio de información de Akaike (AIC) para reducir el conjunto. Se evaluaron tres algoritmos de clasificación: análisis discriminante lineal (LDA), regresión logística (LR) y clasificación de vectores de soporte (SVC).

Para la evaluación de los resultados se desarrollaron dos modelos, el primer TFM, modelo de características tradicionales, y el segundo NFM, modelo de características novedosas. El TFM incluía 9 características seleccionadas, mientras que el NFM incluía 13. Se compararon los desempeños de ambos modelos utilizando métricas como precisión

(accuracy), recuperación, F1 score, AUC y exactitud. Se encontró que el NFM tenía un mejor desempeño en términos de la minimización de pérdidas de información según el criterio del AIC. Se evaluó también el desempeño de algoritmos de detección de picos automatizados.

2.2.3.2. Algoritmos de aprendizaje automático basados en ECG para la clasificación de los latidos del corazón [38]

ECG-based machine-learning algorithms for heartbeat classification

En este estudio se pretende analizar los ECG para determinar y clasificar enfermedades cardíacas. Para ello se usó la base de datos MIT-BIH arrhythmia y la base de datos SPH para obtener los datos de señales de ECG. Se extrajeron latidos de los electrocardiogramas de diferentes registros y se utilizaron como conjuntos de datos de entrenamiento. Se seleccionaron características como intervalos de tiempo entre diferentes picos de ECG y coeficientes del modelo AR (autorregresivo) para representar las señales.

Para la extracción se empleó una metodología que consta de tres pasos, el primer paso consiste en el filtrado de la señal, el segundo en la detección de picos R utilizando la transformada de Fourier de la transformada fraccional (FrFT) y, finalmente el tercer paso, detección de picos P y T utilizando el algoritmo TERMA. Para el filtrado de la señal, se utilizó la transformada wavelet (DWT) basada en el filtro wavelet Daubechies-4 (db4) para eliminar el ruido y los artefactos de la señal. Por otro lado, para la detección de picos R, se aplicó la FrFT a la señal libre de ruido y se utilizaron técnicas de umbralización y promedio móvil para mejorar la detección. Para la detección de picos P y T, se aplicó el algoritmo TERMA y se utilizó también una técnica de umbralización simplificada después de la eliminación de los picos R. De la propia procesado se obtuvieron intervalos de tiempo entre picos y coeficientes del modelo AR, para utilizar como entradas en los algoritmos de clasificación.

Respecto a los resultados, se logró una detección precisa de los picos R, P y T en las señales de ECG. Es importante incidir en que se empleó la clasificación supervisada de aprendizaje automático con algoritmos de máquinas de vectores de soporte (SVM) y perceptrón multicapa (MLP) para la propia clasificación de las señales.

Se ha concluido que las señales de ECG y la metodología propuesta para diagnosticar enfermedades cardiovasculares de forma automática posee potencial para futuros estudios.

Como conclusión del apartado electrocardiogramas, en la mayoría de los estudios se han observado que se han empleado análisis espectrales y modelos de algoritmos inteligentes para extraer las características con el fin de obtener patrones y, consecuentemente, poder procesar junto con el resto de las señales para obtener una emoción.

3. DESARROLLO

El desarrollo consta de varios capítulos los cuales agrupan los procesos esenciales para llevar a cabo el trabajo de fin de grado. Es importante destacar que el ensayo ha sido llevado a cabo por el Grupo de Investigación en Interfaces Avanzadas AFFECTIVE LAB del Instituto de Investigación en Ingeniería de Aragón en el cual se me permite tomar parte para su puesta en marcha.

3.1. RECOPIACIÓN DE DATOS

El desarrollo comienza por la obtención de los datos, para ello se realiza un ensayo de investigación que pone a prueba la combinación del sistema de MsWH (Multi-sensor Wearable Headband) [39] y un conjunto de sensores como la Muse 2 [40] y la pulsera empática E4 [41] reacondicionados al proyecto. Es importante incidir en que esta investigación se orienta exclusivamente al área de las emociones y a los estados del individuo.

3.1.1. Entorno

La investigación toma lugar dentro de una de las aulas de streaming de la EUPLA, la cual es aislada acústicamente. Dentro de la sala se dispone de una pantalla táctil, mesas, sillas y ordenadores [Ilustración 8]. Durante el ensayo, en el interior de la sala se encuentra como mínimo un supervisor que se encargue de que la ejecución de la prueba es llevada a cabo de manera adecuada.



Ilustración 8: Ensayo en un entorno controlado en la EUPLA

Dentro de este entorno se desarrolla el estudio en un grupo de individuos jóvenes, de 18 a 25 años. Estos son expuestos a una prueba en la que se les muestran 17 vídeos, cada uno de los vídeos es un trozo de una

película de un género con el propósito de evocar diferentes emociones. Antes de visualizar los vídeos, se instruye de cómo es la prueba, se informa de la privacidad, el manejo de datos y se les entregan los papeles de consentimiento. En caso de estar de acuerdo con continuar con la prueba, el individuo acepta el tratado de datos, pero puede abandonar el experimento en cualquier momento. Por otro lado, en caso contrario, el sujeto se marcha de la sala.

3.1.2. Sistema de sensores

El ensayo, como se menciona en la introducción del apartado 3.1. consta de varios módulos que conforman un sistema combinado y reacondicionado para la detección de las señales fisiológicas. Este sistema consta de varios módulos:

- MsWH [39] [Ilustración 9], es una plataforma de registro de señales fisiológicas las cuales son la temperatura de la piel (ST), saturación de oxígeno en la sangre (SpO2), frecuencia cardíaca (HR), actividad electrodérmica (EDA), movimiento y posición del usuario (IMU). La plataforma también posee integrada una cámara angular para capturar las expresiones faciales OV2640, la cual permite capturar imágenes y grabar videos a una frecuencia de 12.5 fotogramas por segundo [42].



Ilustración 9: MsWH [39]

Es importante destacar que los sensores de ST, HR, IMU y SpO2 de la plataforma ya se encuentran calibrados por los propios fabricantes.

- Muse 2 [40] [Ilustración 10] es un dispositivo portátil que combina la medición de la EEG junto con otras señales fisiológicas brindando la capacidad de obtener datos relacionados con la salud y el estado

del individuo en diferentes planos, tanto el físico como el mental y, consecuentemente, emocional.



Ilustración 10: Muse 2 [40]

- Empatica E4 [41] [Ilustración 11] es un dispositivo de pulsera portátil para medir y registrar señales fisiológicas del individuo portador. La E4 recopila información de GSR, ST, aceleración y HR.



Ilustración 11: Empatica E4 [41]

Para los ensayos se realiza un cambio. Puesto que la Muse se coloca sobre la frente, la plataforma hardware de MsWH se tiene que readaptar desplazando la cámara hacia el frente del sujeto en una ubicación estática fija, la cual enfoca de manera frontal al individuo.

3.1.2.1. Colocación de los sensores

Posteriormente se colocan los dispositivos sobre el participante. La cámara se coloca de manera frontal fija sobre un soporte fijo, que se menciona anteriormente, que permite regular acorde a la persona a la cual se graba y la Muse 2, que se coloca sobre la frente. [Ilustración 12]

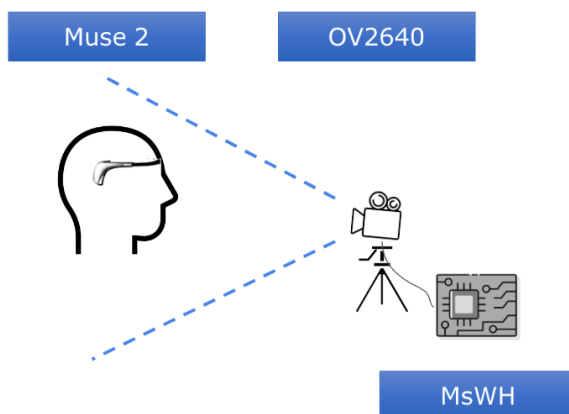


Ilustración 12: Puntos de colocación de Muse 2 y OV2640

La empática E4 se ajusta sobre la muñeca. Tanto la MsWH como la empática emplean electrodos de Ag/AgCl que se colocan sobre las yemas de los dedos de la mano izquierda para capturar las variaciones en la conductancia eléctrica causadas por la actividad de las glándulas sudoríparas. La frecuencia a la cual se graban los datos de ambos dispositivos de EDA (empática E4 y MsWH) es de 4 Hz. [Ilustración 13]

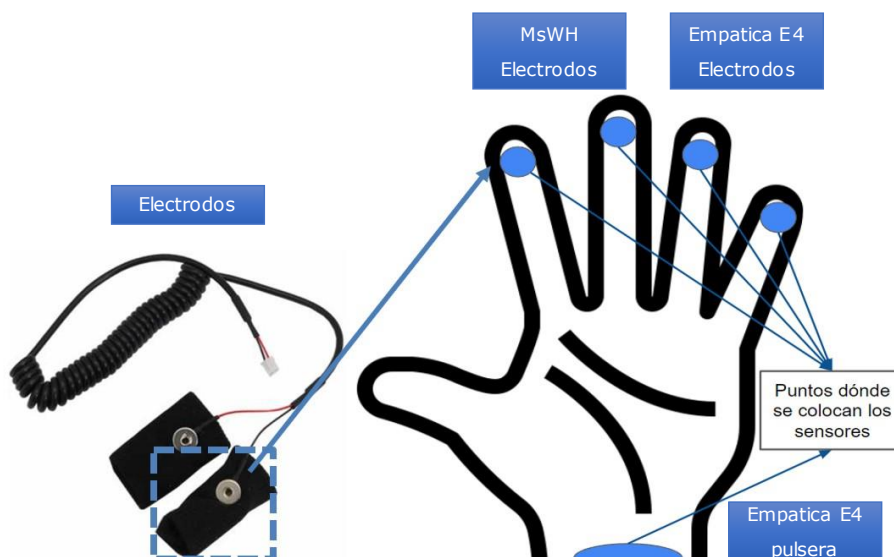


Ilustración 13: Puntos de ubicación de sensores sobre la mano y muñeca izquierdas

Se coloca sobre la mano izquierda puesto que existe una conectividad fisiológica, es decir, al encontrarse más cercana al corazón, se considera teóricamente que, de esta forma, las señales se capturan de una manera más directa. Véase el apartado 2.1.2.1. "Actividad electrodérmica – EDA".

3.1.2.2. Ejecución del ensayo

El conjunto de dispositivos se encuentra integrado en un sistema en el cual se emplean conexiones inalámbricas como el Bluetooth y el Wi-Fi. Para poder realizar la integración de la manera adecuada se emplean softwares para cada uno de los dispositivos comerciales (Muse 2 y empática E4) y se reprograma el módulo de la MsWH para transmitir los datos en bruto. Para la recopilación de estos datos se emplea la aplicación de LabRecorder [43], este permite obtener los datos grabados en formato de XDF. Para ello se abre el programa y se visualiza el flujo de dispositivos que se graban, estos son cada uno de los softwares de los propios dispositivos. A continuación, se actualiza la lista en caso de que falte alguno por conectarse. Posteriormente se seleccionan los flujos deseados de datos marcando las casillas. Luego, se configura la ubicación y el nombre del archivo de salida de la grabación. Finalmente se inicia la grabación. Una vez se finaliza el ensayo, se detiene la grabación de manera manual. La propia estructura del archivo XDF [44] se basa en la especificación detallada del repositorio principal. Los lectores, escritores y herramientas relacionadas se encuentran en la organización XDF-modules. Para poder abrir estos archivos se enlazan submódulos como Python, Matlab y EEGLAB en el repositorio principal donde se tenga planeado manejar los datos. Por tanto, el LabStreamingLayer (LSL)[45] es un protocolo que permite transmitir en tiempo real datos desde diversos dispositivos a través de una red aparte de sincronizar los datos. Para guardar estos datos se usa el LabRecorder que permite la grabación de dichas señales.

Antes de comenzar con la prueba se deja un tiempo para que los sensores se estabilicen y la grabación de los datos no se vea perjudicada por ruidos entre otros factores. Este tiempo también se emplea para explicar al individuo de qué es lo que se hace y se rellena durante el ensayo.

3.1.2.2.1. Descripción de la prueba

La selección de vídeos se basa en un estudio [46] el cual se fundamenta en evaluar la eficacia de los estímulos musicales y cinematográficos en el reconocimiento afectivo. Al igual que en el estudio, se emplean clips de películas de diferentes géneros para poder posteriormente realizar análisis de correlaciones entre las respuestas fisiológicas y las evaluaciones de las películas. Estas evaluaciones se basan en realizar una encuesta después de cada uno de los vídeos que contienen emoción. Es importante incidir en que la duración de los vídeos y la cantidad de las muestras se reducen puesto que muchas emociones se repiten en el estudio de base [47].

Dentro de la encuesta se realizan 2 preguntas, qué emoción se siente, dando a escoger entre las 6 emociones, omitiendo la neutra, y con qué intensidad en una escala del 1 al 9 [Ilustración 14]. Para ello se tiene que marcar una opción de SAM [Ilustración 14], siendo la de la izquierda la opción más calmada y la de la derecha la más excitada.

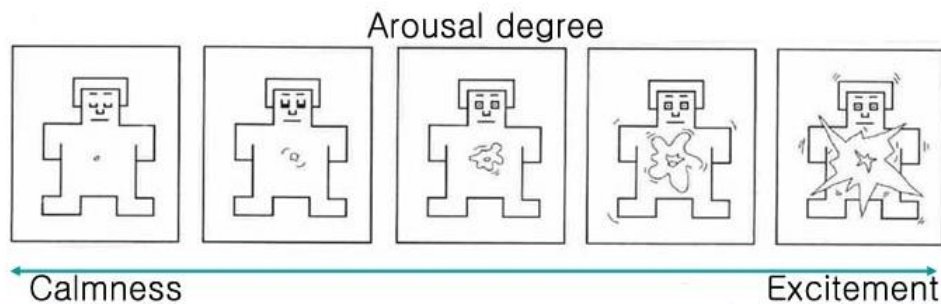


Ilustración 14: Sam grado de activación [46]

Para evitar la influencia entre vídeos y sus correspondientes emociones, se muestra un vídeo de unos pocos segundos de una nube para neutralizar la emoción anterior.

Es importante incidir en que también se graba el identificador de emoción que predomina en cada uno de los clips de películas de acuerdo con su marca de tiempo para posteriormente poder identificar los vídeos y poder comparar los resultados al insertar la propia etiqueta de los cortos en la red neuronal.

3.2. SELECCIÓN DE DATOS

Para la selección de datos, y asegurar la integridad de dichos, se revisan todos los conjuntos de datos y se seleccionan los conjuntos que se consideren apto acorde a los siguientes criterios:

- Ventana de tiempo amplia para poder trabajar con las señales que se emplean en este trabajo. Esto es, que la conexión de los dispositivos no haya fallado durante el grabado de datos.
- Según la presencia de ruido. Si existían señales que poseían demasiadas anomalías; como por ejemplo depresiones en la señal; estas se descartaban. Los ruidos demasiado grandes se debían a causa de movimientos de los participantes durante el ensayo.

3.3. INSTALACIÓN DE SOFTWARES

Se ha decantado por el uso de PyCharm. Este es un entorno de desarrollo integrado, IDE, perteneciente a JetBrains [48] para el lenguaje de alto nivel Python. La versión para usar es la de Community (comunitaria) puesto que es de código abierto y permite la creación de proyectos de Python puros.

Una vez instalado PyCharm, se instala el Python 3.10, principalmente porque proporciona ventajas como la posibilidad de empleo de la mayoría de las bibliotecas y, puesto que no es la versión más nueva, se encuentra más depurada de errores de versión y bugs. También es importante incidir en que los scripts completos se encuentran en los Anexos, por lo que aquí únicamente se exponen los diagramas de flujos de datos y tan solo trozos relevantes para la explicación del desarrollo.

3.4. PROCESAMIENTO DE DATOS

El procesamiento primario es esencial para poder obtener las señales lo más limpias, sincronizadas y veraces posible, puesto que la presencia de ruidos, anomalías, etc, se presentan como una distorsión de los datos y, consecuentemente, alteran los resultados y la realidad. Para ello, en función del tipo de señal, se aplica un análisis determinado y ajustado también a la calidad de la señal. A continuación, se presenta un diagrama de flujo de datos en el cual se expone de manera gráfica el esquema general que sigue el desarrollo de este trabajo. [Ilustración 15]

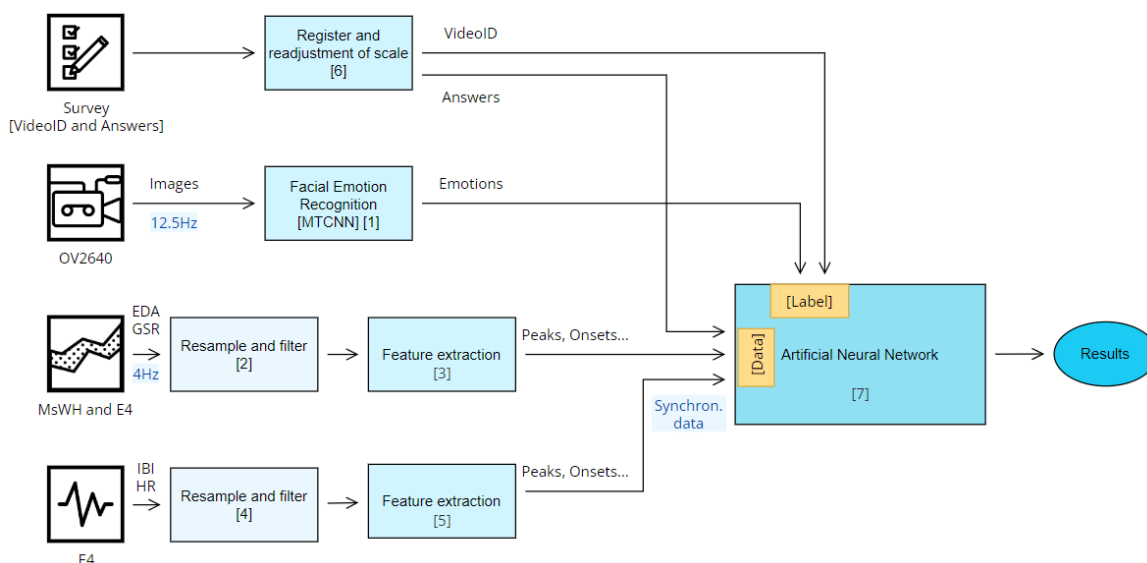


Ilustración 15: Diagrama de flujo de datos

Se observan varios tipos de bloques en función del color. El color más claro representa el procesamiento primario de los datos (bloques [2] y [4]). Posteriormente, les siguen los correspondientes a la extracción de características (bloques [1], [3], [5] y [6]). Finalmente, un bloque grande correspondiente a la Red Neuronal Artificial que se aplica en este trabajo, la BLSTM, que tiene como datos de salida los “Resultados” que corresponden tanto al estudio de las pérdidas de la red neuronal como a los resultados de predicción de emoción (bloque [7]) en función del tipo de datos de entrada.

Todos estos bloques y apartados se explican a continuación en función del tipo de señal de entrada.

Se comienza por las imágenes capturadas a 12.5 fps, las imágenes, primero, se clasifican en función de si existe imagen o no y posteriormente se introducen en la Red Neuronal Convolutiva de Cascada Multitarea (MTCNN), véase apartado 2.1.3.5.3. para ver el funcionamiento de las CNN, la cual decide qué emoción reconoce en la imagen, véase en Anexo, 1. Procesamiento y extracción de emociones de imágenes (MTCNN). [Ilustración 16] La MTCNN extrae la información de cada una de las emociones básicas, más el neutro con sus porcentajes. Estos porcentajes van en una escala de 0 a 1.

```

else:
    fer_emo_detector = FER(mtcnn=True)
    class_emo = fer_emo_detector.detect_emotions(img)

```



```
predom_emo, score_emo =
fer_emo_detector.top_emotion(img)
```

Posteriormente, la información extraída por la MTCNN se guarda en un archivo de tipo CSV para poder realizar un análisis de manera conjunta una vez terminadas de procesar todas las señales en bruto y no sobrecargar el software lanzando todos los scripts a la vez.

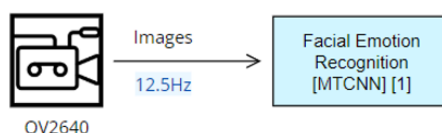


Ilustración 16: Facial Emotion Recognition (MTCNN) (1)

Por otro lado, las señales EDA en bruto grabadas a 4Hz se procesan de la siguiente manera. [Ilustración 17]

Primero se realiza un remuestreo hacia arriba, subiendo la frecuencia de muestreo a 100 Hz, esto permite aumentar la cantidad de datos, y posteriormente se realiza un filtrado con un filtro no causal que usa coeficientes de Butterworth, el filtro usa 12.5 Hz como frecuencia de corte, véase el apartado 2.1.3.3. de Procesamiento y extracción de características de las señales. Es importante destacar que se usa el tiempo de grabado, de cuándo se registra cada dato para remuestrear de una manera adecuada, respetando la sincronización y el tiempo de guardado del dato.

```
def func_upsample_prefilter(filtered, series, stamp):
    (upsample_signal, upsample_t) =
    scipy.signal.resample(filtered,
    len(np.array(series.flat))*25, t=stamp, axis=0)
    fs = 100
    fc = 12.5
    w = fc * 2 / fs
    b, a = scipy.signal.butter(4, w, 'low')
    filtered_signal =
    scipy.signal.filtfilt(b, a, np.array(upsample_signal).flat)
    return filtered_signal
```

Posteriormente, se disminuye la cantidad de muestra para tener la señal de EDA a la misma frecuencia a la cual se encuentran las imágenes. Se emplea esta frecuencia (12.5Hz) a modo de frecuencia de referencia, puesto que para poder tener los datos sincronizados de manera adecuada y que sean correlacionados es necesario tenerlos a la misma frecuencia o en un múltiplo o divisor de esta.

```
def func_corr_function(x, y):
    N = max(len(x), len(y))
    n = min(len(x), len(y))
    if N == len(y):
        lags = np.arange(-N + 1, n)
    else:
        lags = np.arange(-n + 1, N)
    correl = correlate(x / np.std(x), y / np.std(y), 'full')
    return lags, correl, n
```

El inconveniente de por qué no se ha escogido un múltiple o un divisor de la frecuencia de 12.5 Hz es porque la cantidad de imágenes que se obtienen es fija y generar nuevas imágenes intermedias entre cada una de las muestras grabadas crea problemas innecesarios para este estudio.

Después de obtener la señal de EDA a 12.5 Hz se realiza de nuevo un filtrado de tipo no causal, la frecuencia de corte cambia a 0.75 Hz, puesto que esta es la frecuencia arbitraria que elimina las componentes de alta frecuencia y ruidos.

Posteriormente, se calcula la correlación entre cada una de las señales de actividad electrodérmica, GSR (empática E4) y EDA (MsWH) y se alinean acorde a la cantidad de desvío que se haya generado al ejecutar un remuestreo hacia arriba. Finalmente, los datos son segmentados acorde a una ventana y normalizados.

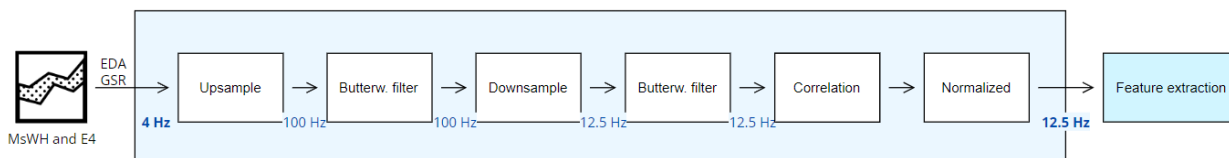


Ilustración 17: Resample and filter (2)

Esta señal procesada pasa a través de unos algoritmos programados pertenecientes a una biblioteca existente denominada neurokit, la cual integra conjuntos de algoritmos y funciones que extraen características buscadas.

```
def func_valores_tonic_phasic(signal):
    t,p = signal["EDA_Tonic"].values,
    signal["EDA_Phasic"].values
    return (t,p)
(t0, p0) = func_valores_tonic_phasic(e4_data)
(t1, p1) = func_valores_tonic_phasic(mod_data)
```

```
e4_data['EDA_Raw'] = e4_signal
mod_data['EDA_Raw'] = mod_signal
```

La señal pasa por un extractor de la componente física, posteriormente se calcula el umbral (threshold), con el umbral se determina dónde se encuentran los onsets y, consecuentemente, los picos. De estos picos se extraen también la amplitud y el ancho del pico. Véase el script completo en Anexos, 2. Procesamiento y extracción de características de la señal de actividad electrodérmica.

```
def func_peaks_onset_AmHt(p, num):
    signal_findpeak = nk.eda_findpeaks(p, sampling_rate=num)
    calc_threshold = ((signal_findpeak['SCR_Height'].max() -
    signal_findpeak['SCR_Height'].min())/2)*0.01
    peaks_array = nk.eda_peaks(p, sampling_rate=num,
    method='kim2004', amplitude_min=calc_threshold) #neurokit,
    gamboa2008, kim2004, vanhalem2020, nabian2018
    final_peaks =
    nk.eda_fixpeaks(peaks_array[0]["SCR_Peaks"],
    onsets=peaks_array[0]["SCR_Onsets"])
    return peaks_array, final_peaks,
    peaks_array[0]["SCR_Onsets"]
```

Todos estos datos finalmente son guardados en un archivo CSV para que, al igual que en el caso de las imágenes, se pueda realizar posteriormente un análisis conjunto. [Ilustración 18]

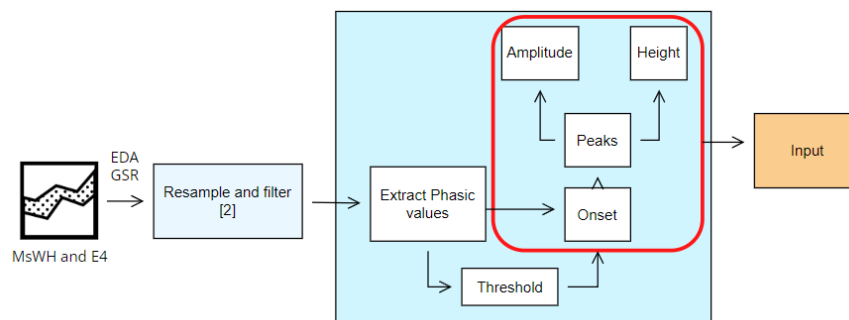


Ilustración 18: Feature extraction (3)

Por el lado de la frecuencia cardíaca, el procesamiento comienza de una manera similar a la de la señal EDA. Este proceso de remuestreo acorde al tiempo y filtrado se realiza para poder sincronizar y depurar correctamente los datos. Finalmente, los datos son segmentados [Ilustración 19], o recortados para una ventana para no tener tantos datos y poder manejar el conjunto de una manera más cómoda.

Es importante incidir que esto se realiza también para poder recortar las anomalías presentes en las señales.

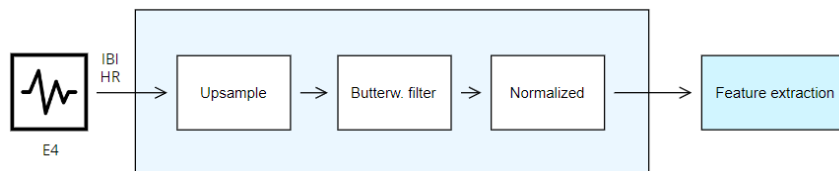


Ilustración 19: Resample and filter (4)

Para la extracción de las características de la señal se calcula la desviación estándar y la media para finalmente obtener los picos de las señales de la frecuencia cardíaca [Ilustración 20].

```

def func_Standard_Deviation(signal0):
    signal = np.array(np.array(signal0).flat)
    for i in range(len(signal) - 1):
        values = np.std([signal[i], signal[i+1]])
        SD_2_array.append(values)
    return SD_2_array

def func_mean_HR(signal, num_data_points):
    return
np.sum(np.array(np.array(signal).flat))/num_data_points
  
```

Es fundamental incidir en que en el desarrollo se emplean tanto la IBI como la HR, puesto que una es inversa de la otra y dividida entre 60. Al igual que en los casos anteriores, los datos se guardan en un CSV para posteriormente realizar un estudio o un análisis estadístico.

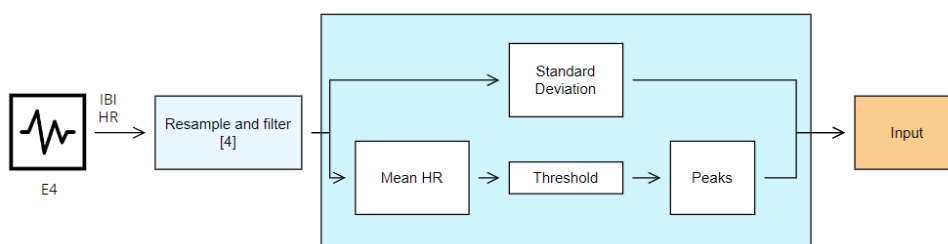


Ilustración 20: Feature extraction (5)

Finalmente, el último conjunto de datos a procesar son los correspondientes a los identificadores de los vídeos y las respuestas de cada uno de los individuos. [Ilustración 21]

Los identificadores se encuentran registrados con dos dígitos, el primer dígito se encuentra en función del tipo de emoción que se asocia al vídeo, si la emoción se repetía, se guardaba un identificador de emoción repetida +1 en el segundo dígito del registro. Esta escala, al no corresponder con la escala de emociones establecida, se reajusta. Por otro lado, las respuestas se sobrescriben con los identificadores de vídeo, para este problema, se emplea el mismo conjunto de datos y se reajusta acorde a las respuestas apuntadas de manera manual durante cada uno de los ensayos. Cada uno de los conjuntos de datos se guardan de manera separada, puesto que posteriormente, los datos de los identificadores de vídeo reajustados son etiquetas para la red neuronal, véase script de Anexos 4. Readaptación de señales.

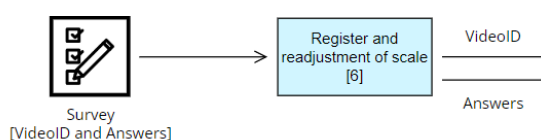


Ilustración 21: Register and readjustment of scale (6)

En este caso, se expone a un problema abstracto, puesto que la relación entre las emociones y las características de los datos de entrada no es lineal, también es importante incidir en que las etiquetas para los identificadores de emociones son discretas, por lo que es necesario recurrir a una red neuronal. La red neuronal, por otro lado, se diseña acorde a la cantidad de datos disponibles, la complejidad de la propia red y la disponibilidad de recursos computacionales (RAM).

Para continuar con los datos, las etiquetas ('labels') y las características ('features') se reagrupan para conformar cada una un único conjunto de datos que sirven de inputs para la red neuronal. En el apartado de Anexos se encuentran los scripts en los cuales se conforma cada conjunto y se almacena como CSV.

Posteriormente, se establece la arquitectura de la red [Ilustración 22], se decanta por la BLSTM (Bidirectional Long Short-Term Memory), se determinan los hiperparámetros que son los que establecen cómo se comporta la red. Los scripts correspondientes se encuentran en Anexos con las entradas 5. Red Neuronal Bidireccional de memoria a largo y corto plazo(manual) y 5.1. Optimización automatizada.

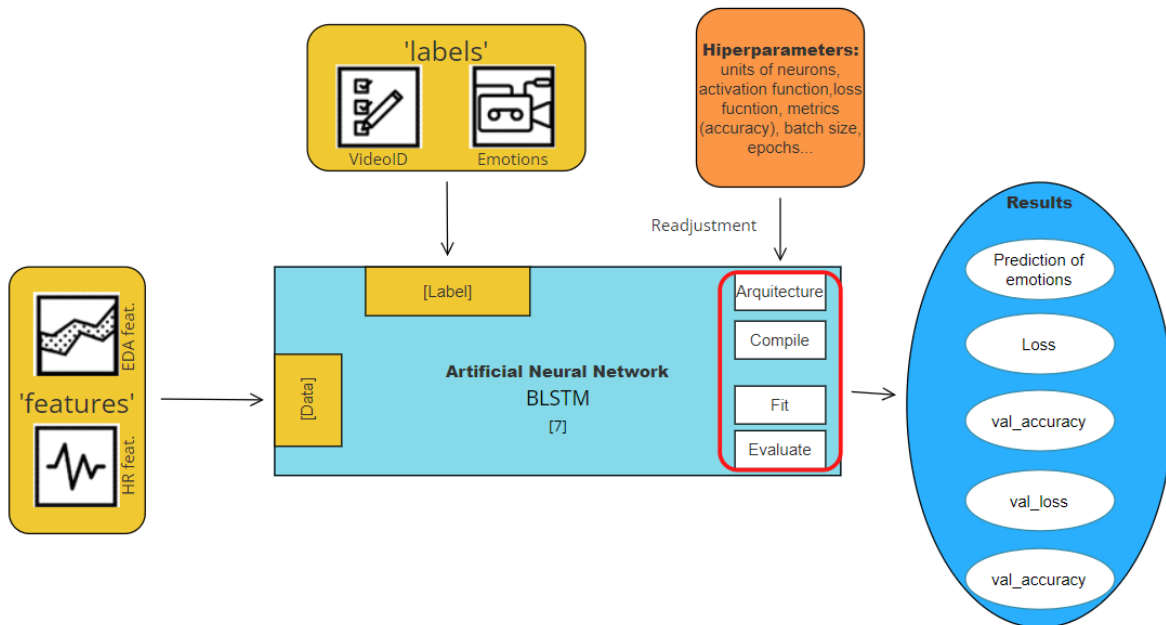


Ilustración 22: Artificial Neural Network (7)

A continuación, tanto labels como features se dividen en 20-80. 20% de los dos conjuntos son para el testeo o la evaluación y el otro 80% se emplea para el entrenamiento de la propia red neuronal.

Se introducen los conjuntos del 80% en la red neuronal y se testea o evalúa con el conjunto del 20% para asegurarse de que la red neuronal es capaz de aprender de los errores.

Finalmente, para poder obtener los resultados se indica a la BLSTM que prediga los resultados en función de los features de testeo y que devuelva valores de las predicciones de emociones, pérdidas y precisión del modelo en el conjunto de entrenamiento y, por último, de las pérdidas y precisión del modelo en el conjunto de validación por cada una de las iteraciones realizadas.

En el apartado "4. Resultados" se estudian las diversas optimizaciones de la red neuronal desde varias perspectivas, como la cantidad de etiquetas a introducir en la red o como tipo de optimización a utilizar para los hiperparámetros de la red neuronal.

4. RESULTADOS

Los datos presentados a continuación se fundamentan en las características extraídas de las señales de actividad electrodérmica y frecuencia cardíaca en función de los estímulos proporcionados por los vídeos correspondientes. De los 15 participantes que se sometieron al estudio, 3 cumplen con los criterios de inclusión al presentar datos con menor cantidad de ruido en las señales. Sin embargo, es importante resaltar que el estudio se ve afectado por la desconexión de la empática E4 con su API (véase apartado "3.1.2. Sistema de sensores"), y, como consecuencia, se excluye a un número considerable de usuarios.

Para poder comprender la evolución del estudio, se presentan las tablas: Tabla 1, Tabla 2 y Tabla 3. Estas tablas contienen información sobre la duración de cada uno de los vídeos, el número de puntos de respuesta galvánica, período y frecuencia de las señales de EDA y HR en cada una de las ventanas temporales definidas en segundos. Además, se incluyen gráficas de los usuarios seleccionados para el estudio.

Se observa una marcada variabilidad en el número total de picos de SCR [Ilustración 23] durante la duración de los vídeos para los tres sujetos que se analizan.

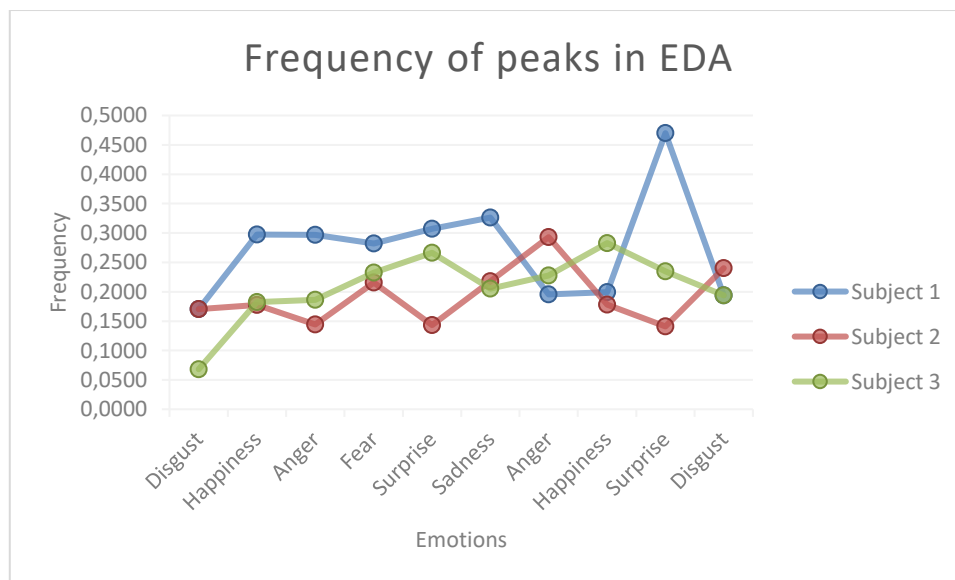


Ilustración 23: Frecuencia de picos en EDA de 3 sujetos

Por otro lado, la frecuencia de picos R-R [Ilustración 24] también exhibe una alta variabilidad, lo que dificulta la identificación visual del patrón de comportamiento que cada individuo presenta durante la visualización de

los vídeos. Con el objetivo de abordar esta problemática, se procede a realizar un análisis parcial e individualizado de cada sujeto.

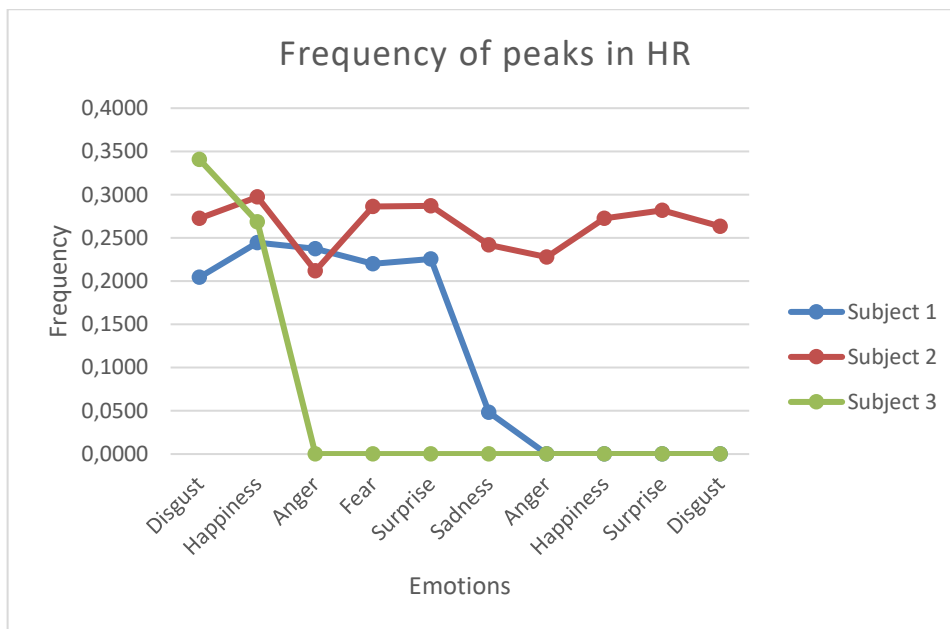


Ilustración 24: Frecuencia de picos en HR de 3 sujetos

Duration (s)	EDA S1	EDA Freq. S1	EDA Per. S1	HR S1	HR Freq. S1	HR Per. S1
29,36	5	0,1703	5,8720	6	0,2044	4,8933
208,56	62	0,2973	3,3639	51	0,2445	4,0894
118,00	35	0,2966	3,3714	28	0,2373	4,2143
240,88	68	0,2823	3,5424	53	0,2200	4,5449
48,80	15	0,3074	3,2533	11	0,2254	4,4364
82,72	27	0,3264	3,0637	4	0,0484	20,6800
61,44	12	0,1953	5,1200	0	0,0000	0,0000
95,44	19	0,1991	5,0232	0	0,0000	0,0000
21,28	10	0,4699	2,1280	0	0,0000	0,0000
129,04	25	0,1937	5,1616	0	0,0000	0,0000

Tabla 1: Sujeto 1, características de HR y EDA

El primer individuo [Tabla 1], identificado en azul [Ilustración 23], muestra un aumento inicial en sus datos, lo que sugiere un nivel de interés o atención hacia los vídeos y la prueba ejecutada durante la captación de las señales. Posteriormente, se observa una disminución en los niveles de enfado y felicidad [Ilustración 25]. En contraste, la frecuencia cardíaca [Ilustración 24, Ilustración 25] se mantiene estable durante la duración de los cortos, hasta llegar al vídeo relacionado con la emoción de tristeza. Esta estabilidad se debe a que el sistema nervioso

autónomo no es tan sensible a los cambios emocionales, lo que resulta en una respuesta cardíaca menos pronunciada frente a las emociones.

Es importante señalar y recordar que una gran parte de los datos no se registran debido a un fallo en el dispositivo de Empatica E4 durante la grabación. Como resultado, a partir del momento de la emoción de tristeza, no se pueden analizar los resultados en el caso del primer participante [Ilustración 25].

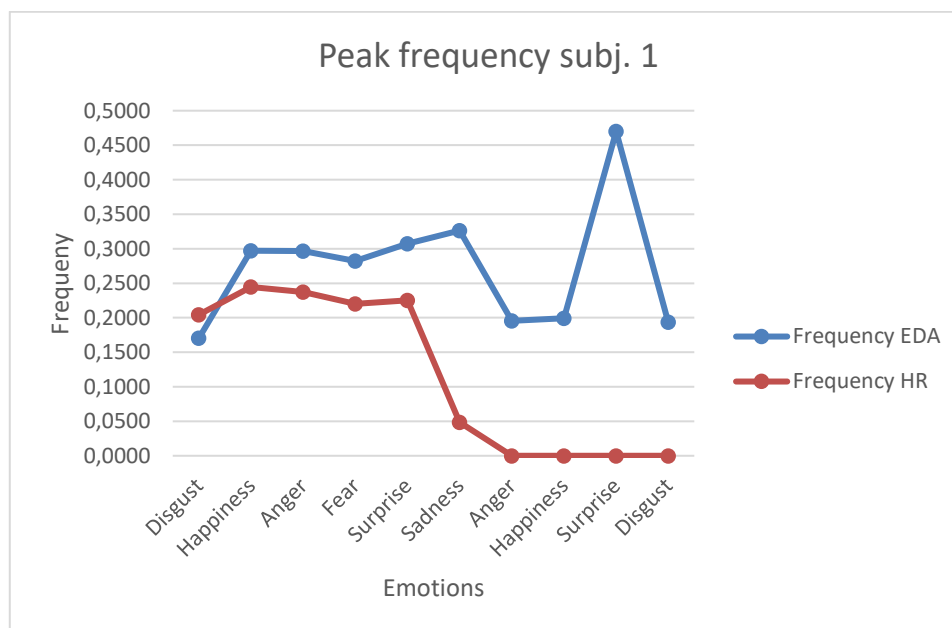


Ilustración 25: Sujeto 1: frecuencia de picos en EDA y HR

Duration (s)	EDA S2	EDA Freq. S2	EDA Per. S2	HR S2	HR Freq. S2	HR Per. S2
29,36	5	0,1703	5,8720	8	0,2725	3,6700
208,56	37	0,1774	5,6368	62	0,2973	3,3639
118,00	17	0,1441	6,9412	25	0,2119	4,7200
240,88	52	0,2159	4,6323	69	0,2864	3,4910
48,80	7	0,1434	6,9714	14	0,2869	3,4857
82,72	18	0,2176	4,5956	20	0,2418	4,1360
61,44	18	0,2930	3,4133	14	0,2279	4,3886
95,44	17	0,1781	5,6141	26	0,2724	3,6708
21,28	3	0,1410	7,0933	6	0,2820	3,5467
129,04	31	0,2402	4,1626	34	0,2635	3,7953

Tabla 2: Sujeto 2, características de HR y EDA

Por otro lado, en el segundo caso [Tabla 2], representado en rojo [Ilustración 23], se observa una tendencia lineal media pero variable en la duración de los eventos de actividad electrodérmica, con períodos mínimos de 4.5956 y máximos de 7.0933. Se destacan varios momentos en los que la SCR aumenta considerablemente en comparación con otras emociones, como el miedo, el enfado y el asco. Se percibe que la emoción de enfado es la que genera más actividad de SCR en este sujeto. Esto se debe a la activación del sistema nervioso autónomo del individuo, lo que resulta en un aumento en los niveles de sudoración. En el vídeo se muestra una situación injusta, lo que lleva a una evaluación cognitiva negativa.

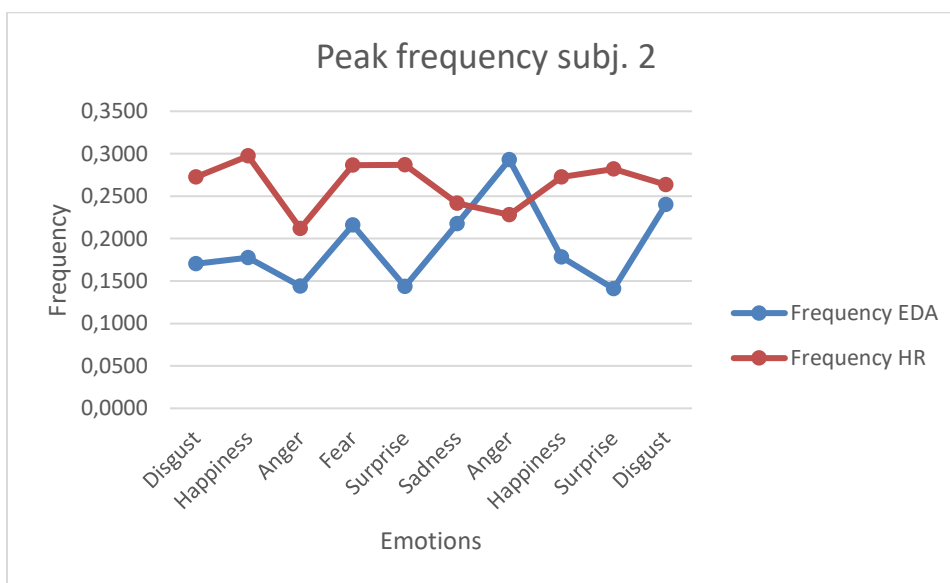


Ilustración 26: Sujeto 2: frecuencia de picos en EDA y HR

Al observar la frecuencia cardíaca [Ilustración 24], se aprecia que las mayores frecuencias se encuentran en las emociones de alegría, miedo y sorpresa. Estas tres emociones están estrechamente relacionadas con el sistema nervioso simpático, ya que implican una respuesta inmediata que regula actividades como la frecuencia cardíaca, la dilatación de los bronquios para redistribuir el flujo sanguíneo hacia los músculos y, consecuentemente, un aumento en la liberación de adrenalina.

Duration (s)	EDA S3	EDA Freq. S3	EDA Per. S3	HR S3	HR Freq. S3	HR Per. S3
29,36	2	0,0681	14,6800	10	0,3406	2,9360
208,56	38	0,1822	5,4884	56	0,2685	3,7243
118,00	22	0,1864	5,3636	0	0,0000	0,0000
240,88	56	0,2325	4,3014	0	0,0000	0,0000
48,80	13	0,2664	3,7538	0	0,0000	0,0000
82,72	17	0,2055	4,8659	0	0,0000	0,0000
61,44	14	0,2279	4,3886	0	0,0000	0,0000
95,44	27	0,2829	3,5348	0	0,0000	0,0000
21,28	5	0,2350	4,2560	0	0,0000	0,0000
129,04	25	0,1937	5,1616	0	0,0000	0,0000

Tabla 3: Sujeto 3, características de HR y EDA

Y, finalmente, en el tercer caso [Tabla 3], en verde [Ilustración 23], se observa una tendencia gradual ascendente en la respuesta de actividad electrodérmica, lo cual sugiere un aumento progresivo en la activación fisiológica y emocional a medida que se avanza en la visualización de los vídeos [Ilustración 27]. Esto indica un incremento en el interés y/o la atención, lo que a su vez refleja un aumento en la excitación emocional.

La actividad electrodérmica se destaca especialmente en la aparición de altos niveles respuestas de conductancia de la piel durante las emociones positivas, como la alegría y la sorpresa [Ilustración 23]. En comparación con los otros dos sujetos, esta persona muestra una mayor actividad y respuesta en situaciones relacionadas con emociones positivas.

Por otro lado, en lo que respecta a la frecuencia de los picos R-R, se observa un aumento durante la primera emoción, lo cual se debe a que la emoción de asco es una emoción negativa que activa el sistema nervioso autónomo simpático como respuesta a estímulos visuales que pueden poner en riesgo la integridad física. Es importante destacar que la frecuencia de los picos R-R disminuye posteriormente, debido a que la persona se adapta al entorno y entra en un estado de calma, lo que indica un cambio fisiológico hacia la relajación.

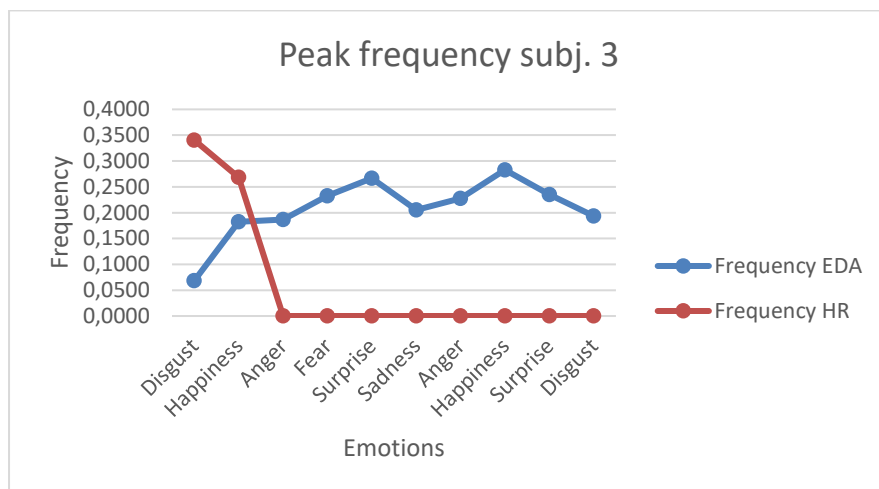


Ilustración 27: Sujeto 3: frecuencia de picos en EDA y HR

En general se observa que las emociones más influyentes en la actividad electrodérmica son el enfado [Ilustración 28], la sorpresa [Ilustración 29], la alegría [Ilustración 30] y el miedo [Ilustración 31]. Es fundamental destacar que el incremento en la sudoración, y por ende en la actividad electrodérmica, está principalmente asociado con la emoción de miedo, lo que indica una activación del sistema nervioso simpático.

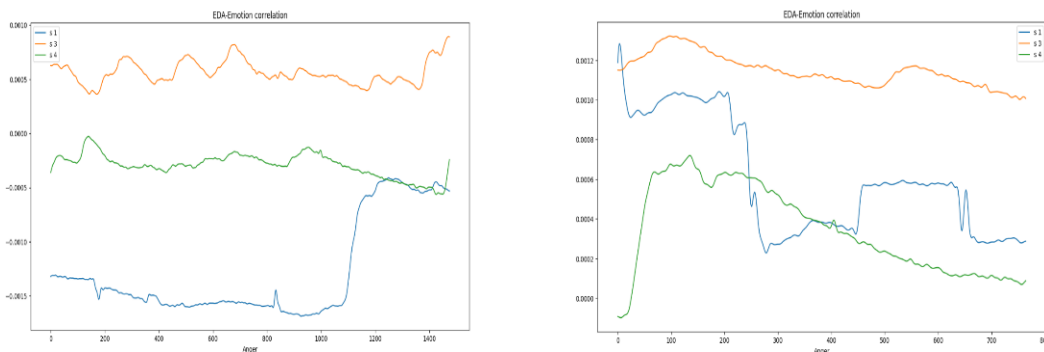


Ilustración 28: Emoción: Enfado - cambios bruscos en señal EDA

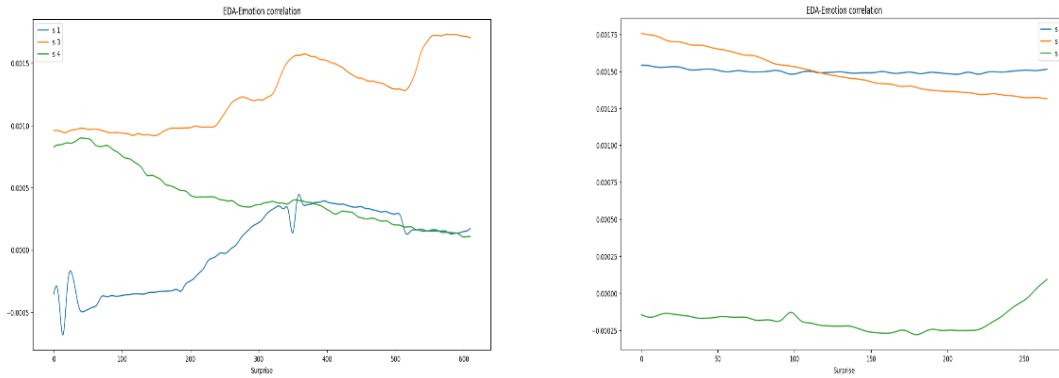


Ilustración 29: Emoción: Sorpresa - cambios bruscos y medios en señal EDA

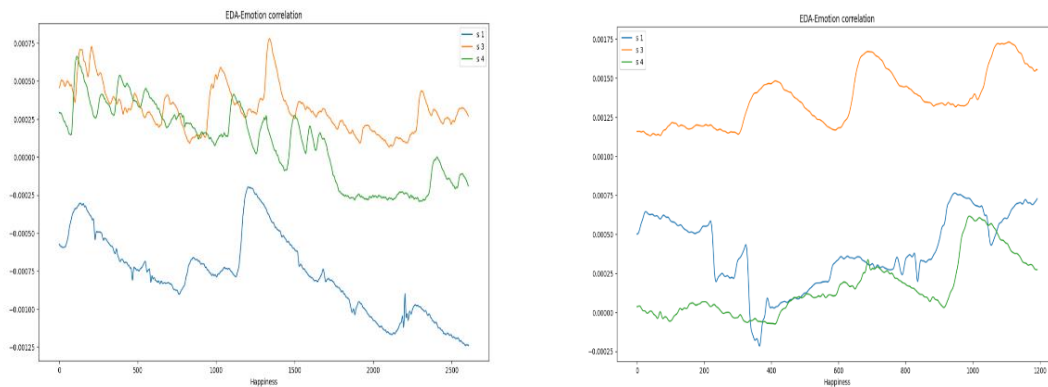


Ilustración 30: Emoción: Alegría - cambios inestables en señal EDA

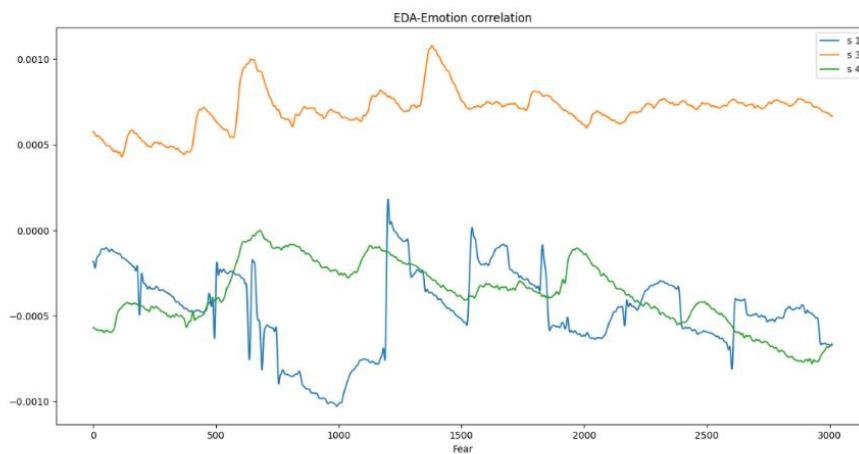


Ilustración 31: Emoción: Miedo - cambios bruscos, variedad de picos en señal EDA

En relación con la frecuencia cardíaca [Ilustración 24], se destacan las emociones de tristeza y miedo, ya que se asocian con la activación del sistema nervioso simpático, lo que aumenta la frecuencia cardíaca y prepara el cuerpo para un estado de alerta.

Respecto a las amplitudes de las señales de actividad electrodérmica de los tres sujetos [Ilustración 32], se pueden identificar patrones emocionales en cada uno de ellos. En el primer caso, "Subj. 1", se observan amplitudes altas en emociones negativas y neutras como el miedo, la ira y la sorpresa, mientras que la tristeza presenta amplitudes bajas. Por otro lado, el segundo sujeto, "Subj. 2", destaca con amplitudes altas en el disgusto y la sorpresa, nuevamente, la tristeza presenta amplitudes bajas. Finalmente, el tercer individuo, "Subj. 3" exhibe amplitudes altas en el disgusto y la felicidad, y, como en los casos anteriores, la tristeza predomina con amplitudes relativamente bajas.

Por lo que en general se observa que el conjunto de datos revela un patrón común de amplitudes altas para el disgusto, miedo, ira y sorpresa, con amplitudes moderadas a la felicidad y bajas en la tristeza. Es relevante señalar que estos patrones son específicos, pero para obtener resultados más concluyentes y detallados, se requiere de un mayor número de conjuntos de datos.

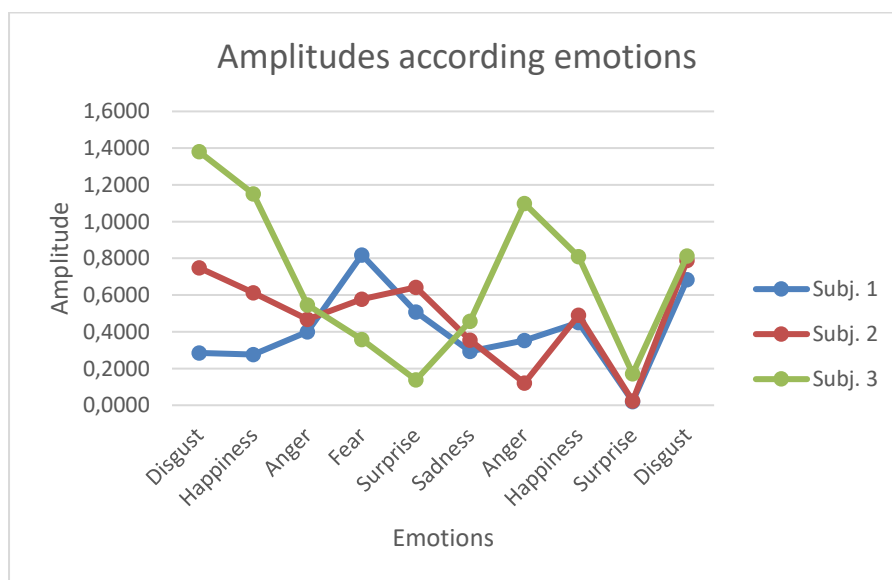


Ilustración 32: Amplitudes de EDA en función de las emociones de 3 sujetos

4.1. HIPÓTESIS DE RED NEURONAL CON DOBLE ETIQUETA

A continuación, se expone el estudio de la red neuronal propuesta con doble etiqueta. La doble etiqueta, como se indicó en el apartado de desarrollo, consta de las emociones determinadas por el identificador de video y por las emociones determinadas por la MTCNN de imágenes.

Epoch	Loss	Accuracy	Val_Loss	Val_Accuracy
1	9,249	0,865	14,6221	0,8586
2	8,2089	0,8662	16,9942	0,8586
3	6,8088	0,8662	19,464	0,8586
4	6,1134	0,8662	21,7821	0,8586
5	5,005	0,8662	24,3459	0,8586
6	3,1616	0,8662	27,1924	0,8586
7	3,0294	0,8662	29,6815	0,8586
8	1,702	0,8662	32,3824	0,8586
9	0,8931	0,8662	35,0894	0,8586
10	0,1038	0,8662	37,6766	0,8586

Tabla 4: Etiqueta doble. Entrenamiento y testeo de red, pérdidas y precisiones.

En la tabla anterior [Tabla 4], se presentan los datos correspondientes a la red neuronal en la cual el modelo es entrenado durante 10 épocas, y se registran cambios en la función de pérdida ("loss") y la precisión ("accuracy") en el conjunto de entrenamiento y el conjunto de validación ("val_loss" y "val_accuracy", respectivamente).

A medida que el modelo avanza en las épocas de entrenamiento, se observa una disminución gradual en la pérdida tanto en el conjunto de entrenamiento como en el conjunto de validación. Esto sugiere que el modelo aprende y mejora su capacidad para hacer predicciones precisas a medida que se le proporcionan más datos. Es importante señalar que, para evaluar su desempeño en datos no vistos durante el entrenamiento, se utiliza un conjunto de validación creado a partir de un segmento de los datos destinados para el testeo, y estos datos se alternan de manera aleatoria.

En cuanto a la precisión, se mantiene constante en 0.8662 en todas las épocas, tanto en el conjunto de entrenamiento como en el de validación. Esto indica que el modelo alcanza un nivel de precisión consistente y no mejora más allá de ese punto.

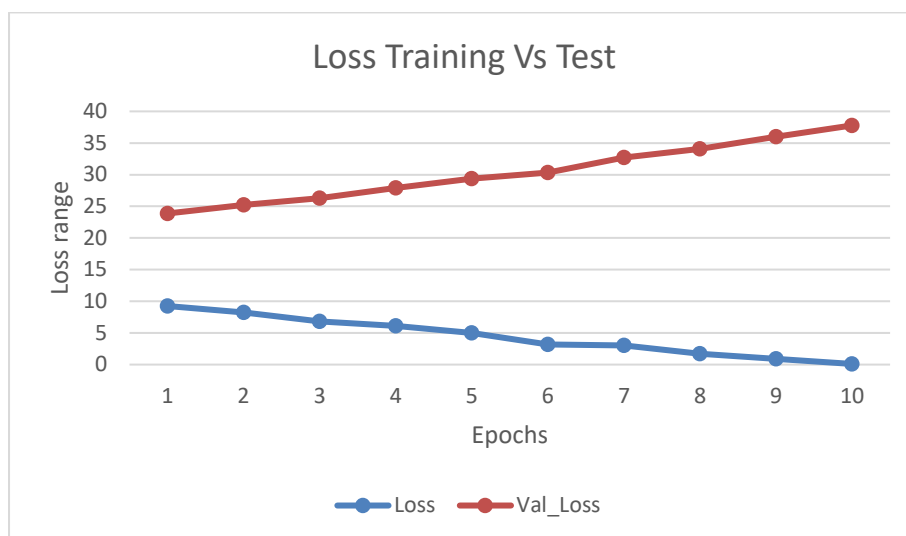


Ilustración 33: Etiqueta doble. Entrenamiento y validación de red: pérdidas.

Al analizar la gráfica [Ilustración 33] se observa una tendencia descendente a medida que avanzan las épocas de entrenamiento. Esto indica que el modelo logra reducir la función de pérdida tanto en el conjunto de entrenamiento como en el de validación.

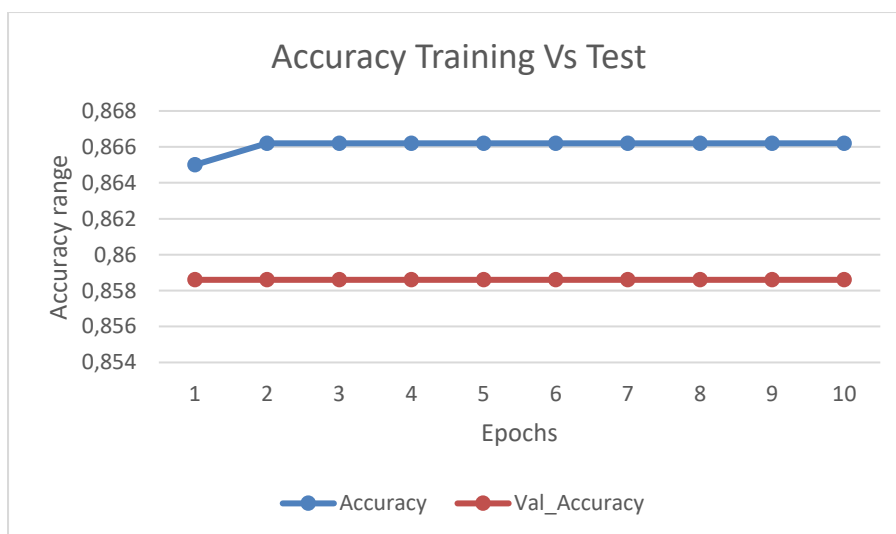


Ilustración 34: Etiqueta doble. Entrenamiento y validación de red: precisión.

Por otro lado, la gráfica de precisión vs. precisión de validación [Ilustración 34] muestra una línea horizontal en 0.8662 en todas las épocas. Este resultado respalda lo que se menciona previamente sobre la precisión constante alcanzada por el modelo en ambos conjuntos de datos.

4.2. HIPÓTESIS DE RED NEURONAL CON UNA ÚNICA ETIQUETA

Para analizar la influencia de las etiquetas sobre la red neuronal, se presentan a continuación los ensayos con la BLSTM para una única etiqueta.

4.2.1. Emociones según MTCNN

En este apartado, se exponen los dos métodos de optimización a usar para la determinación de hiperparámetros en las redes neuronales. Cada uno de estos métodos incluye una tabla y dos gráficas que muestran las pérdidas y las precisiones en función del tipo de datos utilizados. Estos resultados se comparan con los obtenidos en el apartado "4.1. Hipótesis de Red Neuronal con doble etiqueta" [Tabla 4].

Es relevante destacar que este apartado emplea como etiqueta única las predicciones realizadas por la MTCNN. Estas predicciones son las 6 emociones, cada una representada en una columna y con un porcentaje que refleja la cantidad de emoción que muestra el individuo en un determinado momento a partir de la imagen de entrada recibida.

4.2.1.1. Optimización manual MTCNN

Durante el proceso de entrenamiento de la red neuronal "BLSTM", que utiliza características de la señal de actividad electrodérmica y la frecuencia cardíaca como "features" y emociones con sus porcentajes como "labels", se genera la Tabla 5 que muestra las pérdidas y precisiones obtenidas en 10 épocas, es decir, en 10 iteraciones.

Para configurar adecuadamente la red acorde a sus características, se seleccionan 64 neuronas para cada uno de los modelos, tanto el de la propagación hacia delante como el de la propagación hacia atrás. Se establece que la función de activación es la de softmax, el optimizador es Adam, y la función de pérdida es la correspondiente a "categorical_crossentropy".

Epoch	Loss	Accuracy	Val_Loss	Val_Accuracy
1	0,8375	0,6659	1,1919	0,836
2	0,6883	0,6686	1,234	0,836
3	0,5048	0,6686	1,1822	0,836
4	0,3094	0,6686	1,134	0,836
5	0,1027	0,6686	1,0246	0,836
6	-0,1076	0,6686	0,9829	0,836
7	-0,3064	0,6686	0,941	0,836
8	-0,504	0,6686	0,8636	0,836
9	-0,704	0,6686	0,7999	0,836
10	-0,9014	0,6686	0,706	0,836

Tabla 5: Etiqueta única: MTCNN. Entrenamiento y testeo de red, pérdidas y precisiones. Optimización manual.

En cuanto a la función de pérdida [Ilustración 35, "Loss"], se observa una disminución gradual a medida que avanza el proceso de entrenamiento.

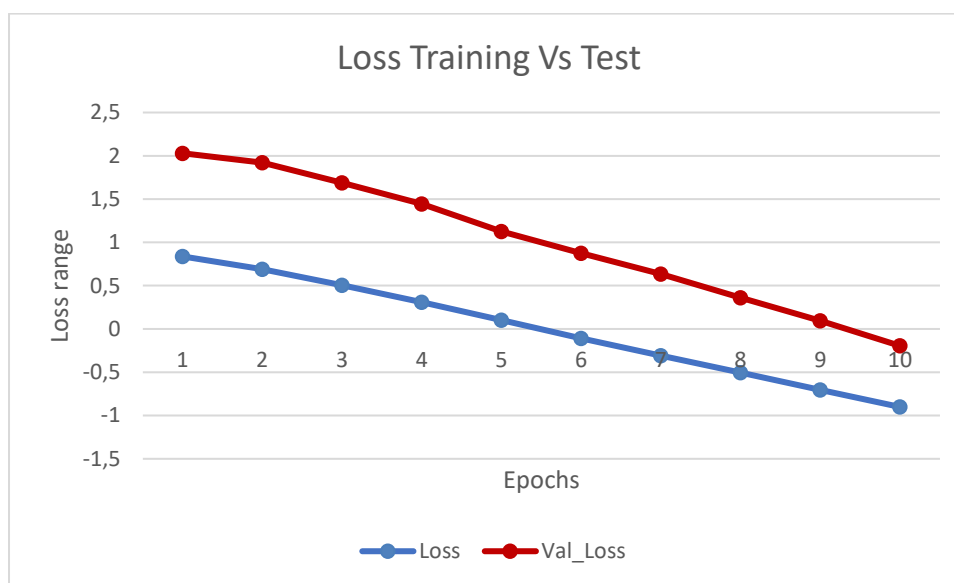


Ilustración 35: Etiqueta única: MTCNN. Entrenamiento y validación de red: pérdidas. Optimización manual.

En la primera época, la función de pérdida se sitúa en 0.8375, y a medida que se progresa, disminuye hasta alcanzar un valor negativo en las últimas épocas, llegando a -0.9014 en la décima época. Es relevante destacar que la presencia de valores negativos en la función de pérdida indica un sobreajuste para el conjunto de datos de entrenamiento.

Por otro lado, se observa que en el conjunto de testeo se alcanza una pérdida de 0.706 [Ilustración 35, "Val_loss"], lo que indica que el sobreajuste influye principalmente en conjuntos grandes de datos o cuando la red se aplica a conjuntos de datos pequeños.

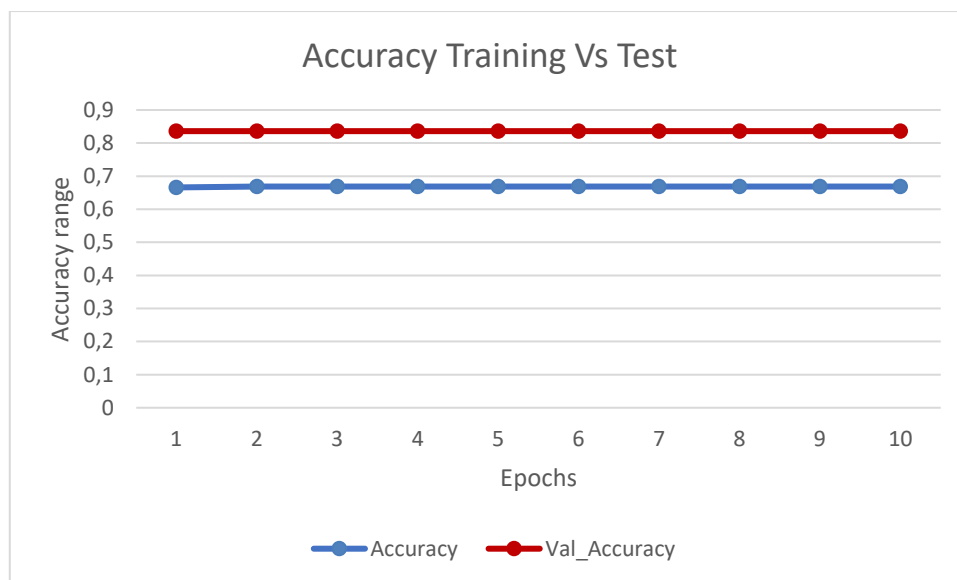


Ilustración 36: Etiqueta única: MTCNN. Entrenamiento y validación de red: precisión. Optimización manual.

Este mismo comportamiento se puede ver en la precisión, ya que esta se mantiene constante a lo largo de todas las épocas [Ilustración 36, Accuracy]. Tanto en los datos de entrenamiento como en los de validación, la precisión se estabiliza en un valor de 0.6686 con el conjunto de datos de entrenamiento. La precisión de validación [Ilustración 36, "Val_accuracy"] también se mantiene constante en 0.836 en todas las épocas, lo que sugiere que el modelo alcanza un límite en su capacidad para mejorar su rendimiento.

4.2.1.2. Optimización automatizada MTCNN

Así como en el apartado "4.2.1. Emociones según MTCNN", la red neuronal recibe las mismas características. Los parámetros utilizados para su entrenamiento se determinan mediante KerasClassification, que modela el comportamiento de GridSearchCV.

Después de varias iteraciones, se determinan 16 neuronas para cada uno de los modelos, tanto el de la propagación hacia delante como hacia atrás. Se establece que la función de activación es de ReLu, el optimizador se mantiene como Adam, y la función de pérdida también se

mantiene en "categorical_crossentropy". La evolución del entrenamiento con estos hiperparámetros se muestra en Tabla 6, donde se detallan las pérdidas y precisiones en cada época.

Epoch	Loss	Accuracy	Val_Loss	Val_Accuracy
1	2,6462	0,5273	1,4601	0,836
2	1,7396	0,6653	3,9488	0,836
3	1,8629	0,6599	3,9426	0,836
4	1,0096	0,6655	1,4631	0,836
5	0,9463	0,6683	1,2991	0,836
6	0,7223	0,6686	1,234	0,836
7	0,7501	0,6686	1,2169	0,836
8	0,7068	0,6686	1,2093	0,836
9	0,6991	0,6679	1,2078	0,836
10	0,6978	0,6685	1,2081	0,836

Tabla 6: Etiqueta única: MTCNN. Entrenamiento y testeo de red, pérdidas y precisiones. Optimización GridSearchCV.

En la primera época del conjunto de entrenamiento [Tabla 6], se observa un valor inicial de pérdida de 2.6462 [Ilustración 37] y una precisión de 0.5273 [Ilustración 38].

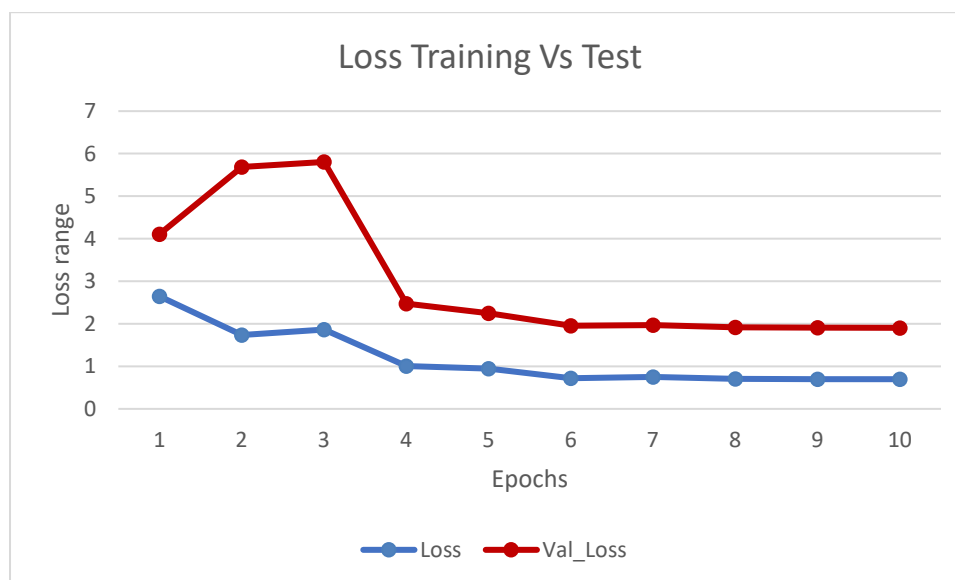


Ilustración 37: Etiqueta única: MTCNN. Entrenamiento y validación de red: pérdidas. Optimización GridSearchCV.

A medida que el entrenamiento progresa, la pérdida disminuye progresivamente, alcanzando un valor de 0.6978 en la décima época.

Por otro lado, la precisión se mantiene constante en 0.6685 a lo largo de todo el proceso de entrenamiento. Esto indica que el modelo mejora su capacidad para hacer predicciones a medida que avanza el entrenamiento, pero no son del todo precisas.

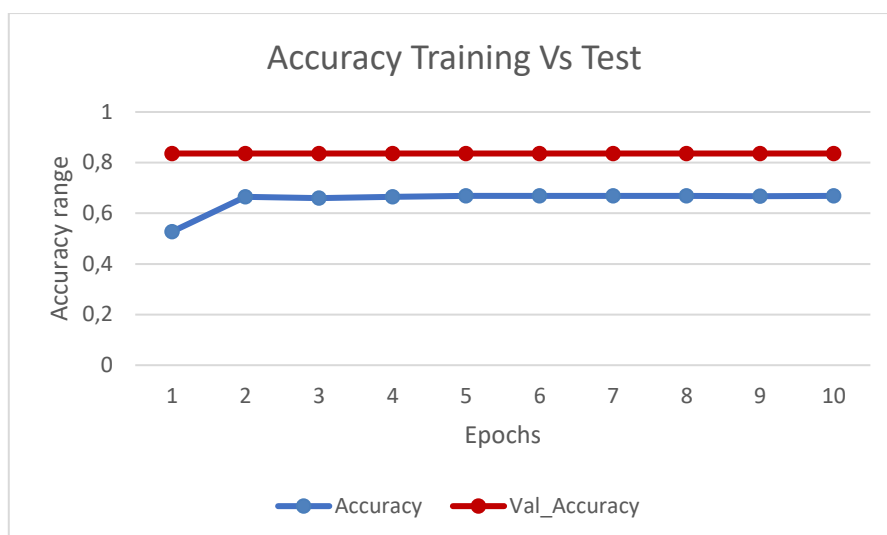


Ilustración 38: Etiqueta única: MTCNN. Entrenamiento y validación de red: precisión. Optimización GridSearchCV.

Por otro lado, la función de pérdida de validación [Ilustración 37] inicia en 1.4601 y disminuye hasta alcanzar 1.2081 en la décima época. En cuanto a la precisión de validación [Ilustración 38], se mantiene constante en 0.836, al igual que la precisión en el conjunto de entrenamiento.

El modelo es capaz de predecir con bastante precisión, pero tiene un nivel más bajo en aciertos. Sin embargo, con un mayor volumen de datos, se observa una reducción en el valor de pérdida, lo que indica que el modelo tiene la capacidad de predecir datos no vistos previamente.

4.2.2. Emociones según VideoID

Al igual que en el apartado "4.2.1. Emociones según MTCNN", a continuación, se presentan los dos métodos de optimización a usar para la determinación de hiperparámetros en las redes neuronales.

Es relevante destacar que este apartado emplea como etiqueta única las identificaciones de cada una de las emociones que más se encuentran presentes en cada uno de los vídeos. Por tanto, la red neuronal recibe solamente una única etiqueta de una emoción de cada conjunto de características.

4.2.2.1. Optimización manual VideoID

La Tabla 7 presenta los resultados del entrenamiento y validación de la BLSTM con la configuración de hiperparámetros realizada de manera manual. En esta configuración, se determina que la cantidad de neuronas es de 64 unidades, la función de activación es de softmax, el optimizador queda en Adam y la función de pérdida es "categorical_crossentropy".

Epoch	Loss	Accuracy	Val_Loss	Val_Accuracy
1	1,6327	0,2686	2,2387	0,00036643
2	1,6014	0,3096	2,4539	0,101
3	1,5803	0,3265	2,4215	0,118
4	1,571	0,3344	2,5385	0,101
5	1,5666	0,3331	2,6181	0,1367
6	1,5616	0,3375	2,5489	0,09
7	1,5582	0,3378	2,5783	0,0999
8	1,5531	0,341	2,5953	0,1233
9	1,549	0,345	2,5839	0,0907
10	1,5457	0,3469	2,5348	0,124

Tabla 7: Etiqueta única: VideoID. Entrenamiento y testeo de red, pérdidas y precisiones. Optimización manual.

Al analizar los resultados de la Tabla 7, se observa que a medida que avanza el número de épocas, la función de pérdida en el conjunto de entrenamiento [Ilustración 39] disminuye gradualmente, lo que indica que el modelo aprende y se adapta a los datos proporcionados.

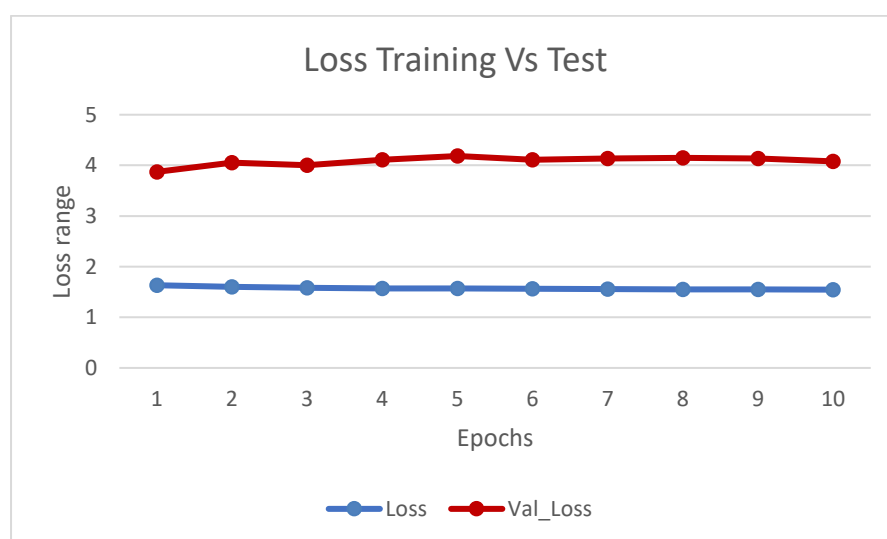


Ilustración 39: Etiqueta única: VideoID. Entrenamiento y validación de red: pérdidas. Optimización manual.

La precisión en el conjunto de entrenamiento también aumenta progresivamente con cada época [Ilustración 40].

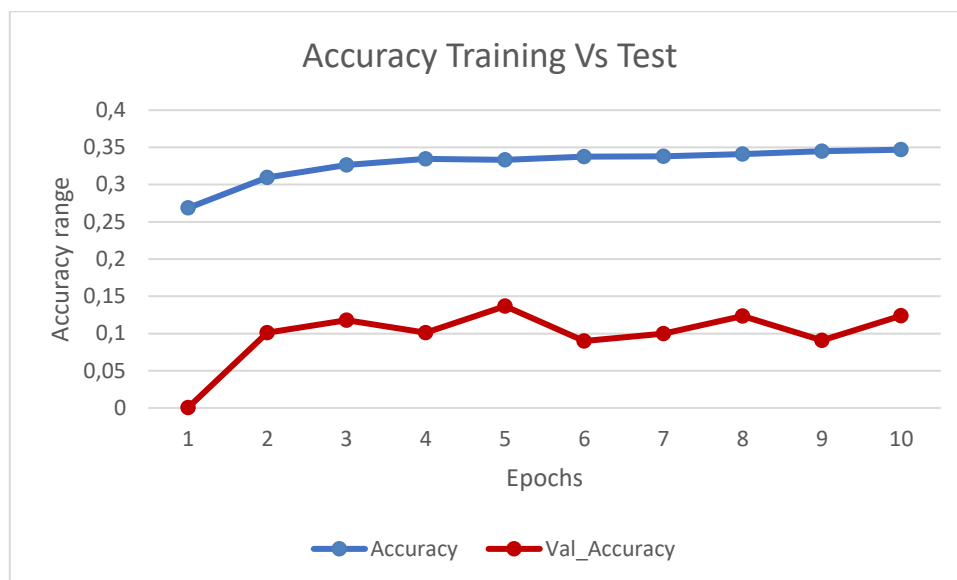


Ilustración 40: Etiqueta única: VideoID. Entrenamiento y validación de red: precisión. Optimización manual.

Sin embargo, al examinar el conjunto de validación de la Tabla 7, se observa que la función de pérdida en el conjunto de validación tiende a aumentar después de las primeras épocas, lo que sugiere que el modelo experimenta un sobreajuste y, consecuentemente, no generaliza adecuadamente para nuevos datos. La precisión en el conjunto de validación muestra fluctuaciones y no alcanza niveles óptimos, lo que indica que el modelo tiene dificultades para clasificar correctamente las emociones.

Por esto, se puede determinar que el modelo muestra progreso en el conjunto de entrenamiento, pero es necesario mejorar la generalización en el conjunto de validación.

4.2.2.2. Optimización automatizada VideoID

En esta sección, los parámetros se configuran mediante el uso de KerasClassification y GridSearchCV como en el apartado "4.2.1.1 Optimización automatizada MTCNN". Los resultados del entrenamiento se presentan en la Tabla 8, la cual muestra las funciones de pérdida y precisiones obtenidas.

Epoch	Loss	Accuracy	Val_Loss	Val_Accuracy
1	0,5743	0,2551	9,6709E-06	0,4104
2	9,6709E-06	0,2553	9,6709E-06	0,4104
3	9,6709E-06	0,2553	9,6709E-06	0,4104
4	9,6709E-06	0,2553	9,6709E-06	0,4104
5	9,6709E-06	0,2553	9,6709E-06	0,4104
6	9,6709E-06	0,2553	9,6709E-06	0,4104
7	9,6709E-06	0,2553	9,6709E-06	0,4104
8	9,6709E-06	0,2553	9,6709E-06	0,4104
9	9,6709E-06	0,2553	9,6709E-06	0,4104
10	9,6709E-06	0,2553	9,6709E-06	0,4104

Tabla 8: Etiqueta única: VideoID. Entrenamiento y testeo de red, pérdidas y precisiones. Optimización GridSearchCV.

Al analizar los resultados, se destaca que las funciones de pérdida [Ilustración 41] se mantienen en valores cercanos a cero, lo que indica que el modelo logra ajustarse adecuadamente a los datos de entrenamiento y realiza predicciones cercanas a los valores reales, en otras palabras, el modelo aprende de manera adecuada los patrones y la estructura de datos de entrenamiento.

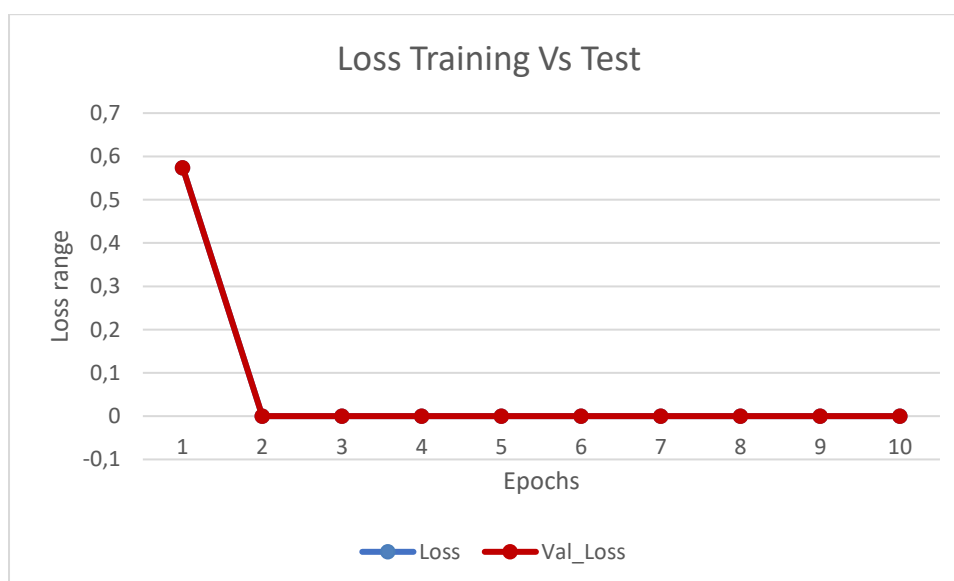


Ilustración 41: Etiqueta única: VideoID. Entrenamiento y validación de red: pérdidas. Optimización GridSearchCV.

En cuanto a la precisión [Ilustración 42], se observa que tanto en el conjunto de entrenamiento como en el de validación, la red neuronal logra mantener una precisión constante a lo largo de todas las épocas.

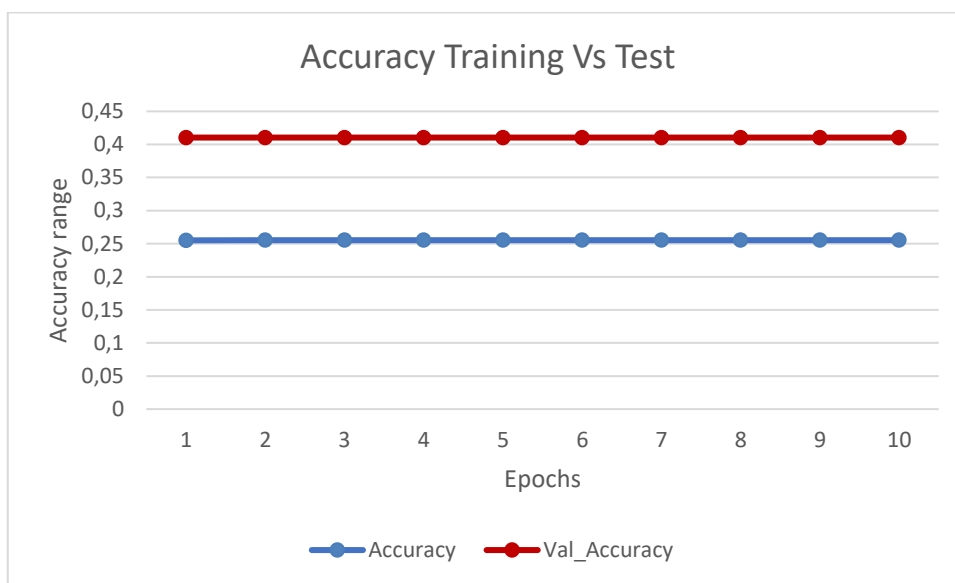


Ilustración 42: Etiqueta única: VideoID. Entrenamiento y validación de red: precisión. Optimización GridSearchCV.

Específicamente, la precisión en el conjunto de entrenamiento es de 0,2553, mientras que en el conjunto de validación es de 0,4104.

No obstante, es importante tener en cuenta que estos valores de precisión son relativamente bajos. Esto puede indicar a que el modelo de la red BLSTM no logra aprender completamente la relación entre las características de entrada y las emociones objetivo.



5. CONCLUSIONES

El estudio realizado utilizando una red neuronal con doble etiqueta arroja resultados alentadores en la disminución de la pérdida y la mejora en la precisión durante el proceso de entrenamiento. La combinación de emociones determinadas tanto por el identificador de vídeo como por MTCNN de imágenes demuestra ser eficaz en este avance. Por otro lado, se identifica que las emociones que más influyen en la actividad electrodermica son el enfado, la sorpresa, la felicidad y el miedo. Estas emociones muestran niveles elevados de sudoración, lo que indica una fuerte asociación con el sistema nervioso simpático.

Durante el análisis de las características extraídas de las señales proporcionadas por los participantes, se observa una marcada variabilidad en el número total de picos de respuesta de conducción de la piel y la frecuencia de picos R-R a lo largo de los vídeos estudiados. Esto sugiere que las respuestas emocionales varían significativamente entre los individuos analizados.

En cuanto a los objetivos, el modelo de red neuronal con doble etiqueta alcanza una precisión constante de 0.8662 en el conjunto de entrenamiento y validación, lo que indica que el modelo es capaz de realizar predicciones con bastante precisión. Sin embargo, se nota un nivel más bajo en aciertos, lo que indica al objeto de mejora. Desde la perspectiva de etiqueta única según VideoID, se observa que la función de pérdida en el conjunto de validación tiende a aumentar después de las primeras épocas. Esto propone que existen dificultades en la generalización del modelo para nuevos datos.

Por otro lado, para mejorar la generalización en conjuntos pequeños con etiquetas únicas, se recomienda explorar la optimización de parámetros o aumentar el conjunto de datos de entrenamiento, puesto que estas medidas pueden permitir al modelo adaptarse de manera más efectiva a datos no vistos de forma previa. Esta es una de las limitaciones que se encontraron en el presente trabajo, puesto que aparte de la presencia de ruido, también aparece el problema de desconexión del dispositivo de la empatica E4 con su API, lo que lleva a la situación de exclusión de un número considerable de usuarios en el estudio.

Respecto a los Objetivos de desarrollo sostenible, en particular al ODS 3.4, relacionado con la Salud y el Bienestar, se vincula este objetivo al trabajo al aplicar técnicas de aprendizaje máquina y reconocimiento de emociones. Con esto se busca mejorar la salud mental y emocional,

permitiendo una comprensión desde otra perspectiva de las emociones humanas. Asimismo, este trabajo contribuye al ODS 9.1, promoviendo la innovación y el desarrollo de infraestructuras tecnológicas resilientes usando tecnologías avanzadas en la resolución práctica de problemas como dispositivos de MsWH, que permite realizar una cobertura amplia de registro de señales que puedan ser usadas para la identificación y estudios de emociones, aparte de usar otros dispositivos como la Muse 2 y la empática E4, que se consideran como nuevas tecnologías en el campo de los dispositivos digitales para el monitoreo fisiológico.

5.1. TRABAJOS FUTUROS

Se identifican diversas áreas de mejora y posibles investigaciones futuras para continuar avanzando en este campo de estudio. Unas de estas mejoras es la generalización del modelo en conjuntos pequeños, para ello, se considera beneficioso explorar otras técnicas de selección de hiperparámetros y ajustarlos adecuadamente mediante herramientas y frameworks como Optuna [49] u otros algoritmos de aprendizaje máquina.

Por otro lado, se puede considerar ampliar y diversificar el conjunto de vídeos empleados en el estudio. Esto implica incluir una mayor variedad de situaciones emocionales y duraciones de los vídeos para abordar la repetición de emociones y permitir un análisis más exhaustivo. También sería de gran importante integrar otras señales contextuales que enriquezcan el análisis de las emociones, puesto que esto proporciona una perspectiva más completa y precisa sobre las respuestas emocionales de los sujetos. Sin embargo, también es de gran interés realizar un estudio que involucre situaciones emocionales reales en lugar de vídeos pre grabados. Esto proporcionaría una comprensión más profunda y realista de cómo las señales fisiológicas se relacionan con las emociones en el día a día de las personas.

Finalmente, se propone llevar a cabo una comparativa e investigación exhaustiva de otras arquitecturas de redes neuronales con el fin de evaluar su desempeño y posibles mejoras en las predicciones.

6. ABREVIATURAS

AFEW	Database de Expresiones Faciales Actuadas en la Naturaleza
AIC	Criterio de información de Akaike
API	Interfaz de programación de aplicaciones
AUC	Área bajo la curva
BLSTM	Bidirectional Long Short-Term Memory
BRNN	Red Neuronal Recurrente Bidireccional
CNN	Red Neuronal Convolutacional
db4	Daubechies-4
DBN	Deep Belief Network
DEAP	Marco de computación evolutiva para crear prototipos y testeo
DF	DataFrame
DWT	Transformada de Wavelet
ECG/EKG	Electrocardiograma
EDA	Actividad Electrodermica
EDR	Respuesta Electrodermica
ELCM	Error Logarítmico Cuadrático Medio
ELU	Función de activación, curva logarítmica que define valores negativos
ET	Error Total
EUPLA	Escuela Universitaria Politécnica de La Almunia
FER	Reconocimiento Facial de Emociones

FFT	Transformada Rápida de Fourier
FIR	Filtro de Respuesta de Impulso Finito
Fps	Frames per seconds, fotogramas por segundo
FrFT	Fourier de la Transformada Fraccional
GDA	Análisis Discriminante Generalizado
GridSearchCV	Búsqueda en rejilla (técnica de búsqueda de hiperparámetros en aprendizaje automático)
GSP	Potencial Galvánico de la piel
GSR	Respuesta Galvánica de la piel
HF	Alta Frecuencia
HOG	Histograma de Gradientes Orientados
HR	Heart Rate, Frecuencia Cardíaca
HRV	Variabilidad de Frecuencia Cardíaca
IA / AI	Inteligencia Artificial
IBI	Intervalo de Tiempo entre latidos
IDE	Entorno de Desarrollo Integrado
IIR	Filtro de Respuesta de Impulso Infinito
IMU	Inertial Measurement Unit, Unidad de Medición Inercial
KNN	Algoritmo de Vecinos Más Cercanos
LDA	Análisis Discriminante Lineal
LF	Baja Frecuencia
LOSO	Leave-One-Subject-Out de validación cruzada
LPM	Latidos Por Minuto
LSL	Lab Streaming Layer

LSTM	Long Short-Term Memory
MAE	Error Absoluto Medio
MLDP	Patrón Direccional Local Modificado
MLP	Perceptrón Multicapa
MSCNN	Red Neuronal Convolutiva Multiescalar
MSE	Mean Squared Error, Error Cuadrático Medio
MsWH	Multi-sensor Wearable Headband
MTCNN	Red Neuronal Convolutiva en Cascada Multitarea
NLMS	Mínimos Cuadrados Medios Normalizados
NLP	Procesamiento de Lenguaje Natural
NS.SCR	SCR no específicos
ODS	Objetivos de Desarrollo Sostenible
Optuna	Biblioteca de optimización de hiperparámetros
PCA	Análisis de Componentes Principales
PGR	Reflejo Psicogalvánico
PPG	Fotopleletismografía, técnica para medir cambios en el volumen sanguíneo
PSD	Densidad Espectral de Potencia
RAM	Random Access Memory, Memoria de Acceso Aleatorio
RBM	Máquinas de Boltzmann Restringidas
RLS	Mínimos Cuadrados Recursivos
RNA	Red Neuronal Artificial
RNN	Red Neuronal Recurrente
ROI	Región de interés

RR	Distancia entre dos ondas R(pico)
SAM	Self-Assesment Manikin, Maniquí de autoevaluación
SCL	Nivel de Conductancia de la piel
SCR	Respuesta de conductancia de la Piel
SD HR	Desviación Estándar de la Frecuencia Cardíaca
SDL	Cambios tónicos
SELU	Función de activación, empleada para redes neuronales que están compuestas por capas alta densidad
SIFT	Transformada de Características Invariantes a Escala
SNP	Sistema Nervioso Parasimpático
SNS	Sistema Nervioso Simpático
SpO2	Saturación de oxígeno en la sangre
SSR	Respuesta Simpática de la piel
ST	Temperatura de la piel
SVC	Clasificación de Vectores de Soporte
SVD	Descomposición en Valores Singulares
SVM	Máquina de Vectores de Soporte
TERMA	Dos medias móviles relacionadas con eventos
TFG	Trabajo de Fin de Grado
TFM	Trabajo de Fin de Máster

7. BIBLIOGRAFÍA

- [1] Ekman P. (2004). Emotions revealed recognizing faces and feelings to improve communication and emotional life (1. ed.). Owl Books.
- [2] Ekman P. & Friesen W. V. (2003). Unmasking the face. Malor Books.
- [3] Konami Digital Entertainment. (2009). Silent Hill: Shattered Memories [Video game]. PlayStation 2, Wii, PlayStation Portable.
- [4] Boucsein W. (2012). Electrodermal activity (2nd ed.). Springer Science Business Media LLC.
- [5] Pouyan, M. B., Rani, P., & Yang, H. (2020). EDA con respuestas de conductancia cutánea detectadas (SCR - puntos negros) [Figura descriptiva]. PLOS ONE
- [6] Martini F. Bartholomew E. F. & Hutchings R. T. (2003). Essentials of anatomy & physiology (3rd ed.). Pearson Education.
- [7] Carlson N. R. (2013). Physiology of behavior (11. ed. internat.). Pearson.
- [8] Farnsworth, B. (2018). What is GSR (galvanic skin response) and how does it work?. GSR (Galvanic Skin Response): The Complete Pocket Guide. iMotions.
- [9] Grimnes, S., Jabbari, A., Martinsen, Ø. G., & Tronstad, C. (2011). Electrodermal activity by DC potential and AC conductance measured simultaneously at the same skin site. Skin Research and Technology, 17(1), 26-34.
- [10] Boucsein W. (2012). Electrodermal activity (2nd ed.). Springer Science Business Media LLC. ISBN 978-1-4614-1125-3
- [11] Greco A. & Valenza G. & Pasquale Scilingo E.(auth.) (2016). Advances in Electrodermal Activity Processing with Applications for Mental Health: From Heuristic Methods to Convex Optimization (2nd ed.). Springer.
- [12] Cacioppo J. T. Tassinary L. G. & Berntson G. G. (2007). Handbook of psychophysiology (3rd ed.). Cambridge University Press.

- [13] Graham B. & McGowan K. (2011). Mind performance projects for the evil genius: 19 brain-bending bio hacks. McGraw Hill Professional.
- [14] Prabal K. Chattopadhyay, Alyson J. Bond, Malcolm H. Lader (1975). Characteristics of galvanic skin response in anxiety states. Journal of Psychiatric Research, Volume 12, Issue 4. ISSN 0022-3956
- [15] Matté James Allan. (2000). Examination and cross-examination of experts in forensic psychophysiology using the polygraph. J.A.M. Publications.
- [16] Kamath M. V. Watanabe M. & Upton A. (2016). Heart rate variability (HRV) signal analysis clinical applications. CRC Press.
- [17] Annunzio G. (2017) Masterarbeit. Die HRV-Analyse als Präventionsinstrument gegen Stress der Mitarbeitenden im Unternehmen. Zürcher Hochschule für Angewandte Wissenschaften
- [18] Nai't-Ali Amine. (2009). Advanced biosignal processing. Springer Berlin Heidelberg.
- [19] Liang H. Bronzino J. D. & Peterson D. R. (2013). Biosignal processing: principles and practices. CRC Press/Taylor & Francis.
- [20] Physiolab Guide to Convert IBI to Kubios HRV Analysis Instructions Manual
- [21] "Вариабельность сердечного ритма." (s.f.). En WikiBrief.
- [22] Picard R. W. (2000). Affective computing (1. paperback). MIT Press.
- [23] Maglogiannis I. (2007). Emerging artificial intelligence applications in computer engineering real world ai systems with applications in ehealth hci information retrieval and pervasive technologies. IOS.
- [24] Giron-Sierra J. M. & Springer Malaysia Representative Office. (2017). Digital signal processing with matlab examples volume 1 signals and data filtering non-stationary signals modulation (1st ed. 2017). Springer Singapore.
- [25] Datta A. K. Datta M. & Banerjee P. K. (2016). Face detection and recognition theory and practice (Online-Ausg). CRC Press.
- [26] Google Cloud Vision API. (2021). AutoML.
- [27] Tobii Pro. (2021). Tobii Pro Sticky [Software de seguimiento ocular].

- [28] Noldus Information Technology. (2021). FaceReader [Software de análisis de expresiones faciales].
- [29] Pérez Estruch, C. (2016). Emotions Recognition using Deep Learning. <http://hdl.handle.net/10251/77873>
- [30] Navas Ten, R. (2020). Metodologías de Reconocimiento facial en modelos bidimensionales (Bachelor's thesis).
- [31] Uddin, M. Z., Hassan, M. M., Almogren, A., Zuair, M., Fortino, G., & Torresen, J. (2017). A facial expression recognition system using robust face features from depth videos and deep learning. *Computers & Electrical Engineering*, 63, 114-125.
- [32] Vidal, J. H. (2019). Aprendizaje automático en el diseño de un detector de estrés a partir de señales biomédicas. Universidad de La Laguna.
- [33] Liu, Y., Gedeon, T., Caldwell, S., Lin, S., & Jin, Z. (2020). Emotion Recognition Through Observer's Physiological Signals. arXiv preprint arXiv:2002.08034.
- [34] Ganapathy, N., Veeranki, Y. R., Kumar, H., & Swaminathan, R. (2021). Emotion recognition using electrodermal activity signals and multiscale deep convolutional neural network. *Journal of Medical Systems*, 45, 1-10.
- [35] M. Vollmer, "A robust, simple and reliable measure of heart rate variability using relative RR intervals," 2015 Computing in Cardiology Conference (CinC), Nice, France, 2015, pp. 609-612, doi: 10.1109/CIC.2015.7410984.
- [36] Н.Н. Д. & Danilova N. N. (2012). Психофизиология: учебник. Аспект Пресс.
- [37] Arquilla, K., Webb, A. K., & Anderson, A. P. (2022). Utility of the Full ECG Waveform for Stress Classification. *Sensors*, 22(18), 7034.
- [38] Aziz, S., Ahmed, S., & Alouini, M. S. (2021). ECG-based machine-learning algorithms for heartbeat classification. *Scientific reports*, 11(1), 18738.
- [39] Asiain, D., Ponce de León, J., & Beltrán, J. R. (2022). MsWH: A Multi-Sensory Hardware Platform for Capturing and Analyzing Physiological

Emotional Signals. Sensors, 22(15), 5775.
<https://doi.org/10.3390/s22155775>

[40] Muse 2. <https://choosemuse.com/>

[41] Empatica E4 wristband. <https://www.empatica.com/en-eu/research/e4/>

[42] Uctronics. (n.d.). OV2640 Datasheet. Retrieved from https://www.uctronics.com/download/cam_module/OV2640DS.pdf

[43] App-LabRecorder [Repositorio de GitHub]. Recuperado de <https://github.com/labstreaminglayer/App-LabRecorder>

[44] Extensible Data Format (XDF). (s.f.). xdf [Repositorio de GitHub]. Recuperado de <https://github.com/sccn/xdx>

[45] LabStreamingLayer. [Repositorio de GitHub]. Recuperado de <https://github.com/sccn/labstreaminglayer>

[46] Abadi, M. K., Subramanian, R., Kia, S. M., Avesani, P., Patras, I., & Sebe, N. (2015). DECAF: MEG-Based Multimodal Database for Decoding Affective Physiological Responses. *IEEE Transactions on Affective Computing*, 6(3), 209-222. <https://doi.org/10.1109/TAFFC.2015.2392932>.

[47] ResearchGate. Neural Network Based Emotion Estimation Using Heart Rate Variability and Skin Resistance [Figura científica de SAM].

[48] JetBrains. (s.f.). PyCharm.

[49] Optuna. (s.f.). Optuna: A hyperparameter optimization framework. Recuperado el 8 de julio de 2023, de <https://optuna.org/>

Relación de documentos

(X) Memoria	109	páginas
(_) Anexos	24	páginas

La Almunia, a 05 de 09 de 2023



Firmado por IEMBERGENOVA
ILMIRA - ****0426* el día
02/09/2023 con un
certificado emitido por
AC FNMT Usuarios

Firmado: Ilmira Iembergenova