



Universidad Politécnica  
de Madrid

**Escuela Técnica Superior de  
Ingenieros Informáticos**



Grado en Ciencia de datos e inteligencia artificial

Trabajo Fin de Grado

**Análisis de la Confianza mediante  
Redes Neuronales Convolucionales  
sobre Señales de EEG con Diferentes  
Distribuciones de Clase**

Autor: Cristina Fernández Gómez

Tutor(a): Laura Melgar García, Javier Bajo Pérez

Madrid, Enero 2026

Este Trabajo Fin de Grado se ha depositado en la ETSI Informáticos de la Universidad Politécnica de Madrid para su defensa.

*Trabajo Fin de Grado*

*Grado en Ciencia de datos e inteligencia artificial*

*Título:* Análisis de la Confianza mediante Redes Neuronales Convolucionales sobre Señales de EEG con Diferentes Distribuciones de Clase

Enero 2026

*Autor:* Cristina Fernández Gómez

*Tutor:* Laura Melgar García, Javier Bajo Pérez

Departamento de Inteligencia Artificial

Escuela Técnica Superior de Ingenieros Informáticos

Universidad Politécnica de Madrid

# Agradecimientos

Lo primero, a mi hermano Alejandro, por estar ahí siempre para mí, apoyándome cuando más perdida estaba y para animarme a hacer lo que más me apasionaba.

A mis padres, sin ellos no habría llegado donde estoy ni podría haber hecho tantas cosas estos últimos años, sin vosotros habría sido imposible.

A mi mejor amiga Claudia, la que más me ha escuchado y la persona que mejor me conoce, con quien llevo más de 16 años. Te quiero amiga.

A mis primos, Lucas y Paloma, que son como hermanos para mí y con quienes he compartido absolutamente todo.

A Álvaro, con quien más he compartido muchos momentos difíciles en la universidad y quien más conseguía hacerme reír a pesar de la situación.

A Irene, que siempre me ha ayudado a ver el lado positivo aunque ella estuviese viéndolo incluso peor que yo, eres la persona más feliz y optimista que conozco.

A David, aunque desde la distancia, es una de las personas que más ha estado ahí desde que le conozco hace pocos años, y uno de los mejores amigos que puedo tener.

A los amigos que he hecho durante esta etapa en la universidad –Adonis, Luis, Fátima, Álex...–, gracias por hacerme las clases más amenas y por conseguir que incluso durante los trabajos interminables sacáramos algunas risas.

A mis tutores Laura y Javier, por guiarme durante este largo proceso y mostrarme más del mundo de la neurociencia.



# Resumen

El uso creciente de sistemas automatizados en procesos de toma de decisiones hace necesaria la evaluación de la confianza en este tipo de sistemas. Medir la confianza de manera objetiva supone un reto debido a su naturaleza subjetivo y cognitiva. En este contexto, las señales EEG (electroencefalografías) representan una buena opción para estimar los niveles de confianza en escenarios de interacción humano-máquina.

Este Trabajo de Fin de Grado aborda el desafío de predecir el estado cognitivo de confianza de sistemas automatizados mediante el análisis de señales electroencefalográficas. Haciendo uso de un conjunto de participantes que han realizado el test de Stroop, se ha llevado a cabo un análisis de las señales en crudo, analizando distintas arquitecturas neuronales especializadas en el procesamiento de señales EEG, como EEGNet, EEGNeX y modelos basados en mecanismos de atención, evaluadas de manera dependiente del sujeto. Este enfoque permite estudiar en detalle la elevada variabilidad intersujeto propia de estas señales y su impacto en el rendimiento de los modelos.

Uno de los principales retos abordados es el desbalance de clase presente en algunos participantes. Para mitigar el sesgo hacia la clase mayoritaria, se han implementado y comparado rigurosamente estrategias de balanceo a nivel de datos, como el oversampling aleatorio y SMOTE, así como adaptaciones a nivel de algoritmo mediante el uso de la función de pérdida Focal Loss y la asignación de pesos ponderados. La evaluación del rendimiento se ha llevado a cabo mediante métricas como el f1-score, recall o ROC-AUC, seleccionadas en base al grado de desbalance de los participantes. Asimismo, se emplean matrices de confusión y otro tipo de gráficas para analizar el efecto de estas técnicas sobre el criterio de decisión de los modelos.

Finalmente, este trabajo pone de manifiesto las limitaciones y oportunidades del uso exclusivo de señales EEG para la estimación de la confianza en sistemas automatizados. Los resultados obtenidos permiten comprender mejor el impacto de la variabilidad intersujeto y de la calidad de la señal en los modelos de deep learning aplicados, así como identificar los principales factores que condicionan su desempeño en escenarios de interacción humano-máquina.

## Capítulo 0. Resumen

---

**Palabras clave:** EEG en crudo, variabilidad intersujeto, confianza, deep learning, desbalance de clases, interacción humano-máquina.

# Abstract

The growing use of automated systems in decision-making processes makes it necessary to evaluate trust in this type of system. Measuring trust objectively is challenging due to its subjective and cognitive nature. In this context, EEG (electroencephalogram) signals are a good option for estimating levels of trust in human-machine interaction scenarios.

This Final Degree Project addresses the challenge of predicting the cognitive state of trust in automated systems by analysing electroencephalographic signals. Using a group of participants who have taken the Stroop test, an analysis of the raw signals has been carried out, analysing different neural architectures specialised in EEG signal processing, such as EEGNet, EEGNeX and models based on attention mechanisms, evaluated in a subject-dependent manner. This approach allows for a detailed study of the high inter-subject variability inherent in these signals and its impact on model performance.

One of the main challenges addressed is the class imbalance present in some participants. To mitigate the bias towards the majority class, balancing strategies at the data level, such as random oversampling and SMOTE, have been rigorously implemented and compared, as well as adaptations at the algorithm level through the use of the Focal Loss function and the assignment of weighted weights. Performance evaluation has been carried out using metrics such as f1-score, recall, or ROC-AUC, selected based on the degree of imbalance among participants. Likewise, confusion matrices and other types of graphs are used to analyse the effect of these techniques on the decision criteria of the models.

Finally, this work highlights the limitations and opportunities of using EEG signals exclusively to estimate confidence in automated systems. The results obtained provide a better understanding of the impact of inter-subject variability and signal quality on the deep learning models applied, as well as identifying the main factors that condition their performance in human-machine interaction scenarios.

**Key words:** raw EEG, inersubject variability, trust in automation, deep learning, class imbalance, human-machine interaction.



# Tabla de contenidos

<b>Agradecimientos</b>	<b>i</b>
<b>Resumen</b>	<b>iii</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Descripción general del trabajo . . . . .	2
1.2. Lista de tareas . . . . .	3
<b>2. Estado del arte</b>	<b>5</b>
2.1. Herramientas y tecnología . . . . .	5
2.2. Machine Learning . . . . .	6
2.2.1. Aprendizaje supervisado . . . . .	6
2.2.2. Aprendizaje no supervisado . . . . .	7
2.2.3. Aprendizaje por refuerzo . . . . .	7
2.3. Deep Learning . . . . .	8
<b>3. Desarrollo</b>	<b>11</b>
3.1. Experimento y descripción de los datos . . . . .	11
3.1.1. Etapa preexperimental . . . . .	11
3.1.2. Etapa de prueba . . . . .	12
3.1.3. Etapa de automatización . . . . .	12
3.1.4. Descripción de los datos . . . . .	12
3.2. Implementación . . . . .	13
3.2.1. Preprocesado . . . . .	13
Remuestreo . . . . .	13
Segmentación en épocas . . . . .	13
3.2.2. Estrategia de partición de datos: Método Trial-Wise . . . . .	14
3.2.3. Técnicas Utilizadas . . . . .	14
Focal Loss . . . . .	15
Oversampling . . . . .	15
Normalización . . . . .	16
Dropout . . . . .	16
Ponderación de clases . . . . .	16
Early Stopping . . . . .	17
Selección de características . . . . .	17
3.2.4. Arquitecturas de las redes neuronales . . . . .	17

## TABLA DE CONTENIDOS

---

Red EEGNet . . . . .	17
Red EEGNeX . . . . .	19
AttentionBaseNet . . . . .	20
3.3. Selección de métricas según la distribución de los datos . . . . .	22
<b>4. Resultados</b>	<b>23</b>
4.1. Modelos base . . . . .	24
4.1.1. Participantes balanceados . . . . .	25
Evolución del F1-Score en la etapa de entrenamiento . . . . .	26
4.1.2. Participantes desbalanceados . . . . .	28
Evolución del f1-score en la etapa de entrenamiento . . . . .	29
4.2. Oversampling y Focal Loss . . . . .	31
4.2.1. Impacto en la arquitectura EEGNet . . . . .	32
4.2.2. Impacto en la arquitectura EEGNeX . . . . .	34
4.2.3. Impacto en la arquitectura AttentionBaseNet . . . . .	37
4.3. Pesos Ponderados . . . . .	40
4.3.1. Participantes balanceados . . . . .	40
Arquitectura EEGNet . . . . .	41
Arquitectura EEGNeX . . . . .	44
Arquitectura AttentionBaseNet . . . . .	46
4.3.2. Participantes desbalanceados . . . . .	49
Impacto en la red EEGNet . . . . .	49
Impacto en la red EEGNeX . . . . .	51
Impacto en la red AttentionBaseNet . . . . .	52
4.4. Selección de características . . . . .	54
4.4.1. Impacto en la arquitectura EEGNet . . . . .	54
Participantes balanceados . . . . .	54
Participantes desbalanceados . . . . .	56
4.4.2. Impacto en la arquitectura EEGNeX . . . . .	58
Participantes balanceados . . . . .	58
Participantes desbalanceados . . . . .	60
4.4.3. Impacto en la arquitectura AttentionBaseNet . . . . .	61
Participantes balanceados . . . . .	61
Participantes desbalanceados . . . . .	63
4.5. Resultados finales . . . . .	66
4.6. Análisis espectral . . . . .	68
4.7. Comparación con enfoques multimodales basados en Machine Learning . . . . .	69
<b>5. Experimento</b>	<b>71</b>
5.1. Equipamiento . . . . .	71
5.1.1. Casco EEG . . . . .	71
5.1.2. Software y experimento . . . . .	72
<b>6. Conclusiones y trabajo futuro</b>	<b>75</b>
<b>7. Análisis de impacto</b>	<b>77</b>
7.1. Impacto personal . . . . .	77

7.2. Impacto social . . . . .	78
7.3. Impacto económico . . . . .	78
7.4. Impacto medioambiental . . . . .	79
7.5. Impacto cultural . . . . .	79
7.6. Objetivos de desarrollo sostenible . . . . .	79
7.6.1. ODS 3: Salud y bienestar . . . . .	80
7.6.2. ODS 8: Trabajo decente y crecimiento económico . . . . .	80
7.6.3. ODS 9: Industria, innovación e infraestructura . . . . .	80
<b>Bibliografía</b>	<b>83</b>
<b>Anexos</b>	<b>89</b>
<b>A. Anexo</b>	<b>89</b>



# Capítulo 1

## Introducción

La inteligencia artificial ya se ha establecido en nuestra sociedad. En muy poco tiempo, la capacidad de estos programas informáticos ha aumentado de forma exponencial. Ya no solo los utilizamos para hacer consultas sencillas, sino que cada vez están más presentes en muchos trabajos en los que su colaboración es necesaria. Por tanto, no es que nosotros simplemente les hagamos preguntas, ahora trabajamos junto a ellas y les cedemos ciertas tareas, desde tareas administrativas rutinarias a llegar a salvar vidas en los hospitales. Para que la interacción humano-máquina sea lo más efectiva posible, hay que tener en cuenta un factor esencial: la confianza.

Cuando trabajamos en equipo con otras personas, depositamos nuestra confianza en ellas para que no tengamos que preocuparnos por su trabajo y por el nuestro al mismo tiempo, lo que nos libera una gran carga. Lo mismo sucede cuando se trabaja con estas máquinas. Para que la colaboración entre ambos sea satisfactoria, las personas tenemos que ganar confianza en ellas, porque es inevitable observar cómo están cada vez más presentes en nuestra vida diaria.

Podemos decir que confiaremos en una máquina si actúa como esperamos, es decir, si sabemos qué acciones va a llevar a cabo, cómo las va a hacer y cuál es la meta que va a alcanzar. Anticipamos el comportamiento deseado de la inteligencia artificial y confiamos en que la máquina hará lo que queremos, dejando una pequeña parte expuesta al riesgo. La confianza se puede definir de la siguiente manera: «El acto en el que uno se expone voluntariamente a externalidades positivas y negativas provocadas por las acciones de los demás»[1].

Si una persona confía en la máquina a la que le cede parte de su trabajo, podrá tomar mejores decisiones. Para ello, es necesario que pueda entender cómo realiza su labor, por qué lleva a cabo ciertas acciones y cómo piensa. Para lograr una buena coordinación entre una persona y una máquina, la confianza es imprescindible [2].

## Capítulo 1. Introducción

---

En este contexto, poder medir este nivel de confianza de las personas hacia las máquinas nos permite saber con mayor exactitud hasta qué punto estamos dispuestos a aceptar decisiones automáticas. Podemos diferenciar dos tipos de confianza:

- **Intrínseca:** es estar de acuerdo en el funcionamiento interno del modelo, en cómo razona y toma sus decisiones.
- **Extrínseca:** la confianza basada en su rendimiento general, es decir, lo que el usuario observa: el desempeño, aunque no comprenda cómo funciona por dentro.

Los impulsos cerebrales nos pueden ayudar a distinguir estas confianzas, de manera que sepamos cuándo está basada en la calidad del sistema o en expectativas, lo que puede ayudar a mejorar el sistema. Si somos capaces de identificar qué componentes internos o externos afectan a la confianza, se podrán diseñar sistemas más adaptables, transparentes y seguros, de forma que las personas se sientan cómodas utilizándolos [3].

Como han demostrado varios estudios, el procesamiento de la información en el cerebro durante la toma de decisiones basadas en la confianza implica la activación diferencial de distintas áreas cerebrales [4]. Gracias a la información que nos ofrecen los datos EEG (electroencefalogramas) podemos medir el nivel de confianza.

En este estudio, se utilizará un conjunto de datos proporcionado por un equipo de investigación de la Universidad Politécnica de Madrid para analizar, mediante aprendizaje profundo, la confianza de las personas en las máquinas a partir de sus datos EEG obtenidos durante su interacción con una máquina.

### 1.1. Descripción general del trabajo

El objetivo de este trabajo es implementar una red neuronal a la que aplicar datos del tipo EEG para medir la confianza de una persona en una máquina. Para ello, se utilizarán una o varias redes con una arquitectura específica para tratar con este tipo de datos. El trabajo se dividirá en dos partes: creación de la red a partir de datos ya recogidos por el equipo de investigación, y la adquisición de mis propios datos con el casco EEG. Los objetivos para alcanzar son:

- Implementación de redes neuronales para el estudio de la confianza
- Comparación intersujeto e intrasujeto
- Evaluación de las redes neuronales
- Toma y estudio de datos reales

### 1.2. Lista de tareas

Las tareas que se llevarán a cabo para la implementación de la arquitectura de la red son las siguientes:

- Exploración y preprocesado de los datos recibidos por el equipo de investigación
- Diseño de la(s) red(es) neuronal(es)
- Comparación intersujeto e intrasujeto
- Evaluación de su rendimiento
- Comparación entre los distintos modelos
- Selección de la mejor arquitectura
- Toma de mis propios datos
- Conclusiones y posibles aplicaciones futuras



## Capítulo 2

# Estado del arte

En esta sección se ha llevado a cabo un estudio sobre distintas investigaciones y metodologías más relevantes en el ámbito del análisis de las señales electroencefalográficas (EEG). El estudio abarca la evolución de la inteligencia artificial aplicada a la neurociencia, estructurándose en torno a: aprendizaje supervisado, aprendizaje no supervisado, aprendizaje por refuerzo y deep learning.

### 2.1. Herramientas y tecnología

Para adquirir las señales EEG, la mayoría de los experimentos han utilizado un casco inalámbrico como el EEG16 canales de BitBrain [5], que cuenta con 16 electrodos. Se coloca en el cuero cabelludo de los participantes y, conectado a un ordenador, se guardan las señales eléctricas emitidas por el cerebro durante la interacción del participante con la actividad en cuestión. Dependiendo del tipo de investigación, el número de electrodos y su ubicación pueden variar.

Para probar la confianza de las personas en las máquinas, en las tareas experimentales se han utilizado distintos métodos interactivos: desde videojuegos como OverCooked AI o el juego del laberinto, hasta juegos inmersivos en 3D e interacciones con robots reales o coches autónomos.

Para el preprocesado de las señales es habitual el uso de MATLAB, cuyas herramientas (como EEGLAB) facilitan mucho las tareas de filtrado de ruido o la segmentación de épocas. Para el entrenamiento de modelos de machine learning o Deep learning, se ha hecho uso de entornos como Python, donde podemos encontrar librerías muy útiles como Pytorch, Tensorflow o scikit-learn.

## 2.2. Machine Learning

### 2.2.1. Aprendizaje supervisado

Dado que la inteligencia artificial está cada vez más presente en diversas áreas y la confianza es clave para su correcto funcionamiento, diversos estudios han abordado este tema desde distintas áreas, como por ejemplo el aprendizaje supervisado.

En este estudio [6] enfatizan en la importancia de estudiar esta confianza durante la interacción humano-máquina, es decir, en tiempo real, y no solo mediante encuestas posteriores que solo ofrecen una pequeña parte de lo que fue esa interacción. Para ello, diseñaron un juego colaborativo, dividido en 7 secciones, en el que el robot debía representar una palabra seleccionada y el jugador debía adivinarla. El robot comienza actuando correctamente y hablando de forma natural, pero en una de las fases comete errores deliberados para así comprobar cómo afecta a la confianza del participante. Para llevar a cabo la clasificación de la confianza, plantearon el problema como una clasificación binaria: High trust y Low Trust. Para la extracción de características, se centraron en la banda alfa del cerebro, la que está asociada a estados de relajación. Las señales EEG emitidas por 14 sensores las recogieron utilizando un casco EEG, y, para cada participante, calcularon la media de esos valores. A partir de ventanas de 128Hz, calcularon la media, pico máximo, mediana, desviación estándar y curtosis, eliminando finalmente la mediana. A la hora de etiquetar los datos, utilizaron el cuestionario Multi-Dimensional Measure of Trust (MDMT) [7], el cual es utilizado como base para medir la confianza, usando el promedio de las respuestas como umbral para poder clasificarla. Este cuestionario lo rellenaban al final de cada sección del experimento, cada respuesta se daba a una escala de 0 a 7. Una vez etiquetados los datos, utilizaron tres modelos de machine learning para el entrenamiento: SVM (Support Vector Machine), k-Nearest Neighbors (kNN) y Random Forest. Obtuvieron los mejores resultados con el Random Forest, logrando un accuracy del 72.64%. Con ello, lograron demostrar que, efectivamente, con las señales EEG se puede estimar el nivel de confianza. Sin embargo, la mayor limitación fue la falta de fluidez en el robot, como se comunicaba de forma lenta y muy natural, eso hizo que el participante confiara muy fácilmente en que sus acciones eran las correctas. Para futuros experimentos proponen que el robot lleve a cabo acciones con más riesgo y utilizar otros datos fisiológicos además de las señales EEG.

En [8] los participantes juegan a un juego inmersivo en 3D que simula la experiencia de ir dentro de un vehículo autónomo. Durante el trayecto, el vehículo se enfrenta a una serie de obstáculos de diferente riesgo y el participante puede decidir si no interviene porque confía en la IA o si lo hace. Durante el experimento, se recopilan datos de EEG, EDA (respuesta galvánica de la piel) y EMG (electromiografía, que mide la actividad eléctrica de los músculos). Para clasificar la confianza, entrenaron distintos modelos: perceptrón multicapa (MLP), máquina de soporte vectorial (SVM), análisis discriminante, Naive Bayes gaussiano, análisis discriminante cuadrático y SVM regularizado. A todos estos modelos

les aplicaron diferentes métodos de selección de características: filtro, wrapped e híbrido. Se obtuvieron resultados más óptimos con el modelo híbrido. Este enfoque permitió detectar qué bandas EEG están más correlacionadas con la confianza: gamma, alfa y theta. Los modelos híbridos superan a los filtros, con una precisión del 68,1% y una puntuación F1 del 76,6%, lo que sugiere que, en modelos con poca cantidad de datos, la selección de características es casi tan determinante como el diseño del modelo en sí.

### 2.2.2. Aprendizaje no supervisado

Dentro del área de la inteligencia artificial, uno de los métodos más clásicos que hay es el aprendizaje no supervisado, donde no se hace uso de etiquetado.

Aunque en este estudio [9] no se centran específicamente en la confianza, sí que trabajan con señales EEG, tratando de encontrar patrones en estas señales que se pueden aplicar en distintas áreas, como podría ser la propia confianza. Hacen especial énfasis en lo siguiente: a pesar del avance del aprendizaje profundo con redes neuronales, explican que han llevado a cabo este estudio porque el hecho de necesitar muchos ejemplos etiquetados es muy costoso, lo que hace que no sean tan prácticos como los no supervisados. En lugar de utilizar el método de clustering clásico, en el que es necesario introducir una serie de parámetros a mano (incluido el propio número de clusters), proponen un algoritmo adaptativo que se compone de tres partes:

- Preprocesamiento de EEG, usando CEEMDAN (Complete Ensemble Empirical Mode Decomposition with Adaptive Noise) para descomponer la señal y CWT (Continuous Wavelet Transform) para filtrar el ruido.
- Extracción de características multivariadas
- Clustering adaptativo mejorado: usando DBSCAN pero con el algoritmo de búsqueda del gorrión, lo que ajusta automáticamente el número de clusters.

Utilizaron los datos de diez pacientes con epilepsia, de los cuales extrajeron doce características, y los algoritmos mencionados anteriormente para establecer el número de clústeres. Para evaluar los resultados, midieron el rendimiento final mediante un coeficiente de variación, que se calcula con varias métricas internas del algoritmo, y lo compararon con los de otros algoritmos de aprendizaje no supervisado, como KNN, K-Medoids o DBSCAN clásico. Observaron que su algoritmo obtuvo los mejores resultados, con un coeficiente de variación de 0.679.

### 2.2.3. Aprendizaje por refuerzo

El aprendizaje por refuerzo es un método de aprendizaje automático en el que un agente interactúa con el entorno y toma decisiones basándose en la información de retorno recibida tras cada acción. Básicamente, consiste en un ensayo de prueba y error continuo en el que, si el agente realiza una acción correcta,

recibe una recompensa para que la repita [10].

En [11] tampoco miden exactamente la confianza, pero sí utilizan las señales EEG con un propósito muy interesante: un agente aprende a jugar a un juego mientras una persona le observa y mejora a partir de las reacciones cerebrales que debe interpretar para determinar si son de aprobación o no. Es decir, la persona no proporciona ningún tipo de respuesta ni lleva a cabo ninguna acción; el agente solo aprende a partir de las señales EEG que emite su cerebro mientras ve a la máquina jugar. Para ello, desarrollaron tres juegos sencillos: uno en 1D y los otros en 2D. Utilizan lo que se llama error potentials (ErrP), señales cerebrales que se generan cuando una persona percibe un error. Estas señales son las que utilizan como retroalimentación para que el agente aprenda. Tras cada acción, si el agente detecta ErrP, se aplica una penalización negativa. Llevaron a cabo varios enfoques.:

- Naive Approach: El agente recibe feedback tras cada acción.
- Zero-shot learning: Entrena al agente en un juego y luego lo prueba en otro sin volverlo a entrenar.
- Robust Reward Shaping: Recibe feedback de los ErrP en las trayectorias iniciales, no tras cada acción, lo que hace que sea menos costoso.

Antes de entrenar a los agentes, comprobaron que el sistema detectaba correctamente los ErrP a partir de las señales EEG, con una precisión del 73% y del 70% sin ErrP. Para medir el rendimiento de los distintos modelos, utilizaron el AUC como métrica principal. Con el enfoque naïve obtuvieron un AUC de 0.89, 0.83 y 0.77 en los juegos de Maze, Catch y Wabble, respectivamente. En el caso del aprendizaje de cero, obtuvieron un AUC del 0.80 al transmitir lo aprendido del juego Catch al juego Maze. Finalmente, con el Robust Reward Shaping utilizaron otras métricas para medir su rendimiento basadas en la velocidad de convergencia y el número de consultas (errores del usuario). Se logró que el agente aprendiera el doble de rápido que en el caso sin retroalimentación y redujera en un 75% el número de consultas al usuario respecto al enfoque Naive. Por lo tanto, aunque no utilice las señales para medir la confianza específicamente, se pueden analizar y procesar las señales EEG con el aprendizaje por refuerzo, por lo que también se podría aplicar al estudio de la confianza.

### 2.3. Deep Learning

Aunque se han obtenido resultados aceptables con modelos tradicionales de aprendizaje automático, el deep learning (aprendizaje profundo) está cobrando cada vez más protagonismo en este campo. Estos modelos se basan en redes neuronales, que nos permiten detectar patrones complejos en los datos que un modelo tradicional podría pasar por alto [12]. Dado que los datos EEG son complejos y presentan mucho ruido, este tipo de arquitecturas son ideales para su estudio, ya que en este tipo de aprendizaje la selección de características no es tan decisiva y se pueden encontrar relaciones no lineales de forma bastante óp-

tima.

En este estudio [13] recogen los datos EEG durante el proceso de cooperación humano-robot, es decir, en tiempo real, utilizando una representación espacial 3D para codificar relaciones topológicas entre las señales EEG. El experimento se hizo utilizando el juego Overcooked-AI, en el que el jugador trabajará junto con agentes autónomos pre-entrenados para saber controlar a los chefs. Este juego consiste en llevar a cabo una serie de pedidos de comida, los jugadores deberán coordinarse para terminar los platos de la forma más rápida y correcta posible. Los agentes estaban entrenados de maneras diferentes, de manera que, uno tenía mucha habilidad, otro media y el último muy baja. Además de ello, probaron niveles con distintas dificultades para ver cómo iba variando la confianza del jugador. En cada partida, mezclan la habilidad del agente con la dificultad del nivel para observar cómo varía la confianza. Una vez finalizado el nivel, el jugador rellena un cuestionario [14] para obtener la etiqueta de trust o distrust del participante. Tras adquirir los datos EEG de 64 canales, los filtran y segmentan en ventanas de 1 segundo para, finalmente, extraer la Diferencial Entropy (la complejidad de la señal cerebral) por canal y mapear los canales para obtener una topología espacial, obteniendo una imagen cerebral en la que cada píxel representa la actividad cerebral de esa zona del cerebro. Esa imagen la introducen en un Vision transformer, una red especializada en analizar imágenes, de manera que aprenda cómo cambian las distintas zonas del cerebro cuando la persona confía o no en el robot. Compararon su modelo con otros de machine learning: Naive Bayes, KNN y SVM, además de una red convolucional CNN. Llevaron a cabo un análisis intrapersonal, cambiando entre la validación cruzada por ventanas (slice-wise-cross-validation) y la validación por ensayos (trial-wise), separando por completo los ensayos. En el primer caso, obtuvieron una precisión entre un 65 % y 90 %, con un promedio de 75 %, y la CNN logró un 73.79 % de f1-score. En el segundo, una precisión media de 62.00 %, pero la CNN apenas logró reconocer la confianza. Además, observaron que, con la representación espacial, la precisión y el F1 aumentaban considerablemente. Como conclusión, pudieron diferenciar los niveles de confianza de manera más que correcta, pero los resultados variaban bastante entre los diferentes sujetos.

Para el ámbito del sector público, en [15] llevan a cabo la implementación de una red neuronal usando datos EEG y GSR (respuesta galvánica de la piel) para identificar las características que están correlacionadas con la confianza. Al igual que en el estudio anterior, vuelven a recalcar la importancia de tomar los datos durante la interacción humano-robot, y de no utilizar simplemente encuestas después del experimento. La idea parte del trabajo que llevan a cabo en los call-centers, donde los empleados deben confiar en un sistema informático que les sugiere respuestas o decisiones a tomar. Los datos provienen de 45 participantes. Dado que tienen muchas características, se utilizaron técnicas especializadas para cada participante, ya que los datos EEG varían mucho de una persona a otra. Entrenaron dos modelos de redes neuronales con distintas arquitecturas. Consiguieron lograr una precisión máxima de un 75 %, con 10

## Capítulo 2. Estado del arte

---

características seleccionadas. En el estudio pudieron observar que las características de la zona central del cerebro eran las más relevantes para predecir la confianza.

Cuadro 2.1: Resumen de estudios sobre la confianza a partir de señales EEG

Tipo de estudio	Experimento	Algoritmos	Resultados
Aprendizaje Supervisado [6]	Juego cooperativo donde el robot representa una palabra seleccionada y la persona debe adivinarla.	KNN, Random Forest, SVM	Precisión del 72,64 %
Aprendizaje Supervisado [8]	Juego inmersivo 3D donde los participantes van en un coche autónomo.	MLP, SVM, Naive Bayes, Análisis Discriminante	Precisión del 68,1 % y F1-Score del 76,6 % con selección híbrida de características.
Aprendizaje No Supervisado [9]	Señales EEG de pacientes con epilepsia	Clustering adaptativo	Coefficiente de variación de 0.679
Aprendizaje por refuerzo [11]	Tres juegos sencillos 1D y 2D	Naive Approach, Zero-Shot y Robust Reward	Naive Approach: AUC del 0.89, Zero-Shot: AUC del 0.81 y Robust Reward aprendizaje el doble de rápido y reducción del 75 % en consultas.
Deep Learning [13]	Juego Overcooked-AI.	Modelo Vision Transformer	Precisión del 65 al 90 %; mejora con la representación espacial 3D.
Deep Learning [15]	Robot sugiere respuestas en centros de atención al público.	Dos redes neuronales con distintas arquitecturas	Precisión del 75,5 % con selección de características.

## Capítulo 3

# Desarrollo

En este capítulo se describe el proceso de experimentación empleado para la recogida de datos, lo cual incluye las fases del procedimiento y el protocolo seguido con los participantes. A continuación, se aborda la implementación práctica del sistema, que incluye las etapas de preprocesamiento de las señales EEG, las estrategias de partición de los datos y las técnicas aplicadas para la optimización del rendimiento de los modelos. Posteriormente, se exponen las distintas arquitecturas de redes neuronales evaluadas, acompañadas de una explicación detallada de sus características y la justificación de su elección para este experimento. Finalmente, se presenta una justificación sobre las métricas empleadas para la evaluación de los modelos.

### 3.1. Experimento y descripción de los datos

El experimento que se realizó para la recolección de datos de los participantes fue el **test de Stroop**. En este test, se muestran nombres de colores escritos en un color diferente. A modo de ejemplo, puede aparecer la palabra «rojo» escrita en azul. Los participantes deben seleccionar la opción del color de la letra (azul en este caso), no el color que está escrito. La prueba consta de 3 etapas: una preexperimental, una práctica y una de automatización

#### 3.1.1. Etapa preexperimental

En esta etapa, los participantes firman el consentimiento para el procesamiento de sus datos, se les proporciona una explicación detallada sobre el funcionamiento del test y se les colocan los sensores correspondientes. Cuando hayan facilitado todos los datos necesarios, pasarán a realizar los tres cuestionarios de propensión a la confianza.

- Cuestionario Merrit: para comprobar la confianza general hacia las máquinas [16].
- Cuestionario Complacency-Potential Rating Scale (CPRS): mide hasta qué

## Capítulo 3. Desarrollo

---

punto una persona tiende a confiar en exceso en la automatización, reduciendo su supervisión o atención.

- Cuestionario extra para relacionar la confianza en la automatización con la respuesta emocional.

### 3.1.2. Etapa de prueba

La siguiente fase del experimento es la práctica. Los participantes realizan la prueba para familiarizarse con ella sin ninguna automatización. Se guardan los resultados obtenidos y el tiempo que les ha llevado realizar el test, de manera que se pueda obtener un rendimiento base para cada participante sin ningún tipo de asistencia automática.

### 3.1.3. Etapa de automatización

En la tercera y última etapa se introduce la automatización, la cual se encuentra dividida en cuatro bloques. En cada uno, a los participantes se recomienda un color para responder, pero los participantes son libres de elegir la respuesta recomendada u otra que ellos decidan. La fiabilidad es del 100% en los dos primeros bloques, del 50% en el segundo y del 100% en el último.

Tras cada bloque se recopila la misma información que en la etapa anterior, así como las señales EEG, ECG, GSR y las señales del movimiento de los ojos, junto con la marca de tiempo tras cada ensayo. Cada bloque está compuesto por 10 ensayos, y tras cada uno, se les solicita a los participantes que expresen su nivel de confianza en la automatización. Una vez finalizado el bloque, deberán responder a una serie de preguntas sobre su experiencia al completarlo. Además, se les presentan dos preguntas con el objetivo de identificar la emoción que mejor representa su estado.

### 3.1.4. Descripción de los datos

Los datos recogidos en este trabajo consisten en señales EEG en crudo procedentes de 19 electrodos. Estos datos representan la actividad cerebral en el dominio de tiempo antes de cualquier filtrado digital profundo. Cada fila está vinculada a una marca de tiempo (latency) y a la etiqueta de la variable 'Binary Trust': Clase 0 (Low Trust) o Clase 1 (High Trust), dependiendo de la respuesta del usuario. El trust proviene de las respuestas de los usuarios al cuestionario que se les presenta después de cada tarea en el test de Stroop. En el test, las respuestas se presentan en una escala del 1 al 5. Sin embargo, para este estudio, se ha transformado esta variable mediante el cálculo de la media de cada participante. En consecuencia, si el valor de trust de cada instancia de ese participante es menor que la media, se le asigna el valor 'Low\_trust', si es mayor, corresponde a 'High\_trust'.

### 3.2. Implementación

Para realizar el análisis de la confianza en los sujetos participantes, se emplearán exclusivamente señales EEG. Este enfoque posibilita que las arquitecturas de aprendizaje profundo implementadas, tales como EEGNet y EEGNeX, se especialicen en la extracción de características espaciotemporales de la señal EEG.

#### 3.2.1. Preprocesado

##### Remuestreo

Tras la recolección de los datos, se observó que la frecuencia de muestreo de algunos participantes no era constante. Para garantizar que todos tuvieran una frecuencia de 256 Hz (algo esencial para poder llevar a cabo un preprocesado efectivo), se hizo lo siguiente. Se segmentaron las señales en intervalos de un segundo y se clasificaron cada segmento según el número de muestras:

- Segmentos completos: contienen 256 muestras, por lo que no es necesaria ninguna modificación.
- Segmentos vacíos: contienen menos de 52 muestras. Su relleno dependerá de su posición específica dentro de la secuencia. En el caso de que la posición sea inicial, se duplica el siguiente segmento completo. Por otro lado, si es final se duplica el anterior. Finalmente, si es intermedia, cada muestra faltante se calcula como la media aritmética de las correspondientes muestras de los segmentos anterior y posterior.
- Segmentos incompletos: contienen entre 52 y 256 muestras. La interpolación se realiza según la posición de las muestras faltantes: si son las primeras se replica la siguiente muestra original; si son las últimas se replica la anterior; y si son intermedias, se aplica una interpolación lineal entre las muestras adyacentes.

##### Segmentación en épocas

En esta etapa, se divide la señal EEG en intervalos de tiempo discretos asociados a eventos específicos. De este modo, podremos estudiar la evolución temporal de la actividad neuronal inmediatamente antes y después de un evento relevante.

Para identificar los momentos exactos en los que ocurrían estos eventos, se definieron una serie de marcadores temporales. Los eventos incluían:

- La presentación de la palabra.
- La sugerencia de la automatización.
- La respuesta del participante.
- La aparición de la pregunta de autoconfianza.
- La valoración subjetiva aportada por el sujeto.

De esta manera, se establecieron cinco épocas por ensayo. Las épocas pueden tener tanto duraciones fijas como variables, para poder gestionar estas diferencias de manera adecuada se hizo lo siguiente:

- Si dos marcadores se generaban con menos de 0.003 segundos de diferencia, se descartaba el primero y la época se extendía hasta el instante del segundo marcador.
- Si una época contenía menos de 10 muestras, se excluía del análisis.

La información correspondiente a las épocas descartadas no se eliminó, sino que se añadió a la época inmediatamente anterior, con el fin de conservar la mayor cantidad de datos posible.

### 3.2.2. Estrategia de partición de datos: Método Trial-Wise

Para entrenar el modelo, la forma en que se divide la información es casi tan importante como el propio proceso de entrenamiento. En los experimentos con señales EEG, dichas señales se organizan en segmentos denominados «trial», donde cada trial representa una muestra completa de la actividad cerebral registrada durante un evento experimental.

Dado que las redes neuronales no pueden procesar un ensayo completo de una sola vez, las señales se dividen en ventanas temporales (slices). No obstante, todas las ventanas extraídas de un mismo segmento se mantienen siempre juntas dentro del mismo conjunto de datos (entrenamiento, validación o prueba). Esto significa que las ventanas de un mismo ensayo nunca se mezclan entre diferentes conjuntos, lo que evita la fuga de información y garantiza que la evaluación del modelo refleje su verdadera capacidad de generalización ante ensayos completamente nuevos.

Reforzando esta integridad metodológica, el análisis realizado en este trabajo es de carácter intra-sujeto (individual). Esto implica que se entrena, valida y prueba un modelo específico para cada participante con sus propios datos. La segmentación de los datos de acuerdo a esta metodología es de vital importancia, dado que, en caso de que se presenten ventanas de un mismo ensayo en los conjuntos de prueba y de entrenamiento, el modelo aprenderá patrones o características de ese sujeto, lo que conducirá a una percepción errónea de precisión elevada, cuando en realidad está reconociendo al participante y no al fenómeno. Numerosos estudios han señalado la manera en que la fuga de datos, ocasionada por una deficiente división de los ensayos, puede influir en los resultados [17] [18].

### 3.2.3. Técnicas Utilizadas

Con el objetivo de desarrollar modelos con buena capacidad de generalización, estabilidad y robustez, se han aplicado diversas estrategias para enfrentarse a distintos desafíos intrínsecos del procesamiento de señales EEG. Estas técnicas

permiten personalizar el entrenamiento para cada participante, abordando específicamente dos problemas críticos: la prevención de sobreajuste y la corrección del sesgo de clase.

### Focal Loss

Uno de los principales desafíos identificados durante el desarrollo de este estudio es el desbalance de clases. Alguno de los participantes contaba con una gran diferencia en la cantidad de valores de 'high trust' en comparación con las de 'low trust', lo que repercutía de manera notable en el rendimiento del modelo. Este problema constituye un desafío recurrente en el ámbito del aprendizaje automático, tanto en machine learning como el deep learning. Diversos estudios han abordado esta cuestión de múltiples maneras [19] [20].

Una de las técnicas más utilizadas es el **Focal Loss** [21], cuya idea principal consiste en reducir la importancia de los ejemplos fáciles y enfocarse en los difíciles. Es decir, al tener un modelo con las clases desbalanceadas, a la red le va a ser más sencillo diferenciar y detectar los ejemplos de la clase mayoritaria al tener muchas más muestras que de la clase minoritaria. Focal Loss es una función de pérdida que se define:

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (3.1)$$

Donde:

- $p_t$  probabilidad de que el modelo asigne a la clase correcta
- $\gamma$  parámetro que controla cuánto se reduce la pérdida de los ejemplos fáciles
- $\alpha_t$  factor de peso para las distintas clases. Permite compensar el desbalance entre la clase mayoritaria y la minoritaria, asignando mayor importancia a los errores cometidos en esta última.

Si el modelo falla en un ejemplo difícil ( $p_t$  pequeña), la pérdida se mantiene alta y el ejemplo recibirá especial atención durante el entrenamiento, lo que hará que la red se centre en aprender los casos difíciles.

### Oversampling

Otra técnica utilizada para tratar el desbalance de clases es el **oversampling** [22]. Consiste en aumentar artificialmente el número de ejemplos de la clase minoritaria, de manera que ambas clases queden equilibradas y el modelo no quede sesgado hacia la mayoritaria. Para este estudio, se han abordado dos variantes de esta técnica:

- **Oversampling aleatorio:** es una de las técnicas más básicas para abordar con el desbalance de clases. El procedimiento consiste en replicar muestras de la clase minoritaria mediante un proceso aleatorio, con el propósito de alcanzar una proporción cercana al número de instancias de la clase

## Capítulo 3. Desarrollo

---

mayoritaria. Aunque es simple, esta estrategia ha demostrado ser efectiva en diversos escenarios [23].

- SMOTE (Synthetic Minority Oversampling Technique) [24]: se implementa mediante la generación de réplicas sintéticas para la clase minoritaria. Este procedimiento se lleva a cabo mediante la extracción de una muestra real y la identificación de sus  $k$  vecinos más cercanos. Selecciona uno de esos  $k$  vecinos y traza un segmento hasta la muestra aleatoria elegida, creando un nuevo punto sintético en cualquier lugar a lo largo de dicha recta.

### Normalización

La **normalización** [25] es una técnica de preprocesamiento de datos que consiste en escalar las características para que todas tengan la misma importancia, ayudando a que el modelo consiga unos resultados más estables y eficientes.

### Dropout

El **dropout** [26] es una técnica de regularización muy utilizada en deep learning. Consiste en 'apagar' de manera aleatoria un porcentaje de neuronas durante el entrenamiento para evitar que la red dependa más de unas neuronas que de otras. Al apagar algunas neuronas, la red aprende representaciones más robustas y evita el sobreajuste hacia los datos de entrenamiento. Para la fase de test se utilizan todas las neuronas.

### Ponderación de clases

Otro método para abordar el desequilibrio de clases es la **ponderación de clases** mediante el uso de pesos ponderados [27]. Cuando se presenta este problema, el modelo tiende a ignorar la clase minoritaria. Este método asigna un peso mayor a las clases minoritarias en la función de pérdida, de modo que los errores sobre estas clases tienen más impacto en la fase de entrenamiento. En este caso, los pesos se calculan de forma inversamente proporcional a la frecuencia de cada clase en el conjunto de entrenamiento:

$$w_c = \frac{1}{n_c} \quad (3.2)$$

Donde  $n_c$  es el número de muestras perteneciente a la clase  $c$ .

Con el objetivo de mantener una escala estable en la función de pérdida, los pesos se normalizan de la siguiente manera:

$$\tilde{w}_c = \frac{w_c}{\sum_c w_c} \quad (3.3)$$

Estos pesos normalizados  $\tilde{w}_c$  se incorporan a la función de pérdida, penalizando en mayor medida los errores cometidos sobre la clase minoritaria y reduciendo así el sesgo del modelo hacia la clase mayoritaria.

Se ha utilizado la misma fórmula para cada participante, aunque cada uno presenta un número de muestras y desbalance diferente, el uso de la misma fórmula permite aplicar un criterio común que se adapta automáticamente a la distribución específica de cada sujeto. De este modo, se garantiza la coherencia metodológica y la comparabilidad de los resultados entre participantes.

### Early Stopping

El **early stopping** [28] es otra técnica de regularización que controla la fase de entrenamiento de la red y detiene el proceso antes de que se produzca un sobreajuste a los datos de entrenamiento. Durante esta fase, en cada iteración se calcula la pérdida y una serie de métricas (como el accuracy, recall, f1-score. . .). Si la pérdida de validación no disminuye durante un número determinado de iteraciones, (o si una de las métricas mencionadas no mejora) el entrenamiento se detiene. De esta manera, en el momento en el que la red deja de ajustarse a los datos de entrenamiento, se detiene, evitando así el overfitting y disminuyendo el tiempo de entrenamiento.

### Selección de características

La **selección de características** [29] consiste en elegir las variables más útiles para entrenar el modelo; en este caso, los canales que más información aporten y eliminar los más redundantes. De este modo, reducimos el tiempo de entrenamiento, mejoramos la precisión e identificamos cuáles son las características más importantes. Este proceso se ha llevado a cabo desde dos enfoques:

- Métodos de filtro: donde seleccionan las características basándose únicamente en propiedades estadísticas de los datos. Se ha usado la prueba ANOVA, que evalúa si la media de cada canal difiere mucho entre clases, y la Información Mutua (MI), que mide la dependencia entre cada canal y la variable objetivo, de manera que los que tengan más MI serán más relevantes.
- Método híbrido (filtro+wrapper): primero se aplica un método de filtro, como alguno de los mencionados anteriormente, y después se aplica un wrapper que evalúa iterativamente el rendimiento del modelo agregando o quitando características según su contribución a la clasificación. En este caso se ha combinado el Mutual Information con Sequential Feature Selector.

### 3.2.4. Arquitecturas de las redes neuronales

#### Red EEGNet

Una de las redes utilizadas en este trabajo es la red convolucional compacta **EEGNet** [30], diseñada específicamente para la clasificación de señales EEG en interfaces cerebro-computadora (BCI).

La estructura de la red está organizada en tres bloques:

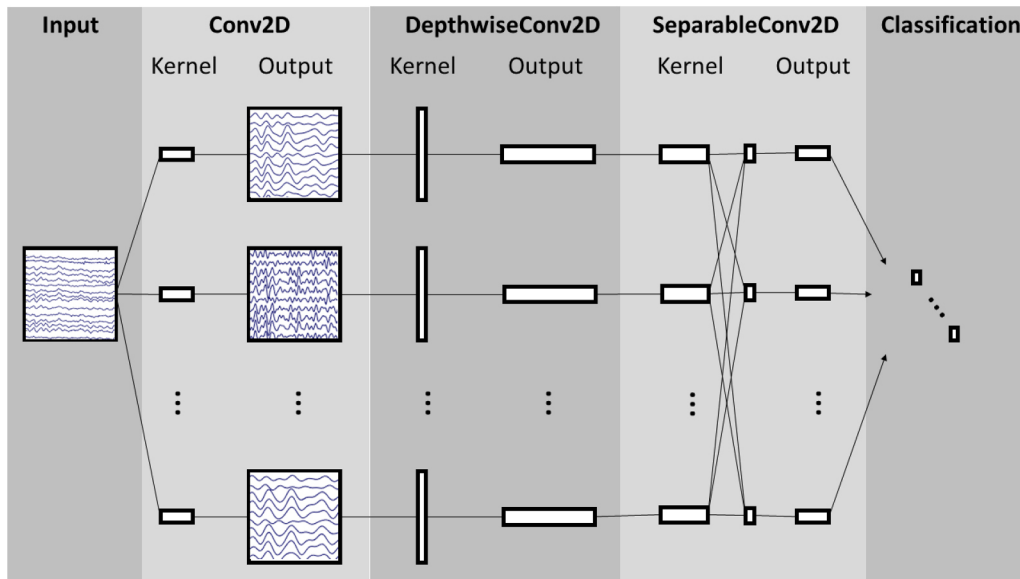


Figura 3.1: Arquitectura red EEGNet [31]

- El **primer bloque** se centra en la aplicación de dos convoluciones secuenciales, un proceso que se utiliza para la extracción de características tanto temporales como espaciales. En primer lugar, se implementa una convolución 2D, diseñada para capturar variaciones en las series temporales del EEG relacionadas con distintas bandas de frecuencia. En segundo lugar, se implementa una **DepthWise convolution** para los filtros espaciales, lo que permite la modelización de patrones específicos sin la necesidad de añadir parámetros de manera innecesaria.
- El bloque 2 se constituye mediante una convolución separable, compuesta por una **Depthwise convolution** seguida de una **Pointwise convolution**. La primera de estas etapas se encarga de resumir la información temporal aprendida en el bloque anterior, mientras que la segunda etapa combina la información de los mapas de características para obtener representaciones de mayor nivel. Este tipo de arquitectura reduce drásticamente la complejidad computacional en comparación con las convoluciones estándar.
- El **tercer bloque** corresponde a la etapa de clasificación. Una vez aplicadas las capas de pooling y regularización, las características finales son proyectadas a una capa softmax. El sistema EEGNet se caracteriza por su capacidad para evitar el sobreajuste y mantener la transparencia, mediante la omisión de capas densas intermedias en su configuración.

La principal aportación de **EEGNet** ha sido la demostración de que, mediante el uso de una única red convolucional, se ha logrado aplicar distintos datasets de señales EEG sin necesidad de rediseñar la arquitectura. Este modelo se caracteriza por su capacidad de funcionar con una cantidad limitada de datos y parámetros, lo que lo hace altamente eficiente y práctico en diversas situaciones. En el ámbito de la neurociencia, se ha observado que el sujeto en cuestión no

solo alcanza una precisión más que correcta, sino que también evidencia una capacidad de aprendizaje de patrones cerebrales reales, acordes a los conocimientos actuales en el campo.

### Red EEGNeX

**EEGNeX** [32] es una arquitectura convolucional diseñada como una extensión de EEGNet, cuyo objetivo principal es fortalecer la extracción de información espacial y temporal en señales EEG mediante una serie de ajustes en sus bloques convolucionales. De esta manera, se podrá observar si aumentar la complejidad de la red EEGNet es necesario para obtener mejores resultados para la clasificación.

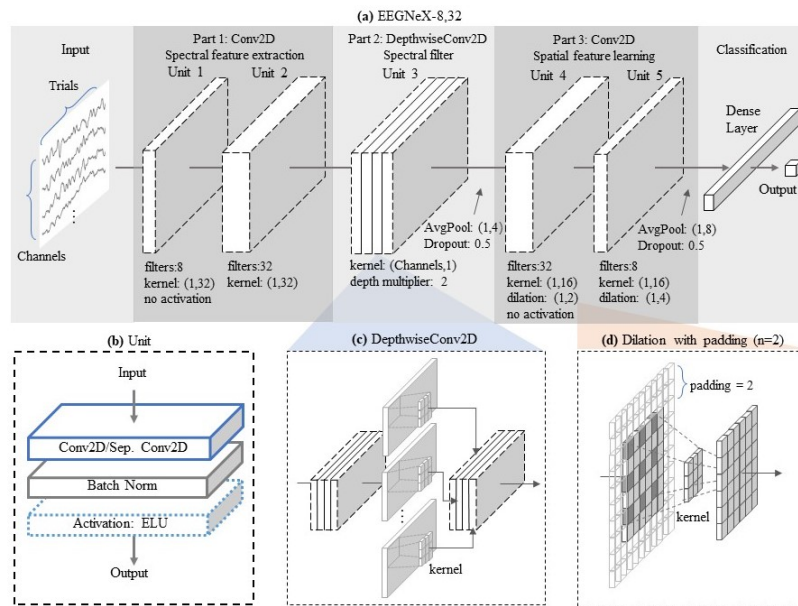


Figura 3.2: Arquitectura red EEGNeX [33]

Para la implementación de esta red neuronal, se probaron distintos modelos, tanto redes convolucionales CNN como redes recurrentes, para ver cuál combinación es la que daba mejores resultados. Pudieron observar que las redes convolucionales superaban a las recurrentes y que las convoluciones 2D superaban a las 1D. Con los resultados obtenidos en las pruebas anteriores, EEGNeX se organiza en cuatro modificaciones principales:

- **Refuerzo de la extracción espacial en el primer bloque.** El primer bloque de EEGNet, compuesto por una única convolución 2D para capturar información temporal y espacial, se amplía duplicando este bloque. Esta ampliación permite aumentar la capacidad de red para aprender patrones espaciales entre canales, manteniendo una profundidad moderada adecuada para datasets EEG.
- **Sustitución de la convolución separable por convoluciones 2D estándar.** En el segundo bloque de EEGNet, la combinación de convolución

depthwise y separable se reemplaza por dos Convoluciones 2D completas. Aunque las convoluciones depthwise reducen parámetros, en EEGNeX se amplía la capacidad representacional eliminando la convolución separable —que pierde utilidad tras la comprensión de canales— e introduciendo una convolución 2D adicional con un mayor número de filtros. Este cambio aumenta la habilidad del modelo para aprender características temporales más complejas.

- **Uso de inverted bottleneck.** EEGNeX también incorpora un bloque de cuellos de botella invertidos, donde la dimensionalidad del canal se expande primero y se reduce después. Este diseño permite a la red aprender representaciones intermedias más ricas sin incrementar de forma desproporcionada el número de parámetros.
- **Dilatación temporal y reducción de activaciones.** Con el propósito de fortalecer la obtención de dependencias temporales de mayor alcance, el tercer bloque incorpora convoluciones dilatadas con factores de 2 y 4. La dilatación permite ampliar el campo receptivo sin aumentar el número de parámetros ni la profundidad de la red. Además, EEGNeX reduce el número de funciones de activación entre bloques, utilizando una única activación ELU por cada parte de la arquitectura. Este enfoque implica una reducción en el número de transformaciones no lineales, lo que promueve una mayor estabilidad y conservación de la información a lo largo del flujo de la red.

En conjunto, EEGNeX constituye una arquitectura completamente convolucional que combina extracción local mediante operaciones de convolución de profundidad, integración global mediante convoluciones dilatadas y bloques de cuello de botella invertidos para mejorar la capacidad representacional sin aumentar excesivamente la complejidad. El diseño de este sistema se centra en la eficiencia, la capacidad de generalización y la capacidad de adaptación a las características particulares de los datos EEG.

### AttentionBaseNet

Si bien AttentionBaseNet [34] fue originalmente desarrollada para tareas de motor imagery (una técnica neurocientífica en la que el individuo imagina mentalmente la ejecución de un movimiento sin llegar a realizarlo), su arquitectura generalista y su capacidad para integrar mecanismos de atención basados en canales la hacen igualmente adecuada para cualquier tarea de clasificación de señales EEG. Se basa en convoluciones espaciales, temporales y generales, así como en mecanismos de atención canal-wise aplicables a cualquier dominio cognitivo. Por lo tanto, su aplicación para predecir la confianza es totalmente apropiado.

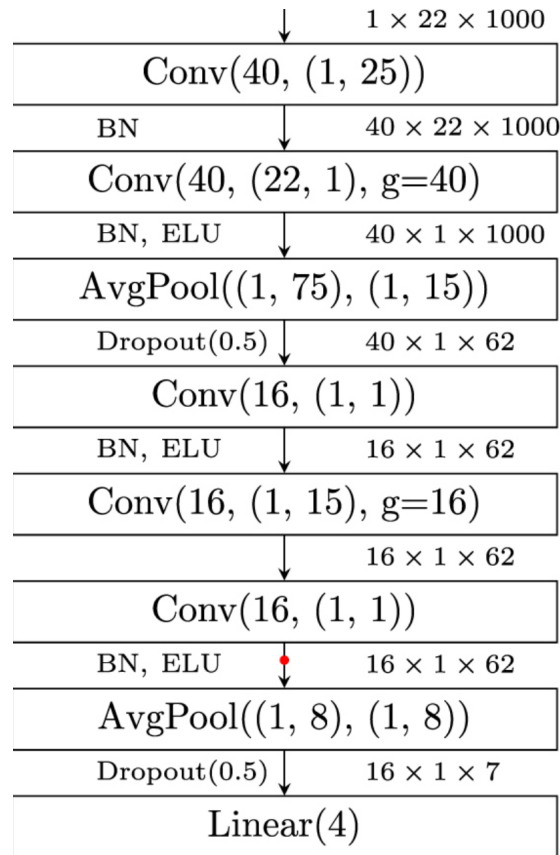


Figura 3.3: Arquitectura red AttentionBaseNet [35]

**AttentionBaseNet** mantiene la estructura clásica de las redes de motor imagery. Su arquitectura es simple y ligera, pero muy efectiva en contextos con señales ruidosas. Está compuesta por:

- Una primera capa que consiste en un stem-block, compuesto por una convolución temporal para capturar patrones a lo largo del tiempo, una convolución espacial para extraer correlaciones cercanas, batch normalization para estabilizar y acelerar el entrenamiento, una función de activación ELU para favorecer el flujo de gradientes y pooling para reducir la dimensionalidad y mantener las características importantes.
- El segundo bloque consiste en un **depthwise separable convolution**, el cual está compuesto de:
  - Una capa **depthwise convolution** para capturar patrones temporales dentro de cada canal EEG,
  - Una capa **pointwise convolution** que combina los resultados de la capa anterior para generar feature channels que mezclan información de distintos canales.
  - Un **1x1 Channel projection** que desacopla los canales EEG físicos de los canales de características, permitiendo que el mecanismo de

atención opere sobre representaciones de alto nivel.

- El tercer bloque trabaja sobre el mecanismo de atención, el cual se encarga de calcular un conjunto de pesos de atención que recalibran la importancia relativa de cada canal de características. Este proceso permite que la red refuerce los canales informativos y suprima los irrelevantes. Es un bloque crucial de la arquitectura porque prioriza las características más importantes para la tarea, reduce el ruido de los canales poco informativos y mejora la generalización del modelo sin incrementar su complejidad.
- El cuarto y último bloque se centra en la clasificación, mediante la cual se aplanará el tensor de atención para que adopte una estructura unidimensional. Posteriormente, se implementa una capa lineal para llevar a cabo la clasificación final, basada en las características ponderadas.

### 3.3. Selección de métricas según la distribución de los datos

Dada la naturaleza del experimento, se han observado dos escenarios claros en cuanto a distribución de las muestras de los participantes: balanceados y desbalanceados. Para asegurar una evaluación objetiva, estos grupos se evaluarán por separado, seleccionando unas métricas principales para el análisis en base a ello [36]:

- Participantes desbalanceados: Se utilizará el f1-score macro y el ROC-AUC. El f1-score garantiza que ambas clases, independientemente de la variación en el número de muestras, tienen una contribución equivalente al éxito del modelo. Por otro lado, el ROC-AUC evaluará la capacidad del modelo para separar las clases independientemente del umbral.
- Participantes balanceados: Se mantendrá el f1-score como medida de equilibrio, pero se sustituirá el ROC-AUC por el recall macro, que ofrece una medida muy adecuada para la interpretación de la sensibilidad del modelo.

## Capítulo 4

# Resultados

En este capítulo se exponen los resultados obtenidos tras la aplicación de las tres arquitecturas de redes neuronales profundas —EEGNet, EEGNeX y AttentionBaseNet— para la predicción de la confianza a partir de señales EEG. El propósito de este estudio es evaluar y comparar el rendimiento de los diferentes modelos para los 12 participantes, considerando diversos factores como el desbalance de clases y la variabilidad individual. Este último aspecto constituye un factor de gran importancia, dado que la actividad cerebral de cada persona presenta una variabilidad más que notable, lo que puede funcionar a la perfección con uno puede no ser correcto para otro. En consecuencia, es necesario implementar modelos específicos para cada participante.

Los resultados se estructuran en diversas secciones, iniciando con los modelos base, donde las redes se entrenan sin aplicar técnicas de balanceo ni ajustes especiales, permitiendo así que se pueda establecer una referencia de rendimiento inicial. A continuación, se muestran los efectos de aplicar estrategias de mejora, como oversampling, focal loss, pesos ponderados o selección de características. Se destaca cómo cada técnica influye en los modelos para mejorar (o empeorar) su capacidad de generalización.

Finalmente, se ofrece un resumen comparativo de los mejores modelos para cada participante, acompañado de tablas y gráficas que facilitan la visualización de diferencias entre arquitecturas y técnicas de entrenamiento. Además, se proporciona un análisis de la densidad espectral de la señal, así como una comparación con algunos modelos de machine learning mediante un análisis multimodal.

### 4.1. Modelos base

En esta sección se presentan los resultados obtenidos tras entrenar y evaluar los modelos sin aplicar ninguna técnica adicional. De esta manera, tendremos una referencia base sobre la que comparar cómo las distintas técnicas aplicadas mejoran o empeoran el rendimiento de los modelos. Todos los modelos se entrenaron de manera individual por participante, siguiendo el mismo protocolo de entrenamiento:

- `Drop_prob=0.5`: Las señales EEG tienen una alta variabilidad intra-sujeto, un `drop_out` elevado es habitual en modelos profundos con el objetivo de mitigar el sobreajuste.
- Épocas: 30-40
- `Weight_decay=1e-3`, con el fin de limitar la complejidad del modelo y mejorar la generalización.
- `Learning_rate=1e-3`: El valor estándar en combinación con el optimizador Adam.
- Optimizador: Adam [37], de gran capacidad para manejar gradientes ruidosos.

A partir de esta configuración base, estos hiperparámetros fueron ajustados de forma progresiva en función del rendimiento observado en los distintos experimentos. Dichos ajustes se realizaron atendiendo a las métricas de validación y al comportamiento de las curvas de pérdida.

Dado que la distribución de clases no es uniforme para todos los participantes, se decidió separar los sujetos en dos grupos:

- Participantes con datos balanceados: aquellos cuyo número de muestras por clase es relativamente similar, manteniendo una proporción inferior al 60% / 40%.
- Participantes con datos desbalanceados: aquellos con un número de muestras por clase desigual.

Como se discutirá más adelante, esta diferencia afecta considerablemente el rendimiento de los modelos, haciendo que el impacto de la aplicación de ciertas técnicas sea muy diferente entre ambos grupos.

## 4.1.1. Participantes balanceados

Participante	Modelo	Accuracy	Recall (macro)	F1-score (macro)	ROC-AUC
P4	EEGNet	0.63	0.64	0.59	0.636
	EEGNeX	0.69	0.69	0.68	0.703
	AttBaseNet	0.67	0.67	0.67	0.753
P7	EEGNet	0.59	0.53	0.49	0.572
	EEGNeX	0.59	0.59	0.58	0.358
	AttBaseNet	0.61	0.61	0.60	0.224
P8	EEGNet	0.61	0.66	0.60	0.868
	EEGNeX	0.82	0.84	0.82	0.893
	AttBaseNet	0.85	0.86	0.85	0.863
P9	EEGNet	0.57	0.58	0.54	0.493
	EEGNeX	0.53	0.53	0.52	0.521
	AttBaseNet	0.53	0.52	0.52	0.414
P17	EEGNet	0.59	0.57	0.55	0.685
	EEGNeX	0.60	0.61	0.60	0.655
	AttBaseNet	0.54	0.55	0.54	0.617
P18	EEGNet	0.36	0.49	0.32	0.450
	EEGNeX	0.45	0.50	0.45	0.452
	AttBaseNet	0.46	0.48	0.45	0.465
P21	EEGNet	0.45	0.50	0.33	0.419
	EEGNeX	0.75	0.75	0.75	0.781
	AttBaseNet	0.68	0.69	0.68	0.738
P22	EEGNet	0.57	0.51	0.51	0.434
	EEGNeX	0.59	0.54	0.54	0.334
	AttBaseNet	0.61	0.62	0.60	0.453

Cuadro 4.1: Métricas promedio macro para los participantes balanceados y sus modelos.

Como se observa en el Cuadro 4.1, se exponen las métricas resultantes de los modelos base para cada participante del conjunto balanceado. En términos generales, no se identifica la presencia de un modelo que destaque de manera consistente entre todos los sujetos. El rendimiento varía según el participante, lo que sugiere una significativa dependencia intersujeto.

En algunos casos, como los participantes 8 y 21, se observa un rendimiento notable con las arquitecturas EEGNeX y AttentionBaseNet, alcanzando valores considerablemente altos, lo que sugiere que estas redes podrían ser las más adecuadas para capturar patrones en los datos.

Para el resto de los participantes, la variación es más pronunciada: algunos muestran resultados deficientes, como el participante 18, mientras que otros, como P4, P7 o P17, presentan valores intermedios en los que la diferencia entre los tres modelos es mínima, indicando que encontrar un patrón específico resulta más complejo con esas señales.

En términos generales, EEGNeX parece ofrecer el mejor rendimiento en la mayoría de los casos, aunque sin generar una diferencia significativa frente a las demás arquitecturas en todos los participantes.

## Capítulo 4. Resultados

### Evolución del F1-Score en la etapa de entrenamiento

El f1-score es una métrica clave, puesto que representa la media armónica entre la precisión y el recall. Este parámetro resulta especialmente ventajoso a la hora de evaluar el rendimiento en clasificación de clases desbalanceadas [38]. Este enfoque proporciona una medida equilibrada del rendimiento. En contraste con el accuracy, que solo refleja la proporción de predicciones totales correctas, el f1-score macro trata a todas las clases por igual, evitando que las clases con más muestras dominen la evaluación.

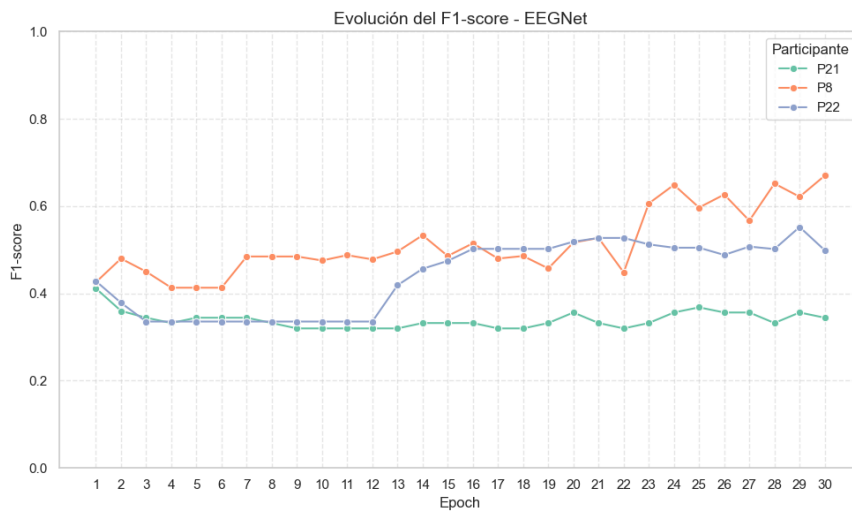


Figura 4.1: Evolución del f1-score en EEGNet

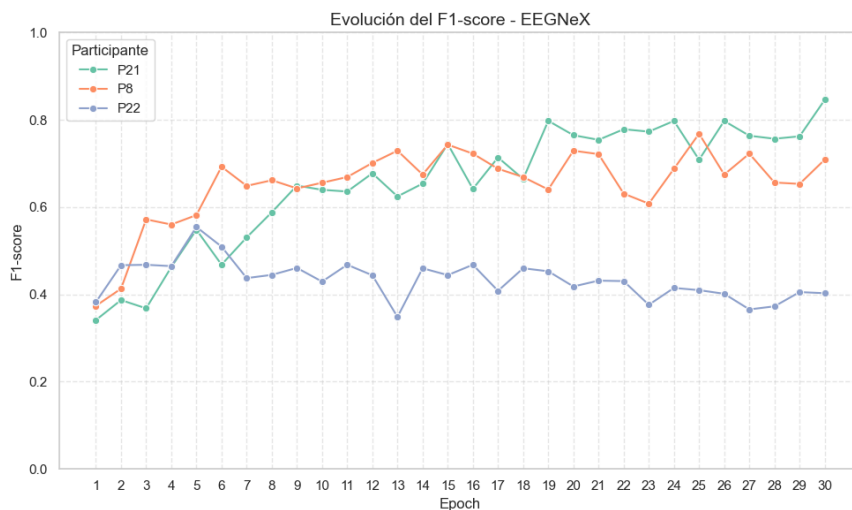


Figura 4.2: Evolución del f1-score en EEGNeX

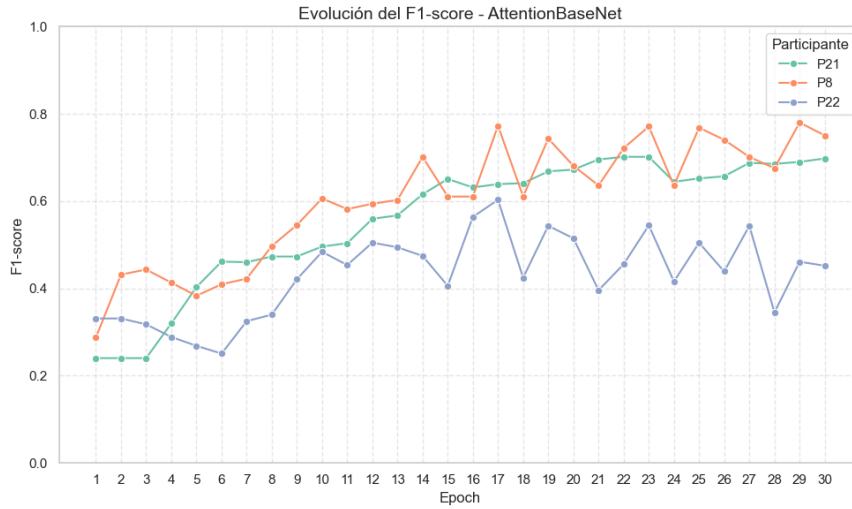


Figura 4.3: Evolución del f1-score en AttentionBaseNet

Como se puede observar en la Figura 4.1 **EEGNet** exhibe el rendimiento más bajo en comparación con las otras dos arquitecturas consideradas. Parece ser la menos efectiva para capturar las características relevantes de estos participantes. El análisis de los datos revela que el único participante que llega a superar el umbral de 0.5 es el participante 8.

A la vista de la Figura 4.2, **EEGNeX** muestra un rendimiento superior a EEGNet y similar a AttentionBaseNet. El participante 21 logra un muy buen rendimiento en las etapas finales del entrenamiento. El 8, por su parte, a mitad de entrenamiento parecía ofrecer buenos resultados, pero termina descendiendo al final. El 22 presenta el peor rendimiento, manteniéndose alrededor de los valores 0.35 y 0.45.

Finalmente, en la Figura 4.3, se puede observar que **AttentionBaseNet** es la que muestra un rendimiento más alto de las tres en estos modelos base, todos los participantes logran un f1-score superior al 0.6 (por encima del azar). El participante 8 es el que alcanza el mejor resultado, mientras que el 22 es el que presenta mayor volatilidad y un rendimiento más bajo, lo que podría sugerir que el modelo tiene dificultad para generalizar con los datos de este participante. Una observación clave es la gran variabilidad entre los participantes que se puede observar en las gráficas, lo que subraya la naturaleza personalizada del procesamiento de señales EEG.

## Capítulo 4. Resultados

### 4.1.2. Participantes desbalanceados

Participante	Modelo	Accuracy	Recall (macro)	F1-score (macro)	ROC-AUC
P3	EEGNet	0.58	0.49	0.49	0.767
	EEGNeX	0.59	0.45	0.44	0.896
	AttBaseNet	0.61	0.48	0.48	0.558
P6	EEGNet	0.71	0.51	0.51	0.437
	EEGNeX	0.68	0.61	0.59	0.709
	AttBaseNet	0.65	0.55	0.54	0.625
P14	EEGNet	0.84	0.60	0.62	0.777
	EEGNeX	0.68	0.65	0.58	0.595
	AttBaseNet	0.55	0.68	0.52	0.652
P19	EEGNet	0.74	0.44	0.44	0.407
	EEGNeX	0.76	0.47	0.47	0.618
	AttBaseNet	0.76	0.45	0.45	0.571

Cuadro 4.2: Métricas promedio macro para los participantes desbalanceados y sus modelos.

La Tabla 4.2 muestra el grupo de participantes con clases desbalanceadas. No se observa ninguna arquitectura sobresaliente en comparación con las demás.

**EEGNeX** es la solución más robusta, es el que logra alcanzar un f1-macro más alto. Por su parte, EEGNet presenta un rendimiento más bajo en cuanto a f1-macro, a excepción del participante 14. No obstante, logra altos valores de ROC-AUC, lo que sugiere una capacidad notable para ordenar las probabilidades de clase. Por su parte, AttentionBaseNet sobresale por lograr el mejor recall-macro; sin embargo, exhibe una inconsistencia en términos de ROC-AUC.

Es fundamental destacar la **poca fiabilidad de la métrica Accuracy** en la evaluación de conjuntos de datos desbalanceados. El Accuracy mide la proporción de aciertos totales, lo que resulta en un sesgo estadístico en favor de la clase mayoritaria [39]. Por ejemplo, en el participante P19 con la arquitectura EEGNeX, se observa un Accuracy de 0.87. Esta cifra, aparentemente elevada, podría sugerir un excelente rendimiento del modelo. No obstante, al observar el f1-macro de 0.56, se evidencia que el rendimiento real es notablemente bajo. La discrepancia en los resultados se atribuye a la tendencia del modelo a favorecer a la clase con mayor número de muestras. Al haber muchas más instancias de la clase mayoritaria, el modelo acierta una alta proporción de las predicciones simplemente apostando por ella de forma consistente, lo que resulta en un aumento artificial del Accuracy. Para contrarrestar este efecto, el f1-score (macro) se establece como la métrica clave, ya que calcula el f1-score de cada clase de forma independiente y luego promedia esos resultados. De esta manera, se garantiza que el rendimiento de la clase minoritaria reciba el mismo peso que el de la clase mayoritaria.

**Evolución del f1-score en la etapa de entrenamiento**

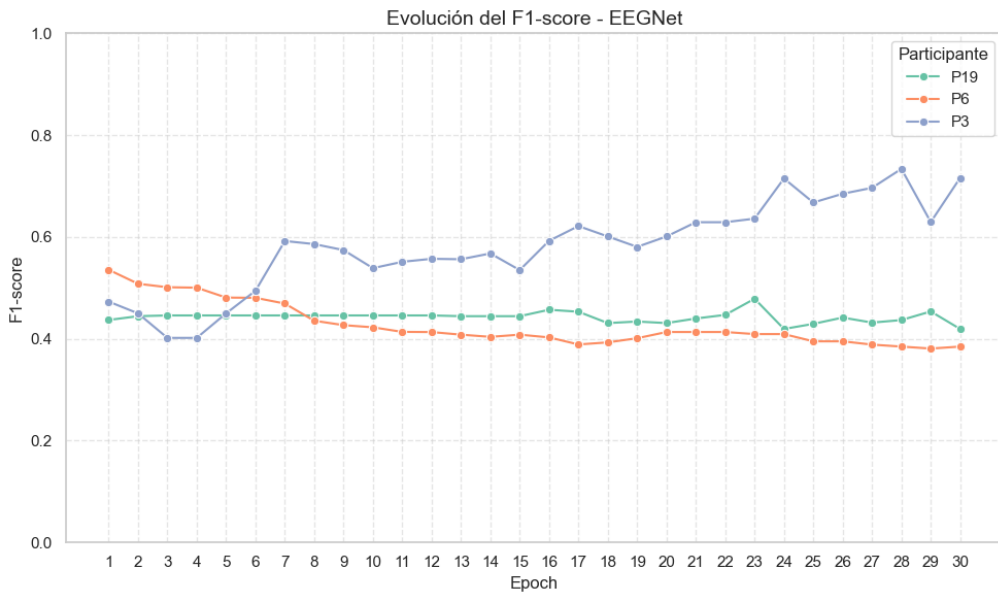


Figura 4.4: Evolución del F1-score en EEGNet

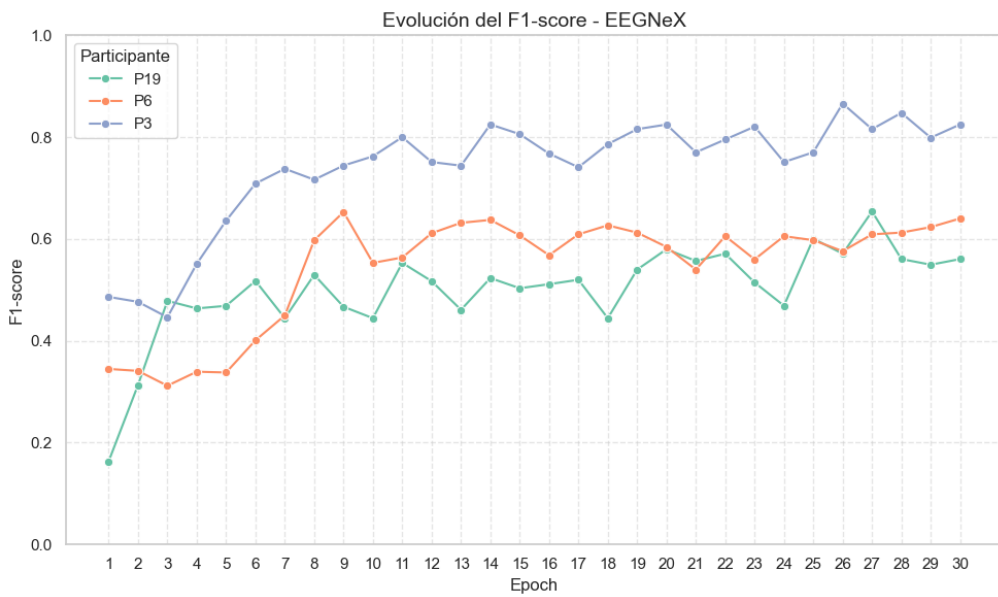


Figura 4.5: Evolución del F1-score en EEGNeX

## Capítulo 4. Resultados

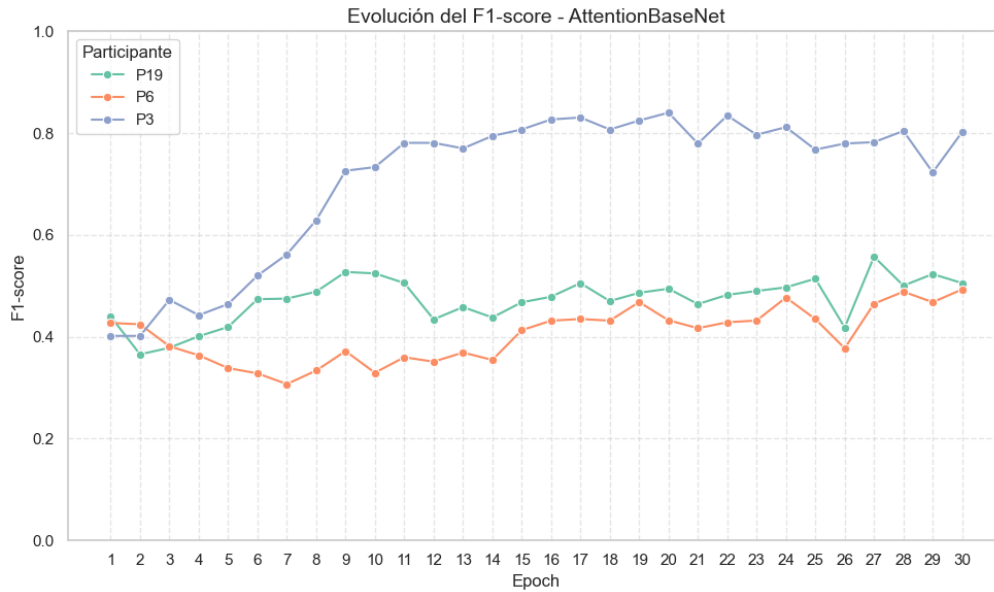


Figura 4.6: Evolución del F1-score en AttentionBaseNet

Como se puede observar en la Figura 4.4, EEGNet es la que muestra el rendimiento más bajo. El análisis estadístico revela que el participante número 3 es el único que logra mantenerse por encima del umbral del 0,6. El resto de los participantes exhiben resultados que se encuentran alrededor del valor aleatorio.

Por su parte, en la Figura 4.5 EEGNeX muestra un rendimiento general más potente. El participante 3 logra un rendimiento excepcional, manteniendo de manera consistente un valor por encima del 0.75, lo que sugiere un alto potencial de separación de clases. Los otros participantes muestran un rendimiento modesto.

Finalmente, a la vista de la Figura 4.6, AttentionBaseNet tiene unos resultados intermedios a las dos anteriores: mayor estabilidad que EEGNet, pero menos destacable que EEGNeX. Una vez más, el participante 3 ofrece una mayor estabilidad y mejores resultados, mientras que los otros tienen un rendimiento inferior.

## 4.2. Oversampling y Focal Loss

El desbalanceo inherente de clases ha representado un gran desafío al momento de implementar las distintas arquitecturas de las redes neuronales. Esta disparidad entre las dos clases de confianza provocó que los modelos exhibieran una tendencia hacia la clase mayoritaria. El análisis se centró en los cuatro participantes que mostraron un mayor desequilibrio: P3, P6, P14 y P19. El número de muestras de cada participante se puede observar en la Figura 4.7. En esta sección, se analizará el impacto en cada arquitectura tras aplicar las dos técnicas de oversampling junto con Focal Loss.

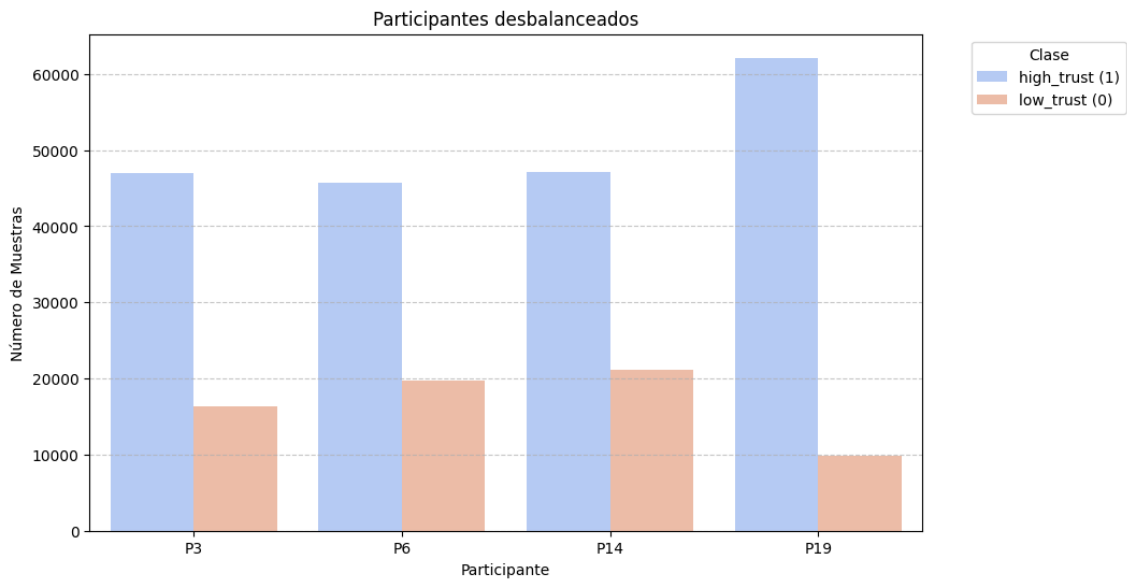


Figura 4.7: Gráfico de barras del desbalanceo de clases

Como se muestra en el análisis de la Figura 4.7, se observa una diferencia muy significativa entre la clase *high\_trust* y *low\_trust*, sobre todo con el participante 19, donde la diferencia es notablemente extrema. Para abordar este problema, se implementaron dos técnicas de oversampling: aleatorio y SMOTE, complementadas también con la técnica de Focal Loss. El objetivo principal será tratar de encontrar la mejor combinación para la clasificación de los datos EEG.

## Capítulo 4. Resultados

### 4.2.1. Impacto en la arquitectura EEGNet

Participante	Oversampling	Accuracy	Recall (macro)	F1-score (macro)	ROC-AUC
P3	Aleatorio	0.58	0.64	0.57	0.554
	SMOTE	0.60	0.60	0.57	0.637
P6	Aleatorio	0.48	0.57	0.52	0.583
	SMOTE	0.44	0.45	0.41	0.392
P14	Aleatorio	0.63	0.74	0.58	0.741
	SMOTE	0.52	0.61	0.48	0.673
P19	Aleatorio	0.60	0.61	0.57	0.561
	SMOTE	0.64	0.54	0.50	0.549

Cuadro 4.3: Comparación de técnicas de oversampling usando EEGNet.

A la vista del Cuadro 4.3, EEGNet exhibe un rendimiento altamente variable entre sujetos. Es importante destacar que la eficacia de una técnica en un sujeto puede no ser necesariamente indicativa de sus resultados en otro, lo que subraya la alta variabilidad de los datos entre los participantes.

Como se puede observar en la Figura 4.8, los participantes 3 y 19 exhiben mejoras significativas en los resultados del modelo base al implementar métodos de oversampling, pasando de un 0.47 a 0.57 en caso del 3, y de un 0.49 a 0.57 en el caso del 19. En contraste, el participante 6 muestra una mejora marginal en su rendimiento con el método aleatorio. Por su parte, el participante 14 resulta notablemente interesante, ya que logra alcanzar el f1-score más elevado de todos con su modelo base, siendo SMOTE el método más perjudicial.

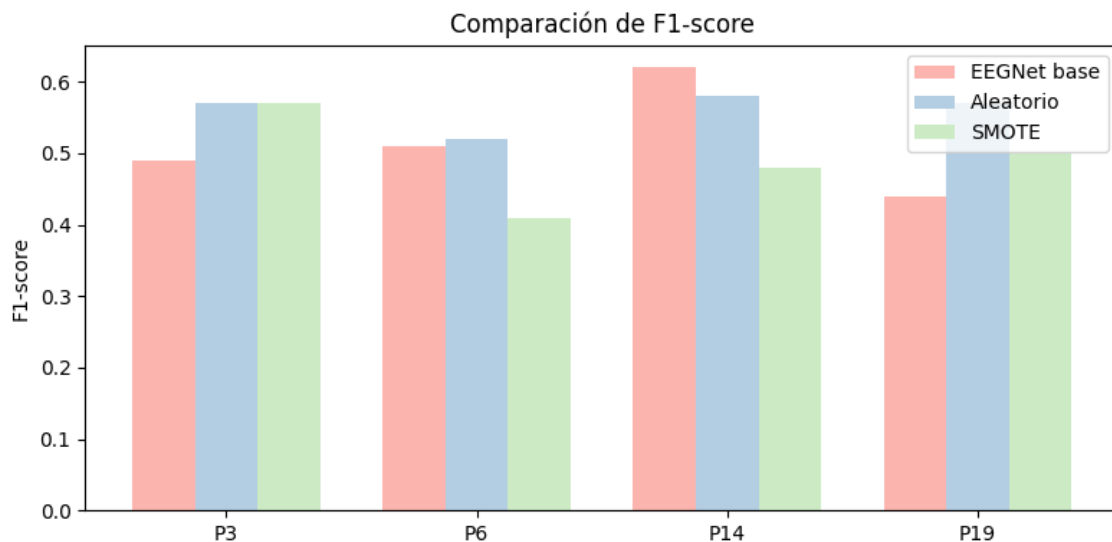


Figura 4.8: Gráfico de barras del f1-score tras aplicar oversampling con EEGNet

El ROC-AUC constituye una métrica de gran interés para participantes con clases desbalanceadas, puesto que nos permite observar la capacidad intrínseca del modelo para separar las clases, es decir, para apreciar su calidad discriminativa [40]. El valor general de esta métrica es superior al f1-score, principalmente en los participantes 3, donde se alcanza un valor de 0.786 con el método aleatorio, y el 14, donde se logra un valor de 0.777 con el modelo base una vez más.

A pesar de que SMOTE es un método más avanzado, en la siguiente Figura 4.9, se puede observar que la técnica que ha mostrado tener mejores resultados es el oversampling aleatorio. No obstante, los valores alcanzados de f1-score son significativamente mejorables. La incorporación de Focal Loss permite al modelo priorizar el aprendizaje de características críticas en la frontera de decisión. El parámetro  $\alpha$  de la Focal Loss controla el peso asignado a la clase minoritaria. En este trabajo se evaluaron distintos valores de  $\alpha$  (en un rango de [0.3, 0.7]) en función del comportamiento observado durante el entrenamiento y la validación, manteniendo fija la arquitectura EEGNet y el oversampling aleatorio previamente seleccionado. Por otro lado, para el parámetro  $\gamma$  se probaron los valores 1.5 y 2, ya que en [21] se probó empíricamente que ofrecen el rendimiento más robusto al balancear la importancia de ejemplos fáciles y difíciles.

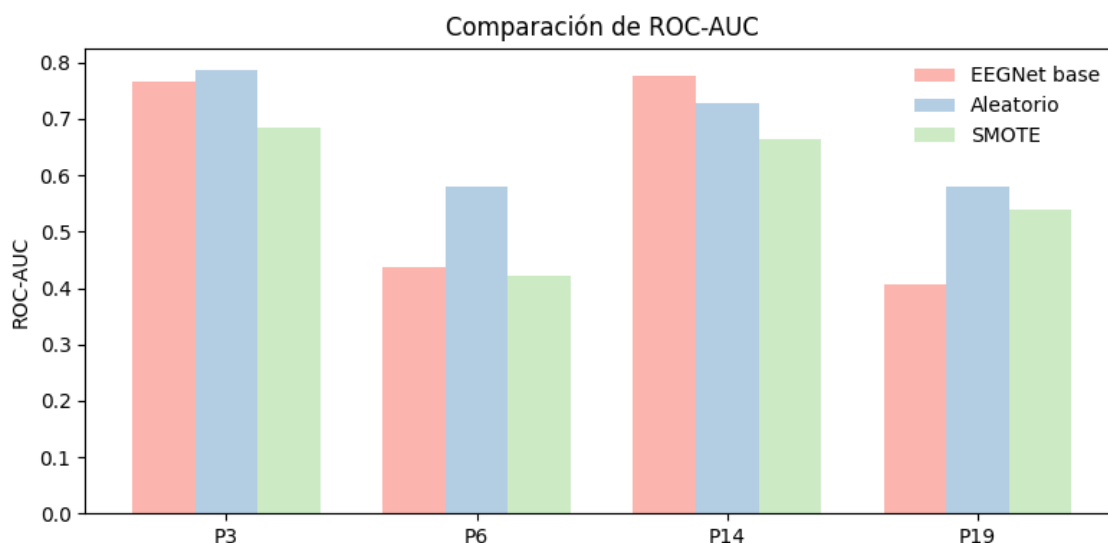


Figura 4.9: Gráfico de barras del ROC-AUC tras aplicar oversampling

A la vista de la Tabla 4.4, el sujeto que más destaca como el caso de más éxito tras la configuración de  $\alpha = 0,3$  y  $\gamma = 2$  es, indudablemente, el participante 19, que logró mejorar un 0.49 en el f1-score a 0.61. De manera simultánea, ha fortalecido su capacidad discriminativa del ROC-AUC, que experimenta un aumento de 0.561 a 0.651. Por otra parte, el participante 14, bajo una configuración de  $\alpha = 0,6$  y  $\gamma = 2$ , exhibe una mejora significativa. En el caso del f1-score, se observó un incremento notable de 0.58 a 0.69, además de una mejora marginal en el ROC-AUC. Por su parte, P6 sufre una disminución en su f1-score, aunque simultáneamente logra una ligera optimización en el ROC-AUC. No obstante, el

## Capítulo 4. Resultados

participante 3 evidencia una significativa pérdida de rendimiento en ambas métricas tras la implementación de esta técnica, experimentando una caída desde un sólido 0.786 de ROC-AUC hasta un 0.541, lo que sugiere que la integración de la Focal Loss ocasionó una perturbación en la frontera de decisión.

Participante	Configuración	F1-score (macro)	ROC-AUC
P3	Oversampling aleatorio	0.57	0.786
	Oversampling + Focal Loss	0.54	0.541
P6	Oversampling aleatorio	0.52	0.583
	Oversampling + Focal Loss	0.45	0.638
P14	Oversampling aleatorio	0.58	0.741
	Oversampling + Focal Loss	0.69	0.768
P19	Oversampling aleatorio	0.49	0.561
	Oversampling + Focal Loss	0.61	0.651

Cuadro 4.4: Impacto de la Focal Loss sobre EEGNet con oversampling aleatorio en participantes desbalanceados.

En consecuencia, estos resultados sugieren que se debe aplicar la Focal Loss con precaución y no confiar en la observación de una mejora en el entrenamiento. Su efectividad se ve influida en gran medida por el participante y de una adecuada selección del parámetro  $\alpha$ .

### 4.2.2. Impacto en la arquitectura EEGNeX

En tabla Tabla 4.5 podemos observar el impacto en la arquitectura EEGNeX tras la aplicación de las dos técnicas de oversampling. En el caso concreto de los participantes 6 y 19, ambos han experimentado un beneficio notable como resultado de la aplicación del método aleatorio. El único sujeto que muestra beneficio por parte de SMOTE es el participante 14 en cuanto a las métricas de f1-score y ROC-AUC, aunque de forma marginal. Finalmente, el sujeto 3 presenta resultados similares con ambos métodos, teniendo diferencias marginales en todas las métricas.

Participante	Oversampling	Accuracy	Recall (macro)	F1-score (macro)	ROC-AUC
P3	Aleatorio	0.61	0.51	0.51	0.632
	SMOTE	0.63	0.50	0.54	0.612
P6	Aleatorio	0.70	0.68	0.63	0.771
	SMOTE	0.61	0.56	0.53	0.639
P14	Aleatorio	0.62	0.62	0.53	0.583
	SMOTE	0.65	0.60	0.55	0.714
P19	Aleatorio	0.81	0.64	0.62	0.666
	SMOTE	0.80	0.55	0.55	0.740

Cuadro 4.5: Comparación de técnicas de oversampling usando EEGNeX.

## 4.2. Oversampling y Focal Loss

Como se puede observar en la Figura 4.10, el método de oversampling aleatorio supera al SMOTE en tres de los cuatro participantes. Si bien SMOTE constituye una técnica más compleja y avanzada, la mejora respecto al modelo base es marginal. En el caso del participante 6, el f1-score aumenta de 0.52 a 0.53 con SMOTE, mientras que con el método aleatorio se incrementa a 0.63, al igual que con los participantes 6 y 19. Esto sugiere que la generación de datos sintéticos puede haber introducido ruido a la señal, afectando negativamente a la capacidad de generalización de EEGNeX. El único caso en el que observa una mejora en relación con el aleatorio es con el participante 14, aunque esta mejora es mínima en comparación con el modelo base.

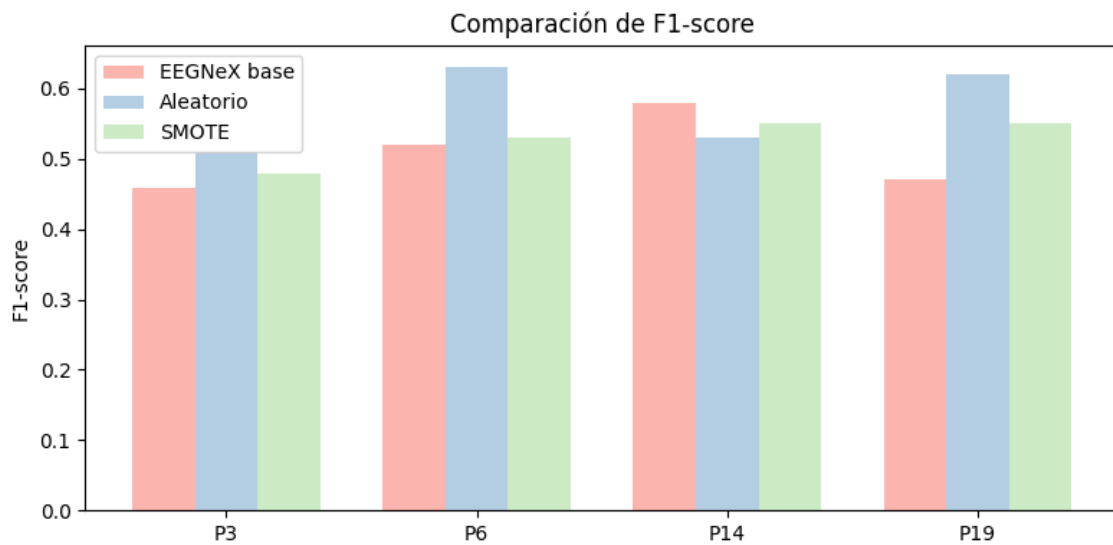


Figura 4.10: Gráfico de barras del f1-score tras aplicar oversampling en EEGNeX

En contraste, en el caso de la Figura 4.11, los resultados obtenidos mediante el ROC-AUC son más heterogéneos. En el caso del participante 6, hay una clara superioridad del método aleatorio, lo que resulta en un incremento notable en la métrica de 0.54 a 0.77. Por su parte, el 19 muestra una mayor mejora con el método SMOTE, superando tanto al modelo base como al aleatorio. El participante 14 exhibe un resultado atípico, dado que el mejor ROC-AUC se logra con el modelo base (0.799), y con una diferencia importante. Esto sugiere que su señal ya era de calidad y aplicar oversampling la distorsionó, provocando una caída de aproximadamente un 20%. Finalmente, con P3 las tres configuraciones dan un resultado similar, con una pequeña desventaja para SMOTE.

## Capítulo 4. Resultados

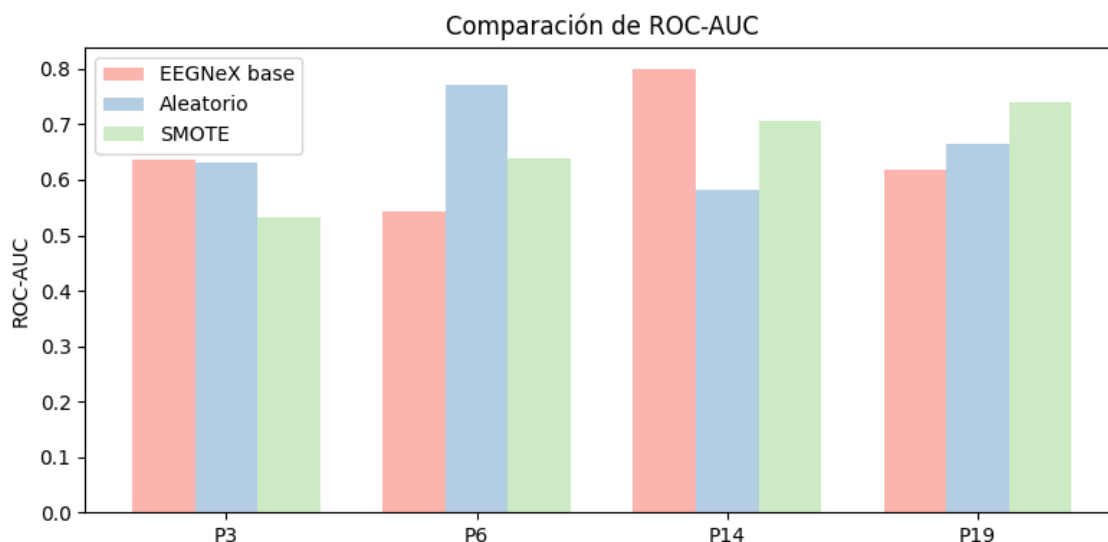


Figura 4.11: Gráfico de barras del ROC-AUC tras aplicar oversampling en EEG-NeX

Aunque hemos podido comprobar que el oversampling aleatorio ha mejorado la capacidad del modelo, se observa que presenta limitaciones significativas. A pesar de haber logrado resultados por encima del azar (0.5), podrían ser todavía más óptimos.

Como se puede observar en la Tabla 4.6, la muestra presenta resultados heterogéneos entre los participantes. En el caso del participante P3, se puede observar una mejora marginal en el f1-score, con una variación de 0.56 a 0.57 con  $\alpha = 0,7$  y  $\gamma = 1,5$ , así como un ligero incremento en el ROC-AUC. El caso de este sujeto nos sugiere que, en escenarios de desbalanceo, introducir Focal Loss con el ajuste adecuado de  $\alpha$  y  $\gamma$  permite incrementar la sensibilidad hacia la clase minoritaria.

No obstante, para los participantes 6, 14 y 19, la combinación oversampling + Focal Loss no solo no mejoró los resultados, sino que en algunos casos provocó una degradación del f1-score macro y del ROC-AUC. Tal y como se observa en el análisis de los datos, se evidencia una disminución significativa en los valores de las métricas correspondientes a los participantes 6 y 19.

Es relevante señalar que, si bien en determinados casos la Focal Loss puede reducir de forma más rápida la pérdida durante el entrenamiento y la validación, dicha optimización no necesariamente se traduce en una mejora de las métricas de evaluación en el conjunto de test. Esto implica que la optimización de la función de pérdida no implica una mejor generalización del modelo.

## 4.2. Oversampling y Focal Loss

Participante	Configuración	F1-score (macro)	ROC-AUC
P3	Oversampling aleatorio	0.56	0.656
	Oversampling + Focal Loss	0.57	0.663
P6	Oversampling aleatorio	0.63	0.771
	Oversampling + Focal Loss	0.45	0.398
P14	Oversampling aleatorio	0.59	0.708
	Oversampling + Focal Loss	0.52	0.606
P19	Oversampling aleatorio	0.62	0.666
	Oversampling + Focal Loss	0.45	0.559

Cuadro 4.6: Impacto de la Focal Loss sobre EEGNeX con oversampling aleatorio en participantes desbalanceados.

### 4.2.3. Impacto en la arquitectura AttentionBaseNet

A la vista de la Tabla 4.7, a diferencia de las arquitecturas anteriores, la técnica de SMOTE ofrece mejores resultados que el oversampling aleatorio, es decir, los mecanismos de atención son capaces de extraer patrones útiles de la densidad adicional de los datos sintéticos generados por SMOTE. No solo mejora el f1-score y el ROC-AUC, sino que también logra alcanzar un buen valor de recall en algunos casos, lo que quiere decir que el modelo minimiza los falsos negativos y es bueno detectando la clase minoritaria.

Participante	Oversampling	Accuracy	Recall (macro)	F1-score (macro)	ROC-AUC
P3	Aleatorio	0.61	0.56	0.55	0.603
	SMOTE	0.67	0.64	0.62	0.685
P6	Aleatorio	0.65	0.61	0.58	0.656
	SMOTE	0.73	0.67	0.64	0.701
P14	Aleatorio	0.66	0.66	0.57	0.749
	SMOTE	0.65	0.71	0.59	0.775
P19	Aleatorio	0.82	0.68	0.65	0.761
	SMOTE	0.77	0.69	0.61	0.709

Cuadro 4.7: Comparación de técnicas de oversampling usando AttentionBaseNet.

Como se observa en la Figura 4.12, en el caso de AttentionBaseNet, la técnica predominante es la SMOTE. El único caso donde el método aleatorio supera ambas técnicas es con el participante 19, quien logra alcanzar el valor más alto de f1-score, partiendo de un 0.45 con el modelo base hasta alcanzar un 0.65, lo que supone un incremento importante. El participante 6 también muestra una mejora más drástica, pasando de un valor de f1-score de 0.54 en el modelo base a 0.64 tras la aplicación de SMOTE. Por su parte, el sujeto 3 también logra una mejora interesante, pasando de un f1-score peor que el azar (0.48) a 0.62 con SMOTE, lo que supone una mejora muy importante. En todos los casos, el modelo base es el que muestra un rendimiento más bajo, al contrario que en las otras dos arquitecturas, donde en el caso del participante 14 era con el que se

## Capítulo 4. Resultados

---

lograba un mejor rendimiento.

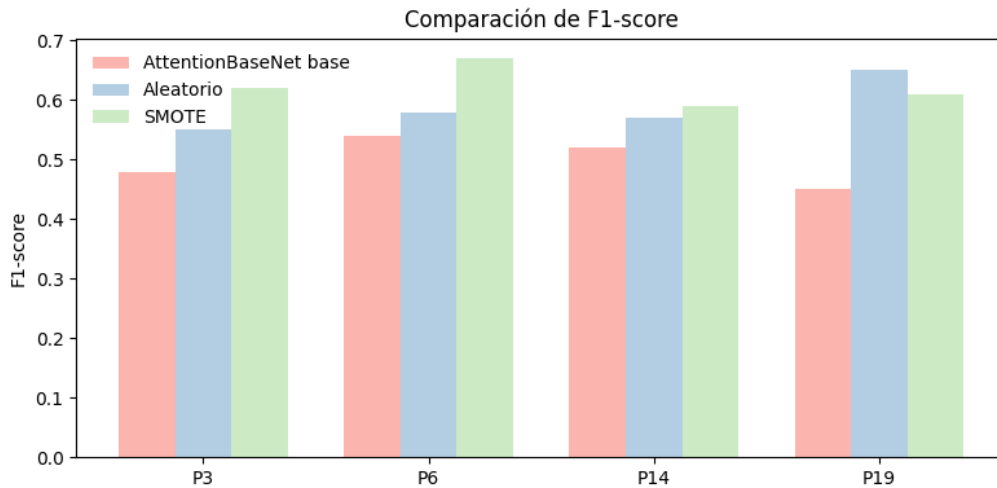


Figura 4.12: Gráfico de barras del ROC-AUC tras aplicar oversampling con AttentionBaseNet

Como se puede observar en la Figura 4.13, al igual que con el f1-score, SMOTE es la técnica predominante en este caso, salvo para, una vez más, el participante 19, quien logra el incremento más importante de todos, alcanzando un sólido 0.761. El participante 3 muestra una progresión significativa, incrementando su valor de ROC-AUC desde un nivel base de 0.56 —muy cercano al umbral del azar— hasta un notable 0.68 tras la aplicación de SMOTE. Esto indica que adquirió sensibilidad técnica para diferenciar de manera fiable las distribuciones de baja y alta confianza. Los participantes 6 y 14 son los que reflejan mejores resultados, alcanzando un ROC-AUC de 0.701 y 0.775, respectivamente. Estos participantes confirman que la combinación de mecanismos de atención con muestras sintéticas permite consolidar fronteras de decisión robustas.

## 4.2. Oversampling y Focal Loss

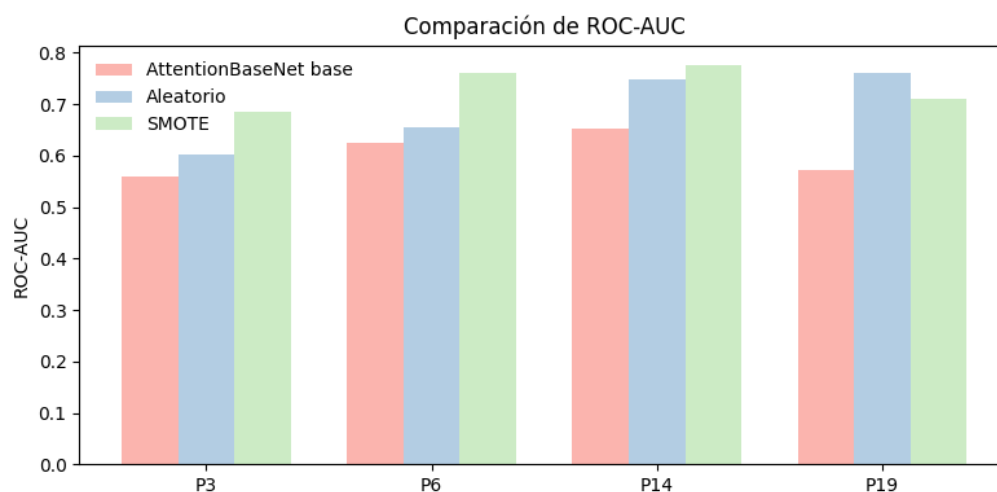


Figura 4.13: Gráfico de barras del ROC-AUC tras aplicar oversampling con AttentionBaseNet

Tras haber comprobado que la técnica SMOTE es la estrategia de balanceo más eficaz para la arquitectura, salvo para el sujeto 19, se procede a evaluar la incorporación de la función de pérdida Focal Loss para tratar de optimizar más los resultados obtenidos en el conjunto de test.

Participante	Configuración	F1-score (macro)	ROC-AUC
P3	SMOTE	0.62	0.685
	SMOTE + Focal Loss	0.62	0.687
P6	SMOTE	0.67	0.760
	SMOTE + Focal Loss	0.64	0.739
P14	SMOTE	0.59	0.774
	SMOTE + Focal Loss	0.58	0.779
P19	Aleatorio	0.65	0.761
	Aleatorio + Focal Loss	0.63	0.713

Cuadro 4.8: Impacto de la Focal Loss sobre AttentionBaseNet con oversampling SMOTE en participantes desbalanceados.

Como se puede observar en la Tabla 4.8, la introducción de Focal Loss no ha resultado favorable para ninguno de los participantes evaluados. En todos los casos, tanto el f1-score como el ROC-AUC han experimentado reducciones, afectando negativamente al rendimiento global del modelo en comparación con el uso de SMOTE bajo una pérdida de entropía cruzada convencional. Focal Loss, al interactuar con la complejidad de las capas de atención, acaba por sobreajustar el modelo, independientemente del valor de  $\alpha$  y  $\gamma$ .

En conclusión, para la arquitectura AttentionBaseNet, el uso SMOTE con la entropía cruzada convencional se consolida como la estrategia más robusta.

## Capítulo 4. Resultados

Sujeto	Arquitectura	Estrategia Óptima	F1-score	ROC-AUC
P3	EEGNeX	Aleatorio + Focal Loss	0.62	0.671
P6	Attention	SMOTE	0.64	0.701
P14	EEGNet	Aleatorio + Focal Loss	0.69	0.768
P19	Attention	Aleatorio	0.65	0.761

Cuadro 4.9: Resumen de los mejores resultados obtenidos para los participantes con alto desbalance de clases.

### 4.3. Pesos Ponderados

La optimización de la función de pérdida mediante el uso de pesos ponderados resultó ser un factor muy importante para algunos modelos donde, en un principio, colapsaban más hacia la clase mayoritaria, aunque las clases no estuviesen muy desequilibradas.

#### 4.3.1. Participantes balanceados

Si bien los sujetos exhiben una distribución de clases balanceada, se ha observado que las arquitecturas profundas pueden presentar sesgos durante el proceso de entrenamiento debido a factores como la complejidad del modelo o la presencia de sobreajuste. Por ello, se ha llevado a cabo un análisis exhaustivo del impacto del uso de pesos ponderados en la función de pérdida, con el objetivo de evaluar su efecto real sobre la estabilidad y capacidad de generalización del modelo.

Para representar el balanceo general de este grupo de participantes, se presenta la Figura 4.14 de un subconjunto elegido representativo sobre cómo es la proporción general de clases.

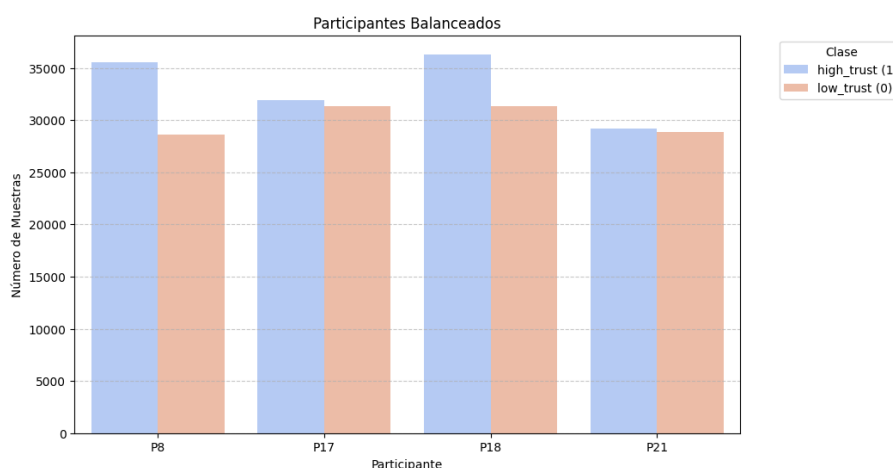


Figura 4.14: Balanceo de los participantes balanceados

En todos los sujetos analizados se observa un predominio de la clase *high\_trust*, sin embargo, la diferencia no es tan significativa como en el caso de los sujetos

desbalanceados. Se ha observado que el participante que muestra disparidades más notables es el número 8, lo que justifica aún más la aplicación de los pesos ponderados en su caso específico. No obstante, esta situación no llega a ser tan crítica como para requerir la aplicación del oversampling.

### Arquitectura EEGNet

Con el objetivo de analizar el impacto de los pesos ponderados en el criterio de decisión del modelo, se estudia en primer lugar un caso representativo en detalle.

El participante 8 fue seleccionado como caso de estudio por presentar una distribución de clases balanceada y un rendimiento estable en los experimentos previos, lo que permite aislar el efecto de los pesos ponderados sin la influencia del desbalance de clases.

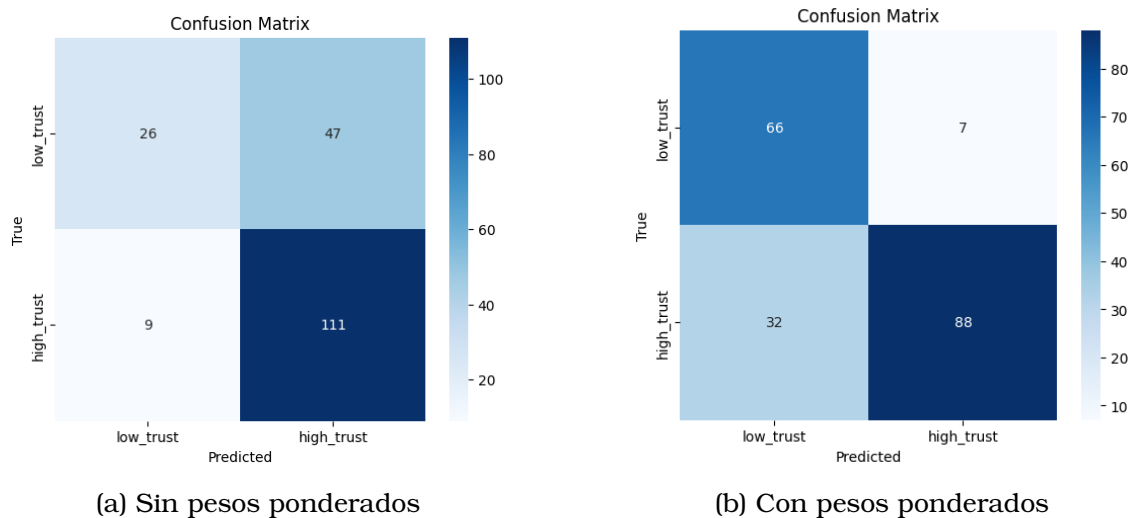


Figura 4.15: Comparación de matrices de confusión para el participante 8.

Como se puede observar en la Figura 4.15, en la primera matriz la clase 0 es prácticamente ignorada, lo que resulta en la clasificación de la mayoría de las muestras como pertenecientes a la clase mayoritaria. Este fenómeno también se observa en los valores del recall, con un valor de 0.36 en la clase minoritaria y de 0.93 en la mayoritaria. Este comportamiento sugiere que el modelo prioriza minimizar el error global a costa de ignorar la clase minoritaria.

Por otro lado, tras introducir los pesos, se observa un reparto mucho más equilibrado de las predicciones. El modelo incrementa considerablemente su capacidad para detectar la clase minoritaria, reduciendo de forma sustancial el número de falsos negativos. El recall aumentó a 0.90 en la clase 0 y se redujo ligeramente a 0.73 en la clase mayoritaria, logrando alcanzar un equilibrio. Esta mejora conlleva un ligero aumento de los errores cometidos en la clase mayoritaria, reflejando el compromiso inherente entre sensibilidad y precisión cuando

## Capítulo 4. Resultados

---

se corrige el desbalance de clases.

Con estos resultados podemos observar que no solo es útil para la mejora de la precisión global, sino para modificar activamente el criterio de decisión del modelo, forzándolo a prestar más atención a la clase infrarepresentada.

Para cuantificar este efecto, la Tabla 4.10 resume las principales métricas obtenidas en ambos escenarios.

Sujeto	Arquitectura	Configuración	Recall C0	Recall C1	F1-macro	ROC-AUC
P8	EEGNet	Sin pesos ponderados	0.36	0.93	0.64	0.638
P8	EEGNet	Con pesos ponderados	0.90	0.73	0.80	0.82

Cuadro 4.10: Comparación del rendimiento de EEGNet para el participante 8 con y sin uso de pesos ponderados.

Una vez analizado en detalle el impacto en este participante específico, se procede a evaluar si el comportamiento se mantiene de forma consistente en otros participantes. Para ello, se realiza con análisis comparativo sobre un subconjunto de participantes con los datos balanceados. Se han escogido 4 sobre los 8 totales.

Como se evidencia en la Figura 4.16, la implementación de pesos ponderados en la función de pérdida ha resultado en una mejora del rendimiento para la mayoría de los participantes. Este ajuste permite que el modelo individual de cada sujeto compense el desbalance de clases intrínseco a sus señales. El participante 8 destaca como el caso con mejores resultados absolutos, logrando elevar su f1-score desde un sólido 0.71 a 0.76. Por su parte, el participante 21 muestra el crecimiento relativo más significativo, pasando de un mal rendimiento de 0.43 a estar en la barrera de 0.60. El único caso donde no ha habido mejora y, de hecho, se ha mantenido el resultado, es con el participante 17, cuyo rendimiento se mantiene estático en 0.55. Finalmente, aunque el participante 18 experimenta una mejora, subiendo de 0.42 a 0.55, su rendimiento final es el más reducido del grupo. El hecho de que apenas supere el umbral del azar sugiere que la señal de este participante presenta desafíos que van más allá del balanceo de clases.

### 4.3. Pesos Ponderados

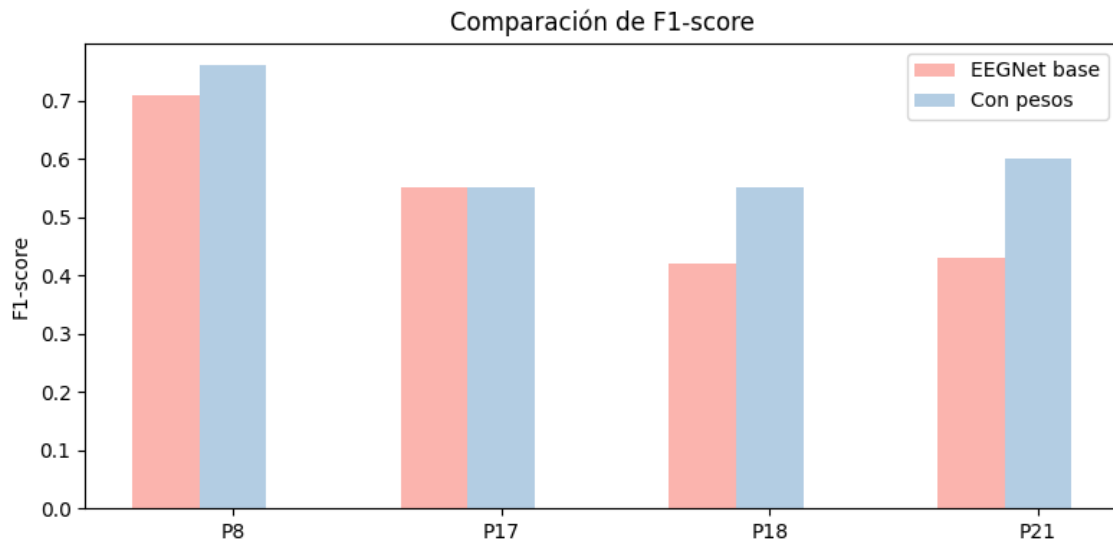


Figura 4.16: F1-score con y sin pesos

A la vista de la Figura 4.17, el recall macro se ha visto incrementado en todos los sujetos analizados. Esto indica que la introducción de pesos ponderados es efectiva para reducir el sesgo del modelo hacia la clase mayoritaria. Al igual que con el f1-score, el participante con mejor rendimiento es el 8, alcanzando un valor que roza el 0.80. En cuanto al impacto relativo, el ganador es el participante 21, logrando pasar de un recall ligeramente superior al azar a alcanzar un correcto 0.62. Finalmente, tanto el participante 17 como el 18 muestran mejoras marginales, manteniendo el recall en rangos moderados, lo que podría ser indicador de una señal de mala calidad.

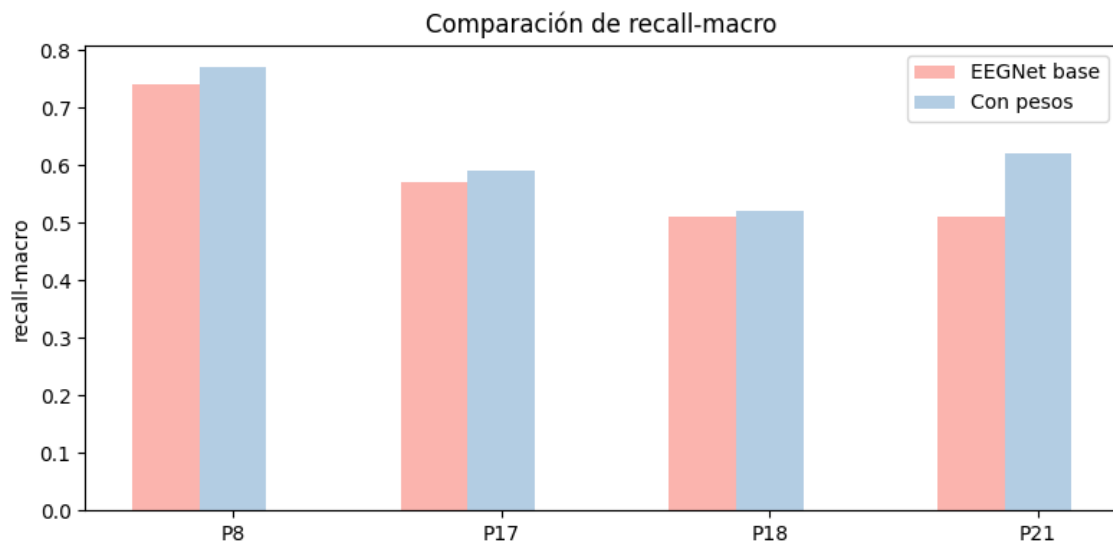


Figura 4.17: Recall-macro con y sin pesos

## Capítulo 4. Resultados

### Arquitectura EEGNeX

En la Figura 4.18 se puede observar que, al contrario que en EEGNet, el modelo inicial sin los pesos ponderados ya muestra un rendimiento muy alto. Esta red es capaz de extraer características más complejas, lo que permite alcanzar un alto rendimiento. El nivel de aciertos ya es muy alto en ambas clases, demostrando que la arquitectura es capaz de aprender la frontera de decisión de forma natural sin ayuda externa. El modelo inicial es tan eficiente que forzar a la función de pérdida a penalizar más los errores ha hecho que su rendimiento general descienda ligeramente.

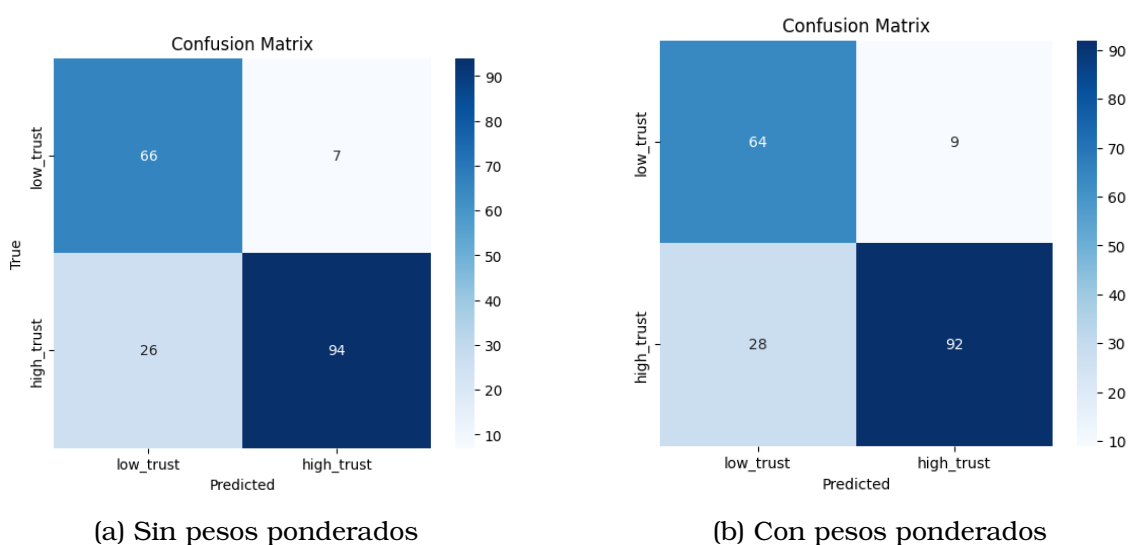


Figura 4.18: Comparación de matrices de confusión para el participante 8.

Al observar con más detalle las métricas obtenidas en la Tabla 4.11, podemos observar que al aplicar los pesos ponderados, hay una reconfiguración en la sensibilidad del modelo. Los falsos negativos han disminuido pero los falsos positivos han aumentado, sacrificando parte de la especificidad. Finalmente, el ROC-AUC ha aumentado ligeramente, lo que se traduce en una mejora en la separación de las clases.

Sujeto	Arquitectura	Configuración	Recall C0	Recall C1	F1-macro	ROC-AUC
P8	EEGNeX	Sin pesos ponderados	0.84	0.82	0.82	0.810
P8	EEGNeX	Con pesos ponderados	0.70	0.91	0.80	0.885

Cuadro 4.11: Comparación del rendimiento de EEGNeX para el participante 8 con y sin uso de pesos ponderados.

En cuanto al f1-score, en la Figura 4.19 el participante 21 es el caso donde el impacto es más evidente. El f1-score logra alcanzar pasar de un correcto pero mejorable 0.65 a un elevado 0.88, lo que significa que para este modelo esta arquitectura y la aplicación de los pesos logra una clasificación mucho más ro-

busta. El participante 17 logra una mejora marginal, logrando alcanzar un valor de 0.63. Este resultado supone un avance cualitativo, ya que el modelo transita desde un rendimiento cercano al azar con EEGNet a una clasificación más discriminativa con EEGNeX. Esto sugiere que el aumento de la complejidad en esta red es necesaria para detectar patrones en sujetos cuyas señales son más complejas. El participante 8, a partiendo de un sólido 0.82, ha experimentado una disminución marginal. El modelo inicial es tan eficiente por naturaleza que forzar una penalización mayor en la clase minoritaria lo desestabiliza ligeramente, haciendo que su f1-score descienda a 0.80. Finalmente, el participante 18 no mejora su rendimiento, obteniendo resultados por debajo del azar. Independientemente de la técnica y arquitectura utilizada, se mantiene como el sujeto con menor rendimiento. Estos resultados podrían sugerir que el problema podría estar en la calidad intrínseca de los datos o en la falta de patrones discriminativos en la señal.

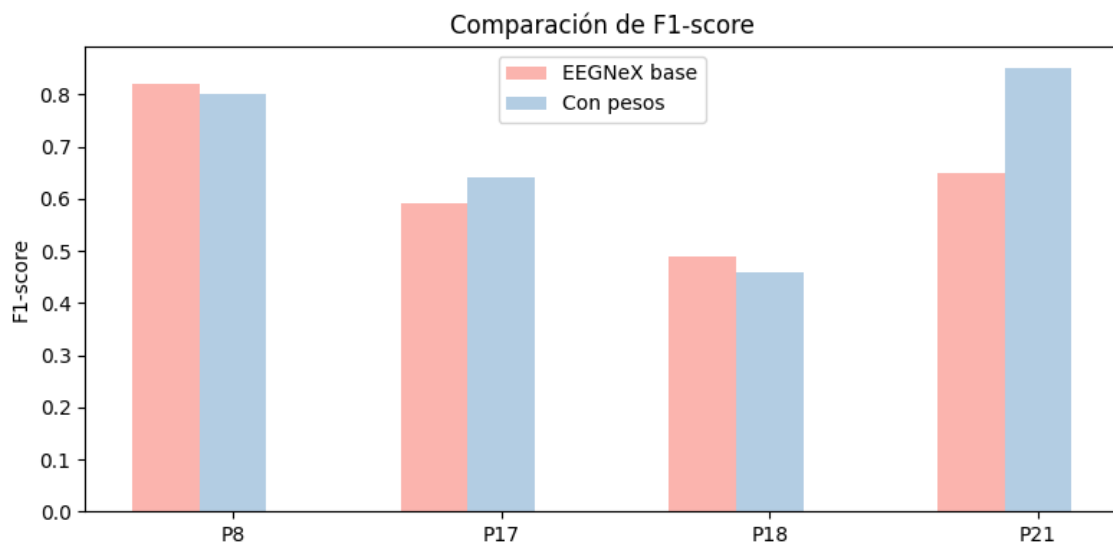


Figura 4.19: F1-score con y sin pesos

Al igual que con el f1-score, en la Figura 4.20 se puede observar que el participante 21 vuelve a ser el punto más destacado de la gráfica, pasando de un recall de 0.64 a superar el 0.80 tras la aplicación de los pesos. El participante 8, por su parte, ya mostraba un rendimiento muy alto antes de la aplicación de la técnica, experimentando una disminución en el recall, aunque aún manteniendo un rendimiento alto. Por otro lado, el participante 17 logra una mejora marginal, pasando de 0.61 a 0.63, lo que supone una mejora respecto a la arquitectura anterior. Finalmente, el participante 18 se mantiene estancado en el umbral del azar, tanto con pesos como sin ellos. Lo que demuestra, una vez más, lo importante que es tener en consideración la variación intra-sujeto.

## Capítulo 4. Resultados

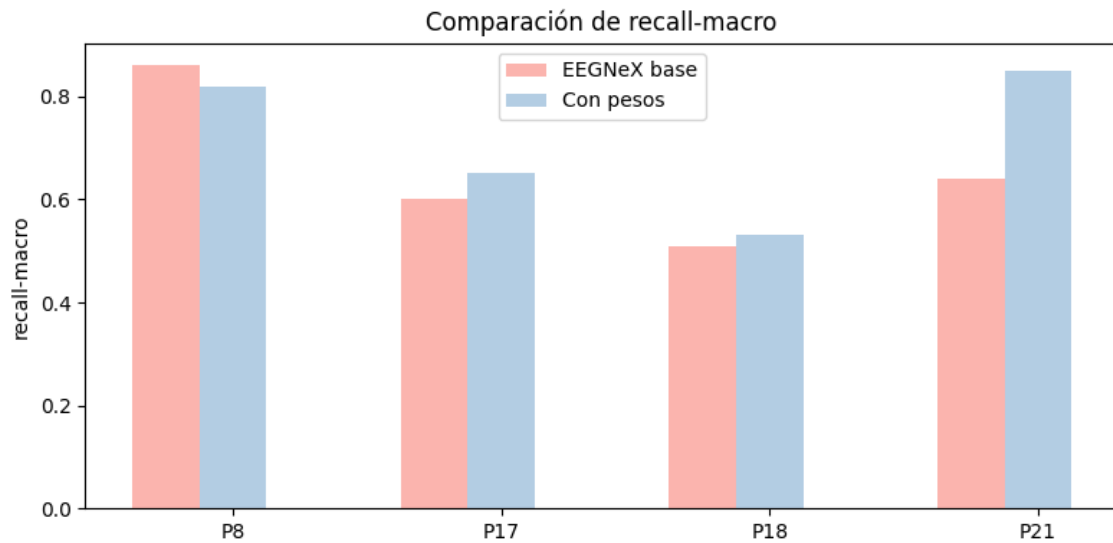


Figura 4.20: Recall macro con y sin pesos

### Arquitectura AttentionBaseNet

En línea con lo observado en la arquitectura EEGNeX, en la Figura 4.21 se puede observar que el participante 8 ya mostraba métricas sobresalientes antes de añadir los pesos ponderados. No obstante, la inclusión de esta técnica permitió un refinamiento adicional a la clasificación. Se observa que los aciertos de la clase minoritaria ascienden de 66 a 79, logrando una reducción significativa de los falsos positivos. El rendimiento de la clase mayoritaria se mantiene prácticamente intacto.

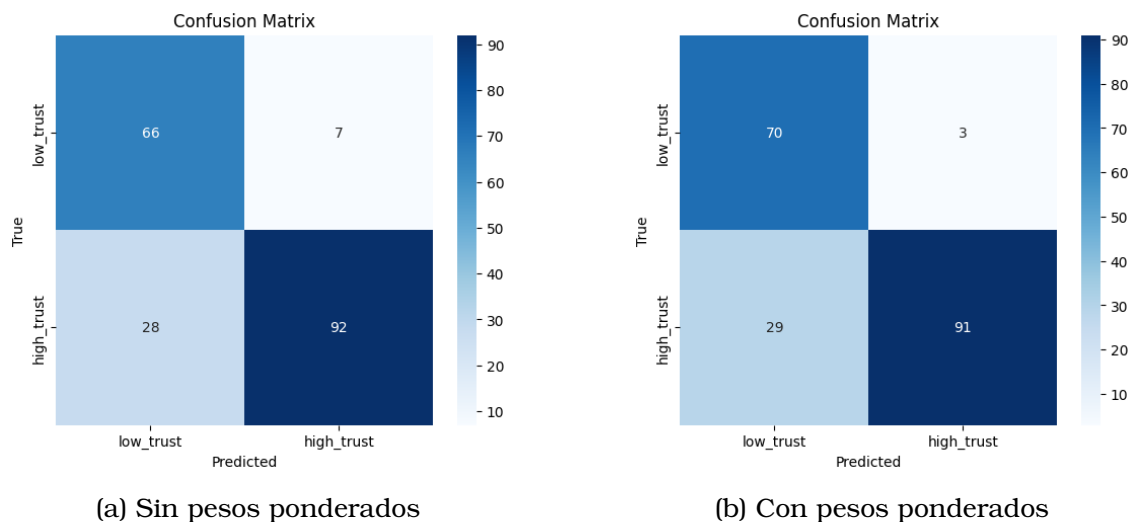


Figura 4.21: Comparación de matrices de confusión para el participante 8.

A pesar de incrementar la prioridad de la clase minoritaria mediante la función de pérdida teniendo ya un buen rendimiento inicial, el modelo demuestra esta-

### 4.3. Pesos Ponderados

bilidad. A la vista de la Tabla 4.12, las métricas mantienen valores similares, manteniendo la estabilidad del modelo.

Sujeto	Arquitectura	Configuración	Recall C0	Recall C1	F1-macro	ROC-AUC
P8	AttentionBaseNet	Sin pesos ponderados	0.92	0.78	0.82	0.893
P8	AttentionBaseNet	Con pesos ponderados	0.96	0.76	0.83	0.901

Cuadro 4.12: Comparación del rendimiento de EEGNeX para el participante 8 con y sin uso de pesos ponderados.

Como se podrá observar en la Figura 4.22, el participante 8 es el que logra los mejores resultados globales de la muestra. Partiendo ya de un valor elevado de 0.82, alcanza un incremento ligero tras la introducción de los pesos ponderados, lo que muestra que la arquitectura de atención es capaz de refinar su frontera de decisión sin desestabilizar el modelo. Por su parte, el participante 21 es el que experimenta un salto de rendimiento más notable en términos de esta métrica, incrementando su f1-score de 0.72 a 0.80. A diferencia de los resultados obtenidos en la primera red, este participante es el que ha logrado una más significativa. Aunque el rendimiento del participante 17 es menor, se observa una progresión sólida de 0.54 a superar el 0.60, lo que demuestra una mejora del modelo para identificar patrones en la señal al prestar más atención a la clase minoritaria. Finalmente, al igual que con el resto de modelos, el participante 18 se mantiene estancado, llegando apenas al umbral del azar.

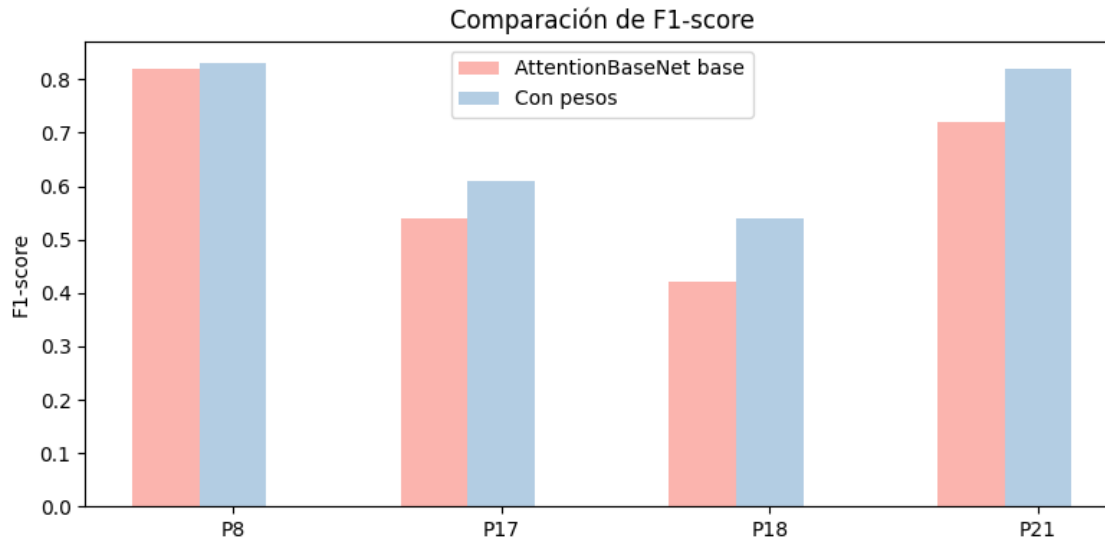


Figura 4.22: F1-score con y sin pesos

Al igual que en métricas anteriores, en la Figura 4.23 se observa que el participante 8 se mantiene líder en los resultados, manteniendo un rendimiento estable y sobresaliente, logrando un ligero aumento en el recall-macro hasta alcanzar el 0.86. El participante 21 logra una mejora sustancial, logrando aumentar su recall-macro hasta 0.82. Este incremento es coherente con la tendencia vista en el f1-score. Por otro lado, el participante 17 vuelve a mostrar un ligero aumento

## Capítulo 4. Resultados

en la métrica, superando el 0.6, un resultado estable. La adición de los pesos ha logrado rescatar información relevante que en un principio se estaba perdiendo por la falta de atención prestada a la clase minoritaria. Por último, el participante 18 muestra una ligera mejora, alcanzando un valor de 0.56, manteniéndose en la línea vista a lo largo del análisis..

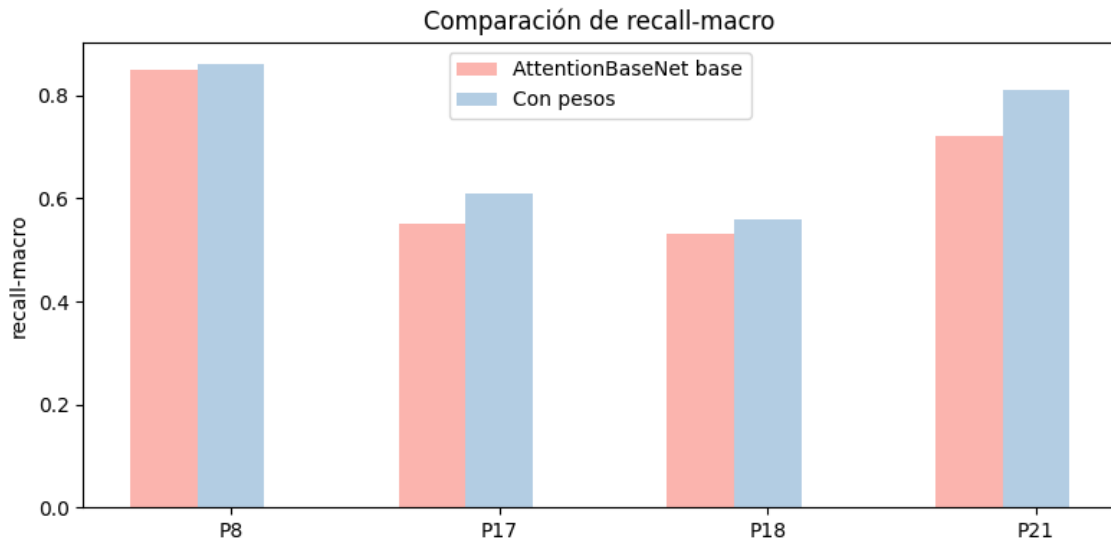


Figura 4.23: Recall macro con y sin pesos

A continuación, en la Tabla 4.13 se muestran los resultados de aplicar la técnica de pesos ponderados a todos los participantes balanceados, incluidos los que no se han mostrado a lo largo del análisis:

Participante	Modelo	Pesos Ponderados	Recall (macro)	F1-score (macro)
P4	EEGNeX	Aplicados	0.71	0.71
P7	EEGNeX	No aplicados	0.68	0.66
P8	EEGNeX	No Aplicados	0.83	0.82
P9	EEGNet	Aplicados	0.58	0.57
P17	EEGNeX	Aplicados	0.63	0.63
P18	AttBaseNet	Aplicados	0.56	0.54
P21	EEGNeX	Aplicados	0.89	0.88
P22	AttBaseNet	No aplicados	0.69	0.68

Cuadro 4.13: Impacto de los pesos ponderados en participantes balanceados.

Los sujetos 8 y 21 obtuvieron los resultados más destacados, empleando la arquitectura EEGNeX y la aplicación de pesos ponderados. El 4 exhibe un rendimiento notable, equiparable al rendimiento del 17 y del 22, aunque con una ligera disminución en su desempeño. En el análisis realizado, los participantes 9 y 18 exhiben una actitud crítica. A pesar de implementar pesos y evaluar diver-

sas arquitecturas, no han logrado alcanzar el 0.60 en ninguna de las métricas evaluadas. Si bien EEGNet es la opción más simple, ha demostrado ser la más efectiva en términos de estabilidad para el participante 18, cuya señal ha presentado una complejidad significativa en su análisis, limitando así su capacidad para alcanzar un rendimiento óptimo. Tanto el f1-score como el recall macro se mantienen estables por igual, lo que sugiere que el modelo no adivina simplemente, sino que logra un equilibrio real entre la detección de ambas clases.

### 4.3.2. Participantes desbalanceados

Para los participantes desbalanceados también se probó a añadir la técnica de los pesos ponderados, forzando a la red a prestar mayor atención a la clase minoritaria. Para evaluar el rendimiento del modelo con la aplicación de esta técnica, se procederá a la comparación de los mejores resultados obtenidos para cada participante en la sección de oversampling + focal loss. Este procedimiento permite discernir qué configuración de la función de pérdida resulta más óptima en cada perfil de datos.

Teniendo en cuenta que tanto los pesos ponderados como la focal loss modifican la función de pérdida, estas técnicas no se han combinado. Utilizar ambas técnicas al mismo tiempo podría resultar redundante, provocando una sobrecorrección en el modelo del sesgo de la clase. En consecuencia, el gradiente de la clase minoritaria quedaría saturado, provocando un sobreajuste que impediría la generalización del modelo en datos reales.

### Impacto en la red EEGNet

Como se evidencia en la Gráfica 4.24, el modelo base presenta un rendimiento superior de manera consistente frente a los modelos a los que se les ha aplicado los pesos ponderados. En todos los participantes se observa una degradación significativa del f1-score, lo que sugiere que la simple ponderación de la función de pérdida no es suficiente para sujetos con clases desbalanceadas.

## Capítulo 4. Resultados

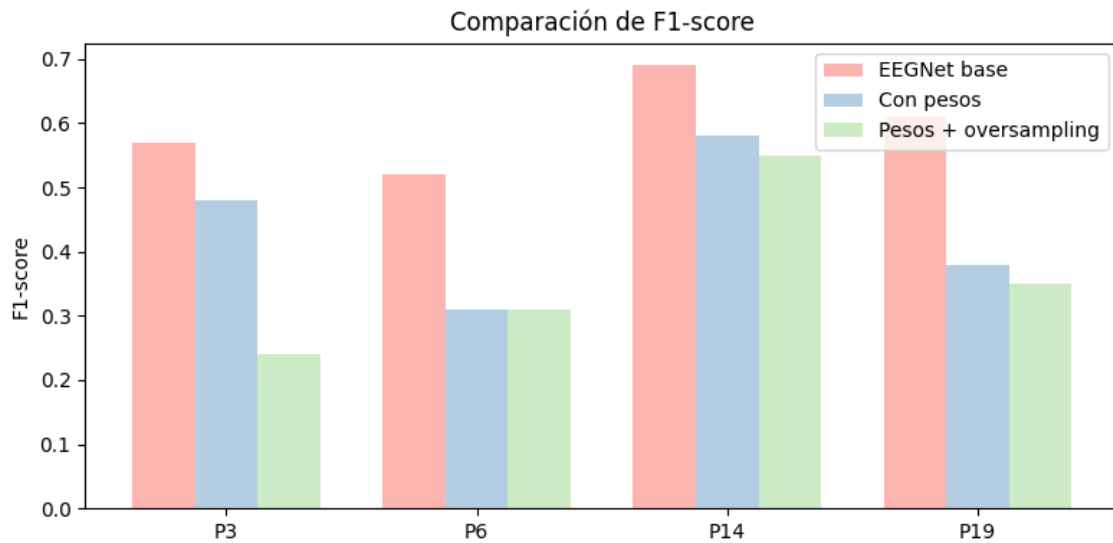


Figura 4.24: Impacto en el f1-score con pesos ponderados para participantes desbalanceados con EEGNet

En cuanto a la métrica de ROC-AUC, en la Gráfica 4.25, en ninguno de los casos la combinación de pesos + oversampling supera al modelo base. Las técnicas de oversampling aplicadas (con o sin Focal Loss) superan en todos los casos a los pesos ponderados.

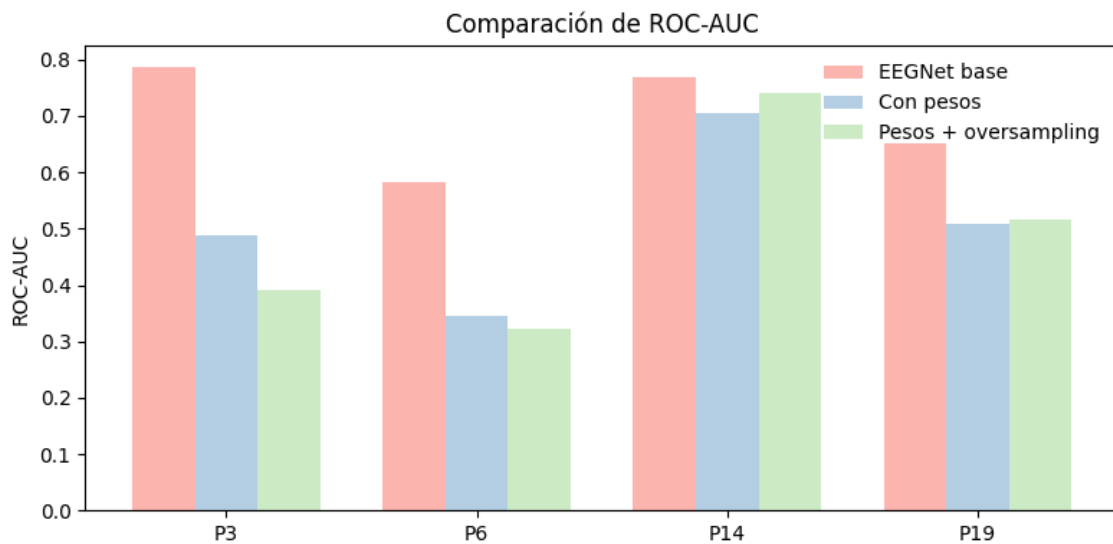


Figura 4.25: Impacto en el ROC-AUC con pesos ponderados para participantes desbalanceados con EEGNet

### Impacto en la red EEGNeX

En el caso de la arquitectura EEGNeX, en la Gráfica 4.26 se evidencia una tendencia similar a la observada en EEGNet, aunque sin presentar caídas tan drásticas. El modelo base sigue siendo superior a los que se les ha aplicado los pesos ponderados, lo que indica que la introducción de pesos introduce un sesgo que perjudica la precisión global del sistema.

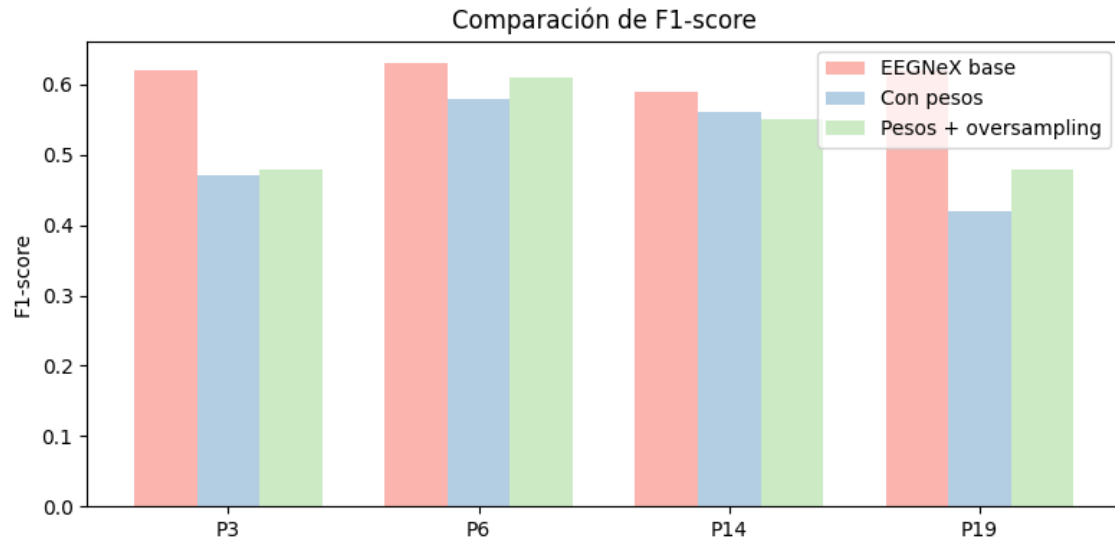


Figura 4.26: Impacto en el f1-score con pesos ponderados para participantes desbalanceados con EEGNeX

Al igual que pasaba con EEGNet, en la Gráfica 4.27 se puede observar que el participante 14 es el único que se ha beneficiado de los pesos ponderados junto con el oversampling. Esta tendencia puede sugerir que, en determinadas distribuciones de datos, el refuerzo de la clase minoritaria bajo esta técnica puede mejorar la separabilidad entre clases, aunque esto conlleve en un aumento de los falsos positivos.

## Capítulo 4. Resultados

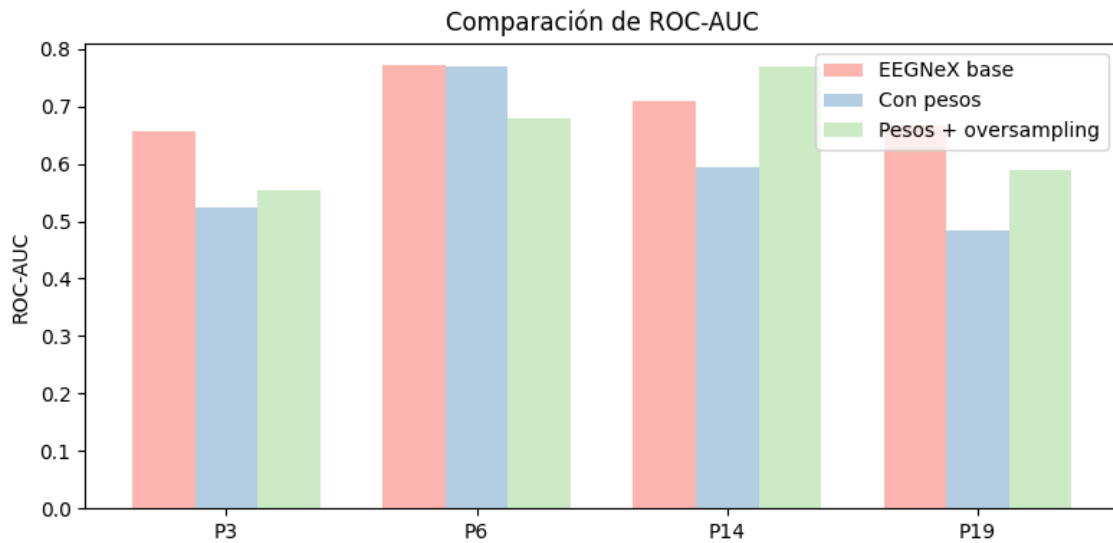


Figura 4.27: Impacto en el ROC-AUC con pesos ponderados para participantes desbalanceados con EEGNeX

### Impacto en la red AttentionBaseNet

Finalmente, en la arquitectura AttentionBaseNet, en la Gráfica 4.28 queda patente que el impacto es mayoritariamente negativo. Los modelos base superan a los pesos ponderados, lo que indica que los mecanismos de atención de la red parecen verse perjudicados por la penalización de las clases a través de ese método.

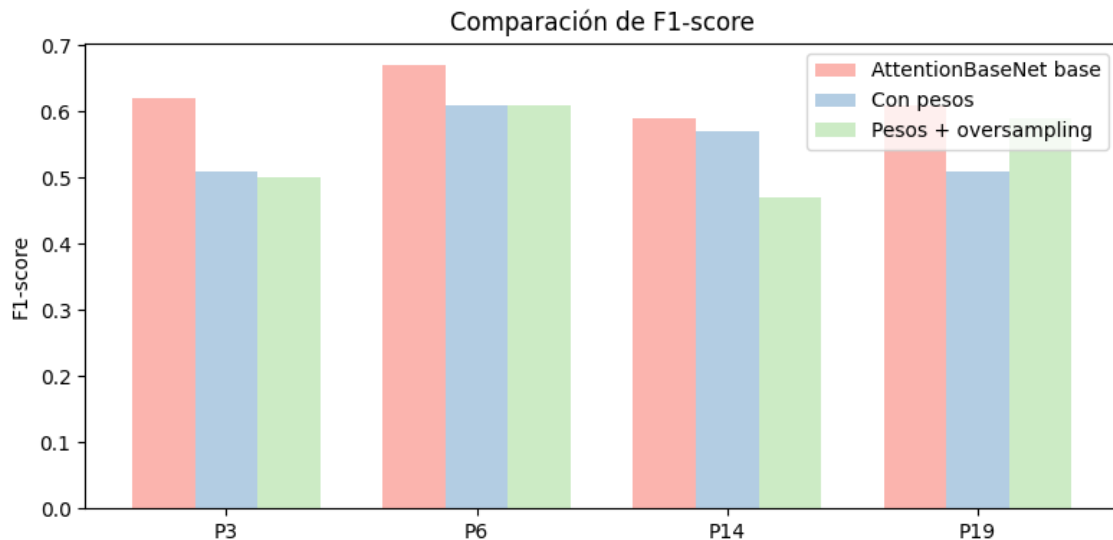


Figura 4.28: F1-score pesos

En cuanto al ROC-AUC, se puede observar en la Gráfica 4.29 una tendencia similar a la métrica del f1-score. Todos los participantes obtienen un rendimiento

menor por la aplicación de los ponderados.

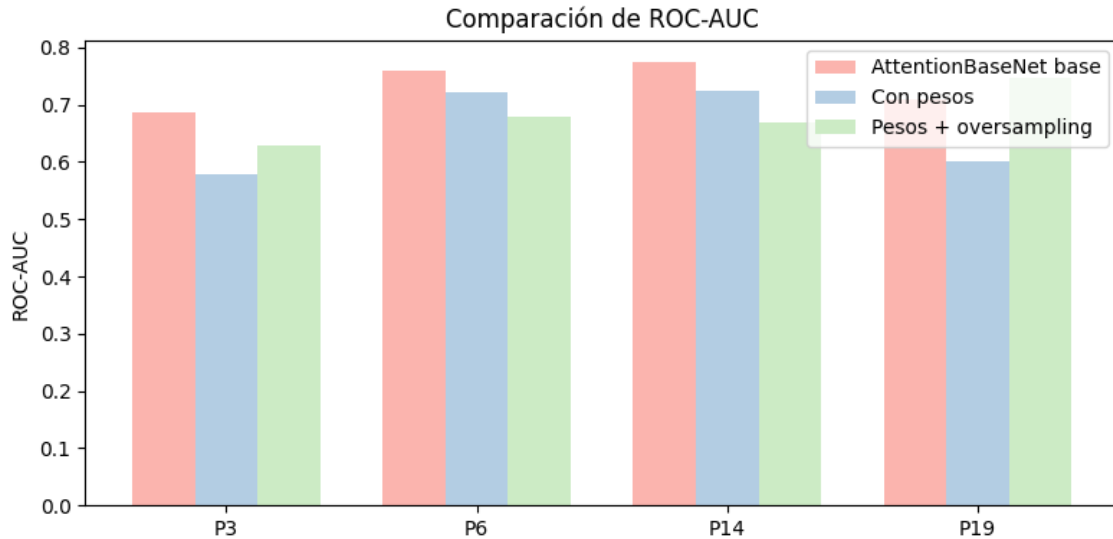


Figura 4.29: F1-score pesos

El descenso generalizado en todas las arquitecturas y todos los participantes indica que la técnica de los pesos ponderados y el oversampling están sesgando el modelo. Los ponderados aplican una penalización estática a las muestras de la clase minoritaria, obligando al modelo a prestar más atención a esa clase. Sin embargo, en escenarios con un desbalance severo, el uso conjunto de pesos ponderados y oversampling puede favorecer el sobreajuste a la clase minoritaria, ya que un conjunto reducido de ejemplos concentra una influencia excesiva en la función de pérdida. Por otro lado, la focal loss reduce la influencia de aquellas muestras que el modelo clasifica correctamente, concentrando el aprendizaje en los ejemplos difíciles. Es decir, los pesos ponderados penalizan las clases; el focal loss prioriza la dificultad, lo que permite un entrenamiento más estable y explica mejor el rendimiento obtenido con esta técnica.

### 4.4. Selección de características

Una vez estabilizados los modelos mediante técnicas de balanceo y de optimización de las funciones de pérdida, se procedió a evaluar su rendimiento mediante la reducción del espacio de entrada. Para ello, se seleccionaron los canales más informativos que más permitan mejorar la generalización. El objetivo es determinar si la reducción de los 19 canales originales a un subconjunto más informativo de 10 canales permite mejorar la generalización de los modelos.

Para cada grupo de participantes, se presenta un gráfico de barras que compara el f1-score obtenido por el modelo base y el alcanzado tras el proceso de selección de características. Además, para el grupo de balanceados también se presentan las gráficas del recall y para los desbalanceados el ROC-AUC. En estas gráficas, para cada sujeto, sobre la barra que representa el resultado optimizado, se incluye una etiqueta que identifica el método de selección con mejor desempeño: **'A'** para ANOVA, **'MI'** para Mutual Information y **'H'** para el enfoque híbrido (Mutual Information y Sequential Feature Selection). El modelo base constituye el punto de partida para la obtención de resultados óptimos, tras la implementación de las técnicas pertinentes. Este enfoque se centra en la evaluación del rendimiento relacionado con la eliminación de canales, permitiendo una evaluación precisa y objetiva.

Esta sección se ha estructurado agrupando en base a la arquitectura empleada, de manera que se pueda comprobar cómo cada arquitectura filtra la señal al eliminar algunos de los canales.

#### 4.4.1. Impacto en la arquitectura EEGNet

##### Participantes balanceados

En el caso de la arquitectura EEGNet, como se puede observar en la Gráfica 4.30 los modelos sufren tras la aplicación de los distintos métodos de selección de características.

El caso más crítico corresponde al participante 8, donde el f1-score experimenta una disminución significativa después de la aplicación de la selección ANOVA (A), pasando de un valor razonable de 0.77 a un valor mucho menor de 0.45, lo cual es incluso inferior al azar. De manera similar, el participante 18 evidencia una ligera disminución, mucho menos drástica. Este resultado sugiere que, en sujetos relativamente estables, la pérdida de algunos canales puede estar asociada con una disminución en la capacidad de generalización. El participante 17 experimenta una ligera mejora gracias a la técnica híbrida, lo que sugiere que al eliminar algunos de los canales, se ha eliminado ruido de la señal, permitiendo al modelo alcanzar un mejor rendimiento. Por último, el sujeto 21, muestra una mejora en el resultado de la aplicación de esta técnica, experimentando un ligero incremento de 0.60 a 0.65 mediante el método Mutual Information.

#### 4.4. Selección de características

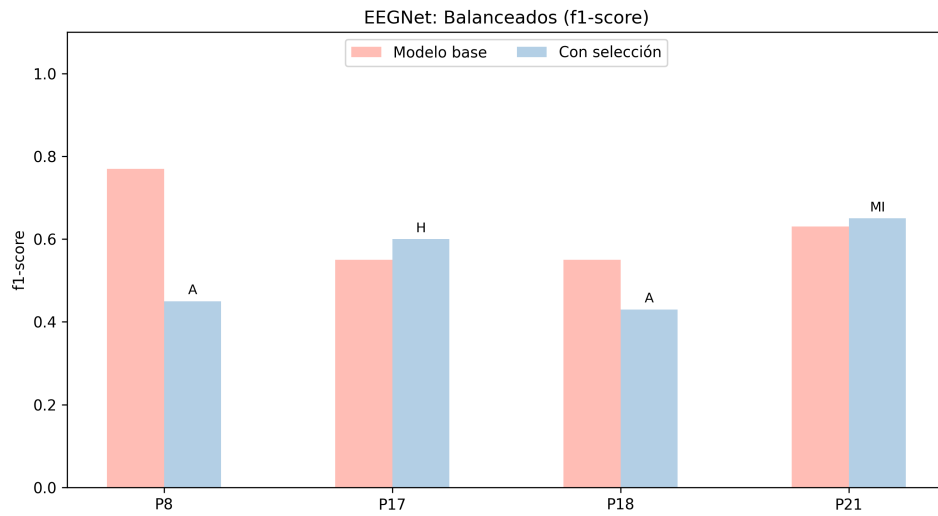


Figura 4.30: f1-score para participantes balanceados con selección de características

En cuanto al recall, en la Gráfica 4.31 se aprecia que el participante 8 sigue experimentando la caída más drástica con la métrica recall. Esto corrobora que el modelo pierde por completo la capacidad de identificar patrones distintivos de cada clase, aumentando drásticamente los falsos negativos. En cuanto al participante 17, el recall-macro muestra un leve incremento mediante la técnica híbrida, ascendiendo de un 0.57 a 0.62. Este hallazgo sugiere que la técnica ayudó al modelo a identificar mejor las clases. El 18 se mantiene estancado en valores alrededor del 0.50, similar al f1-score. Finalmente, el participante 21 es el que continúa viendo experimentando los efectos positivos de esta técnica, alcanzando un valor de 0.75. Junto al 17, son los únicos casos en los que la selección de características logra limpiar la señal al eliminar canales más redundantes.

## Capítulo 4. Resultados

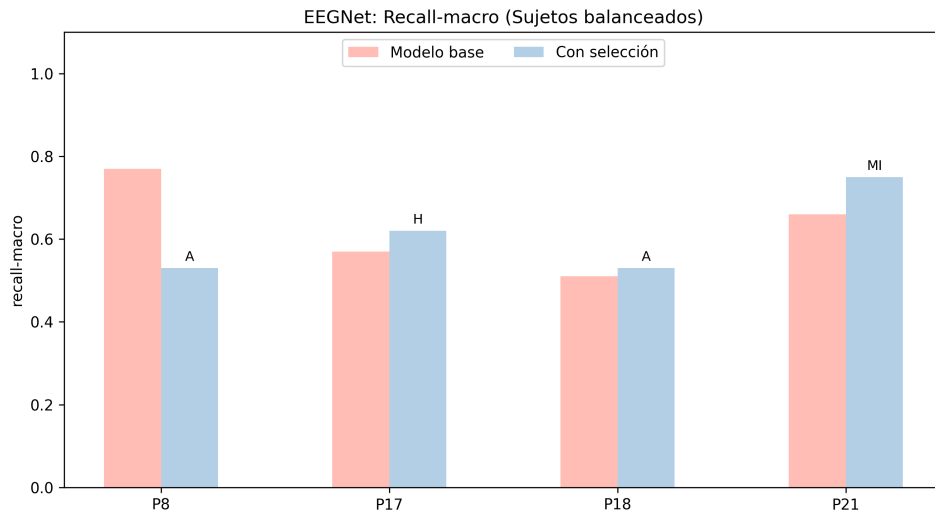


Figura 4.31: Recall para participantes balanceados con selección de características

### Participantes desbalanceados

En el caso del grupo de los desbalanceados, como se evidencia en la Figura 4.32, la selección de características no logra generar mejoras notables. El participante 3 muestra una evidencia de incremento marginal a través del método híbrido. En el caso del sujeto 6, se observa una disminución del rendimiento con Mutual Info, lo que sugiere que para este sujeto, la selección de características es completamente contraproducente. Para el participante 14, partiendo de un valor inicial de 0.69, la aplicación del método ANOVA resultó en una disminución hasta 0.45. En consecuencia, para este sujeto también es preferible mantener todos los canales. En última instancia, el participante 19 experimentó una disminución en su rendimiento con la implementación de la aplicación mutualinfo, situándose incluso por debajo del umbral de aleatoriedad establecido.

#### 4.4. Selección de características

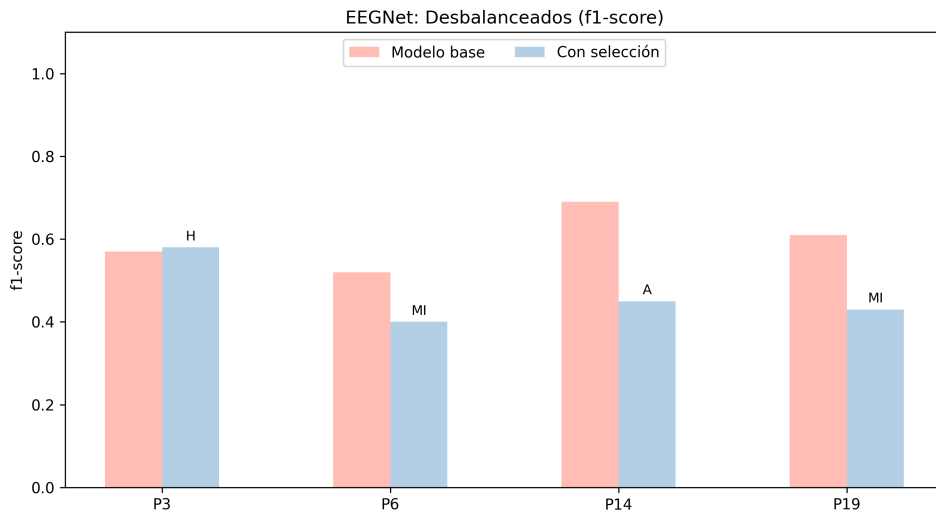


Figura 4.32: f1-score para participantes desbalanceados con selección de características

Mientras que el f1-score se mantuvo estable para el participante 3, en la Figura 4.33 se aprecia que el ROC-AUC sufrió una caída desde un notable 0.79 hasta un 0.54 con el método híbrido. Esta discrepancia entre métricas sugiere que, si bien el modelo mejoró ligeramente su precisión global, redujo de manera significativa la capacidad de la red para distinguir entre clases. Por su parte, el participante 6 sufre una disminución de rendimiento comparable a la del f1-score, situándose por debajo del umbral de la aleatoriedad. En última instancia, el participante 19, es el único que experimenta una ligera mejora en el ROC-AUC, volviéndose el modelo más robusto para la separación de clases a costa de perder parte de la precisión global.

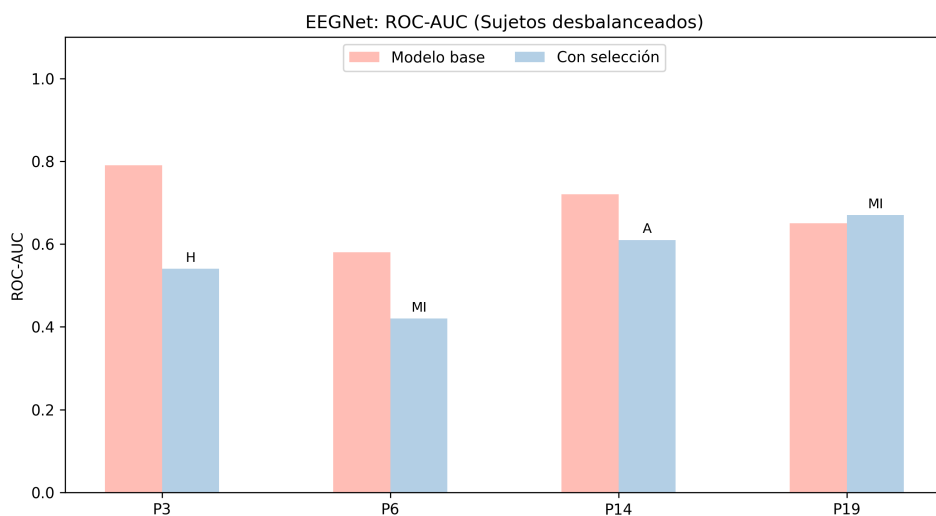


Figura 4.33: ROC-AUC para participantes desbalanceados con selección de características

## Capítulo 4. Resultados

En consecuencia a los resultados obtenidos, es posible afirmar que EEGNet posee una alta dependencia del uso de todos los canales para funcionar correctamente. En los dos grupos analizados, se observó una tendencia generalizada a la disminución del rendimiento. En los casos en los que se observó alguna mejora, fue muy ligera. Este análisis sugiere que, dado que EEGNet es una red compacta y eficiente, imponerle una selección de características suele resultar en una pérdida de rasgos discriminativos esenciales, dado que no es capaz de gestionar adecuadamente la reducción de dimensionalidad, principalmente con el grupo de participantes desbalanceados.

### 4.4.2. Impacto en la arquitectura EEGNeX

#### Participantes balanceados

Como se evidencia en la Figura 4.34, la arquitectura EEGNeX exhibe unos resultados significativamente más robustos que la red EEGNet. El participante 8 alcanza su máximo rendimiento tras la aplicación de la selección de características ANOVA, a pesar de que parte de un resultado muy alto de un 0.82. logra alcanzar el 0.86, una mejora ligera, pero importante. Esto confirma que para señales de alta calidad, un filtrado lineal sencillo es útil para resaltar los rasgos más discriminativos de la arquitectura. Por su parte, el participante 18 evidencia una disminución marginal haciendo uso del método Mutual Information, manteniendo el ritmo constante que se ha visto a lo largo del análisis. El participante 21 sufre una ligera disminución del f1-score tras la aplicación del método MI. Partía de un valor muy alto de 0.88 y desciende a 0.81. Esto sugiere que, aunque el resultado con la selección sigue siendo alto, la disminución de canales puede estar eliminando pequeñas variaciones espaciales que el modelo base estaba usando para generalizar.

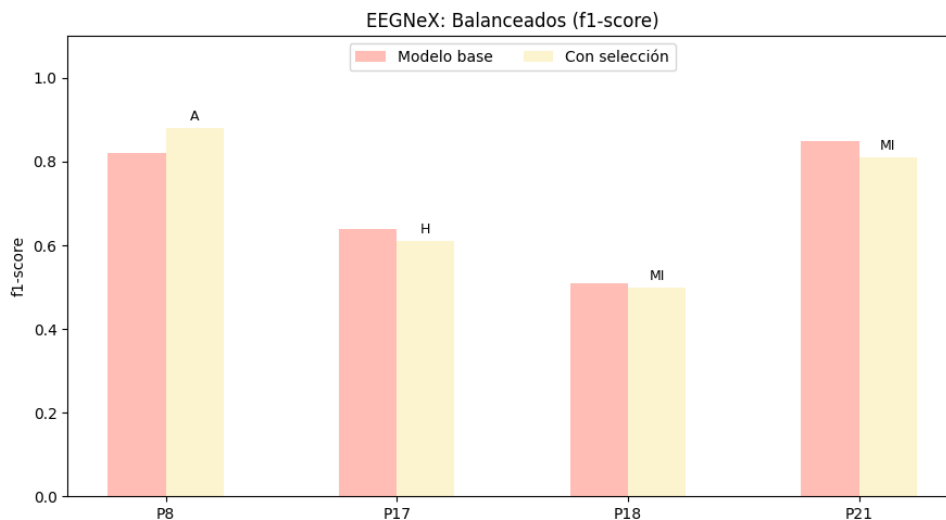


Figura 4.34: f1-score para participantes balanceados con selección de características

#### 4.4. Selección de características

En la Gráfica 4.35 se aprecia que el modelo para el participante 8 no solo ha mejorado su precisión global, sino su capacidad para distinguir la clase minoritaria, alcanzando un recall-macro de 0.87. Para el participante 21, el modelo pierde sensibilidad, disminuyendo su valor de 0.89 hasta 0.82, aumentando los falsos negativos. Finalmente, para los participantes 17 y 18, la variación de esta métrica es similar a la del f1-score: prácticamente nula. Lo que sugiere que la red captura la misma información con menos electrodos sin perder sensibilidad.

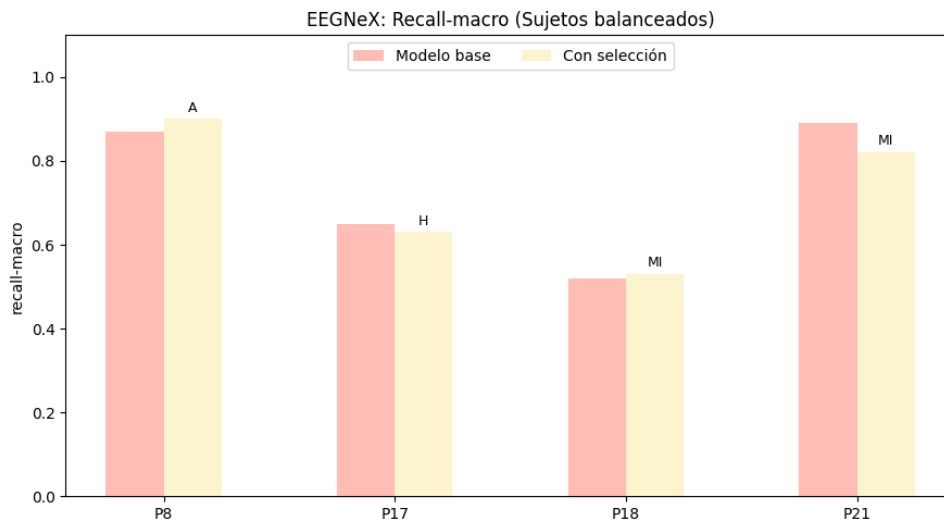


Figura 4.35: Recall para participantes balanceados con selección de características

Al contrario que con EEGNet, con EEGNeX no se han observado caídas drásticas con la aplicación de la selección de características, además de alcanzar un resultado sobresaliente con el participante 8, sugiriendo que a partir de una señal limpia, esta arquitectura aprovecha la eliminación de canales redundantes. El aumento de la complejidad del método no se traduce en una mejora de los resultados. Algunos lograron un incremento de la métrica con la técnica más simple (ANOVA) y otros con la compleja, demostrando una vez más la diferencia inter-sujeto en las señales.

## Capítulo 4. Resultados

### Participantes desbalanceados

En el grupo de sujetos desbalanceados, como se puede observar en la Gráfica 4.36, las mejoras y caídas son marginales. El participante 6 logra su mejor rendimiento hasta ahora, alcanzando un f1-score 0.65, lo que significa un resultado muy positivo para este grupo. El 14 experimenta un aumento similar, pero quedándose en el 0.57. Interesante observar cómo a pesar de la diferencia de complejidad entre ambas técnicas, siendo la híbrida más compleja, no hay una diferencia grande en cuanto a mejora de rendimiento. Por el contrario, tanto el participante 3 como el 19 sufren una ligera caída.

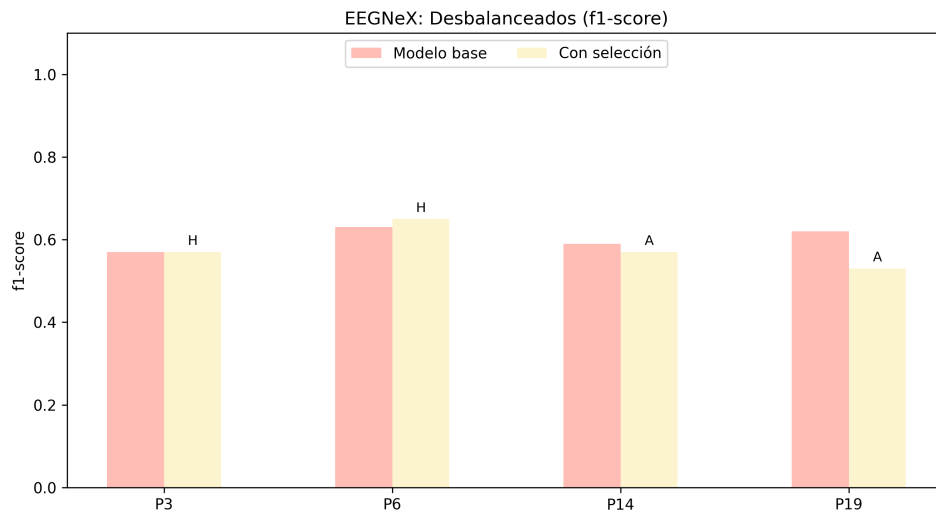


Figura 4.36: F1-score para participantes desbalanceados con selección de características

En el caso del ROC-AUC, en la Gráfica 4.37 se observa que el único participante que no ha empeorado con la selección de características es el sujeto 3, quien ha logrado un incremento marginal. Los demás participantes empeoran su rendimiento con la técnica, especialmente el participante 19, quien sufre una caída drástica, pasando de 0.67 hasta 0.49. Esto sugiere que para estos participantes los modelos requieren la información proveniente de los 19 canales para la diferenciación de los datos, son especialmente sensibles a la pérdida de información.

## 4.4. Selección de características

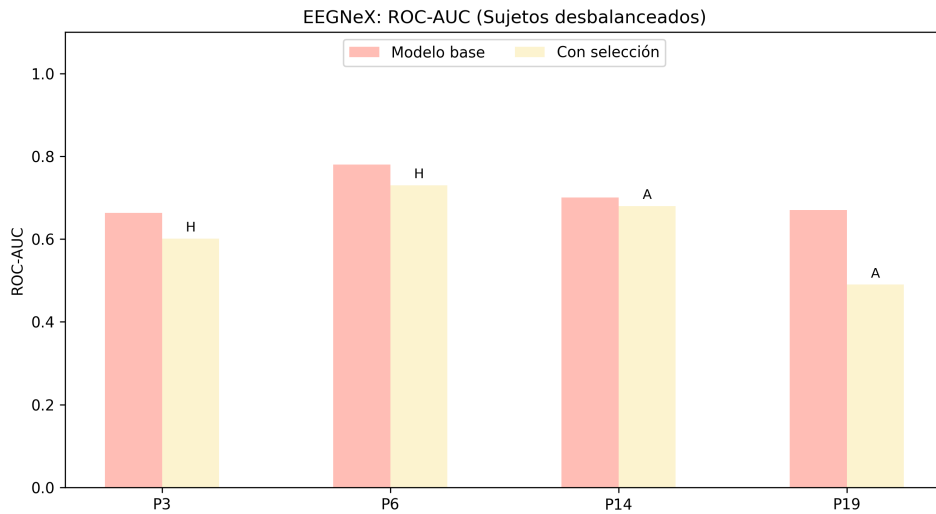


Figura 4.37: ROC-AUC para participantes desbalanceados con selección de características

### 4.4.3. Impacto en la arquitectura AttentionBaseNet

Para concluir con el análisis de la selección de características, se evalúa la arquitectura **AttentionBaseNet** bajo la influencia de esta técnica.

#### Participantes balanceados

La arquitectura AttentionBaseNet muestra en la Gráfica 4.38 una respuesta altamente variable, lo que sugiere una alta dependencia a la señal del sujeto analizado. El participante 8 vuelve a lograr una ligera mejora a pesar de partir de un resultado sobresaliente, alcanzando un f1-score de 0.86. Por su parte, el participante 18 sufre una ligera disminución en esta métrica con la técnica híbrida. La aplicación de la técnica ANOVA en el participante 17 provoca una degradación del rendimiento de ésta, bajando de un 0.62 a 0.57. El participante 21 es el que sufre la caída más drástica, bajando de un f1-score de 0.71 a 0.61, lo que sugiere que para este participante, el modelo necesita de la información de los 19 canales para alcanzar un rendimiento óptimo.

A pesar de la mejora del P8, estos resultados sugieren que AttentionBaseNet funciona más correctamente con todos los canales, aprovechando la información dada por cada uno de ellos.

## Capítulo 4. Resultados

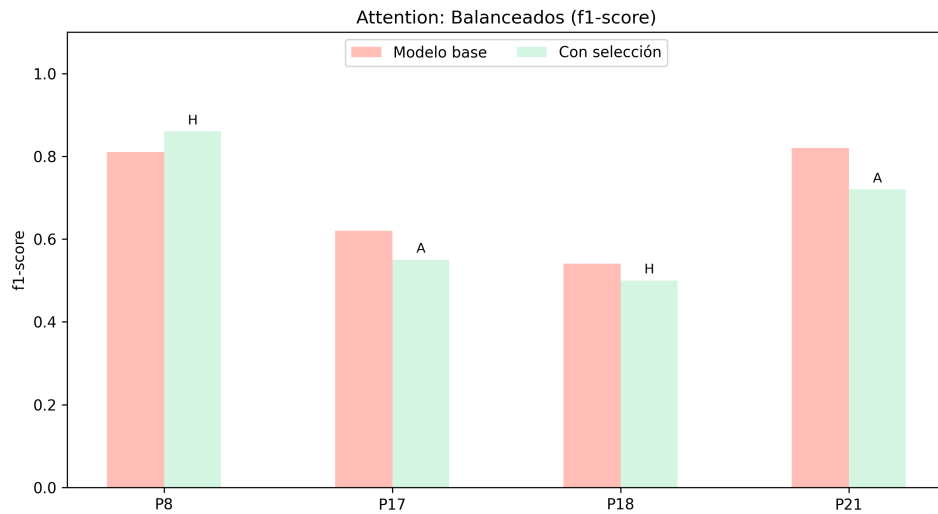


Figura 4.38: F1-score para participantes balanceados con selección de características

En cuanto a la métrica del recall, en la Gráfica 4.39 se observa que el participante 18 evidencia una mejora marginal, pero aún manteniéndose en el intervalo del 0.5, quedándose estancado en el umbral del azar. Los participantes 17 y 18 mantienen un resultado similar al modelo base, ambos sufriendo una ligera caída tras la aplicación de la técnica, por lo que la selección de características no logra extraer un beneficio claro de la reducción de canales. Finalmente, el 21 sufre también un descenso en esta métrica, más drástico que el 17, lo que sugiere que para este participante es necesario aplicar la información de los 19 canales para lograr un rendimiento óptimo.

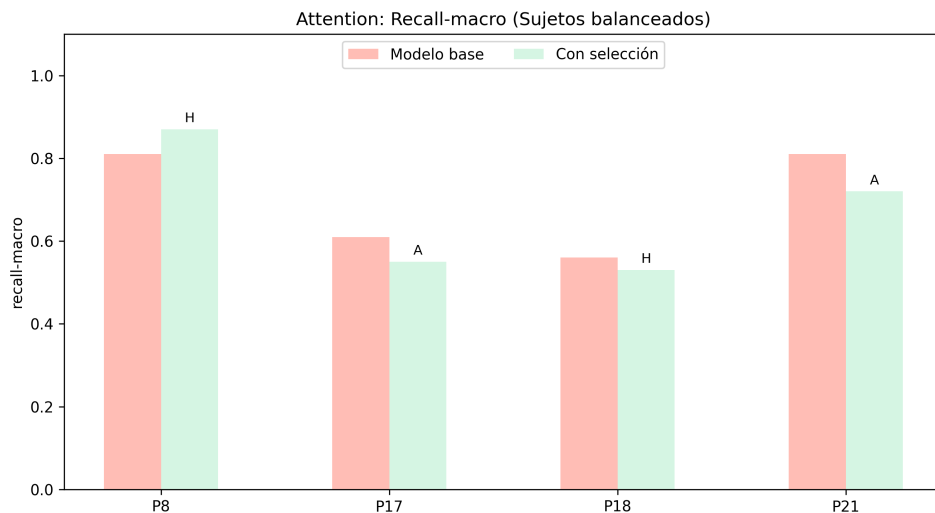


Figura 4.39: Recall para participantes balanceados con selección de características

### Participantes desbalanceados

Dentro de esta arquitectura, como se aprecia en la Gráfica 4.40, el impacto de la selección de características en el grupo de los participantes desbalanceados ha resultado en un mal desempeño total para todos los participantes. A diferencia de las otras arquitecturas, donde alguno de los participantes lograba incrementar su resultado, en este caso ninguno de los modelos base ha visto una mejora con la reducción de canales, lo que sugiere que la eliminación de éstos priva al modelo de información contextual necesaria para gestionar el desbalanceo.

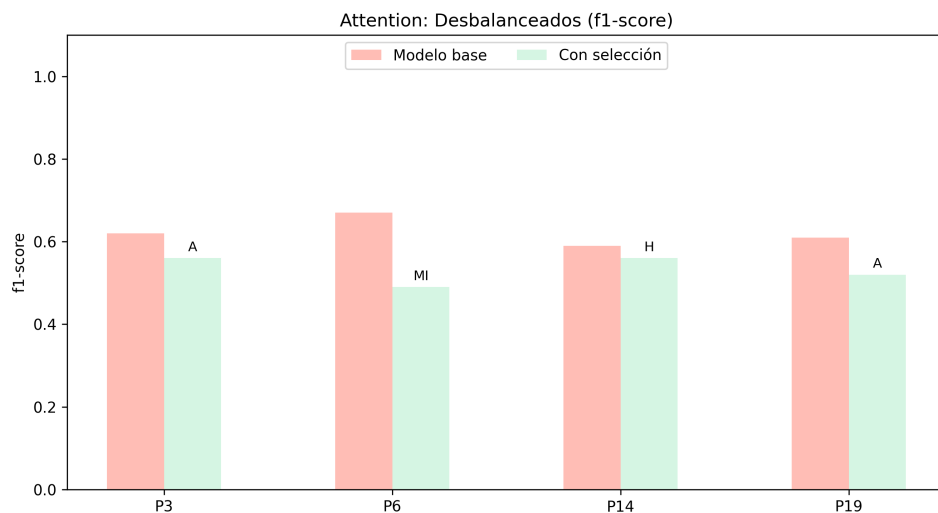


Figura 4.40: f1-score para participantes desbalanceados con selección de características

Finalmente, en la Gráfica 4.41 se puede apreciar que el que menos sufre la aplicación de la selección es el participante 19, pero los demás participantes tienen una caída del rendimiento importante, lo que descarta por completo la aplicación de estas técnicas de selección de características para la arquitectura AttentionBaseNet con los participantes desbalanceados.

## Capítulo 4. Resultados

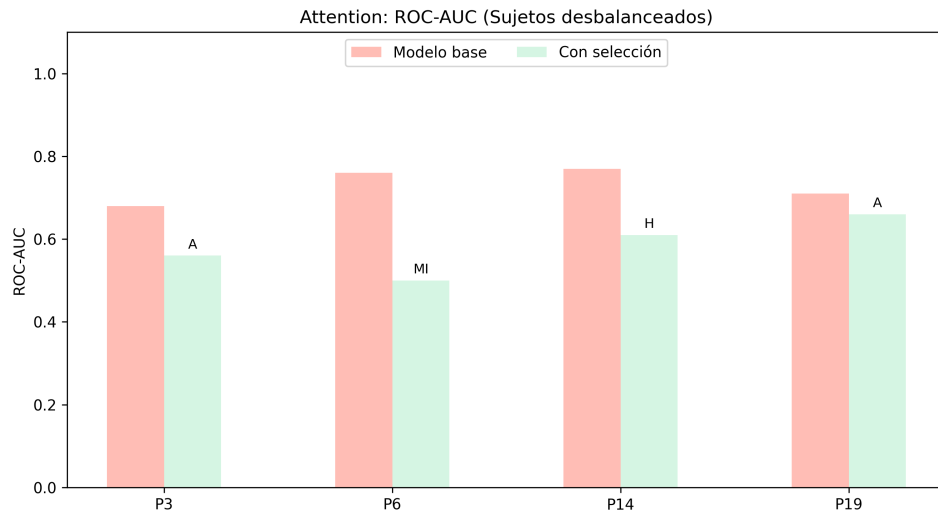


Figura 4.41: ROC-AUC para participantes desbalanceados con selección de características

La Tabla 4.14 muestra un resumen global de la aplicación de la selección de características. Una vez examinado el impacto de la reducción del espacio de características, podemos observar que, si bien se reduce el espacio de características, en la mayoría de casos se manifiesta una significativa pérdida de información crítica que las arquitecturas no logran compensar. Únicamente se observan mejoras considerables en los participantes 4 y 8. En contraste, los participantes 3 y 14 exhiben mejoras marginales en los casos de desequilibrio, aunque estas no resultan significativas. Sin embargo, la tendencia predominante sugiere que la selección de características actúa como un filtro de información que degrada la robustez del clasificador.

#### 4.4. Selección de características

Participante	Estado	Modelo	Selección de Características	F1-score	Recall / ROC
P4	Balanceado	EEGNeX	ANOVA	0.71	0.71
P7	Balanceado	EEGNeX	Ninguna	0.67	0.67
P8	Balanceado	EEGNeX	ANOVA	0.86	0.87
P9	Balanceado	EEGNet	Ninguna	0.57	0.58
P17	Balanceado	EEGNeX	Ninguna	0.64	0.65
P18	Balanceado	AttBaseNet	Ninguna	0.54	0.56
P21	Balanceado	EEGNeX	Ninguna	0.88	0.89
P22	Balanceado	AttBaseNet	Ninguna	0.68	0.69
P3	Desbalanceado	EEGNeX	Ninguna	0.57	0.663
P6	Desbalanceado	AttBaseNet	Ninguna	0.64	0.701
P14	Desbalanceado	EEGNet	Ninguna	0.69	0.768
P19	Desbalanceado	AttBaseNet	Ninguna	0.65	0.761

Cuadro 4.14: Resumen global de métricas. Se muestra el Recall para sujetos balanceados y el ROC-AUC para desbalanceados.

## 4.5. Resultados finales

A continuación, en la Tabla 4.15 se muestra para cada participante el modelo con el que ha obtenido el mejor resultado, con las distintas métricas, arquitectura y técnicas aplicadas. Los resultados recogidos en la tabla muestran una alta variabilidad intersujeto, tanto en el resultado de las métricas como en las arquitecturas y técnicas utilizadas.

Part.	Estado	Arquitectura	Técnicas	Acc.	Rec.	F1	AUC
P3	Desbal.	EEGNeX	Aleatorio + Focal Loss	0.67	0.57	0.57	0.663
P4	Bal.	EEGNeX	Pesos + ANOVA	0.71	0.71	0.71	0.708
P6	Desbal.	AttBaseNet	SMOTE	0.73	0.67	0.64	0.701
P7	Bal.	EEGNeX	Modelo Base	0.66	0.68	0.66	0.692
P8	Bal.	EEGNeX	ANOVA	0.86	0.87	0.86	0.909
P9	Bal.	EEGNet	Pesos	0.55	0.55	0.55	0.515
P14	Desbal.	EEGNet	Aleatorio + Focal Loss	0.86	0.66	0.69	0.768
P17	Bal.	EEGNeX	Pesos	0.63	0.63	0.63	0.629
P18	Bal.	AttBaseNet	Pesos	0.56	0.56	0.54	0.549
P19	Desbal.	AttBaseNet	Aleatorio	0.82	0.68	0.65	0.761
P21	Bal.	EEGNeX	Pesos	0.89	0.89	0.88	0.905
P22	Bal.	AttBaseNet	Modelo Base	0.70	0.69	0.68	0.653

Cuadro 4.15: Resumen consolidado del mejor modelo y configuración técnica por participante.

En los participantes con datos desbalanceados, el uso de oversampling aleatorio para el 3, 14 y 19 y SMOTE para el 6, combinado en algunos casos con Focal Loss, tiende a mejorar el f1-score y el recall de la clase minoritaria. Además, se han alcanzado valores notables de ROC-AUC, lo que facilita que el modelo identifique de manera más efectiva estados de baja confianza.

Por otro lado, en los participantes con datos balanceados, algunos modelos alcanzan buenos resultados incluso sin aplicar ninguna técnica específica, sino mediante ajustes arquitectónicos como el aumento del drop\_out o la disminución del batch\_size, como el participante 22. En otros casos, la introducción de pesos ponderados o selección de características mejora la estabilidad del modelo, logrando una mayor estabilidad en el entrenamiento y un equilibrio entre clases, especialmente un arquitecturas más complejas como EEGNeX o AttentionBaseNet. Junto al participante 8, el único que saca partido de la selección de características es el participante 4, haciendo uso también del método ANOVA.

En cuanto a las arquitecturas, EEGNeX destaca por ofrecer un buen compromiso entre capacidad de estabilidad y representación, alcanzando los mejores resultados globales en varios participantes, especialmente cuando se combina con técnicas de regularización y selección de características. AttentionBaseNet,

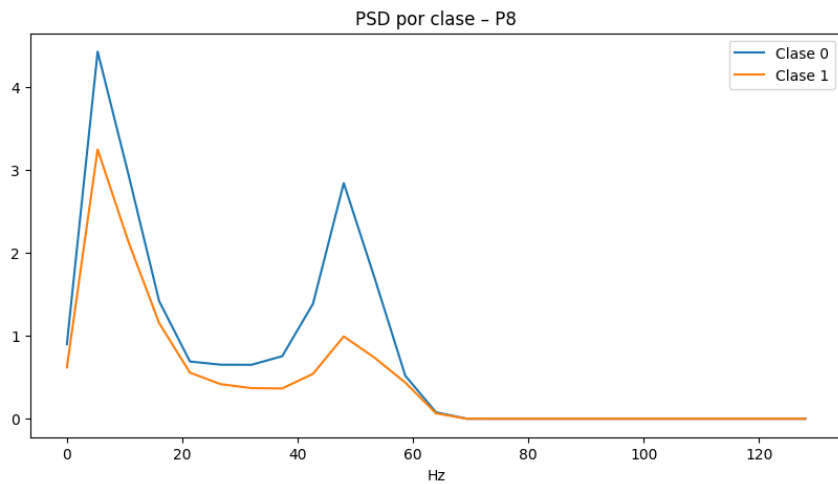
por su parte, muestra un alto potencial en algunos casos, pero también una mayor sensibilidad al ajuste de hiperparámetros y al desbalanceo, lo que suele derivar en modelos sesgados. Por último, EEGNet, aunque más simple, continúa siendo una alternativa competitiva para casos más simples.

Finalmente, el análisis global de la tabla refuerza la idea de que no existe una estrategia óptima para todos los participantes. Los resultados resaltan la necesidad de enfoques personalizados, donde la elección de la arquitectura y de las técnicas de balanceo o selección de características se adapte a las propiedades específicas de los datos de cada sujeto. Este aspecto es especialmente relevante en aplicaciones reales basadas en EEG, donde la robustez y la interpretabilidad son tan importantes como el rendimiento predictivo.

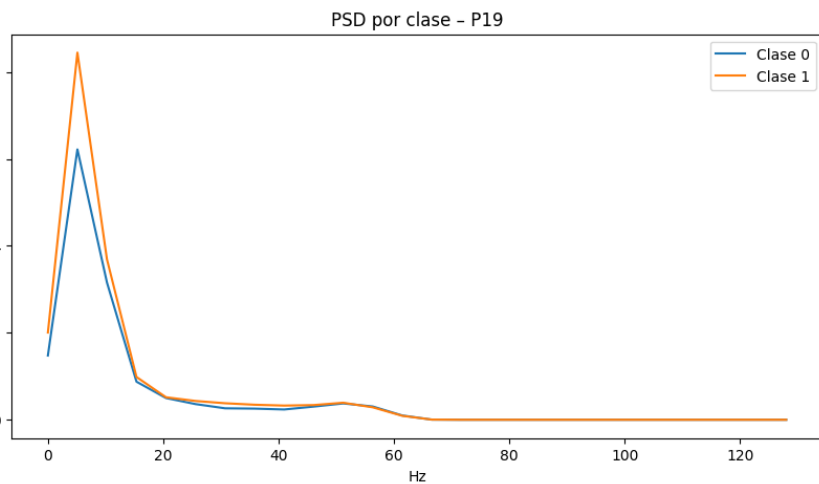
### 4.6. Análisis espectral

Para profundizar en los resultados obtenidos y comprender en mayor profundidad la disparidad de los datos entre los sujetos, se ha realizado un análisis de la densidad espectral de potencia (PSD), como se puede observar en la Figura 4.42. Este análisis permite visualizar la distribución de energía de la señal EEG en el dominio de frecuencia.

Para ello, se han seleccionado dos participantes con una diferencia considerable en la distribución de los datos y en sus resultados: el participante 8, con datos balanceados y muy buenos resultados en el análisis, y el participante 19, con datos desbalanceados y resultados mejorables.



(a) PSD Sujeto P8



(b) PSD Sujeto P19

Figura 4.42: Comparativa de la Densidad Espectral de Potencia (PSD) entre un sujeto con alto rendimiento (P8) y un sujeto con rendimiento moderado (P19).

#### 4.7. Comparación con enfoques multimodales basados en Machine Learning

El participante 8 muestra una potencia máxima de 4.5 con picos de actividad bien definidos y una separación visual significativa entre ambas clases. La presencia de estos picos disjuntos, especialmente en los rangos de 10Hz y 50Hz, proporciona al modelo puntos de referencia estadísticos sólidos para establecer la frontera de decisión. Además, la existencia de picos de elevada claridad representa una de las justificaciones que explican la eficacia de los algoritmos de selección de características como el análisis de la varianza (ANOVA). Estos algoritmos permiten identificar con facilidad qué canales registran tal diferencia espectral.

Por su parte, el participante 19 muestra un solapamiento casi total de las densidades espectrales. Este fenómeno implica que la distribución de energía entre clases es altamente similar, lo que puede atribuirse a la presencia de ruido o, simplemente porque algunas actividades cerebrales responden de manera muy parecida ante distintas condiciones. En consecuencia, el modelo recibe entradas muy similares para etiquetas distintas, lo que provoca un bajo recall y f1-score.

#### 4.7. Comparación con enfoques multimodales basados en Machine Learning

Con el objetivo de contextualizar en un marco más amplio los resultados obtenidos mediante modelos de deep learning basados exclusivamente en señales EEG, se incluye a continuación un Cuadro 4.16 con la comparación entre distintos enfoques de machine learning que emplean señales multimodales: ECG, GSR y Eye Tracking. Estos resultados no forman parte del núcleo experimental del trabajo, han sido obtenidos por el grupo de investigación NeuroTechAI. Permiten situar el rendimiento alcanzado dentro de un marco más amplio.

Modelo	Accuracy	F1-score
KNN	0.6554	0.5273
SVM	0.7311	0.6732
Naïve Bayes	0.5462	0.4411
Random Forest	0.8202	0.8114
XGBoost	0.9172	0.8443

Cuadro 4.16: Resultados obtenidos por los diferentes modelos basados en Machine Learning.

Como se puede observar, los modelos que incorporan información multimodal alcanzan valores superiores, destacando XGBoost con un Accuracy de 0.917 y un f1-score de 0.844. Esto sugiere que la combinación de distintas señales fisiológicas aporta información complementaria que facilita la discriminación entre estados, reduciendo la ambigüedad presente cuando se emplea únicamente

## Capítulo 4. Resultados

---

EEG.

Este aspecto refuerza la hipótesis de que la principal limitación del enfoque propuesto no reside exclusivamente en la arquitectura del modelo, sino en la restricción a una única modalidad de la señal.

# Capítulo 5

## Experimento

Como parte de este proyecto, se optó por generar un conjunto de datos propio mediante el uso del casco EEG16 de BitBrain, haciendo uso de 16 electrodos en lugar de 19 como con el resto de participantes. A continuación se describe el equipamiento y software utilizado, además de las distintas etapas de prueba, la realización del test y la toma de datos.

### 5.1. Equipamiento

#### 5.1.1. Casco EEG

El casco EEG16 de BitBrain [5] es un dispositivo que facilita la grabación de datos en tiempo real. Este casco está equipado con una serie de electrodos activos de tipo semi-seco, los cuales se humedecen mediante el uso de esponjas mojadas y se disponen de acuerdo al sistema internacional 10/20. Sus sensores están desarrollados con contactos estables para permitir una buena monitorización de los datos.

Su colocación resulta sencilla y su diseño garantiza la comodidad del usuario, facilitando sesiones de registro prolongadas sin comprometer la calidad de los datos.

Para el experimento en cuestión, se utilizaron los electrodos de acuerdo con la nomenclatura establecida, como se puede observar en la Figura 5.1. Los puntos correspondientes a la zona frontal son FP1, FP2, F7, F3, Fz, F4 y F8; los de la región temporal son T7 y T8; los de la zona central son C3, Cz y C4; y, por último, los de la región parietal son P7, P3, Pz y P4. En consecuencia, la implementación de esta colocación permite que los electrodos capten la información correspondiente a las funciones ejecutivas, la atención, la integración sensorial y el procesamiento visual.

## Capítulo 5. Experimento

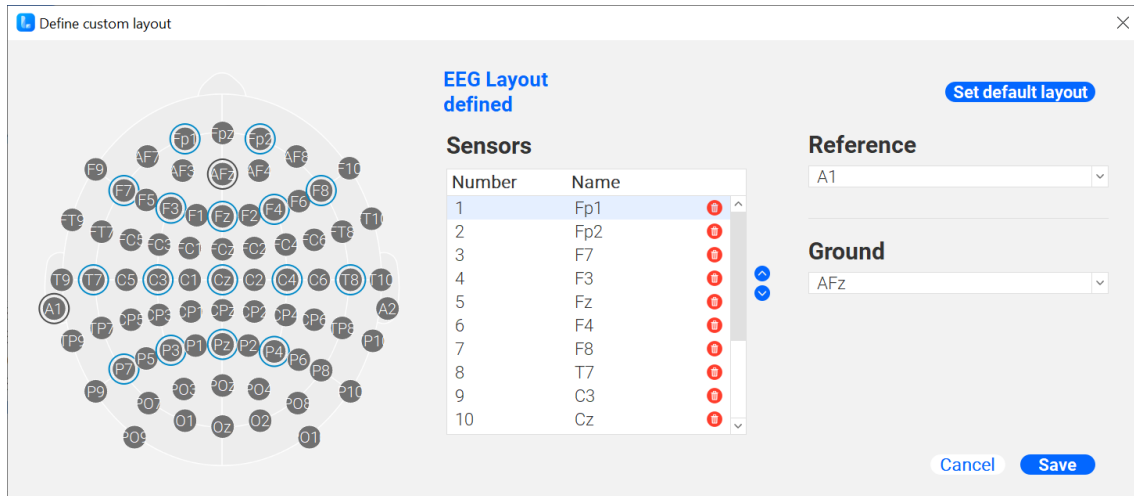


Figura 5.1: Posicionamiento de los electrodos

Para garantizar una buena conductividad, las esponjas de cada electrodo fueron humedecidas con agua corriente antes de la sesión. Durante el experimento, las señales EEG fueron transmitidas al ordenador en tiempo real, donde quedaron registradas para su posterior análisis.

### 5.1.2. Software y experimento

Una vez colocado el casco, se procedió a la realización del experimento:

1. Cuestionarios sobre la confianza: se respondieron a una serie de preguntas relacionadas con la confianza de las personas hacia las máquinas, además sobre la preferencia de algunas imágenes.
2. Explicación sobre la realización del test de Stroop.
3. Realización del test.
4. Cuestionario entre cada bloque de test sobre cómo se ha sentido el sujeto en cuanto a confianza durante el experimento.
5. Cuestionario final

Para la recopilación de los datos se utilizó el Bitbrain Software Kit, denominado SENSE Lite. Este software está especializado en la adquisición de señales neurofisiológicas como EEG en tiempo real, cuyo objetivo principal es registrar las señales, visualizar la señal mientras se graba, gestionar participantes, marcar eventos (triggers) durante el experimento y exportar los datos para su posterior análisis.

En el caso de la Figura 5.2, corresponde al sujeto eeg\_16, identificado con el ID:015. Aunque el software nos da la posibilidad de registrar otro tipo de datos capturados con otros sensores, en este caso solo nos interesan los electroencefalogramas. En la zona central se puede observar la posición aproximada de los

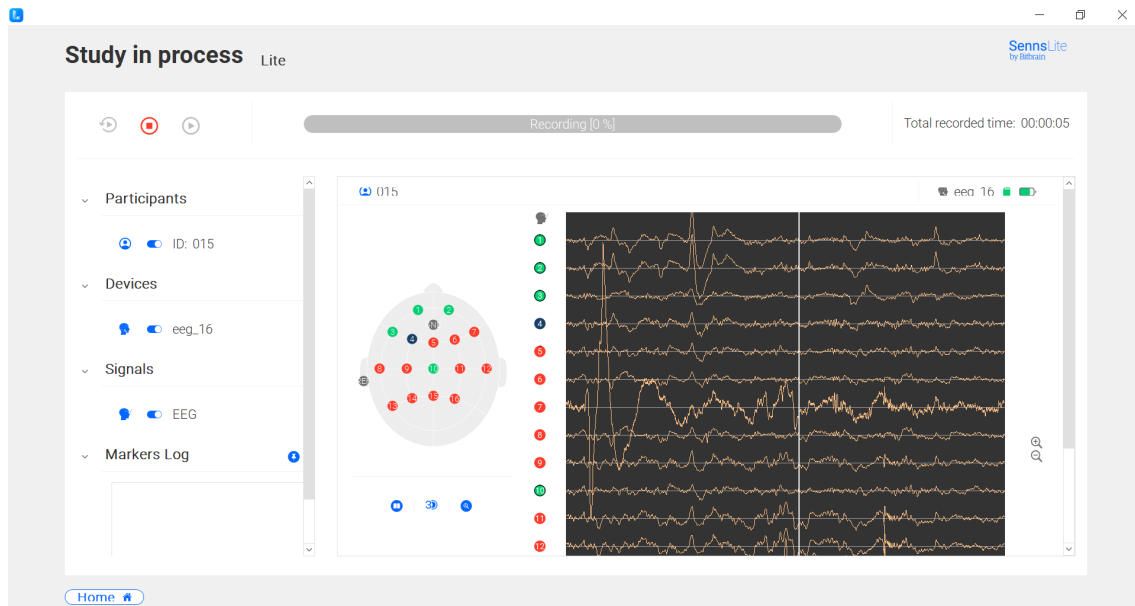


Figura 5.2: Interfaz de la aplicación SenssLite

electrodos, el panel con las señales en crudo, los números que identifican cada canal y los colores que representan el estado de cada electrodo:

- Verde: Muy buena calidad de señal.
- Amarillo: Buena calidad de señal.
- Roja: Mala calidad de la señal.
- Gris: No se detecta la señal.

Cada línea representa un electrodo, donde el eje horizontal se corresponde con el tiempo y el vertical con la amplitud de la señal. A pesar de que algunos electrodos se encuentran en rojo, la señal es bastante estable en su mayor parte. En los casos donde hay picos más abruptos, corresponden a parpadeos o movimientos musculares.

Para verificar la correcta recepción de la información por parte de los electrodos, se realizaron movimientos musculares voluntarios, tales como el pestañeo, el movimiento de la cabeza y el habla. De esta manera, se pudo observar si las oscilaciones cambiaban:

En el caso de la Figura 5.3 , se pueden observar oscilaciones mucho más abruptas, provocadas por los distintos movimientos mencionados, asegurándonos de que los electrodos detectan bien la actividad.

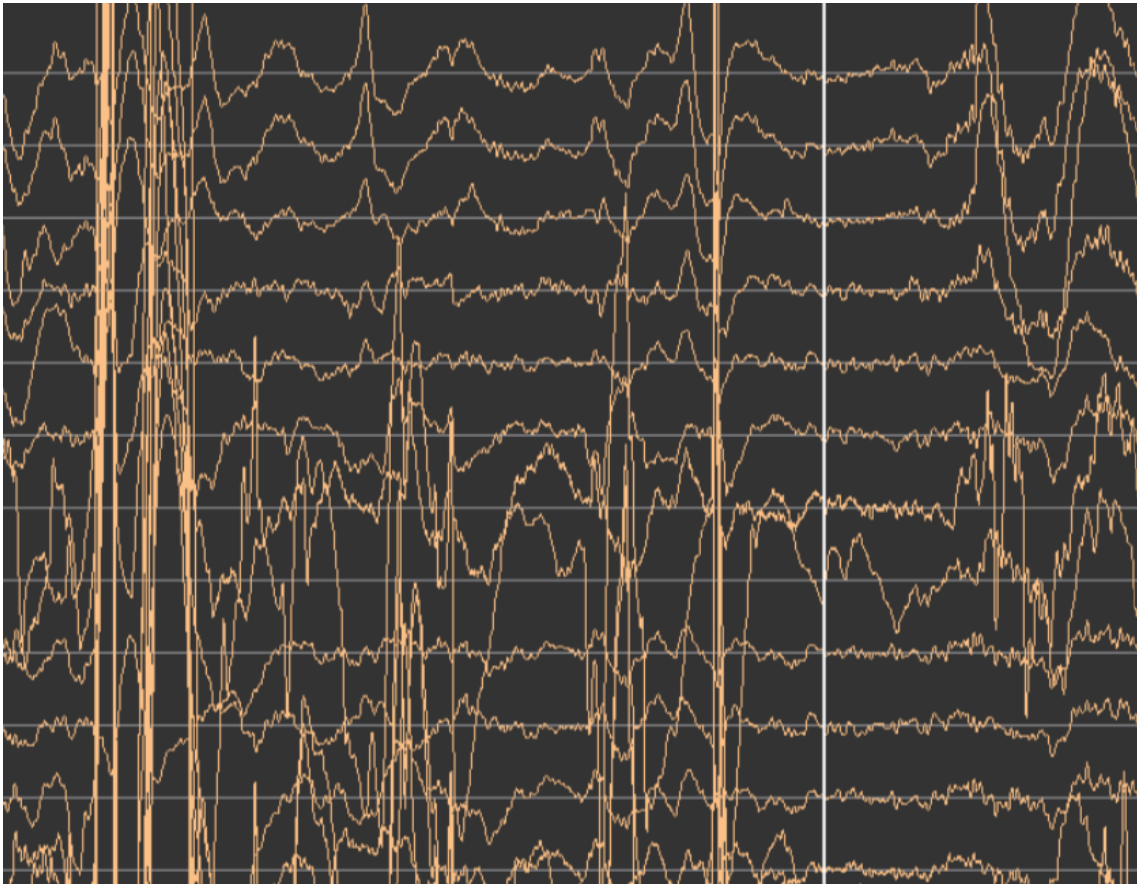


Figura 5.3: Interfaz de la aplicación SenssLite. Comprobación de la detección de la señal.

## Capítulo 6

# Conclusiones y trabajo futuro

Los resultados obtenidos muestran que el rendimiento de los modelos de deep learning basados en señales EEG presenta una alta variabilidad entre participantes, lo que pone de manifiesto la naturaleza altamente dependiente del sujeto de este tipo de señales. En términos generales, los resultados más sobresalientes se obtienen con los participantes con datos balanceados, especialmente haciendo uso de la arquitectura EEGNeX en conjunto con técnicas avanzadas, tales como pesos ponderados y las adecuadas configuraciones de los hiperparámetros de la red neuronal. Esto indica que, bajo condiciones favorables, los modelos son capaces de capturar patrones relevantes asociados a estados de confianza del usuario.

En el caso de los participantes desbalanceados, la implementación de técnicas específicas de oversampling y Focal Loss permite a los modelos optimizar su rendimiento frente a los enfoques de base, aunque con una mayor variabilidad en los resultados y sin asegurar una mejora uniforme en los sujetos. Estos resultados reflejan las limitaciones técnicas de balanceo cuando se aplican a señales EEG, donde el desbalance de clases se ve amplificado por la variabilidad intra-sujeto e inter-sujeto.

El análisis espectral proporciona una explicación complementaria a estos resultados. En los sujetos con mayor rendimiento, se observa una mayor disparidad entre clases en la densidad espectral de potencia, con bandas dominantes más definidas y un menor solapamiento entre señales. En contraste, los participantes que presentan resultados menos favorables, exhiben un espectro de señal caracterizado por un solapamiento significativo y una distribución de potencia menos pronunciada. Esta circunstancia dificulta la extracción de características discriminantes y limita el rendimiento de los modelos. Este análisis subraya la importancia de la calidad y la separabilidad de la señal para poder alcanzar resultados óptimos.

Como principal línea de trabajo futuro, se propone la incorporación de señales

## Capítulo 6. Conclusiones y trabajo futuro

---

multimodales, tales como ECG, GSR y Eye Tracking, para su combinación con las EEG. La fusión de modalidades facilita la captura de dimensiones del estado cognitivo y emocional del usuario, reduciendo la dependencia a una única señal y fortaleciendo la robustez de los modelos. Los resultados obtenidos con los modelos de machine learning sugieren que esta integración podría mejorar de forma significativa la capacidad de generalización y la fiabilidad de los sistemas de evaluación de confianza en entornos reales. Asimismo, otra línea de trabajo futuro interesante sería profundizar en el análisis intersujeto, con el objetivo de evaluar la capacidad de los modelos más allá de configuraciones intrasujeto. En este contexto, el método Leave-One-Subject-Out [41] representa una estrategia muy adecuada, ya que entrena el modelo utilizando los datos de todos los participantes salvo el de uno, proporcionando un escenario más realista donde el modelo debe enfrentarse a usuarios no vistos durante el entrenamiento. A pesar de suponer un desafío en señales EEG debido a la elevada variabilidad intersujeto, al mismo tiempo contribuye a una herramienta clave para identificar modelos y representaciones más robustas y menos dependientes de características individuales.

## Capítulo 7

# Análisis de impacto

En este capítulo se realizará un análisis del impacto potencial de los resultados obtenidos durante la realización del trabajo, en los diferentes contextos para los que se aplique:

- Personal
- Empresarial
- Social
- Económico
- Medioambiental
- Cultural

### 7.1. Impacto personal

En cuanto al impacto personal, el desarrollo de este trabajo me ha sido especialmente útil para saber hacia dónde me gustaría dirigirme en el ámbito laboral. A lo largo de la carrera se han abordado técnicas de deep learning en distintos contextos, pero no en aplicaciones relacionadas con la neurociencia. El estudio de los electroencefalogramas, así como el análisis de la información que puede extraerse de las señales, me ha permitido descubrir un ámbito científico de gran interés en el que me gustaría profundizar en el futuro.

Asimismo, la aplicación de redes neuronales a un problema real y complejo ha sido de gran utilidad para comprender con mayor profundidad su funcionamiento interno, especialmente en lo relativo al comportamiento del modelo frente a datos desbalanceados.

### 7.2. Impacto social

La capacidad de evaluar la confianza de los usuarios hacia sistemas automatizados tiene un impacto directo en la adopción de nuevas tecnologías. La comprensión de la percepción individual sobre la fiabilidad de la automatización, así como la identificación de los contextos en los que esta fiabilidad es evidente y en cuáles no, son aspectos fundamentales para la identificación de posibles barreras cognitivas que podrían limitar su uso efectivo. La comprensión del funcionamiento de la confianza de las personas hacia las máquinas constituye un aspecto esencial para el diseño de sistemas más fiables y óptimos. Estos sistemas deben estar orientados a generar una sensación de comodidad en los usuarios, lo cual resulta fundamental en diversas áreas, tales como la medicina, los sistemas de recomendación y el ámbito aéreo, por mencionar algunos ejemplos.

La implementación de métricas específicas facilita a los desarrolladores la capacidad de adaptar y personalizar dichos sistemas automatizados, conforme a las necesidades individuales de los usuarios. Fomentar una adopción consciente y segura de la automatización contribuye a una sociedad más informada y responsable frente al uso de este tipo de tecnologías, asegurando que la implementación de los sistemas automáticos no comprometan la confianza ni la seguridad de las personas que dependen de ellos.

### 7.3. Impacto económico

La implementación de sistemas automáticos para la medición de la confianza se presenta como una estrategia que permite optimizar la eficiencia operativa, garantizando así una mayor consistencia en los resultados obtenidos. Los usuarios recurrirán con mayor frecuencia a sistemas que les inspiren más confianza, lo que se traduce en un uso más eficaz de los mismos y de una reducción de errores humanos y pérdidas de tiempo. Esto se traduce directamente en una reducción de costes operativos y en un mejor aprovechamiento del tiempo y de los recursos disponibles.

Además, disponer de la información sobre los niveles de confianza permite priorizar inversiones de manera más estratégica. En consecuencia, se podrán dirigir recursos de manera directa hacia aquellos sistemas en los que los usuarios no sientan una confianza óptima, incrementando así el retorno de la inversión al focalizar los esfuerzos en los elementos que realmente generen impacto en esta desconfianza.

En última instancia, un sistema en el que se confía reduce la necesidad de una supervisión continua para garantizar su funcionamiento adecuado, lo que se traduce en ahorro de personal y tiempo.

### 7.4. Impacto medioambiental

La implementación de sistemas fiables permite optimizar el uso de recursos físicos y energéticos, al reducir errores y tareas que generan desperdicios de materiales y energía. En el ámbito industrial y logístico, la implementación de sistemas automatizados y confiables se muestra como una estrategia eficaz para la optimización de la planificación de, transporte, la eficiencia energética y la gestión de recursos. Este enfoque contribuye a la reducción de la huella ecológica, lo que resulta en un mayor impacto medioambiental positivo.

Este tipo de sistemas posibilita la implementación de estrategias más sostenibles y con mayor seguridad. Los usuarios estarían más dispuestos a delegar decisiones más críticas en la automatización, lo que permite mantener procesos optimizados de manera más continua y reducir el impacto medioambiental a largo plazo. De esta manera, se promueve el uso eficiente de recursos en distintos contextos.

### 7.5. Impacto cultural

La confianza depositada en los sistemas automatizados ejerce una influencia significativa en la dimensión cultural, al producir una transformación en la forma en que las personas perciben y utilizan la tecnología. Cuando los usuarios experimentan la confianza necesaria para aceptar la automatización, se favorece la aceptación de soluciones en la vida diaria y entornos laborales, impulsando la innovación y la digitalización.

En este sentido, se promueve una cultura de colaboración humano-máquina, en la cual la tecnología se concibe como un apoyo en el proceso de toma de decisiones, aumentando la confianza en los sistemas automáticos. Además, el desarrollo de herramientas predecibles y seguras, permite a los usuarios adquirir habilidades para comprender y colaborar con tecnologías avanzadas. Este proceso contribuye a la consolidación de una cultura tecnológica más preparada para los retos de la transformación digital.

### 7.6. Objetivos de desarrollo sostenible

La agenda 2030 de desarrollo sostenible establece un marco global orientado a afrontar los principales retos sociales, tecnológicos y económicos de la sociedad mediante diecisiete objetivos de desarrollo sostenible. En este contexto, la implantación de la automatización en la sociedad plantea tanto oportunidades como desafíos, no solo en cuanto a lo técnico, si no principalmente en la interacción entre las personas y este tipo de sistemas.

Este trabajo profundiza en esta problemática, abordando la necesidad de evaluar y comprender la confianza de las personas al trabajar con las máquinas, un factor clave para su aceptación, uso responsable y adaptación en distintos

contextos. Tener la capacidad de medir dicha confianza permite detectar limitaciones, mejorar el rendimiento de los sistemas y reducir el riesgo de un uso inadecuado o de una dependencia excesiva. Este enfoque es realmente importante para contribuir a una integración ética, segura y eficiente en la tecnología en ámbitos de impacto directo con las personas.

### **7.6.1. ODS 3: Salud y bienestar**

Busca garantizar una vida sana y promover el bienestar de las personas. La evaluación de la confianza permite identificar cuándo un sistema es percibido como poco fiable o, por el contrario, excesivamente fiable, facilitando ajustes que mejoren la interacción humano-máquina. Además, el análisis de la variabilidad de rendimiento de los modelos entre distintos participantes pone de manifiesto la importancia de considerar las diferencias individuales en el diseño de sistemas aplicados a la salud. Este enfoque contribuye a desarrollar sistemas adaptativos, capaces de apoyar el bienestar de los usuarios de forma segura y responsable.

### **7.6.2. ODS 8: Trabajo decente y crecimiento económico**

La evaluación de la confianza en sistemas automatizados ayuda a diseñar entornos laborables más confiables. Si los usuarios observan que pueden confiar en estos sistemas, su adopción será más sencilla, lo que permitirá una mejora en la productividad laboral.

En este contexto, la medición de la confianza permite detectar situaciones en las que un sistema puede generar rechazo, desuso o, por el contrario, una dependencia excesiva. Ambos escenarios pueden afectar negativamente a la calidad del trabajo y a la toma de decisiones. Al disponer de métricas que evalúan este aspecto, es posible diseñar sistemas que apoyen al trabajador sin sustituir su criterio, favoreciendo un uso equilibrado de la tecnología.

Desde una perspectiva económica, una adopción más eficiente de sistemas automatizados puede generar mejoras de productividad y optimizar los procesos operativos, sin comprometer las condiciones laborales.

### **7.6.3. ODS 9: Industria, innovación e infraestructura**

En el contexto industrial, la automatización de procesos productivos y de toma de decisiones es cada vez más habitual, especialmente en entornos donde la eficiencia, la seguridad y la continuidad operativa son críticas. La medición de la confianza en sistemas automatizados resulta clave para evaluar si estas tecnologías pueden integrarse de forma fiable en infraestructuras industriales complejas.

Disponer de métricas que permitan analizar el grado de confianza facilita la detección temprana de fallos, comportamientos inconsistentes o respuestas poco interpretables del sistema, aspectos especialmente relevantes en procesos industriales automatizados. Esto contribuye a una implantación más robusta de

## **7.6. Objetivos de desarrollo sostenible**

---

soluciones innovadoras, favoreciendo infraestructuras tecnológicas más seguras, escalables y alineadas con los estándares de la industria, impulsando así una innovación sostenible y responsable.



# Bibliografía

- [1] T. Yamagashi, S. Kanazawa, R. Mashima y S. Terai. «Separating trust from cooperation in a dynamic relationship: prisoner's dilemma with variable dependence», visitado 1 de jun. de 2023. dirección: [https://www.researchgate.net/publication/41818381\\_Separating\\_trust\\_from\\_cooperation\\_in\\_a\\_dynamic\\_relationship\\_prisoner's\\_dilemma\\_with\\_variable\\_dependence](https://www.researchgate.net/publication/41818381_Separating_trust_from_cooperation_in_a_dynamic_relationship_prisoner's_dilemma_with_variable_dependence)
- [2] T. B. Sheridan y R. Parasuraman. «Human-Automation Interaction», visitado 13 de oct. de 2023. dirección: [https://www.researchgate.net/publication/240756917\\_Human-Automation\\_Interaction](https://www.researchgate.net/publication/240756917_Human-Automation_Interaction)
- [3] A. Jacovi, A. Marasovic, T. Miller e Y. Goldberg. «Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI», visitado 13 de oct. de 2023. dirección: <https://arxiv.org/abs/2010.07487>
- [4] J. A. Aimone, D. Houser y B. Weber. «Neural signatures of betrayal aversion: an fMRI study of trust», visitado 16 de oct. de 2023. dirección: <https://pmc.ncbi.nlm.nih.gov/articles/PMC3973250/>
- [5] «Versatile EEG», visitado 9 de ene. de 2026. dirección: <https://www.bitbrain.com/es/productos-neurotecnologia/water-based-eeg/versatile-eeg>
- [6] G. Campagna y M. Rehm. «Trust Assessment with EEG Signals in Social Human-Robot Interaction», visitado 18 de oct. de 2023. dirección: [https://vbn.aau.dk/ws/portalfiles/portal/700401565/Trust\\_Assessment\\_with\\_EEG\\_Signals\\_in\\_Social\\_Human\\_Robot\\_Interaction.pdf](https://vbn.aau.dk/ws/portalfiles/portal/700401565/Trust_Assessment_with_EEG_Signals_in_Social_Human_Robot_Interaction.pdf)
- [7] D. Ullman y B. F. Malle. «What Does it Mean to Trust a Robot?: Steps Toward a Multidimensional Measure of Trust», visitado 18 de oct. de 2025. dirección: <https://dl.acm.org/doi/10.1145/3173386.3176991>
- [8] I. Ajenaghughrure, S. Souse y D. Lamas. «Psychophysiological Modeling of Trust In Technology: Influence of Feature Selection Methods», visitado 19 de oct. de 2025. dirección: [https://www.researchgate.net/publication/351982103\\_Psychophysiological\\_Modeling\\_of\\_Trust\\_In\\_Technology\\_Influence\\_of\\_Feature\\_Selection\\_Methods](https://www.researchgate.net/publication/351982103_Psychophysiological_Modeling_of_Trust_In_Technology_Influence_of_Feature_Selection_Methods)

- [9] Y. Du, G. Li, M. Wu y F. Chen. «Unsupervised Multivariate Feature-Based Adaptive Clustering Analysis of Epileptic EEG Signals», visitado 30 de oct. de 2025. dirección: <https://www.mdpi.com/2076-3425/14/4/342>
- [10] M. Ghasemi y D. Ebrahimi. «Introduction to Reinforcement Learning», visitado 10 de ene. de 2026. dirección: <https://arxiv.org/abs/2408.07712>
- [11] D. Xu, M. Agarwal, F. Fekri y R. Sivakumar. «Accelerating Reinforcement Learning using EEG-based Implicit Human Feedback», visitado 30 de oct. de 2025. dirección: <https://arxiv.org/pdf/2010.08140>
- [12] J. Fan, C. Ma e Y. Zhong. «A selective overview of deep learning», visitado 10 de ene. de 2026. dirección: <https://pmc.ncbi.nlm.nih.gov/articles/PMC8300482/>
- [13] C. Xu, C. Zhang, Y. Zhou, Z. Wang, P. Lu y B. He. «Trust Recognition in Human-Robot Cooperation Using EEG», visitado 24 de oct. de 2025. dirección: <https://arxiv.org/abs/2403.05225>
- [14] A. Xu y G. Dudek. «OPTIMo: Online Probabilistic Trust Inference Model for Asymmetric Human-Robot Collaborations», visitado 24 de oct. de 2025. dirección: <https://dl.acm.org/doi/10.1145/2696454.2696492>
- [15] F. Faruqe, R. Watkins y L. R. Medske. «Monitoring Trust in Human-Machine Interactions for Public Sector Applications», visitado 25 de oct. de 2025. dirección: <https://arxiv.org/abs/2010.08140>
- [16] S. M. Merritt, H. Heimbaugh, J. LaChapell y D. Lee. «Effects of Implicit Attitudes Toward Automation on Trust in an Automated System», visitado 1 de dic. de 2025. dirección: <https://www.studeersnel.nl/nl/document/technische-universiteit-delft/human-robot-interaction/q2-merritt-et-al-2013/111049213>
- [17] G. Bookshire et al. «Data leakage in deep learning studies of translational EEG». dirección: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11099244/>
- [18] F. D. Pup, A. Zanola, L. F. Tshimanga, A. Bertoldo, L. Finos y M. Atzori. «The role of data partitioning on the performance of EEG-based deep learning models in supervised cross-subject analysis: a preliminary study». dirección: <https://arxiv.org/pdf/2505.13021>
- [19] M. Buda, A. Maki y M. A. Mazurowski. «A systematic study of the class imbalance problem in convolutional neural networks», visitado 20 de nov. de 2025. dirección: <https://arxiv.org/pdf/1710.05381>
- [20] C. J. Hellín, A. A. Olmedo, A. Valledor, J. Gómez, M. López-Benítez y A. Tayebi. «Unraveling the Impact of Class Imbalance on Deep-Learning Models for Medical Image Classification», visitado 20 de nov. de 2025. dirección: <https://www.mdpi.com/2076-3417/14/8/3419>
- [21] T. Lin, P. Goyal, R. B. Girshick, K. He y P. Dollár. «Focal Loss for Dense Object Detection», visitado 20 de nov. de 2025. dirección: <https://arxiv.org/pdf/1708.02002>

- [22] A. Vilora, O. B. P. Lezama y N. Mercado-Caruzo. «Unbalanced data processing using oversampling: Machine Learning», visitado 6 de ene. de 2026. dirección: <https://www.sciencedirect.com/science/article/pii/S1877050920316975>
- [23] G. E. A. P. A. Batista, R. C. Prati y M. C. Monard, «A study of the behavior of several methods for balancing machine learning training data», 2004. dirección: [https://www.kdd.org/exploration\\_files/batista.pdf](https://www.kdd.org/exploration_files/batista.pdf)
- [24] K. W. Bowyer, N. V. Chawla, L. O. Hall y W. P. Kegelmeyer. «SMOTE: Synthetic Minority Over-sampling Technique», visitado 22 de dic. de 2025. dirección: <http://arxiv.org/abs/1106.1813>
- [25] P. J. M. Ali y R. H. Faraj. «Data Normalization and Standardization: A Technical Report», visitado 8 de ene. de 2026. dirección: [https://www.researchgate.net/profile/Peshawa-Muhammad-Ali/publication/340579135\\_Data\\_Normalization\\_and\\_Standardization\\_A\\_Technical\\_Report/links/5e91b65d299bf130798fc1bd/Data-Normalization-and-Standardization-A-Technical-Report.pdf](https://www.researchgate.net/profile/Peshawa-Muhammad-Ali/publication/340579135_Data_Normalization_and_Standardization_A_Technical_Report/links/5e91b65d299bf130798fc1bd/Data-Normalization-and-Standardization-A-Technical-Report.pdf)
- [26] I. Salehin y D.-K. Kang. «A Review on Dropout Regularization Approaches for Deep Neural Networks within the Scholarly Domain», visitado 8 de ene. de 2026. dirección: <https://www.mdpi.com/2079-9292/12/14/3106>
- [27] J. He y M. X. Cheng. «Weighting Methods for Rare Event Identification From Imbalanced Datasets», visitado 8 de ene. de 2026. dirección: <https://www.frontiersin.org/journals/big-data/articles/10.3389/fdata.2021.715320>
- [28] Y. Yao, L. Rosaco y A. Caponnetto. «ON EARLY STOPPING IN GRADIENT DESCENT LEARNING», visitado 22 de dic. de 2025. dirección: <https://web.mit.edu/lrosasco/www/publications/earlystop.pdf>
- [29] B. Venkatesh y J. Anuradha. «A Review of Feature Selection and Its Methods», visitado 8 de ene. de 2026. dirección: <https://scienciendo.com/2/v2/download/article/10.2478/cait-2019-0001.pdf>
- [30] V. Lawhern, A. J. Solon, N. Waytowich y S. M. Gordon. «EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces». dirección: [https://www.researchgate.net/publication/310953136\\_EEGNet\\_a\\_compact\\_convolutional\\_neural\\_network\\_for\\_EEG-based\\_brain-computer\\_interfaces](https://www.researchgate.net/publication/310953136_EEGNet_a_compact_convolutional_neural_network_for_EEG-based_brain-computer_interfaces)
- [31] V. Lawhern, A. J. Solon, N. Waytowich y S. M. Gordon. «EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces», visitado 22 de nov. de 2025. dirección: [https://content.cld.iop.org/journals/1741-2552/15/5/056013/revision2/jneaace8cf01\\_hr.jpg](https://content.cld.iop.org/journals/1741-2552/15/5/056013/revision2/jneaace8cf01_hr.jpg)
- [32] X. Chen, X. Teng, H. Chen, Y. Pan y P. Geyer. «TOWARD RELIABLE SIGNALS DECODING FOR ELECTROENCEPHALOGRAM: A BENCHMARK STUDY TO EEGNEXs», visitado 13 de nov. de 2025. dirección: <https://arxiv.org/pdf/2207.12369>

- [33] X. Chen, X. Teng, H. Chen, Y. Pan y P. Geyer. «Toward Reliable Signals Decoding For electroencephalogram: a benchmark study to EEGNeXs», visitado 22 de nov. de 2025. dirección: [https://braindecode.org/dev/\\_static/model/eegnex.jpg](https://braindecode.org/dev/_static/model/eegnex.jpg)
- [34] M. Wimpff, L. Gizzi, J. Zerfowski y B. Yang. «EEG motor imagery decoding: A framework for comparative analysis with channel attention mechanisms», visitado 29 de nov. de 2025. dirección: <https://arxiv.org/abs/2310.11198>
- [35] M. Wimpff, L. Gizzi, J. Zerfowski y B. Yang. «EEG motor imagery decoding: A framework for comparative analysis with channel attention mechanisms», visitado 29 de nov. de 2025. dirección: [https://content.cld.iop.org/journals/1741-2552/21/3/036020/revision2/jnead48b9f2\\_hr.jpg](https://content.cld.iop.org/journals/1741-2552/21/3/036020/revision2/jnead48b9f2_hr.jpg)
- [36] R. Borja-Robalino, A. Monleón-Getino y J. Rodellar. «Estandarización de métricas de rendimiento para clasificadores Machine y Deep Learning», visitado 9 de ene. de 2026. dirección: [https://www.researchgate.net/profile/Antonio-Monleon-Getino/publication/342009715\\_Estandarizacion\\_de\\_metricas\\_de\\_rendimiento\\_para\\_clasificadores\\_Machine\\_y\\_Deep\\_Learning/links/5ede3d0392851cf13869078e/Estandarizacion-de-metricas-de-rendimiento-para-clasificadores-Machine-y-Deep-Learning.pdf](https://www.researchgate.net/profile/Antonio-Monleon-Getino/publication/342009715_Estandarizacion_de_metricas_de_rendimiento_para_clasificadores_Machine_y_Deep_Learning/links/5ede3d0392851cf13869078e/Estandarizacion-de-metricas-de-rendimiento-para-clasificadores-Machine-y-Deep-Learning.pdf)
- [37] D. P. Kingma y J. Ba. «Adam: A Method for Stochastic Optimization», visitado 22 de dic. de 2025. dirección: <https://arxiv.org/abs/1412.6980>
- [38] M. Altalhan, A. Algarni y M. Turki-Hadj Alouane. «Imbalanced Data Problem in Machine Learning: A Review», visitado 11 de ene. de 2026. dirección: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10845793>
- [39] J. Brabec y L. Machlica. «Bad practices in evaluation methodology relevant to class-imbalanced problems», visitado 11 de ene. de 2026. dirección: <https://arxiv.org/abs/1812.01388>
- [40] E. Richardson, R. Trevizani, J. A. Greenbaum, H. Carter, M. Nielsen y B. Peters. «The receiver operating characteristic curve accurately assesses imbalanced datasets», visitado 11 de ene. de 2026. dirección: [https://www.cell.com/patterns/fulltext/S2666-3899\(24\)00109-0](https://www.cell.com/patterns/fulltext/S2666-3899(24)00109-0)
- [41] S. Kunjan et al. «The Necessity of Leave One Subject Out(LOSO) Cross Validation for EEGDisease Diagnosis». dirección: [https://link.springer.com/chapter/10.1007/978-3-030-86993-9\\_50](https://link.springer.com/chapter/10.1007/978-3-030-86993-9_50)

# **Anexos**




# **Apéndice A**

## **Anexo**

Enlace al repositorio de Github donde está alojado el proyecto: <https://github.com/crisfernandez/TFG>

Este documento esta firmado por



<b>Firmante</b>	CN=tfgm.fi.upm.es, OU=CCFI, O=ETS Ingenieros Informaticos - UPM, C=ES
<b>Fecha/Hora</b>	Wed Jan 14 11:45:41 CET 2026
<b>Emisor del Certificado</b>	EMAILADDRESS=camanager@etsiinf.upm.es, CN=CA ETS Ingenieros Informaticos, O=ETS Ingenieros Informaticos - UPM, C=ES
<b>Numero de Serie</b>	561
<b>Metodo</b>	urn:adobe.com:Adobe.PPKLite:adbe.pkcs7.sha1 (Adobe Signature)