



UNIVERSIDAD DE VALLADOLID

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE
TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

GRADO EN INGENIERÍA DE TELECOMUNICACIÓN

**DETECCIÓN DE SPINDLES DEL SUEÑO MEDIANTE
TÉCNICAS DE INTELIGENCIA ARTIFICIAL**

Autor:

Dña. Victoria Pacho Velasco

Tutores:

Dr. D. Gonzalo César Gutiérrez Tobal

Dr. D. Roberto Hornero Sánchez

Valladolid, 27 de Septiembre de 2022

**TÍTULO: Detección de *spindles* del sueño
mediante técnicas de inteligencia
artificial**

AUTOR: Dña. Victoria Pacho Velasco

**TUTORES: Dr. D. Gonzalo César Gutiérrez Tobal
Dr. D. Roberto Hornero Sánchez**

**DEPARTAMENTO: Teoría de la Señal y Comunicaciones
e Ingeniería Telemática**

TRIBUNAL

PRESIDENTE: Dr. D. Roberto Hornero Sánchez

VOCAL: Dr. D. Jesús Poza Crespo

SECRETARIA: Dra. Dña. María García Gadañón

SUPLENTE 1: Dr. D. Calos Gómez Peña

SUPLENTE 2: Dr. D. Mario Martínez Zarzuela

AGRADECIMIENTOS

Transmitir mi más sincero agradecimiento a todos aquellos que me han ayudado de una forma u otra a lo largo de esta etapa y han colaborado en esta investigación.

En primer lugar, a mis tutores Roberto y Gonzalo, los cuales me han guiado en todo momento durante el desarrollo de este TFG, han mostrado un gran interés sobre la evolución del proyecto y me han ayudado a solucionar varias dudas que se han ido surgiendo a lo largo del curso. También agradecer a los miembros del Grupo de Ingeniería Biomédica de la Universidad de Valladolid, por acogerme desde el principio y aportarme nuevos conocimientos.

En segundo lugar, a mis amigos y a mi pareja. Gracias por estar ahí siempre, en las buenas y en las malas, apoyándome y animándome en todo momento a lo largo de estos años.

Finalmente, agradecer a mi familia, mi madre Mercedes, mi padre Jesús y mi hermano Jesús. Gracias por creer en mí siempre, sobre todo durante todos estos años, apoyarme en todo tipo de decisiones y por hacerme ver que todo, con esfuerzo, se puede conseguir.

Gracias a todos de corazón.

RESUMEN DEL TFG

El síndrome de la Apnea Obstructiva del Sueño (AOS) pediátrico es un trastorno muy prevalente (en torno al 25% de la población) en la sociedad occidental, causante de muchas enfermedades asociadas a un deterioro cognitivo que empeora la calidad de vida de los niños afectados. Gracias a los nuevos avances tecnológicos, se ha observado que pequeñas oscilaciones presentes fundamentalmente durante las fases N2 y N3 del sueño entre 11-16 Hz, conocidas como *spindles* del sueño, están íntimamente relacionadas con el proceso cognitivo de las personas en general y de los niños en particular. Este descubrimiento abre una nueva línea de investigación orientada a desarrollar algoritmos que detecten automáticamente *spindles* empleando para ello señales de electroencefalogramas (EEG), y que sirvan para estudiar posibles trastornos en función de su número, densidad y características concretas. Además, estas nuevas técnicas permiten disminuir la carga de trabajo de expertos en la labor de detección de dichas oscilaciones en las señales.

En este Trabajo Fin de Grado se ha diseñado un algoritmo basado en técnicas de *machine learning* que detecta *spindles* del sueño en niños de entre 6 y 9 años con sospecha de AOS. La mayoría de estudios científicos publicados hasta la fecha actual sobre la detección de *spindles* se ha centrado principalmente en pacientes adultos, por lo que parece muy adecuado elaborar clasificadores en niños.

La base de datos utilizada está formada por 9 niños y ha sido proporcionada por el Hospital Universitario Río Hortega de Valladolid. Esta base de datos se separó en distintos grupos: uno de entrenamiento (5 sujetos) y otro de test (4 sujetos). Después de preprocesar la señal, varias características se calcularon tanto en tiempo como en frecuencia. Para el análisis temporal se han obtenido la media absoluta, el valor cuadrático medio, relaciones en distintas bandas y *sigma index*, mientras que para el análisis espectral se han extraído potencias relativas, potencias absolutas, envolventes y frecuencias medias. Todas estas medidas se han estimado en las bandas frecuenciales delta (0-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), sigma (12.5-15 Hz), spindle (8-15 Hz) y beta (12-30 Hz) del EEG. El modelo empleado para la clasificación es un *Random Forest*, ya que es un modelo muy flexible formado por una combinación de árboles de decisión que presenta buenas prestaciones para los problemas de clasificación.

Los resultados obtenidos mediante el modelo original reflejan una alta exactitud del 96.51%, sensibilidad del 96.13% y especificidad del 96.52%, lo que implica que se detectan prácticamente todos los *spindles* presentes en las señales. Sin embargo, este modelo presenta la desventaja de obtener precisiones, o valor predictivo positivo, bajas del 10.10%, implicando que hay un alto número de falsos positivos en la predicción debido al desbalanceo en las clases *spindles/no spindle*, siendo mucho más numerosos los segmentos no *spindle* de manera natural. Con el objetivo de mejorar este último parámetro, se experimentó con nuevos modelos entrenados únicamente con las fases N2 y N3 del sueño, por ser donde aparecen los *spindles*. De esta forma se consiguió mejorar la precisión hasta un 12.56%. Nuestros resultados sugieren que es posible llevar a cabo una detección automática y precisa de los *spindles* del sueño en niños con AOS. No obstante, existen limitaciones que deben salvarse para lograrlo.

PALABRAS CLAVE: *spindles* del sueño, Electroencefalograma (EEG), *machine learning*, Random Forest, apnea del sueño infantil.

ABSTRACT

Paediatric Obstruvtive Sleep Apnoea Syndrome (OSA) is a very prevalent disorder (around 25% of the population) in Western societies, causing many diseases associated with cognitive impairment that decreases the quality of life affected children. Thanks to new technological advances, it has been observed that small bursts present mainly during the N2 and N3 phases of sleep between 11-16 Hz, known as sleep spindles, are closely related to the cognitive process of people in general and children in particular. This discovery opens up a new line of research aimed at developing algorithms that automatically detect spindles using electroencephalogram (EEG) signals, and that can be used to study possible disorders according to their number, density and specific characteristics. In addition, these new techniques make it possible to reduce the workload of experts in detecting these oscillations in the signals.

In this Final Degree Project, it has been designed an algorithm based on machine learning techniques that detects sleep spindles in children within 6 and 9 years old with suspected OSA. Most scientific studies published to date on the detection of spindles have focused mainly on adult patients, so it seems very appropriate to develop classifiers for children too.

The database used consists of 9 children and was provided by the Hospital Universitario Rio Hortega in Valladolid, Spain. This database was separated into several sets: a training set (5 subjects) and a test set (4 subjects). After preprocessing the signal, several features were calculated both in time and frequency domains. For the temporal analysis, the absolute mean, the root mean square, ratios in different bands and sigmaindex were obtained, while for the spectral analysis, relative powers, absolute powers, envelopes and mean frequencies were extracted. All these measures have been estimated in the frequency bands delta (0-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), sigma (12.5-15 Hz), spindle (8-15 Hz) and beta (12-30 Hz). The model used for classification was a Random Forest as it is a very flexible model consisting of a combination of decision trees that performs well for classification problems.

The results obtained using the original model reflect a high accuracy of 96.51%, sensitivity of 96.13% and specificity of 96.52%, which implies that practically all the spindles present in the signals are detected. However, this model has the disadvantage of obtaining precisions, also known as positive predictive value, as low as 10.10%, implying that there is a high number of false positives in the prediction due to the imbalance in the spindle/non-spindle classes, being much more numerous the non-spindle segments in a natural way. In order to improve this last metric, it has been experimented with new models trained only with the N2 and N3 sleep stages, as this is where the spindles occur. In this way, we are able to improve precision by up to 12.56%. Our results suggest that it is possible to perform automatic and accurate detection of sleep spindles in children with OSA. However, there are limitations that need to be overcome.

KEYWORDS: sleep spindles, electroencephalogram (EEG), machine learning, Random Forest, paediatric sleep apnea.

ÍNDICE GENERAL

CAPÍTULO 1: INTRODUCCIÓN	1
SEÑALES BIOMÉDICAS	1
ELECTROENCEFALOGRAMA (EEG)	4
OBTENCIÓN DEL EEG	6
APNEA DEL SUEÑO EN NIÑOS	7
SPINDLES DEL SUEÑO	9
HIPÓTESIS Y OBJETIVOS	11
ESTRUCTURA DEL TFG	12
CAPÍTULO 2: DETECCIÓN AUTOMÁTICA DE SPINDLES DEL SUEÑO	14
INTRODUCCIÓN	14
DETECCIÓN CON MACHINE LEARNING	16
PREPARACIÓN DE DATOS	16
OBTENCIÓN DE BASES DE DATOS	16
PREPROCESADO DE DATOS	17
EXTRACCIÓN DE CARACTERÍSTICAS	18
SELECCIÓN DE CARACTERÍSTICAS	18
NORMALIZACIÓN DE LAS CARACTERÍSTICAS	19
BALANCEADO DE DATOS	20
CLASIFICACIÓN: ALGORITMOS DE MACHINE LEARNING	20
SVM (SUPPORT VECTOR MACHINES)	20
ANN (ARTIFICIAL NEURAL NETWORKS)	21
CNN (CONVOLUTIONAL NEURAL NETWORK)	22
DECISION TREE	23
RANDOM FOREST	24
CAPÍTULO 3: MATERIALES Y MÉTODOS	26
SUJETOS Y SEÑALES	27
POBLACIÓN BAJO ESTUDIO	27
SEÑALES CON LAS QUE SE HA TRABAJADO	27
PREPROCESADO DE DATOS	28
EXTRACCIÓN DE CARACTERÍSTICAS	29
BOXPLOTS DE CARACTERÍSTICAS	32
SELECCIÓN DE CARACTERÍSTICAS	33
ENTRENAMIENTO, VALIDACIÓN Y TEST	35

NORMALIZACIÓN DE LAS CARACTERÍSTICAS	35
BALANCEADO DE DATOS (SMOTE).....	36
ALGORITMO DE CLASIFICACIÓN: RANDOM FOREST	37
MÉTRICAS DE EVALUACIÓN	40
MÉTRICAS DE DISCRIMINACIÓN BASADAS EN UMBRAL.....	40
ERROR CUADRÁTICO MEDIO (MSE)	41
IMPORTANCIA DE CARACTERÍSTICAS	42
<i>CAPÍTULO 4: RESULTADOS</i>	44
ANÁLISIS DEL ESPECTROGRAMA	44
RESULTADOS CON EL MODELO ORIGINAL	45
BOXPLOTS DE CARACTERÍSTICAS	46
CARACTERÍSTICAS EN EL TIEMPO	47
CARACTERÍSTICAS EN FRECUENCIA	48
RESULTADOS CON EL CONJUNTO DE TEST	49
RESULTADOS CON FASES N2 Y N3	49
BOXPLOTS DE CARACTERÍSTICAS	50
CARACTERÍSTICAS EN EL TIEMPO	51
CARACTERÍSTICAS EN FRECUENCIA	53
RESULTADOS CON EL CONJUNTO DE TEST	55
IMPORTANCIA DE CARACTERÍSTICAS	56
EJEMPLO DE PREDICCIÓN	58
<i>CAPÍTULO 5: DISCUSIÓN</i>	60
DISCUSIÓN DEL MODELO ORIGINAL	60
DISCUSIÓN DE LOS MODELOS ESPECÍFICOS	61
COMPARATIVA DE RESULTADOS	63
LIMITACIONES	65
PROBLEMAS Y SOLUCIONES.....	66
<i>CAPÍTULO 6: CONCLUSIONES Y LÍNEAS FUTURAS.....</i>	68
CONTRIBUCIONES AL CAMPO DE INVESTIGACIÓN	68
CONCLUSIONES.....	69
LÍNEAS FUTURAS	70

ÍNDICE DE FIGURAS

CAPÍTULO 1: INTRODUCCIÓN	1
Figura 1: Esquema general de distintas señales biomédicas. En él se muestran: electroencefalograma (EEG), potenciales evocados (EP), electroretinograma (ERG), electrocardiograma (ECG), vectorcardiograma (VCG), electrocardiograma fetal (fECG), electroneurograma (ENG), electrohisterograma (EHG), electrogastrograma (EGG), electromiograma (EMG), electrooculograma (EOG), y electrocorticograma (ECoG) (Martinek <i>et al.</i> , 2021).....	3
Figura 2: Ritmos del electroencefalograma dependiendo del estado del sujeto: (a) excitado, (b) relajado, (c) adormilado, (d) dormido y (e) profundamente dormido (Sörnmo & Laguna, 2005).	5
Figura 3: Sistema internacional 10/20 para EEG.	7
Figura 4: Representación de un <i>spindle</i> del sueño frente a un <i>k-complex</i>	9
Figura 5: Estas imágenes representa la estimación media de la densidad de <i>spindles</i> en función de la frecuencia: (a) banda de frecuencias <i>spindle</i> en función de la edad. Las sombras más oscuras indican una mayor densidad <i>spindle</i> . Se puede ver cómo la densidad va disminuyendo con la edad. (b) En las ordenadas se representa la densidad media de <i>spindle</i> de grupos de personas con la misma edad y en las abscisas se representa la frecuencia. (c) Se representan curvas de densidad relativa agrupadas por frecuencias entre 11-15 Hz en función de la edad (Purcell <i>et al.</i> , 2017).....	10
CAPÍTULO 2: DETECCIÓN AUTOMÁTICA DE SPINDLES DEL SUEÑO	14
Figura 6: Margen de hiperplano de separación: (a) hiperplano de separación no óptimo con márgenes asociados (τ) no máximos, (b) hiperplano de separación óptimo con margen asociado máximo (Carmona, 2016).	21
Figura 7: Estructura general de una red neuronal	22
Figura 8: Estructura general de un árbol de decisión (Charbuty & Abdulazeez, 2021).....	23
CAPÍTULO 3: MATERIALES Y MÉTODOS	26
Figura 9: Resumen general sobre los pasos seguidos para la implementación de la detección automática de <i>spindles</i> del sueño empleando <i>Random Forest</i>	26
Figura 10: Representación del canal original con respecto al canal filtrado entre 0.5-30 Hz	28
Figura 11: Representación de la estructura de un <i>boxplot</i> o diagrama de cajas y bigotes. .	32
Figura 12: Representaciones de nubes de puntos asociadas a dos variables <i>X</i> e <i>Y</i> con su coeficiente de correlación de Pearson (Jorge Dagnino, 2014).....	33

Figura 13: Estructura del clasificador <i>Random Forest</i> . <i>Random Forest</i> es un algoritmo de <i>machine learning</i> que combina varios árboles de decisión aleatorios y agrega sus predicciones promediando (Wei et al., 2022)	37
Figura 14: Procedimiento Bootstrap: probabilidad de no seleccionar una muestra del conjunto de entrenamiento tras N extracciones (C. Zhang & Ma, 2012)	38
Figura 15: Esquema general de un nodo no terminal siendo la condición de separación "c" (C. Zhang & Ma, 2012)	39
CAPÍTULO 4: RESULTADOS	44
Figura 16: Representación de 30 segundos del sujeto COG001 a partir del segundo 19031 de grabación: a) representación del espectrograma, b) representación de las marcas de <i>spindles</i> etiquetadas por un experto y c) canal C3-M2 del EEG	45
Figura 17: <i>Boxplots</i> de características normalizadas en el tiempo pertenecientes al conjunto de entrenamiento para el modelo original: a) valor cuadrático medio, b) <i>sigmaindex</i> , c) amplitud media absoluta, d) amplitud media banda sigma, e) amplitud media banda spindle, f) relación en banda alpha y g) relación en banda spindle.	47
Figura 18: <i>Boxplots</i> de las características normalizadas en frecuencia pertenecientes al conjunto de entrenamiento para el modelo original: a) frecuencia media, b) valor medio de la envolvente en la banda sigma, c) valor medio de la envolvente en la banda spindle, d) potencia absoluta en la banda sigma, e) potencia absoluta en la banda spindle, f) potencia relativa en la banda sigma y g) potencia relativa en la banda spindle.....	48
Figura 19: <i>Boxplots</i> de características normalizadas en el tiempo pertenecientes al conjunto de entrenamiento tras la selección para el modelo N2-N3: a) valor cuadrático medio, b) <i>sigmaindex</i> , c) amplitud media absoluta, d) amplitud media absoluta en banda sigma, e) amplitud media absoluta en banda spindle, f) relación en banda alpha y g) relación en banda spindle.....	51
Figura 20: <i>Boxplots</i> de características adicionales normalizadas en el tiempo pertenecientes al conjunto de entrenamiento para el modelo N2-N3: a) diferencia entre amplitud máxima y mínima, b) <i>curtosis</i> , c) <i>skewness</i> d) amplitud media absoluta en la banda delta, e) amplitud media absoluta en la banda theta, f) amplitud media absoluta en la banda alpha y g) amplitud media absoluta en la banda beta.	52
Figura 21: <i>Boxplots</i> de las características normalizadas en frecuencia del conjunto de entrenamiento tras la selección para el modelo N2-N3: a) frecuencia media, b) valor medio de la envolvente en la banda sigma, c) valor medio de la envolvente en la banda spindle, d) potencia absoluta en la banda sigma, e) potencia absoluta en la banda spindle, f) potencia relativa en la banda sigma y g) potencia relativa en la banda spindle.....	53
Figura 22: <i>Boxplots</i> de características adicionales normalizadas en frecuencia pertenecientes al conjunto de entrenamiento para el modelo N2-N3: a) potencia absoluta en la banda delta, b) potencia absoluta en la banda theta, c) potencia absoluta en la banda	

alpha, d) potencia absoluta en la banda beta,e) potencia relativa en la banda delta, f) potencia relativa en la banda theta, g) potencia relativa en la banda alpha y h) potencia relativa en la banda beta.....	54
Figura 23: Importancia de características obtenida a partir del modelo N2-N3 entrenado con el total de características estimadas.	56
Figura 24: Importancia de las características seleccionadas obtenidas en el artículo original (Wei et al., 2022).....	57
Figura 25: Gráfica en la que se representan las marcas originales en azul y en rojo las predicciones. Esta predicción es la correspondiente a la evaluación del modelo N2-N3 para todas las características con el conjunto de validación.	58
CAPÍTULO 5: DISCUSIÓN	60
CAPÍTULO 6: CONCLUSIONES Y LÍNEAS FUTURAS	68

ÍNDICE DE TABLAS

CAPÍTULO 1: INTRODUCCIÓN	1
Tabla 1: Características esenciales de las fases del sueño	6
CAPÍTULO 2: DETECCIÓN AUTOMÁTICA DE SPINDLES DEL SUEÑO	14
CAPÍTULO 3: MATERIALES Y MÉTODOS	26
Tabla 2: Características de los sujetos participantes en el estudio.....	27
Tabla 3: Matriz de confusión.....	40
CAPÍTULO 4: RESULTADOS	44
Tabla 4: Matriz de confusión obtenida tras evaluar el modelo original con el conjunto de test.....	49
Tabla 5: Métricas de discriminación basadas en umbral obtenidas a partir de la matriz de confusión del conjunto de test.	49
Tabla 6: Error cuadrático medio y coeficiente de correlación de Matthew para el modelo original evaluado con el conjunto de test.....	49
Tabla 7: Matriz de confusión obtenida tras evaluar el modelo N2-N3 con el conjunto de test para las 14 características seleccionadas	55
Tabla 8: Matriz de confusión obtenida tras evaluar el modelo N2-N3 con el conjunto de test para todas las características extraídas (sin selección de características).....	55
Tabla 9: Métricas de discriminación basadas en umbral obtenidas a partir de las matrices de confusión de la evaluación de los modelos N2-N3 con los distintos conjuntos de test..	55
Tabla 10: Error cuadrático medio y coeficiente de correlación de Matthew para ambos modelos N2-N3 evaluados con el conjunto de test.	55
CAPÍTULO 5: DISCUSIÓN	60
Tabla 11: Matriz de confusión para el conjunto de test del artículo original (Wei <i>et al.</i> , 2022).	63
Tabla 12: Matriz de confusión para el conjunto de validación del artículo original (Wei <i>et al.</i> , 2022).	63
Tabla 13: : Métricas de discriminación basadas en umbral obtenidas a partir de la matriz de confusión para el conjunto de test y validación del artículo original (Wei <i>et al.</i> , 2022).	63
Tabla 14: Coeficiente de correlación de Matthew para ambos conjuntos de test y validación obtenidos en el modelo del artículo original (Wei <i>et al.</i> , 2022).	63
CAPÍTULO 6: CONCLUSIONES Y LÍNEAS FUTURAS	68

CAPÍTULO 1: INTRODUCCIÓN

SEÑALES BIOMÉDICAS

El sistema nervioso central se comunica con el resto de los sistemas del cuerpo humano mediante señales de origen eléctrico o químico. El estudio de estas señales, portadoras de información, resulta indispensable ya que permite a expertos diagnosticar distintas patologías asociadas a ellas. Cuando la complejidad de estas señales biomédicas es considerable, el preprocesado es la herramienta más utilizada para eliminar componentes de ruido y conseguir que la información se extraiga con mayor precisión (Sörnmo & Laguna, 2005).

Históricamente, las señales biomédicas eran analizadas de forma visual mediante procedimientos en los que las medidas se debían hacer de forma manual, lo que conllevaba a errores de identificación de patrones. Es por ello, que uno de los objetivos del desarrollo del preprocesado de señales biomédicas se centra en reducir de forma significativa la subjetividad de las medidas introduciendo métodos implementados por ordenador que permitan mejorar considerablemente no solo su precisión sino también su exactitud (Sörnmo & Laguna, 2005).

Otra de las ventajas que ha proporcionado el procesado de señales biomédicas es que permite la extracción de características pudiendo así caracterizar las señales bajo estudio y, de esta forma, comprender la información contenidas en ellas. Uno de los propósitos del diseño de tales métodos de extracción, se centra en obtener detalles de la señal que no pueden ser percibidos mediante una mera inspección visual (Sörnmo & Laguna, 2005). Un ejemplo para ello sería el estudio de oscilaciones conocidas como *spindles*, tema que se desarrollará a lo largo de este Trabajo Fin de Grado (TFG). A pesar de ser deseable que las características obtenidas tengan un significado intuitivo para los especialistas, no siempre es así y se necesita por ello cierto conocimiento de la técnica para poder interpretar los resultados.

Existen muchos tipos de señales biomédicas, todas ellas forman parte de procesos de homeostasis (equilibrio que se produce en un medio interno) dentro de algún órgano o pueden ser invocadas por estímulos externos. Dichas señales se pueden clasificar en función de su generación. Entre ellas destacan (Martinek *et al.*, 2021):

- ❖ **Señales bioacústicas:** conllevan la medición de sonidos que generan algunos órganos del cuerpo debido a movimientos mecánicos o de fluidos del cuerpo. Los sonidos del Pulmón (LS) y los del corazón (HS) son los más utilizados para el diagnóstico médico. Para obtener este tipo de señales se necesita disponer de un equipo básico como un estetoscopio (Druzgalski *et al.*, 2016).

- ❖ **Señales biomecánicas:** son aquellas que se generan como consecuencia de procesos mecánicos que tienen lugar en el cuerpo. Contienen medidas sobre desviaciones en la posición, lugar, aceleración, presiones y velocidades (Martinek et al., 2021).
- ❖ **Señales bioquímicas:** aportan información acerca de la concentración de sustancias en el cuerpo humano y del pH (Martinek et al., 2021).
- ❖ **Señales bioeléctricas:** serán las más importantes en este estudio y, por lo tanto, de las que más se va a profundizar. Estas señales se obtienen gracias a fenómenos eléctricos que tienen lugar en la membrana de las células y suelen provenir principalmente de órganos como el cerebro, corazón y otros músculos. Entre ellas podemos encontrar varios tipos (Martinek et al., 2021) (Sörnmo & Laguna, 2005):
 - **EEG** (*electroencefalograma*): refleja la actividad eléctrica de cerebro y se usa para diagnosticar enfermedades tales como la epilepsia. Se hablará más detenidamente de esta técnica en la siguiente sección ya que las señales extraídas para la realización de este TFG proceden del EEG de varios pacientes (Sörnmo & Laguna, 2005).
 - **ECG** (*electrocardiograma*): refleja la actividad eléctrica del corazón colocando electrodos sobre el pecho, manos y pies. Los impulsos eléctricos generados por el miocardio (músculo del corazón), estimulan el latido del corazón y gracias a ellos se puede determinar el ritmo cardiaco. De este modo se permite investigar a cerca de enfermedades como el infarto de miocardio (Sörnmo & Laguna, 2005).
 - **EP** (*evoked potentials*, potenciales evocados): constituyen manifestaciones eléctricas procedentes del cerebro generadas como respuesta a estímulos externos de origen típicamente visual o acústico. Los electrodos empleados para obtener estas señales son similares a los utilizados para el EEG. Clínicamente son muy útiles ya que permiten diagnosticar trastornos originados en los canales de comunicación visual y el tronco encefálico (Chiappa, 1997).
 - **EMG** (*electromiograma*): este tipo de señal representa la actividad neuromuscular, la cual mide corrientes eléctricas generadas por contracciones y relajaciones musculares. El electromiograma es una señal muy compleja debido a que es controlada por el sistema nervioso y depende anatómica y fisiológicamente de las características musculares del cuerpo (Sadikoglu et al., 2017). Su principal uso reside en la detección de enfermedades como distrofia muscular o inflamación de los músculos.
 - **ENG** (*electroneurograma*): se obtiene como resultado de la estimulación de un nervio periférico con una descarga eléctrica tal que la respuesta a lo largo

del nervio pueda ser medida. La adquisición de la señal se logra por medio de electrodos de aguja y se utiliza para determinar la velocidad de conducción de un nervio resultante de su estimulación. Estimando dicha velocidad, se puede diagnosticar si dicho nervio se encuentra dañado o no (Sörnmo & Laguna, 2005).

- **ERG** (*electroretinograma*): este tipo de señal se emplea para el estudio de potenciales eléctricos generados por la retina durante su estímulo por medio de la luz. Se obtiene colocando un electrodo encapsulado en una lente sobre la córnea del ojo. De esta forma es posible evaluar la respuesta eléctrica de los conos y bastones que componen el fondo de la retina de los ojos (Sörnmo & Laguna, 2005).

Existen muchos más tipos de señales bioeléctricas de las cuales no se va a profundizar en este estudio. Se pueden identificar algunas de ellas en función de su localización alrededor del cuerpo humano en la *Figura 1*.

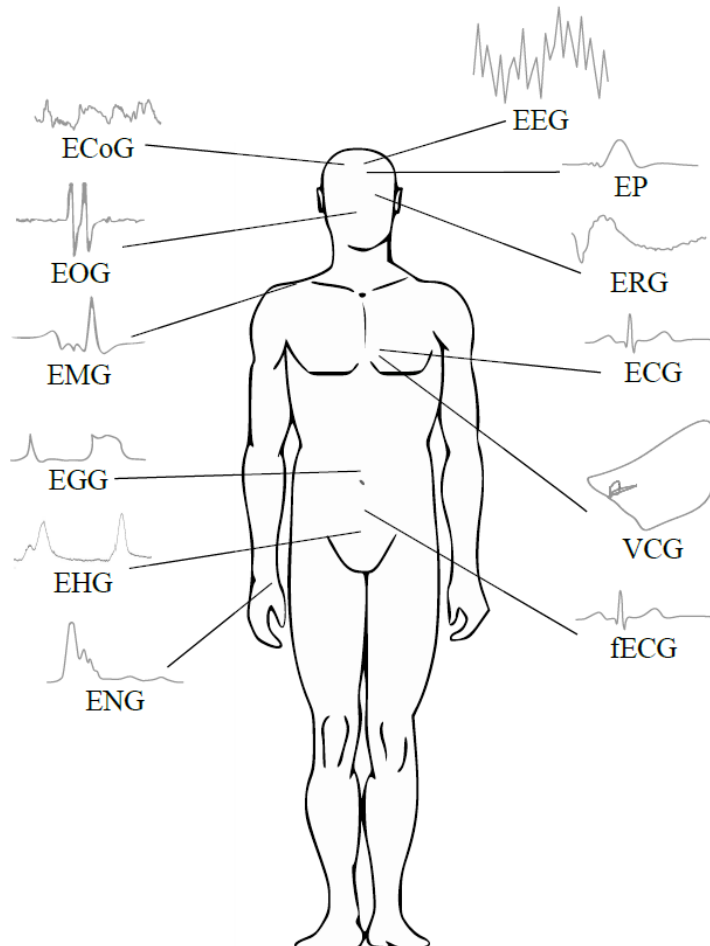


Figura 1: Esquema general de distintas señales biomédicas. En él se muestran: electroencefalograma (EEG), potenciales evocados (EP), electroretinograma (ERG), electrocardiograma (ECG), vectorcardiograma (VCG), electrocardiograma fetal (fECG), electroneurograma (ENG), electrohisterograma (EHG), electrogastrograma (EGG), electromiograma (EMG), electrooculograma (EOG), y electrocorticograma (ECoG) (Martinek *et al.*, 2021)

ELECTROENCEFALOGRAMA (EEG)

El cerebro humano es uno de los órganos más complejos que se conocen. El procesamiento de su información resulta indispensable para poder entender su comportamiento y estudiar su función a lo largo del tiempo. Las técnicas electrofisiológicas son las empleadas para registrar los procesos eléctricos originados en las membranas neuronales que constituyen el cerebro.

El interés que despertó el descubrimiento de fenómenos eléctricos que tenían lugar en los nervios y músculos motivó a varios investigadores a realizar diferentes estudios sobre ellos. Fue el electrofisiólogo inglés Richard Caton en 1874 el pionero de la existencia de la actividad eléctrica del cerebro quien, junto con la influencia de Eduard Hitzig y Gustav Theodor Fritsch, demostraron la evidencia de respuestas motoras eléctricas en el cerebro de otros animales y consiguieron producir convulsiones en dicho órgano después de estimularlo con intensos impulsos eléctricos. Sin embargo, el descubridor del electroencefalograma (EEG) es Hans Berger, el cual, en 1924 realizó el primer registro de las oscilaciones rítmicas a una joven empleando un galvanómetro de cuerda. Gracias a él se inventó el electroencefalograma (Sörnmo & Laguna, 2005), (Palacios, 2002).

Actualmente, el EEG registra la actividad eléctrica de los tejidos cerebrales midiendo, en base a una referencia, el diferencial de potencial existente desde un electrodo explorador (Peralta *et al.*, 2004). La información contenida en el EEG puede ser muy diversa, dependiendo en gran medida del estado en el que se encuentre el sujeto: despierto, dormido, realizando actividad física o relajado (*Figura 2*). Estos ritmos se caracterizan en función del rango de frecuencias en el que se encuentra y en base a su amplitud (Sörnmo & Laguna, 2005).

La amplitud de la señal está íntimamente relacionada con el grado de sincronismo que hay entre la interacción de las neuronas (Sörnmo & Laguna, 2005). La excitación sincrónica de un grupo de neuronas genera amplitudes altas debido a que la señal proviene de la superposición de las señales individuales generadas por las neuronas. La repetición de varias excitaciones sincrónicas neuronales es lo que provoca una señal rítmica en el EEG. Por el contrario, la excitación asincrónica genera amplitudes generalmente bajas, aunque dependiendo de la dispersión temporal, se puede obtener como resultado variaciones en las amplitudes del EEG. (Sörnmo & Laguna, 2005).

En cuanto a la frecuencia o velocidad de oscilación de las distintas señales del EEG, esta se encuentra vinculado a la actividad del tálamo. La información rítmica en forma de patrón se genera gracias a las capacidades neuronales localizadas en esta parte junto con las interacciones que presentan las neuronas de la zona del cortex entre sí (Sörnmo & Laguna, 2005).

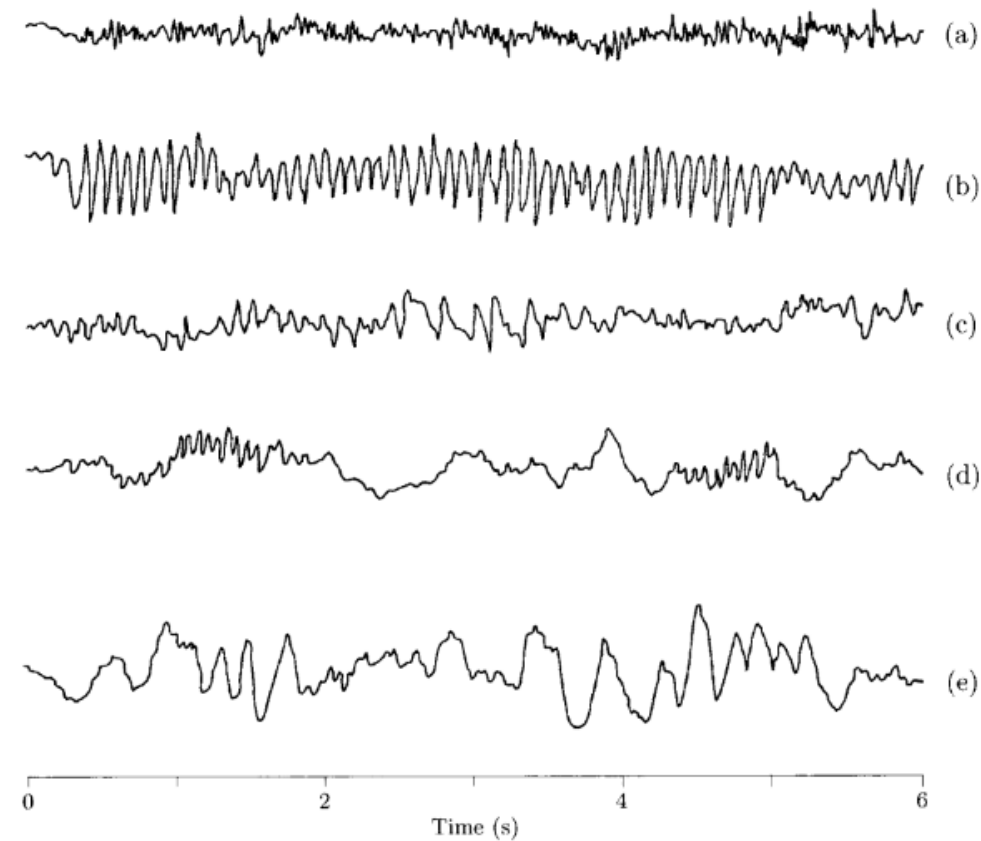


Figura 2: Ritmos del electroencefalograma dependiendo del estado del sujeto: (a) excitado, (b) relajado, (c) adormilado, (d) dormido y (e) profundamente dormido (Sörnmo & Laguna, 2005).

Señales con alta frecuencia y pequeñas amplitudes son propias de situaciones de alerta o de sueño profundo, mientras que señales con baja frecuencia y grandes amplitudes están asociadas a estados de somnolencia o sueño no profundo. Generalmente, el rango de amplitudes en el que se suele encontrar este tipo de señales es de unos pocos microvoltios, aproximadamente $100 \mu\text{V}$. Para el caso de la frecuencia, esta se suele mover entre los rangos de 0.5 a 40 Hz.

En función de la frecuencia registrada del EEG, se pueden distinguir cinco formas de onda categorizadas en sub-bandas:

- ❖ **Delta (< 4 Hz):** caracterizadas por darse en una situación de sueño profundo y presentan amplitudes altas. Estas oscilaciones pueden informar de daños cerebrales o discapacidades en caso de detectarse cuando el sujeto está despierto.
- ❖ **Theta (4-8 Hz):** se caracterizan por darse durante sueño liviano. En los niños, es la principal banda de frecuencia.
- ❖ **Alpha (8-13 Hz):** las oscilaciones detectadas en esta banda se suelen dar en situaciones de un sujeto despierto, pero relajado y con ojos cerrados. Normalmente las amplitudes son mayores en las regiones occipitales. Es la principal banda de frecuencia en los adultos.

- ❖ **Beta (13-30 Hz):** esta forma de onda presenta frecuencias altas y amplitudes bajas, que son asociadas a la actividad del córtex. Se suelen dar en ciertas etapas del sueño y en situaciones de alerta.
- ❖ **Gamma (> 30 Hz):** esta banda se puede observar cuando se realizan se realizan movimientos conscientemente. Está relacionado al estado activo de procesamiento de información del córtex.

Cabe destacar que, dependiendo de la fase del sueño, el EEG puede presentar distintas formas de onda en forma de picos o pequeñas oscilaciones impredecibles con un patrón irregular cuya presencia puede informar a especialistas de la existencia o no de ciertos trastornos o enfermedades.

En la *Tabla 1* se recogen los tipos de oscilaciones más comunes que aparecen en función de las distintas fases del sueño: la etapa NREM, la cual está separada en tres fases distintas (1, 2 y 3), y la etapa REM. Ambas explicadas con mayor detenimiento en apartados posteriores. En este TFG se va a profundizar en el estudio de *spindles* que, como se puede apreciar, aparecen durante las fases 2 y 3 del sueño.

Fase del sueño	Profundidad del sueño	Formas de onda
1	Sonmolencia	Ondas de vértice, ondas alpha
2	Sueño ligero	Ondas de vértice, <i>spindles</i> , <i>k-complex</i>
3	Sueño profundo	<i>K-complex</i> , <i>spindles</i>
REM	Fase REM	Desincronización con frecuencias altas

Tabla 1: Características esenciales de las fases del sueño

OBTENCIÓN DEL EEG

Clínicamente el EEG se obtiene empleando el sistema internacional 10/20 formado por electrodos (*Figura 3*). Este sistema se denomina así debido a que la distancia relativa entre las localizaciones de los electrodos dentro del perímetro craneal está equiespaciada entre el 10% y 20%. Además, su uso ha aumentado considerablemente en los últimos años en aplicaciones cognitivas principalmente en estudios de neurociencia y tratamiento psiquiátrico. En particular, en este sistema 10/20 se emplea un máximo de 21 electrodos unidos en distintos puntos al cuero cabelludo. Para medir la señal se suelen emplear referencias, las cuales pueden ser electrodos cuya posición sea distante con lo que se pretende medir, o el promedio de todos los electrodos del sistema (Sörnmo & Laguna, 2005) (Herwig *et al.*, 2003).

En cuanto al espaciado de los electrodos, éste suele ser bastante disperso. La distancia entre electrodos aproximada para un adulto suele ser de unos 4.5 cm. En caso de querer aumentar la precisión y así conseguir un mayor detalle de las señales, hay estudios que indican que se debe emplear 64 electrodos o más. Esto interesa especialmente cuando el mapeado del cerebro es de gran interés.

En este sistema, los electrodos se identifican con una letra dependiendo de la zona craneal a la que estén conectados. En concreto se emplean las letras F, P, C, T, O y A para identificar las zonas frontal, parietal, central, temporal, occipital y aurícula, respectivamente. Este trabajo se va a realizar a partir de la señal obtenida por el canal C3 localizado en la zona central izquierda como se puede ver en la *Figura 3*.

Una frecuencia de muestreo habitual para el EEG es 200 Hz. Sin embargo, para un mayor detalle en el análisis es habitual emplear una frecuencia de muestreo mayor. Por ejemplo, en el caso de este TFG, las señales obtenidas han sido muestreadas a 500 Hz.

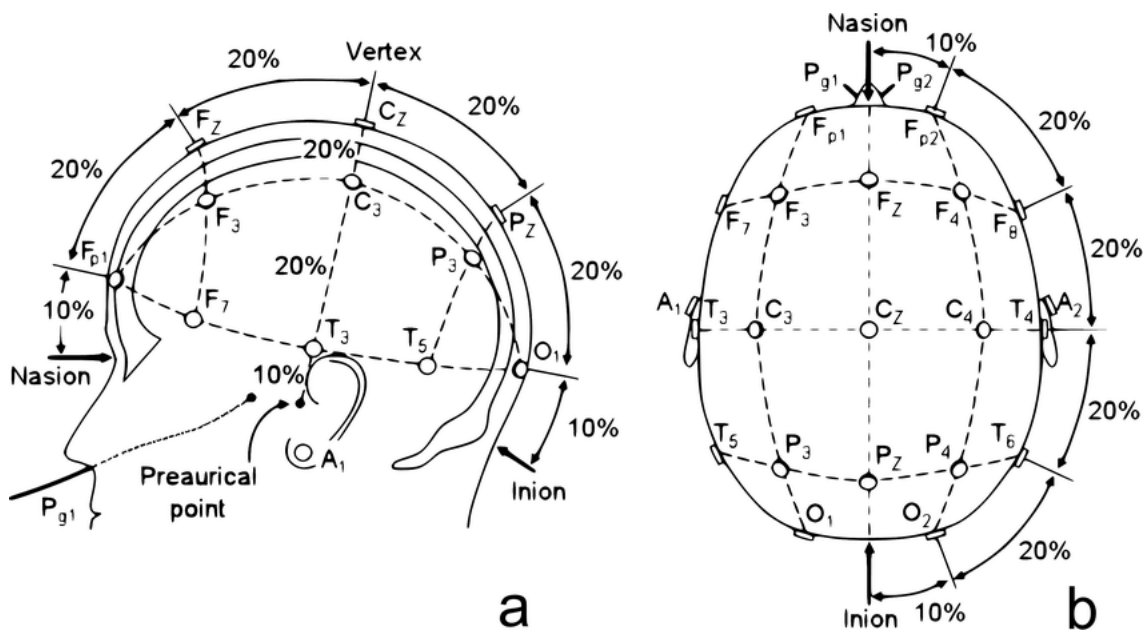


Figura 3: Sistema internacional 10/20 para EEG.

APNEA DEL SUEÑO EN NIÑOS

La apnea del sueño pediátrica (AOS) es un trastorno frecuente, crónico y de evolución progresiva que afecta al 5% de la población adulta (Henry Olivi, 2013) y entre el 1-3% de la población en niños (Uong *et al.*, 2005). Sus rasgos más característicos son episodios repetitivos durante el sueño provocados por la obstrucción completa o parcial de la faringe, que ocasionan hipoxia seguida de re-oxigenación y despertares transitorios, responsable de alteraciones fisiopatológicas (Henry Olivi, 2013). Este trastorno no solo está asociado a la

disminución de la efectividad del sueño, sino también a un empeoramiento de la calidad de vida y un aumento en riesgos cardiovasculares, hipertensiones y metabolismos. Además, el AOS está asociado a un deterioro cognitivo y, es por ello que tiene un gran impacto en patologías como la enfermedad del Alzheimer o Parkinson, en adultos, y en lo que afecta sensiblemente al desarrollo natural y al rendimiento escolar en niños (Uong *et al.*, 2005) (Mohammadi *et al.*, 2021).

Ha sido en las últimas décadas cuando este trastorno ha sido reconocido científicamente y se han llevado a cabo estudios sobre las diferentes enfermedades que conlleva no solo por su alta aparición en la población, sino por el gran impacto que tiene en la salud (Henry Olivi, 2013).

En comparación con los adultos, las causas y características de la apnea del sueño en niños son diferentes, aunque sus consecuencias son igual de perjudiciales para la salud. El AOS en niños puede conllevar a serias complicaciones pulmonares, neuro-cognitivas y disfunciones en la conducta ya que procesos como el crecimiento, la memoria, el desarrollo físico y mental o el desarrollo del cuerpo se realizan durante el sueño principalmente en la etapa de niñez y de adolescencia. Por ello, en los últimos años, se está tratando este tipo de problemas para evitar un efecto negativo que infiera de manera directa a la salud de los niños (Uong *et al.*, 2005). En concreto, para el desarrollo de esta investigación se ha trabajado con varios EEG de niños que presentan este trastorno de edades entre 6 y 9 años.

En niños, uno de los factores de riesgo con mayor relevancia que puede aumentar las probabilidades de padecer apnea al 4,5% es la obesidad. Estudios han comprobado que tanto la obesidad como la apnea del sueño se complementan puesto que el hecho de tener problemas con el sueño favorece un aumento de peso y la obesidad tiende a aumentar la somnolencia (Hermida L, 2016).

Además, el ronquido es otro indicador significativo a la hora de sospechar sobre la existencia de este trastorno en tempranas edades. Presentar este síntoma durante la noche parece un signo importante y por ello se recomienda realizar exámenes clínicos detallados para comprobar la existencia o no de SAO (Hermida L, 2016).

Estudios recientes han encontrado que los *spindles* del sueño están relacionados con procesos cognitivos, emocionales y sociales. La alta densidad de *spindles* infiere en un aumento en la consolidación de la memoria, habilidades lingüísticas y aprendizaje, mientras que el caso contrario suele estar asociado a pacientes con demencia (Mohammadi *et al.*, 2021). Existen diferencias en las características de *spindles* entre la población que sufre AOS y la que no. Se ha demostrado que los pacientes con AOS que están siendo tratados presentan un aumento en la actividad *spindle*. Por este motivo, resulta de gran interés desarrollar algoritmos que detecten estas oscilaciones de forma automática y así poder diagnosticar enfermedades asociadas (Mohammadi *et al.*, 2021).

SPINDLES DEL SUEÑO

Durante el sueño, la actividad neuronal humana va pasando por distintas fases en las que se manifiestan patrones u oscilaciones que están directamente relacionados con estados cerebrales. Se pueden diferenciar dos etapas: NREM (*non rapid eye movement*) y REM (*rapid eye movement*). Dentro de NREM, se pueden localizar tres fases distintas: N1, N2 y N3. Dentro de cada fase del sueño coexisten microeventos que tienen un gran impacto en la actividad neuronal y que se pueden detectar gracias a herramientas como el electroencefalograma (EEG). Existen muchos tipos de oscilaciones como *ripples* o *K-complex*, pero esta sección se centrará únicamente en *spindles* (Figura 4) (Gomez-Pilar et al., 2021).

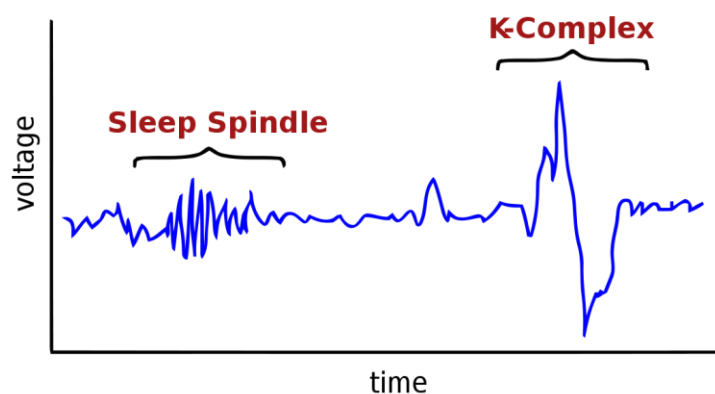


Figura 4: Representación de un *spindle* del sueño frente a un *k-complex*.

El estudio de estas complejas iteraciones resulta importante ya que nos dan información que puede ser útil para detección y prevención de ciertas enfermedades. Debido a ello, en los últimos años, se han llevado a cabo varias investigaciones en las que se demuestra que los *spindles* del sueño desempeñan un importante papel dentro del proceso cognitivo al estar relacionados con la memoria y el aprendizaje. (Gomez-Pilar et al., 2021). La esquizofrenia es un claro ejemplo de trastorno ligado al déficit de *spindles* durante la noche (Manoach et al., 2016).

La mayoría de los expertos caracterizan los *spindles* en función de su duración, número de apariciones por minuto, por su apariencia e incluso en base a su frecuencia de oscilación. Los *spindles* del sueño, generados por la interacción del núcleo reticular talámico y otros núcleos talámicos (Kulkarni et al., 2019), se producen mayoritariamente durante la fase N2 del sueño, aunque también durante N3, entre 11-16 Hz (banda frecuencial sigma) y presentan una duración entre 0.5 y 3 segundos. Además, estos eventos se pueden caracterizar extrayendo características temporales y espectrales (Gomez-Pilar et al., 2021).

Múltiples estudios argumentan distinguir entre dos tipos de *spindles* en función de la frecuencia debido a que presentan distintos rasgos topográficos y funcionales. Los denominados *fast spindles*, por encima de los 13 Hz originados principalmente sobre la zona centroparietal del cerebro, y los *slow spindles* por debajo de los 13 Hz producidos con mayor frecuencia en la zona frontal (Purcell et al., 2017).

La actividad *spindle* puede ser distinta dependiendo del sexo y suele variar con la edad. Es por ello por lo que se necesitan algoritmos distintos según interese detectarlos en adultos o en niños. En adultos, tanto la densidad de *spindle* como la potencia en la banda sigma disminuyen con la edad. Por otro lado, en niños la actividad *spindle* y la potencia en su banda aumentan con la edad alcanzando su pico más alto en la adolescencia. En cuanto a la duración de los *spindles*, progresivamente va disminuyendo desde la infancia hasta la edad adulta. Lo mismo ocurre con la amplitud, aunque se caracteriza por mantenerse estable durante las etapas de niñez y adolescencia. Por último, la frecuencia de oscilación presenta un comportamiento contrario, ya que va aumentando conforme lo hace la edad. Típicamente en niños suele ir de 8 a 15 Hz. Todas estas características se pueden apreciar en la *Figura 5* (Purcell *et al.*, 2017).

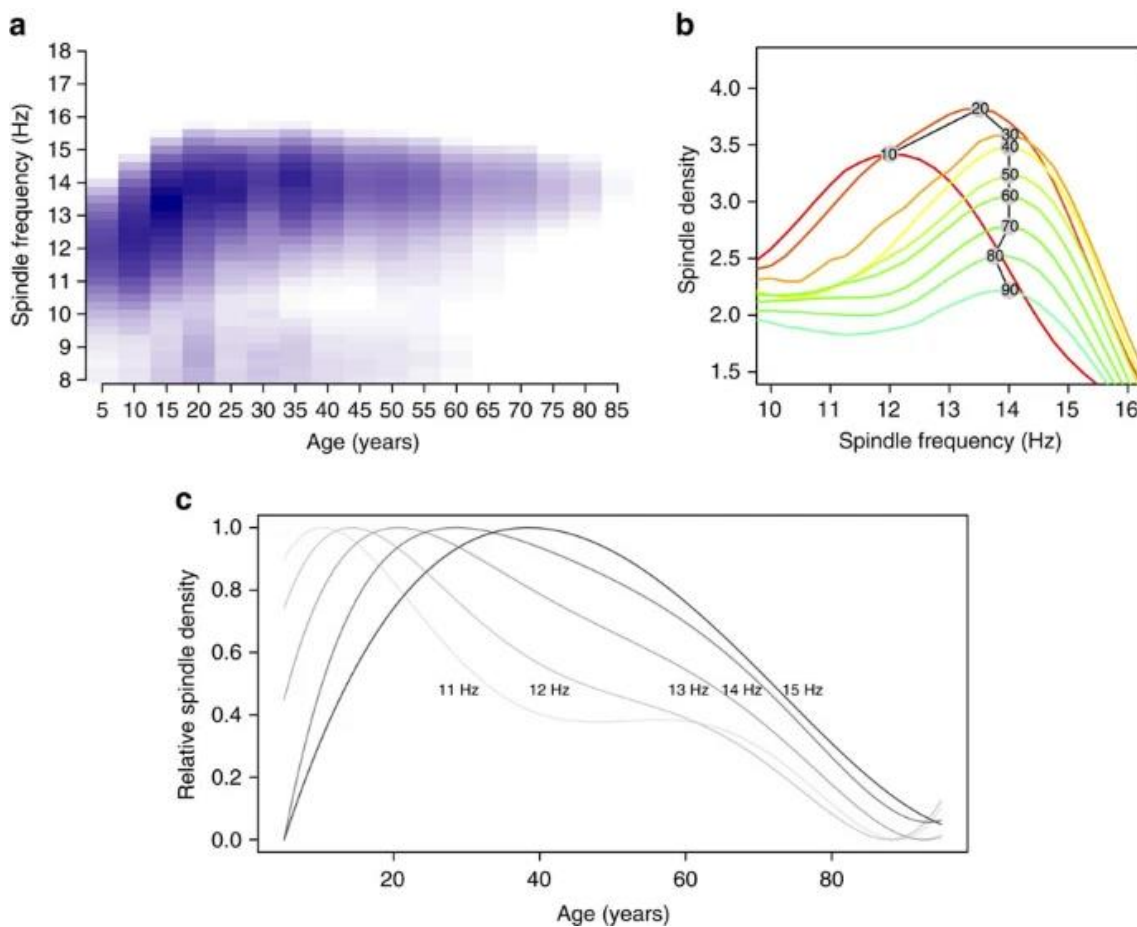


Figura 5: Estas imágenes representan la estimación media de la densidad de *spindles* en función de la frecuencia: (a) banda de frecuencias *spindle* en función de la edad. Las sombras más oscuras indican una mayor densidad *spindle*. Se puede ver cómo la densidad va disminuyendo con la edad. (b) En las ordenadas se representa la densidad media de *spindle* de grupos de personas con la misma edad y en las abscisas se representa la frecuencia. (c) Se representan curvas de densidad relativa agrupadas por frecuencias entre 11-15 Hz en función de la edad (Purcell *et al.*, 2017).

Además, se ha demostrado que hay diferencias en la densidad de *spindles* en función del sexo. El sexo femenino presenta un 0.16 *spindles*/segundo a mayores que el sexo

masculino. Estas diferencias se han observado principalmente en personas con edades en torno al rango de 15-45 años (Purcell *et al.*, 2017).

HIPÓTESIS Y OBJETIVOS

El desarrollo de este TFG parte de la hipótesis de que es posible detectar de manera automática *spindles* del sueño por medio de la información extraída del procesado de la señal de EEG. Persiguiendo esta hipótesis, el objetivo general se centra en desarrollar un algoritmo basado en *machine learning* que sea capaz de detectar estos micro-eventos en niños diagnosticados con apnea del sueño. Para ello, se ha trabajado con registros EEG de niños proporcionados por el Hospital Universitario Río Hortega de Valladolid.

Una vez detallado el objetivo general, se muestran los objetivos específicos para alcanzar este objetivo general:

1. Revisar el estado del arte sobre los algoritmos de *machine learning* que ya han sido publicados en revistas científicas relacionados con la detección de *spindles* del sueño.
2. Seleccionar el algoritmo más apropiado de *machine learning* teniendo en cuenta el tipo de registros *a priori* que se tienen para poder reproducirlo e implementarlo.
3. Estudiar e implementar el algoritmo seleccionado para poder llevar a cabo este trabajo y aportar ideas de mejora o cambios, en caso de ser necesarios.
4. Extraer conclusiones a cerca del estudio desarrollado.

Estos objetivos específicos se pueden dividir en 3 fases, las cuales se han ido siguiendo para completar el trabajo:

Fase 1: Documentación y revisión bibliográfica

- i. Desarrollo del curso introductorio de Python con el objetivo de implementar el algoritmo en este lenguaje de programación.
- ii. Desarrollo del curso introductorio a *machine learning*, en base al cual se ha podido realizar todo el estudio del TFG.
- iii. Revisión bibliográfica de métodos ya implementados con el objetivo de conocer el estado del arte y poder seleccionar el algoritmo que mejor se adapte a las condiciones iniciales.

Fase 2: Estudio del algoritmo e implementación

- i. Información sobre el algoritmo seleccionado para conocer en profundidad su funcionamiento interno y cómo trabajar con él.
- ii. Implementación de este método en Python siguiendo aproximadamente los pasos indicados en la publicación.
- iii. Modificación de ciertos aspectos para la implementación del método que permitan una mejor integración para las señales y población con las que se está trabajando.

Fase 3: Extracción de resultados y conclusiones

- i. Entrenamiento del sistema para obtener resultados a cerca del algoritmo implementado.
- ii. Extracción de conclusiones en base a estos resultados.
- iii. Modificación de la estrategia para tratar de obtener mejores resultados y comparar con otros modelos.
- iv. Redacción del informe final.

ESTRUCTURA DEL TFG

El informe de este TFG consta de un total de 6 capítulos en los cuales se ha ido detallando el proceso a seguir para la realización del estudio. Estos capítulos se dividen en: “Introducción”, “Detección automática de *spindles* del sueño”, “Materiales y métodos”, “Resultados”, “Discusión”, “Conclusiones y líneas futuras”. A continuación, se va a hablar de la estructura que presenta el resto del documento.

En el capítulo 2 se lleva a cabo la revisión del estado del arte dentro del contexto de detección automática de *spindles* del sueño basado en algoritmos de *machine learning*. Este apartado comienza con una explicación en la que se habla de cómo se debe preparar la base de datos con la que se desea trabajar, su preprocesado y cómo se realiza la estimación y selección de características a partir de dichos datos. Por último, se detallan varios métodos utilizados en diversos artículos como clasificadores.

Los sujetos y señales con las que se ha trabajado se describen en el tercer capítulo. Asimismo, se documenta la metodología que se ha utilizado para el desarrollo del TFG comentando todos los pasos seguidos para conseguir el objetivo final de la implementación. Este capítulo finaliza con una serie de métricas de evaluación, las cuales caracterizan los resultados predictivos obtenidos.

Es en el capítulo 4 donde se reflejan los resultados obtenidos para los distintos modelos desarrollados. En concreto, se han diseñado tres modelos dependiendo del número de características con las que se entrena y en función de cómo se ha procesado la señal original. Por cada uno de ellos se presentan *boxplots* de las características extraídas, así como los resultados obtenidos con las métricas de evaluación.

Posteriormente, en el quinto capítulo, se realiza una discusión sobre los valores obtenidos en el capítulo anterior para los distintos modelos desarrollados. Además, se exponen los resultados del artículo seguido para comparar los modelos con los teóricamente esperados. Finalmente, se concluye con una serie de limitaciones con las que se ha topado a lo largo del desarrollo del proyecto.

Finalmente, se expone en el último capítulo las conclusiones extraídas del proyecto junto con posibles líneas futuras y contribuciones al campo de investigación de la Ingeniería Biomédica. Seguidamente la bibliografía, donde se recogen todas las referencias que se han ido consultando durante el transcurso del Trabajo Fin de Grado.

CAPÍTULO 2: DETECCIÓN AUTOMÁTICA DE SPINDLES DEL SUEÑO

A lo largo de esta investigación, ha sido necesaria una revisión del estado del arte con el fin de comprender los avances que se han realizado hasta la fecha. Durante este periodo de documentación, se han revisado artículos publicados en los últimos años y se ha podido analizar cuál de ellos era el más apropiado para replicar en este estudio. A lo largo de ese capítulo se recoge la evolución que ha sufrido la detección de *spindles* hasta la actualidad. Además, se lleva a cabo una explicación en cuanto a las fases de preparación de datos, selección de características, normalización y balanceado. Por último, un apartado dedicado a los distintos clasificadores que se han utilizado en los últimos años y la elección del método a implementar.

INTRODUCCIÓN

Los estudios relacionados con el sueño constituyen una línea de investigación que ha tenido un gran auge en las últimas décadas. El interés que ha despertado el descubrimiento de *spindles* del sueño y su fuerte implicación en el proceso cognitivo de las personas ha hecho que investigadores busquen métodos para poder detectar estas oscilaciones y así ayudar a comprender posibles enfermedades (Parkinson, Alzheimer, etc.) en caso de notificar alguna anomalía en los pacientes.

Tradicionalmente, la identificación de *spindles* se hacía de forma manual mediante la inspección del EEG por expertos en laboratorios clínicos, los cuales eran entrenados en la interpretación de señales polisomnográficas. La detección realizada por expertos es la más segura en cualquiera de los casos, sin embargo, esta labor resulta muy tediosa y conlleva a dedicarle una excesiva cantidad de tiempo, lo cual resulta perjudicial para el interés clínico y biológico. Por ello, este motivo es una de las principales razones por las que se busca automatizar esta detección (Warby *et al.*, 2014).

Hay varias estrategias para la detección automática de *spindles*, y cada una de ellas ha dado lugar a distintos detectores estrechamente relacionados. Uno de los primeros

detectores de *spindles* se desarrolló en el año 1994 y estaba basado en tres etapas: en primer lugar, el EEG se pasaba por un filtro paso banda entre las frecuencias 1.2-16 Hz, en segundo lugar, se estableció un criterio de amplitud respecto a un umbral de 25 μ V y, por último, se calculó la duración del tiempo en épocas en las zonas donde la amplitud superaba dicho umbral (Schimicek *et al.*, 1994).

Este método sirvió de base a muchos posteriores, los cuales introdujeron algunos cambios al respecto como son:

- Cambiar el filtro paso banda estandarizado por uno particular para cada sujeto.
- Reemplazar el filtro paso banda por transformaciones de onda.
- Otras modificaciones empleando la forma del *spindle* para determinar su comienzo y su final en vez de establecer un criterio de umbral para la amplitud.

Gracias al gran avance tecnológico de los últimos años, estos algoritmos de detección automática han evolucionado a técnicas basadas en *machine learning* (ML). Acorde con el pionero de estas herramientas, Arthur Samuel, ML se define como el campo que estudia la habilidad de los ordenadores para aprender a partir de cierta información dada (Batta, 2020). De esta forma las máquinas son capaces de trabajar de manera mucho más eficiente y rápida. El propósito principal reside en aprender de los datos y en base a ellos obtener un resultado con sentido (Batta, 2020). Estas técnicas poco a poco han ido tomando importancia dentro del ámbito del sueño y actualmente es uno de los métodos más utilizados para la detección y clasificación de eventos.

Los primeros artículos que hablan de detección de *spindles* empleando estas técnicas se basan en redes neuronales automáticas (ANN, *Authomatic Neural Network*) (Ventouras *et al.*, 2005). Además, hay publicaciones en las que primeramente emplean una preclasificación para eliminar de la señal elementos *no-spindle* y quedarse únicamente con los elementos *spindle* con los que se trabajará en una etapa posterior, denominada postclasificación. Máquinas de vectores de soporte (SVM, *Support Vector Machine*) o perceptrones multicapa (MLP, *Multilayer Perceptron*) han sido necesarios para implementar esta técnica en el artículo (Acir & Güzeliş, 2004). Con el paso de los años se ha ido evolucionando hacia técnicas de mayor complejidad, que dejan atrás las redes neuronales convencionales con las que se obtienen buenos resultados. Duman *et al.* (2009) utiliza en su publicación un único árbol de decisión que realiza la clasificación en función de tres técnicas propuestas (Duman *et al.*, 2009), mientras que Kinoshita *et al.* optó por una combinación de la transformada SST (*Synchrosqueezed Transform*) y un sobremuestreo aleatorio implementado con técnicas de boosting (Kinoshita *et al.*, 2020). Además, Wei *et al.* (2022) consiguió unos buenos resultados utilizando el algoritmo de *Random Forest*, de forma similar a la metodología que se llevará a cabo en este TFG (Wei *et al.*, 2022). Incluso hay publicaciones en las que, debido a la gran cantidad de datos, se da un paso más hacia *deep learning* con un algoritmo cuyos autores denominan "SpindleNet" compuesta por una red neuronal convolucional (CNN, *Convolutional Neural Network*) y una red neuronal recurrente (RNN, *Recurrent Neural Network*) (Kulkarni *et al.*, 2019).

Todos los procesos basados en *machine learning* convencional presentan una serie de etapas en común, las cuales son: preparación de datos (incluye tanto su obtención como su

preprocesado), extracción de características y, por último, clasificación. En los siguientes apartados se abordan esta serie de aspectos.

DETECCIÓN CON MACHINE LEARNING

PREPARACIÓN DE DATOS

Para poder llevar a cabo una detección aplicando *machine learning*, resulta necesario la preparación de los datos con los que se va a trabajar. En esta parte pueden distinguirse dos fases. La primera de ellas está relacionada con la obtención de bases de datos y, la segunda, relacionada con el pre-procesado de dichos datos.

OBTENCIÓN DE BASES DE DATOS

Trabajar dentro de la rama de investigación dedicada a la apnea del sueño, implica tratar con un gran número de señales biomédicas procedentes de distintos pacientes. Normalmente estas señales se adquieren de bases de datos ya existentes, aunque hay estudios en los que registran y analizan sus propias señales. En los últimos estudios, las bases de datos que más han sido utilizadas son:

- ❖ **MASS** (Montreal, Canadá): Contiene 200 grabaciones polisomnográficas (PSG) de 97 hombres y 103 mujeres de edad comprendidas entre 18 y 76 años. Estas señales se encuentran almacenadas en el formato europeo de datos (EDF, *European Data Format*) (Kulkarni *et al.*, 2019). Las señales de 200 sujetos de todos ellos contienen marcas de *spindles* establecidas por expertos en el canal C3 del EEG.
- ❖ **DREAMS** (Universidad de Bruselas, Bélgica): Contiene grabaciones de 30 minutos de PSG provenientes de 8 sujetos formados por 4 hombres y 4 mujeres de edades comprendidas entre 35 y 50 años. Todas estas señales presentan marcas de *spindles* anotadas por expertos (Kulkarni *et al.*, 2019).
- ❖ **NSRR** (*National Sleep Research Source*): se trata de un sistema que se diseñó con el objetivo de facilitar la investigación de estudios relacionados con el sueño. En él se encuentran 26000 grabaciones PSG y datos clínicos de varios sujetos durante toda la noche (Kulkarni *et al.*, 2019) (G. Q. Zhang *et al.*, 2018).
- ❖ **CHAT** (*Childhood Adenotomyllectomy Trial*): se encuentra almacenado en NSRR. En particular, contiene un total de 1447 grabaciones EEG de los canales C3 y C4 en niños con anotaciones de las fases del sueño marcadas por expertos (Kulkarni *et al.*, 2019).

PREPROCESADO DE DATOS

El pre-procesado de datos engloba aquellas técnicas de análisis de datos que permiten mejorar la calidad de las señales con la finalidad de proporcionar una mayor y mejor información. El ruido o corrupciones dentro de la señal además de posibles interacciones entre canales influyen en gran medida en los resultados obtenidos (Herrera, 2004). Es por ello, que uno de los aspectos más cruciales en el preprocesado de señales biomédicas consiste en adquirir conocimientos sobre el ruido o artefactos presentes en la señal con el objetivo de minimizar su influencia (Sörnmo & Laguna, 2005). Las técnicas más empleadas en este ámbito son:

- ❖ **Referenciado promedio de la señal:** Antes de filtrar la señal en las bandas de interés, es habitual realizar un referenciado promedio de todos los canales del EEG y restarlo a cada elemento de la señal. Esta técnica se suele emplear para eliminar posibles interferencias que haya entre unos canales y otros.
- ❖ **Filtrado digital:** Es conveniente filtrar la señal EEG para eliminar ruido y componentes de información innecesaria. Es una práctica habitual emplear un filtro *Nocht* lineal invariante en el tiempo (filtro elimina banda) para cancelar la interferencia de línea eléctrica que aparece al grabar los registros (50 Hz en Europa y 60 Hz en Estados Unidos). Además, el uso de filtros paso banda permiten trabajar únicamente con las bandas de interés. Esto mejora considerablemente la calidad y precisión de la señal. Aplicar filtrados paso bajo suele ser muy útil para reducir la interferencia de la actividad EMG cuando interesa especialmente frecuencias bajas en el EEG. Por último, es importante llevar a cabo un buen diseño de estos filtros ya que en el caso contrario podrían añadir espurios de ruido a la señal (Sörnmo & Laguna, 2005).
- ❖ **Eliminar *offset* de la señal:** Otra técnica de preprocesado en señales biomédicas consiste en eliminar la componente de continua de la señal para cada canal de EEG extraído. Normalmente esta práctica consiste en hacer el promediado de toda la señal y restárselo a cada elemento (Wei *et al.*, 2022).
- ❖ **División del EEG en fragmentos:** Una vez aplicados los preprocesados anteriores, se divide el EEG en segmentos de la misma duración para posteriormente poder extraer características de ellos. Concretamente, para el estudio de *spindles* del sueño, la mayoría de las publicaciones dividen el EEG en fragmentos entre 0.5 y 3 segundos con un *overlap* de entre 0.25 y 0.5 segundos. Esto es así debido a que la duración de un *spindle* se encuentra comprendida entre esos rangos (Wei *et al.*, 2022), (Duman *et al.*, 2009). Aunque no sea lo habitual, es posible encontrar artículos que dividan la señal en 30 segundos antes de realizar la extracción de características (Güneş *et al.*, 2011).

EXTRACCIÓN DE CARACTERÍSTICAS

La actividad *spindle* presenta características propias que la diferencian del resto de la señal de EEG. La extracción de estas características son las que permiten aplicar un algoritmo de clasificación que permita detectar para cada época la presencia o no de un *spindle* del sueño. Se pueden distinguir dos conjuntos de características en función de si se realiza un análisis temporal o frecuencial de la señal.

- ❖ **Análisis temporal:** Este análisis es el más simple ya que se trabaja con la señal original en el dominio del tiempo, sin necesidad de aplicar ningún tipo de transformación en ella. Para cada época en la que se ha segmentado la señal de EEG se pueden extraer parámetros estadísticos conocidos como son la media, varianza, desviación típica, skewness, kurtosis, simetría de la señal, valor cuadrático medio, etc (Wei *et al.*, 2022).
- ❖ **Análisis frecuencial:** El estudio de las componentes espectrales del EEG proporciona una gran cantidad de información que no podría ser obtenida con un mero análisis temporal. El análisis frecuencial se puede evaluar aplicando la densidad espectral de potencia (PSD) de las señales usando herramientas como la FFT (*Fast Fourier Transform*) o STFT (*Short Time Fourier Transform*). Los métodos más usados para el cálculo de la PSD son el método de Welch y el método de Blackman-Tukey (Gomez-Pilar *et al.*, 2021). Después de realizar este cálculo, se pueden obtener características como son potencias absolutas y relativas para las distintas bandas del EEG (Wei *et al.*, 2022).

SELECCIÓN DE CARACTERÍSTICAS

Para llevar una correcta clasificación, resulta importante conocer aquellas características que presentan una alta correlación con otras, ya que no suelen proporcionar información adicional y se puede prescindir de ellas. Se requiere de técnicas que permitan seleccionar de entre todas las características extraídas, aquellas que realmente aporten contenido relevante. Hay tres técnicas dentro de *machine learning* que se encargan de esta selección: métodos de envoltura, métodos de filtrado y métodos integrados o híbridos (Khalid *et al.*, 2014).

- ❖ **Métodos de envoltura** (*wrapper methods*): Se diferencian con los de filtrado en que el proceso de selección está optimizado en función del clasificador usado. Estos métodos presentan un mayor coste computacional cuando el número de características es muy grande.
- ❖ **Métodos de filtrado** (*filter methods*): Son métodos mucho más rápidos y con menor coste computacional, aunque hay ocasiones en las que presentan una ineficiente fiabilidad en cuanto a la clasificación. Estos métodos son más adecuados para conjuntos de datos de gran tamaño.

- ❖ **Métodos integrados o híbridos** (*embedded/hybrid methods*): Presentan ventajas de ambos métodos ya mencionados. Un método híbrido emplea para cada subconjunto de características una función de prueba y una evaluación de su rendimiento para obtener mejores resultados.

NORMALIZACIÓN DE LAS CARACTERÍSTICAS

Antes de comenzar con la implementación del algoritmo de *machine learning*, conviene realizar una normalización para evitar que unas características destaquen por encima de otras. Hay varios métodos que permiten realizar dicha normalización de los cuales destacan: escalador estándar (*standard scaler*), escalador de valor absoluto máximo (*maximum absolute value scaler*) y la transformación cuantil (*quantile transform*) (Ferreira *et al.*, 2019).

El **escalador estándar (SS)** es un método de normalización que se aplica a cada característica restando su media y dividiendo entre su varianza. Esta técnica, al depender únicamente de su propia varianza y media, presenta ventajas como son una alta escalabilidad, mayor rapidez computacional y linealidad, sin embargo, manifiesta una alta sensibilidad en los datos. En la ecuación (1) se puede ver la normalización estimada de una característica en función de su media \bar{X} y desviación típica σ (Ferreira *et al.*, 2019). Este método será el que se utilizará para la normalización en este TFG.

$$\widehat{X}_i = \frac{X_i - \bar{X}}{\sigma} \quad (1)$$

El **escalador de valor máximo absoluto (MA)**, por su parte, es un método que consiste en normalizar cada muestra dividiéndola por el máximo valor absoluto de la característica. Esta técnica permite manejar datos dispersos además de permitir escalabilidad, linealidad y reversibilidad. Como desventaja, el MA presenta una alta sensibilidad, al igual que el método anterior. En la ecuación (2) se puede ver la estimación de la normalización en función del máximo valor absoluto X_{MA} (Ferreira *et al.*, 2019).

$$\widehat{X}_i = \frac{X_i}{X_{MA}} \quad (2)$$

Por último, la **transformación cuantil (QT)** es un método de escalado de características que emplea una función de distribución acumulativa sobre la información cuantil de cada característica. Presenta la ventaja de que es un método muy robusto para los datos atípicos de una distribución. Sin embargo, un conjunto de test que se encuentre fuera de los límites de la función de distribución del conjunto de entrenamiento, puede generar problemas a la hora de la clasificación (Ferreira *et al.*, 2019).

BALANCEADO DE DATOS

La aparición de *spindles* se distribuye en las fases N2 y N3 del sueño (NREM) con una densidad que varía de unos sujetos a otros. El número y duración de estos eventos es substancialmente menor que los eventos de no *spindles*, generando un problema de desequilibrio entre los datos de gran relevancia para la clasificación ya que hay muchas más muestras de la clase *no-spindle* que de *spindle*. En diversos artículos se emplean técnicas con el objetivo de paliar esta cuestión:

- ❖ **SMOTE** (*Synthetic Minority Oversampling*): Es un método de sobremuestreo, en el cual se crean nuevas características sintéticas de la clase menor hasta conseguir que el número de muestras entre una clase y la otra sea igual (Wei *et al.*, 2022).
- ❖ **RUSBoost**: Es un algoritmo de *boosting* para clasificación binaria el cual puede paliar los problemas relacionados con el desequilibrio de datos empleando submuestreo. La metodología se basa en eliminar muestras de la clase dominante de manera apropiada (Kinoshita *et al.*, 2020).
- ❖ **Aplicando intervalos de tiempo fijo**: El objetivo es generar un conjunto de entrenamiento equilibrado aumentando el tamaño de la clase desfavorable hasta que ambas estén equilibradas seleccionando un intervalo de tiempo fijo antes y después del centro de un *spindle*. Así, se genera una ventana móvil de tamaño T_w de tal forma que, si el 50% de la muestra se solapa con un *spindle*, se incluye dentro de la clase de *spindle* (Kulkarni *et al.*, 2019).

CLASIFICACIÓN: ALGORITMOS DE MACHINE LEARNING

Durante esta última etapa se consigue predecir si el segmento de señal pertenece a un *spindle* o no. En los últimos años este proceso de clasificación se ha llevado a cabo gracias a algoritmos basados en *machine learning* supervisados. Dentro de este ámbito hay diversas técnicas disponibles y muchas de ellas se pueden aplicar a la detección automática de *spindles* del sueño. A continuación, se recoge una breve explicación de varios algoritmos que se han utilizado en artículos que pretenden detectar estas oscilaciones en distintas bases de datos.

SVM (SUPPORT VECTOR MACHINES)

Son algoritmos de aprendizaje supervisado pertenecientes a la categoría de clasificadores lineales, ya que inducen separadores lineales (hiperplanos) entre las dos clases cuya

dependencia carece de linealidad. Para ello, los datos de entrada pueden ser mapeados a otro espacio aplicando herramientas o funciones *kernel*: funciones polinómicas, sigmoideas, multi cuadráticas, funciones de base radial (RBF, *Radial Basis Function*) o gaussianas, entre otras. En esta nueva situación, tras la etapa de transformación, el problema de optimización se centra en escoger de entre todos los hiperplanos posibles, aquel que optimice la distancia entre los puntos más cercanos de cada clase y cuyo margen asociado sea máximo (Figura 6). De este modo, el problema de optimización no lineal se convierte, en cierta medida, en uno lineal. Desde el punto de vista práctico, este tipo de algoritmos evitan los problemas relacionados con el desajuste entre los datos (Carmona, 2016), (Wang, 2005).

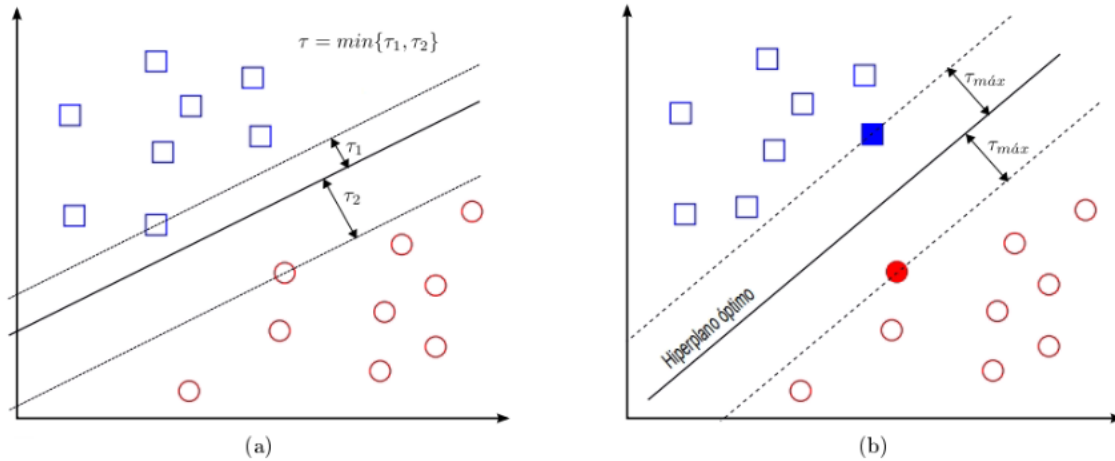


Figura 6: Margen de hiperplano de separación: (a) hiperplano de separación no óptimo con márgenes asociados (τ) no máximos, (b) hiperplano de separación óptimo con margen asociado máximo (Carmona, 2016).

En relación con la detección de *spindles* del sueño, este algoritmo se ha utilizado como post clasificador en la publicación (Acir & Güzeliş, 2004). En esta implementación se utiliza la función de base radial RBF como *kernel*. Posteriormente, en el problema de optimización se emplea el parámetro de regulación “C”, que controla el hiperplano de separación. Además, en este artículo se comparan los resultados con otro post clasificador MLP (*multilayer perceptron*) obteniendo unos mejores resultados con SVM en cuanto a sensibilidad y tasa de falsa detección (Acir & Güzeliş, 2004).

ANN (ARTIFICIAL NEURAL NETWORKS)

Una red neuronal artificial (ANN) es un algoritmo de *machine learning* supervisado, inspirado en las redes neuronales biológicas del cerebro humano. Generalmente, esta arquitectura está compuesta por varios niveles o capas, cada una de ellas formada por una o varias neuronas artificiales, conectadas por medio de unos coeficientes conocidos como pesos. Existen dos capas que conectan un ANN al exterior: una capa de entrada (*input layer*), por medio de la cual se introducen las características extraídas de los datos, y una capa de salida (*output layer*), gracias a la cual se obtiene la predicción o respuesta de clasificación de la red neuronal. El resto de capas intermedias se denominan ocultas (*hidden layer*) (Agatonovic-Kustrin & Beresford, 2000) (Figura 7).

Por otra parte, la optimización del valor de los pesos que interconecta las neuronas se realiza mediante propagación hacia atrás (*backward propagation*) del error cometido durante la fase de entrenamiento. Es decir, una ANN realiza una lectura de los valores de entrada y de salida de esta durante su entreno y va cambiando el valor inicial de los pesos hasta que se consigue minimizar el error entre las predicciones y la realidad.

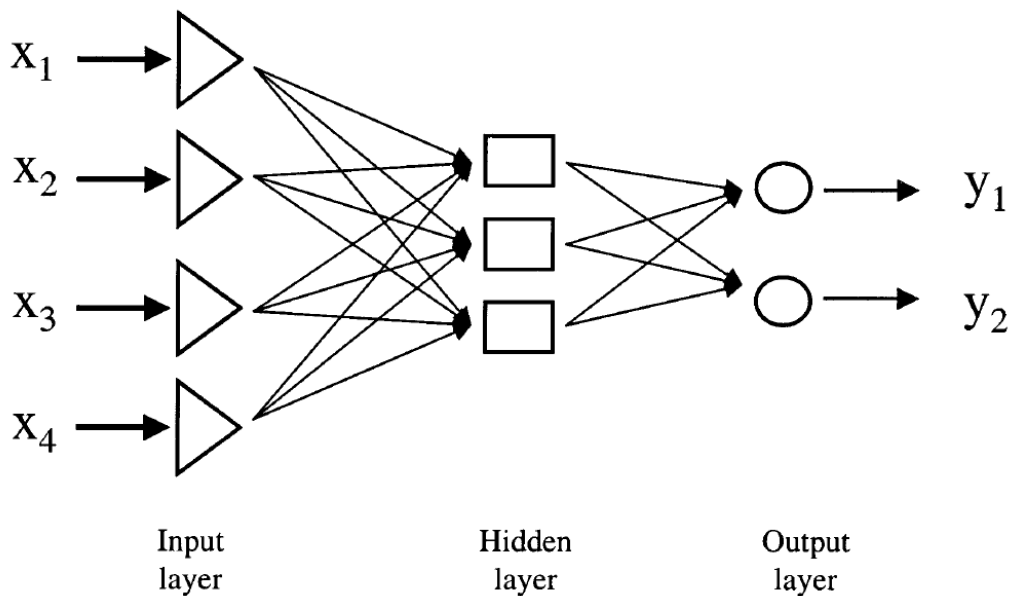


Figura 7: Estructura general de una red neuronal

Dentro del campo de *spindles* del sueño se han encontrado artículos como Ventouras *et al.* (2005) ó Güneş *et al.* (2011), que realizan su detección empleando este tipo de algoritmos. Además, es una práctica habitual el empleo de ANN junto con otros algoritmos de clasificación. Un ejemplo para ello lo podemos encontrar en Acir & Güzeliş (2004), donde se emplea una versión de estos algoritmos para la pre clasificación de tal forma que las salidas obtenidas en la última capa de la red corresponderán a las entradas de los algoritmos utilizados en la post clasificación.

CNN (CONVOLUTIONAL NEURAL NETWORK)

Las redes convolucionales neuronales (CNN) son un tipo de ANN que se emplea en algoritmos de *deep learning*. Toma este nombre de la operación matemática lineal entre matrices conocida como convolución. CNN es un algoritmo multicapa que incluye una capa convolucional, una no lineal, una capa de agrupación y otra de conexión. De todas ellas sólo presentan parámetros la capa convolucional y la de conexión (Albawi *et al.*, 2018).

Para detección de *spindles* del sueño, se ha empleado este algoritmo en Kulkarni *et al.* (2019). Debido al empleo de *deep learning*, resulta necesario disponer de bases de datos muy grandes con el objetivo de tener una masiva cantidad de datos con la que trabajar. En concreto se elabora un algoritmo denominado SpindleNet, el cual está formado por una CNN complementada por una RNN (*Recurrent Neural Network*). Las conclusiones más claras que

se reflejan en este artículo son la obtención de resultados con mayor precisión al emplear *deep learning* en comparación con *machine learning*. Además, se ha testeado el algoritmo con las diferentes bases de datos obteniendo diferencias en valores de especificidad o precisión.

DECISION TREE

Es una de las técnicas de clasificación más amplias y usadas en las últimas décadas. Los árboles de decisión son una herramienta muy potente con aplicaciones en varios campos: *machine learning*, procesado de imagen e identificación de patrones. En cuanto a su estructura, es similar a la de un árbol. Inicialmente presenta un nodo raíz (*root node*), a partir de la cual van saliendo ramificaciones hasta llegar a los nodos hojas (*leaf node*), que contienen el resultado final de clasificación de tipo booleano (*Figura 8*). Cada nodo representa una característica distinta que será evaluada la entrada. Para la clasificación, en este tipo de modelos se lleva a cabo comparando, por cada nodo, el valor de la entrada en función de un umbral. Este tipo de arquitectura resulta mucho más sencilla que el proceso de optimización de pesos llevado a cabo en ANN, y es por ello que se puede encontrar implementada en una gran variedad de campos.

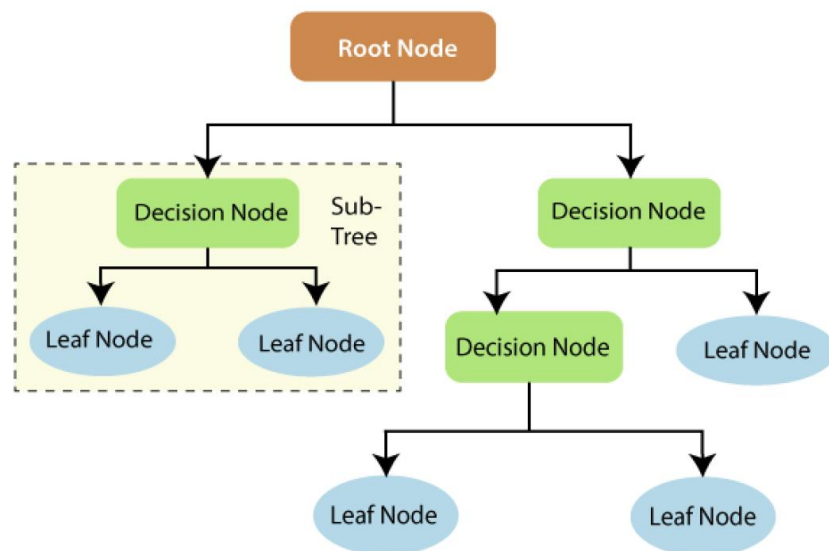


Figura 8: Estructura general de un árbol de decisión (Charbuty & Abdulazeez, 2021)

En la publicación Duman *et al.* (2009) se emplea un árbol de decisión para desarrollar un detector de spindles analizando registros EEG de varios sujetos. Para ello se han evaluado diferentes métodos como son TEO (*Teager Energy Operator*), MUSIC (*Multiple Signal Classification*) o STFT (*Short Time Fourier Transform*). Los resultados obtenidos son muy adecuados tanto de sensibilidad como de especificidad.

RANDOM FOREST

Random Forest es un algoritmo de *ensemble learning* que utiliza múltiples árboles de decisión como clasificadores base. Este método ofrece un mayor rendimiento en comparación con un solo árbol y con otros modelos de clasificación. Además, presenta la ventaja de poder trabajar con grandes bases de datos y varios predictores, no obstante, para mejorar los resultados es conveniente minimizar este número de predictores para mejorar su eficiencia (Speiser *et al.*, 2019). El empleo de *Random Forest* presenta un gran interés no solo para la clasificación, sino también porque permite estimar la importancia de las características extraídas en el *dataset*. En cuanto a la clasificación, cada árbol que forma el *Random Forest* determina la clase más probable a la que pertenece su entrada. Con todas estas salidas se calcula la clasificación final mediante voto por mayoría (Gislason *et al.*, 2006).

Este modelo se ha implementado para la detección de *spindles* en el artículo de Wei *et al.* (2022), el cual se ha tratado de replicar en este estudio debido a los buenos resultados obtenidos y el tipo de población bajo estudio. Los pasos seguidos, así como los resultados obtenidos tras el diseño del modelo vienen descritos en los siguientes capítulos.

CAPÍTULO 3: MATERIALES Y MÉTODOS

En este capítulo se pretende detallar la metodología llevada a cabo en este estudio para conseguir desarrollar un algoritmo que detecte de forma automática *spindles* del sueño en niños. Este procedimiento se ha basado fundamentalmente en el artículo de Wei *et al.*, (2022), en el cual se desarrolla un algoritmo similar centrado en la detección de *spindles* en recién nacidos. Se han realizado algunas modificaciones al respecto debido principalmente al tipo y cantidad de sujetos y señales con las que se ha podido trabajar.

En el primer apartado “sujetos y señales” se describe la población bajo estudio y señales con las que finalmente se ha trabajado del EEG. Posteriormente se detallan todas las fases por las que se ha pasado para lograr el objetivo final: procesado de datos, extracción y selección de características, balanceado de datos y algoritmo de clasificación. Una vez finalizada la clasificación final con el conjunto de test, se ha evaluado el algoritmo final utilizando métricas de evaluación como son el uso de matrices de confusión, importancia de características o precisión (Figura 9).

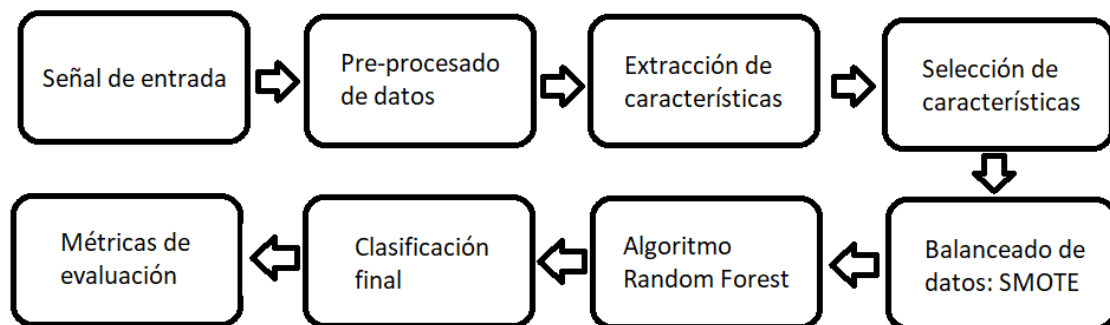


Figura 9: Resumen general sobre los pasos seguidos para la implementación de la detección automática de *spindles* del sueño empleando *Random Forest*.

La implementación de todos estos bloques se ha realizado principalmente con la herramienta Python 3.10, aunque para el procesado de señal y alguna extracción de características se ha utilizado MATLAB R2021b ya que ofrecía una mayor facilidad.

SUJETOS Y SEÑALES

En este apartado se pretende presentar la base de datos que se ha empleado para realizar el estudio, así como presentar al tipo de pacientes sobre los cuales se han recogido los distintos datos. Además, se detalla el canal del EEG con el que se ha trabajado a partir del cual se han podido extraer y seleccionar las distintas características.

POBLACIÓN BAJO ESTUDIO

El objetivo principal de este TFG se centra en desarrollar un algoritmo que detecte de manera automática *spindles* del sueño en niños que sufren la apnea del sueño. Se ha trabajado concretamente con el EEG de 9 sujetos de edades comprendidas entre 6 y 9 años, de los cuales 4 de ellos son de sexo femenino y 5 masculino. Estas señales han sido proporcionadas por el Hospital Universitario Río Hortega de Valladolid.

Las bases de datos no solo contienen las distintas señales de cada sujeto, sino que también presentan algunas características de ellos (*Tabla 2*) y se incorporan las marcas de *spindles*, que han sido colocadas por un experto en el ámbito de la AOS, en los canales C3, C4, Fp1 y Fp2 del EEG.

Sujeto	Edad (años)	Peso (kg)	Altura (cm)	Sexo	Fecha PSG
COG001	6.7	28	124	Femenino	28/01/2020
COG002	7.3	31	130	Femenino	04/02/2020
COG003	8.7	40.5	132	Masculino	25/02/2020
COG004	8.1	36	138	Masculino	26/02/2020
COG005	9.6	33.8	138	Femenino	04/03/2020
COG006	7.6	28	125	Femenino	05/03/2020
COG007	7.5	32	135	Masculino	10/03/2020
COG008	6.4	17	113	Masculino	11/03/2020
COG009	6.6	23	117	Masculino	12/03/2020

Tabla 2: Características de los sujetos participantes en el estudio.

SEÑALES CON LAS QUE SE HA TRABAJADO

Las señales EEG se grabaron, para cada sujeto, empleando los canales C3, C4, F3, F4, O1 y O2 referenciados al mastoide M2. Todos los registros presentan una duración de 10 horas y todas las señales han sido muestreadas a la misma frecuencia de 500 Hz. De todos los canales mencionados, únicamente se ha empleado el C3-M2 para la realización del algoritmo de detección.

PREPROCESADO DE DATOS

Para poder caracterizar los *spindles* del sueño en niños, resulta necesario realizar varias transformaciones a las señales originales con el objetivo de eliminar ruido e interferencias. Por ello se ha filtrado la señal original entre 0.5-30 Hz para eliminar, no solo la componente de ruido procedente de la corriente eléctrica a 50 Hz, sino también para descartar información banal superior a 30 Hz. En la *Figura 10* se representa el canal original junto con el canal filtrado en las frecuencias anteriormente indicadas.

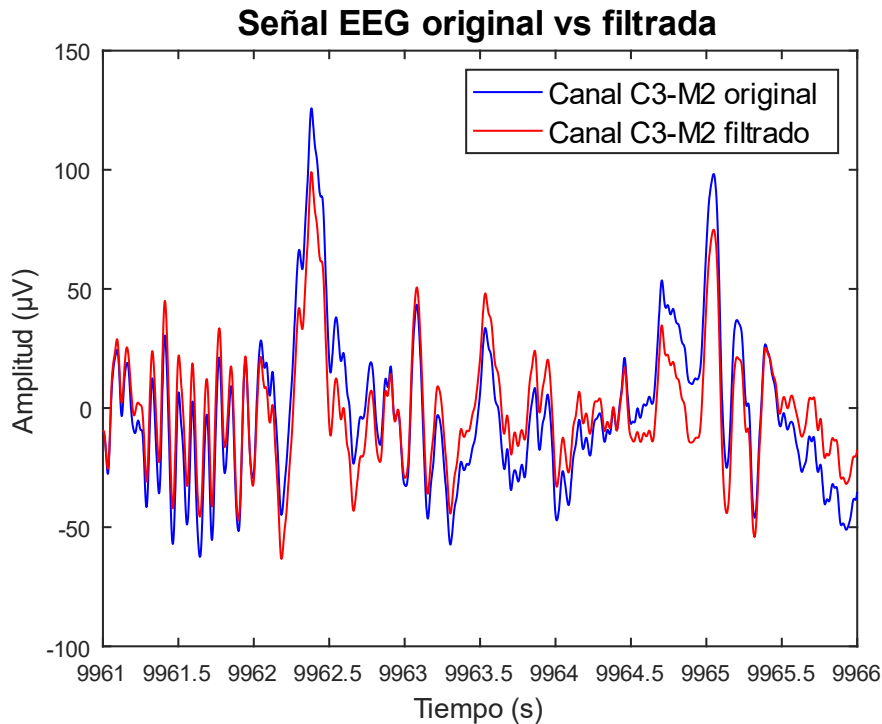


Figura 10: Representación del canal original con respecto al canal filtrado entre 0.5-30 Hz

A continuación, y siguiendo las indicaciones del artículo seleccionado (Wei *et al.*, 2022), se vuelve a filtrar la señal en varias bandas de interés con el propósito de estimar características en cada una de ellas. Las bandas frecuenciales en las que se ha filtrado el canal C3-M2 son:

- Delta (0-4 Hz)
- Theta (4-8 Hz)
- Alpha (8-12 Hz)
- Sigma (12.5-15 Hz)
- Spindle (10.5-16 Hz)
- Spindle en niños (8-15 Hz)
- Beta (12-30 Hz)
- F1 (4-10 Hz)
- F2 (20-40 Hz)
- F3 (12.5-15 Hz)

De todas estas frecuencias, la banda frecuencial *spindle* en niños se ha añadido a mayores puesto que el artículo original habla de recién nacidos, los cuales presentan una frecuencia

spindle distinta. Las bandas F1, F2 y F3 serán utilizadas para calcular alguna característica específica como es sigma index empleada para conocer la relación entre las distintas bandas. El filtrado se ha realizado gracias a la potente herramienta de MATLAB *filterDesign*. Gracias a ella se han podido diseñar los filtros de manera sencilla con las siguientes características: FIR (*Finite Impulse Response*) enventanado con ventana *hamming*, de orden 1000 puntos y frecuencia de muestreo 500 Hz. Dependiendo de la banda frecuencial de interés, el filtro se ha comportado como un paso bajo o paso banda.

El siguiente paso antes de extraer ninguna característica ha sido dividir cada señal filtrada en épocas de duración 0.5 segundos con un *overlap* de 0.25 segundos. La elección de estos tiempos no ha sido al azar, sino que se basa en el principio de que el tiempo mínimo requerido para la duración de un *spindle* es 0.5 segundos y de las especificaciones del artículo (Wei *et al.*, 2022). Esta segmentación se ha realizado en Python empleando la librería “*more_intertools*” para poder utilizar una ventana deslizante que recorra toda la señal. De esta forma se obtiene una matriz de $M \times N$ (M épocas de N muestras cada una) de almacena en archivos con extensión “.mat”.

EXTRACCIÓN DE CARACTERÍSTICAS

Tras obtener las épocas para cada señal filtrada de cada sujeto, se puede pasar a la siguiente etapa de cálculo o extracción de características. Para ello ha servido de base las características propuestas por el artículo (Wei *et al.*, 2022). En total se han extraído 29 parámetros, los cuales se pueden clasificar en dos tipos en función del tipo de análisis utilizado: temporal o frecuencial.

En primer lugar, se va a hablar de las características calculadas para cada época mediante el análisis temporal:

- **Valor cuadrático medio:** El valor cuadrático medio, media cuadrática, raíz de la media cuadrática o RMS (*root mean square*) es una medida estadística que se define como la raíz cuadrada de la media de todos los elementos al cuadrado de una magnitud (*ecuación 3*). Es una medida muy útil cuando interesa calcular la media que toman valores positivos y negativos. Su cálculo se ha llevado a cabo empleando la librería “*numpy*” de Python.

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

- **Amplitud media absoluta:** Es una medida estadística que calcula la media de todos los elementos en valor absoluto (*ecuación 4*). Con el valor absoluto se consigue trabajar con valores positivos y negativos del EEG obteniendo así el valor medio positivo de cada segmento. El cálculo se ha llevado a cabo empleando la librería “*numpy*” de Python.

$$mean_abs = \frac{1}{N} \sum_{i=1}^N |x_i| \quad (4)$$

- **Amplitud media absoluta en cada banda:** Esta característica se calcula de igual modo que la ecuación 4 con la singularidad de que se calcula para las distintas bandas en que la señal del EEG ha sido filtrada. Así pues, la amplitud media absoluta tendrá valores próximos a cero cuando no haya señal entre esas frecuencias y distintos a cero en el caso contrario. Las bandas en las que se ha estimado esta característica son: delta (0-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), sigma (12.5-15 Hz), spindle en niños (8-15 Hz) y beta (12-30 Hz).
- **Skewness:** El *skewness* de una variable X se mide en función de su tercer momento central ($\gamma = \mu_3/\sigma^3$) y viene dado por la ecuación 5 (Yetton *et al.*, 2016). Este parámetro informa a cerca de la asimetría de la distribución de probabilidad de la variable (Author *et al.*, 1984). Su cálculo se ha llevado a cabo gracias a la librería “*scipy.stats*” de Python. Si este parámetro vale cero significa que se trata de una distribución normal, simétrica, mientras que valores más altos implican un mayor peso en la cola derecha de la distribución.

$$skew = \frac{E(X - \mu)^3}{\sigma^3} \quad (5)$$

- **Curtosis:** Se trata de una medida estadística que mide lo achatada que está la distribución de probabilidad de una variable aleatoria en función de su cuarto momento central ($\gamma = \mu_4/\sigma^4$) (ecuación 6) (Yetton *et al.*, 2016). Un valor positivo indica que la distribución se concentra más en el valor de la media, mientras que un valor negativo indica que hay una mayor varianza entre los datos de la muestra. Su cálculo se ha realizado utilizando la librería “*scipy.stats*” de Python empleando el método de Fisher.

$$kurt = \frac{E(X - \mu)^4}{\sigma^4} \quad (6)$$

- **Diferencia entre amplitud max-min:** Se trata del cálculo de la diferencia entre el valor máximo y mínimo de las amplitudes contenidas en cada época segmentada. Este cálculo se ha llevado a cabo mediante la librería “*numpy*” de Python.
- **Sigma index:** Se trata de una medida muy utilizada en artículos debido a sus buenos resultados en detección de *spindles* en adultos. Mide la relación que hay, para cada segmento del EEG, entre las bandas F3 (12.5-15 Hz) y F1 (4-10 Hz) y F2 (20-40 Hz) (Wei *et al.*, 2022). Se puede ver la ecuación implementada en Python para el cálculo de esta ecuación 7.

$$sigmaindex = \frac{mean(|F3|)}{mean(|F1|) + mean(|F2|)} \quad (7)$$

- **Relación en banda alpha:** Este parámetro es el valor cuadrático medio en la banda alpha en comparación con el valor cuadrático medio de la señal completa para cada época. De este modo se puede conocer la proporción de señal que se encuentra localizada en ese rango de frecuencias. El cálculo de esta característica se ha llevado a cabo en Python implementando la *ecuación 8*.

$$\alpha_band_ratio = \frac{rms(\alpha\ band)}{rms(complete\ signal)} \quad (8)$$

- **Relación en banda spindle:** Idem al cálculo anterior particularizando para la banda spindle en niños.

En el análisis frecuencial fueron extraídas las siguientes características:

- **Frecuencia media:** Se calcula la frecuencia media de cada segmento del canal C3 en función de la frecuencia de muestreo de la señal. De esta forma, si la frecuencia media se localiza dentro de la banda *spindle*, será probable que ese fragmento se corresponda a un evento de este tipo. La obtención de esta característica se ha realizado en MATLAB con el comando “meanfreq”.
- **Potencia absoluta en cada banda:** Para el cálculo de la potencia absoluta se calcula la densidad de potencia espectral por cada segmento de 0.5 segundos en los que se divide la señal EEG tras el preprocesado. La obtención de la PSD se ha realizado empleando el método de Welch, muy utilizado en señales EEG (*Güneş et al., 2011*). Este método divide la señal temporal en segmentos, posteriormente calcula su periodograma aplicando un enventanado para cada segmento y, por último, se obtiene el promedio de las funciones de densidad espectrales estimadas. La implementación de la PSD se ha realizado empleando el método “pwelch” de MATLAB con parámetros: ventana *hanning*, 50% de solapamiento entre segmentos y 512 puntos de la DFT.

Una vez calculada la PSD, la potencia absoluta se calcula sumando los valores de las componentes entre las frecuencias deseadas. Las bandas en las que se ha llevado a cabo la estimación de este parámetro son: delta (0-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), sigma (12.5-15 Hz), spindle en niños (8-15 Hz) y beta (12-30 Hz).

- **Potencia relativa en cada banda:** La frecuencia relativa se calcula dividiendo la potencia absoluta en la banda específica entre la suma de la PSD completa. Esta característica se ha calculado empleando la herramienta de MATLAB de forma análoga a la potencia absoluta. Las bandas en las que se ha evaluado esta característica son: delta (0-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), sigma (12.5-15 Hz), spindle en niños (8-15 Hz) y beta (12-30 Hz).
- **Valor medio de la envolvente:** La estimación de esta característica se ha llevado a cabo empleando la transformada de Hilbert de la señal, ya que esta es una herramienta muy útil para describir la envolvente compleja de una señal modulada

con una portadora real. Con el valor absoluto de esta transformación se calcula el valor medio de dicha envolvente. Este proceso se realiza para cada época para las señales filtradas tanto en la banda sigma como en la *spindle*. El valor medio de la envolvente se ha podido obtener gracias a la librería “*scipy.signal*” de Python.

BOXPLOTS DE CARACTERÍSTICAS

Una buena práctica, que se ha llevado a cabo en este proyecto, consiste en realizar *boxplots* de las características calculadas. Un *boxplot*, también conocido como diagrama de “caja y bigotes”, es un tipo de representación que permite identificar valores atípicos y la distribución de un conjunto de valores a través de sus cuartiles. Se pueden detectar rápidamente los siguientes valores (*Figura 16*):

- **Primer cuartil (Q1)**: El 25 % de los valores son menores o iguales a ese valor
- **Mediana o segundo cuartil (Q2)**: Divide la distribución de valores en dos partes iguales de tal forma que el 50% de los valores son menores o iguales a ese punto.
- **Tercer cuartil (Q3)**: El 75% de los valores son menores o iguales a ese valor.
- **Rango intercuartílico (RIC)**: Corresponde a la diferencia de valores entre el tercer y primer cuartil (Q3-Q1).
- **Bigotes**: Se corresponde a las líneas que se extienden desde la caja hasta los valores máximo y mínimo o hasta 1.5 veces el RIC.
- **Valores atípicos**: También conocidos como *outliers*. Representan observaciones anormales o atípicas de la distribución de los datos que se quedan fuera de los límites de la caja y bigotes.
- **Notches**: Muestran el 95% del intervalo de confianza de la mediana. Se representan con forma de V alrededor del valor de la mediana.

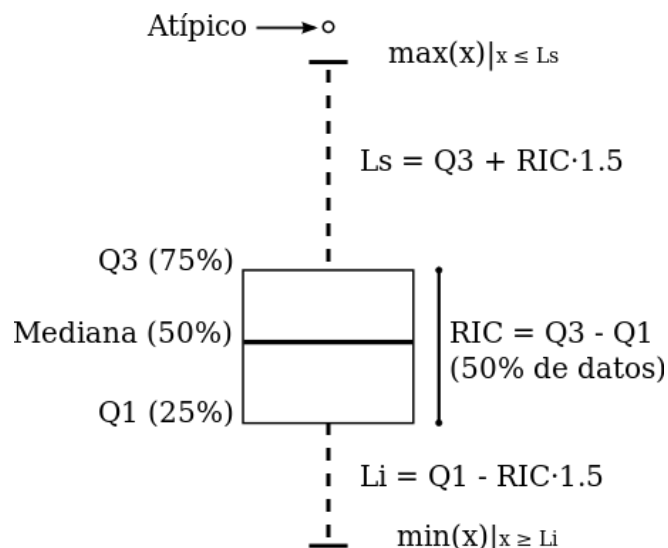


Figura 11: Representación de la estructura de un *boxplot* o diagrama de cajas y bigotes.

SELECCIÓN DE CARACTERÍSTICAS

El siguiente paso, tras el cálculo de las características para cada época, consiste en establecer una selección de todas ellas para evitar la redundancia de información y quedarnos sólo con lo más relevante. Aunque la sección de características no se haya realizado explícitamente en este estudio, a continuación se determina el proceso seguido por el artículo (Wei *et al.*, 2022) para conseguirlo.

En primer lugar, esta selección de características se ha llevado a cabo aplicando un método de envoltura o *wrapper* alrededor del clasificador de *Random Forest*. El cálculo de la pérdida media de precisión y la desviación estándar fue necesario con el objetivo de obtener un ranking de las características. A continuación, se evaluó el coeficiente de correlación de Pearson para identificar la asociación que hay entre las variables. Por último, se reevaluó cada característica por separado utilizando el Coeficiente de Correlación de Matthew (MCC).

El coeficiente de correlación de Pearson (r) mide el grado de asociación lineal entre dos variables tomando un valor comprendido entre -1 y +1. El valor de 0 se alcanza cuando no hay ningún tipo de relación o correlación entre ambas variables, mientras que, si toma un valor de los extremos, -1 o +1, significa que la asociación entre ambas variables, X e Y , es muy fuerte. Antes de tomar cualquier decisión sobre la correlación, conviene representar en una “nube de puntos” la relación entre ambas variables. En la figura 11 se muestran ejemplos en función del valor que toma “ r ” con las variables, de tal manera que cuanto más cerrada es dicha “nube de puntos”, mejor es el valor que toma “ r ”.

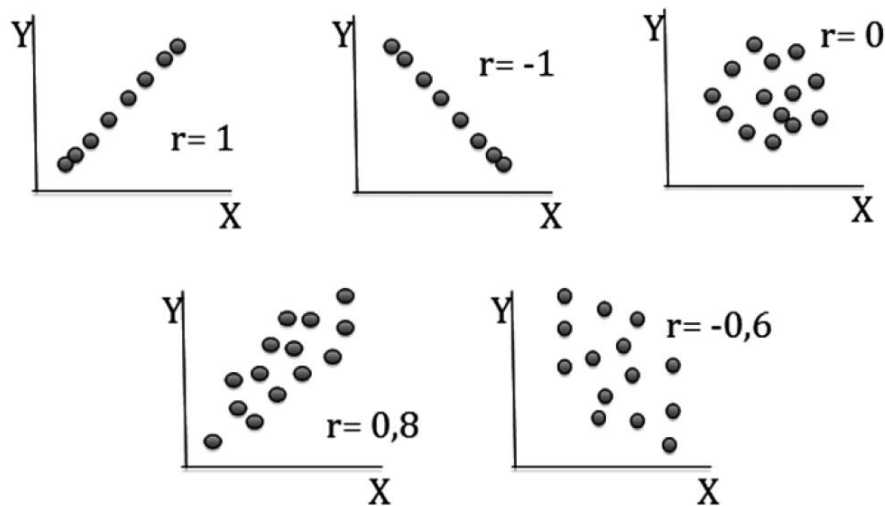


Figura 12: Representaciones de nubes de puntos asociadas a dos variables X e Y con su coeficiente de correlación de Pearson (Jorge Dagnino, 2014)

La ecuación que describe el coeficiente de correlación lineal es la 10, la cual se desarrolla a partir de la raíz del coeficiente de determinación (ecuación 9) para poder conservar el signo.

$$r^2 = \frac{(XY)^2}{(XX)(YY)} \quad (9)$$

$$r = \frac{(XY)}{\sqrt{(XX)(YY)}} \quad (10)$$

El coeficiente de correlación de Pearson fue calculado por cada par de características para identificar cuáles de ellas presentaban un valor alto, mayor de 0.8. Aquellas variables que presentaron un valor $r > 0.8$ se eliminaron de tal forma que el conjunto de características final quedó redujo a 20.

Finalmente, se aplicó la métrica MCC individualmente a cada característica después de aplicar el algoritmo de *Random Forest* con el conjunto de validación. Esta herramienta se puede considerar una discretización en dos variables del coeficiente de correlación de Pearson y su cálculo, ecuación 11, se realiza en base a la matriz de confusión: falso positivo (FP), falso negativo (FN), verdadero positivo (TP) y verdadero negativo (TN) (Boughorbel et al., 2017). Tras aplicar la medida MCC, se contó con un total de 14 características finales, eliminando aquellas que presentaban un coeficiente menor que 0.0001 (Wei et al., 2022).

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FN)(TP + FP)(TN + FP)(TN + FN)}} \quad (11)$$

Finalmente, se seleccionaron las siguientes 14 características:

1. Amplitud media absoluta.
2. Valor cuadrático medio.
3. Frecuencia media.
4. Sigma index.
5. Relación en banda alpha.
6. Relación en banda spindle.
7. Amplitud media absoluta en banda sigma.
8. Valor medio de la envolvente en banda sigma.
9. Potencia relativa en banda sigma.
10. Potencia absoluta en banda sigma.
11. Amplitud media absoluta en banda spindle.
12. Valor medio de la envolvente en banda spindle.
13. Potencia relativa en banda spindle.
14. Potencia absoluta en banda spindle.

En el estudio realizado para este TFG, se ha implementado el algoritmo *Random Forest* con estos 14 parámetros para poder comparar en el capítulo 5 de discusión las diferencias obtenidas en la implementación con respecto al artículo original (Wei et al., 2022). Además, también se ha evaluado el algoritmo con todas las características iniciales para posteriormente conseguir la importancia de cada una de ellas en el conjunto de entrenamiento y ver si existe coherencia entre los diferentes resultados obtenidos.

ENTRENAMIENTO, VALIDACIÓN Y TEST

La aplicación de algoritmos basados en *machine learning* es una práctica muy habitual en los últimos años en los que los ordenadores aprenden y predicen en función de unos datos iniciales. Es una práctica habitual dividir estos datos en tres bases de datos con distintas finalidades: conjunto de entrenamiento (*training set*), conjunto de validación (*validation set*) y conjunto de test (*test set*). Estos grupos se forman de tal manera que cada sujeto sólo pertenezca a uno de ellos, por lo que los datos que contienen cada uno de ellos son independientes entre sí (Eagan et al., 2019).

El **conjunto de entrenamiento** se emplea para entrenar al sistema. Por ello, es importante que los datos que forman este grupo sean de calidad y estén bien preparados, ya que el modelo irá aprendiendo en base a ello y generará las predicciones. Además, el *training set* es el que presenta una mayor cantidad de datos en comparación con los otros. El **conjunto de validación**, por su parte, se utiliza para evaluar el modelo entrenado y así poder ajustar el valor de los hiperparámetros del modelo hasta optimizarlos al máximo. Normalmente se emplea también este grupo de validación para evitar el problema del sobreajuste en los modelos modificando los parámetros. Por último, el objetivo del **conjunto de test** reside en evaluar el algoritmo final después de ser entrenado y ajustado (Eagan et al., 2019).

La estrategia que se ha seguido para dividir la base de datos en los distintos grupos ha sido repartir los sujetos de tal forma que el conjunto de entrenamiento contenga la mayor cantidad, y que tanto el grupo de validación como el de test presenten el mismo número de sujetos con la idea de adaptarse de la mejor manera posible al artículo original (Wei et al., 2022). En este artículo se contiene un total de 141 sujetos de niños recién nacidos, los cuales 81 de ellos forman el conjunto de entrenamiento y 30 tanto el de validación como el de test. De esta forma el 58% de los sujetos pertenecerían al *training set* y el 21% a cada uno de los grupos restantes. Sin embargo, como en este estudio no ha sido necesaria una optimización de los hiperparámetros dado que se han utilizado los mismos que en el artículo, se ha prescindido del grupo de validación trabajando finalmente con sólo dos grupos: grupo de entrenamiento y grupo de test.

Teniendo en cuenta que en nuestro estudio se cuenta con un total de 9 sujetos, la forma de repartirlos entre los sets de entrenamiento y test ha sido 5 y 4, respectivamente. El *training set* ocupará por tanto el 55.5% de los datos, mientras que el grupo de test ocupará el 44.5% restante. Es muy importante que un sujeto pertenezca exclusivamente a uno de los conjuntos de datos y no estar repartido entre varios, ya que de hacer lo contrario, podría incurrirse en sesgos.

NORMALIZACIÓN DE LAS CARACTERÍSTICAS

Antes de comenzar con la implementación del algoritmo de *Random Forest* y después de realizar la separación de características en los distintos conjuntos de entrenamiento,

validación y test, conviene realizar una normalización para evitar que unas características destaquen por encima de otras. De entre todos los métodos posibles, se ha implementado el escalador estándar explicado en el capítulo anterior.

La metodología seguida se ha basado en calcular tanto la media como la desviación típica de cada característica únicamente del conjunto de entrenamiento. Con estos datos, se estima la estandarización de la variable restándole a cada muestra la media y dividiendo entre la desviación estándar anteriormente calculadas. Este proceso se realiza tanto para el grupo de entrenamiento como para la validación y test, teniendo en cuenta que las medias y desviaciones pertenecen en todo momento a las características de entrenamiento.

BALANCEADO DE DATOS (SMOTE)

El desequilibrio entre datos es un problema frecuente en aplicaciones biomédicas que ocurre cuando el tamaño de las clases de datos presenta una distribución desigual. La mayoría de algoritmos basados en *machine learning* trabajan bien cuando los datos están balanceados, sin embargo, cuando no lo están tienden a estar sesgados hacia la clase mayoritaria provocando que sus predicciones sean ineficientes (Boughorbel *et al.*, 2017).

En este estudio se ha trabajado con una base de datos que presenta un 1.15% de clase positiva (*spindle*) y un 98.85% de clase negativa (*no spindle*), por lo que se puede ver que el problema del desequilibrio entre los datos es muy significativo y conducirá a problemas en las predicciones.

Para paliar esta cuestión, se ha utilizado la técnica del SMOTE (*Sythetic Minority Oversampling Technique*) al igual que se hace en el artículo replicado (Wei *et al.*, 2022) antes de realizar el entrenamiento del sistema. Este método sobremuestra la clase minoritaria (la clase minoritaria es la clase *spindle*) para aumentar el número de muestras creando otras sintéticas a partir de segmentos provenientes de los “*k vecinos*” más cercanos de los ejemplos originales. Estos “*k vecinos*” se eligen de forma aleatoria por el algoritmo para generar muestras distintas. Este enfoque fuerza a que la región de decisión de la clase sobremuestreada sea más general (Kovács *et al.*, 2020).

SMOTE ha sido implementado sobre el conjunto de entrenamiento utilizando la librería “*imblearn.over_sampling*” de Python. Para ello son necesarios dos parámetros de configuración: *k_neighbors* y *random_state*. El primero de ellos indica el número de vecinos implicados en la construcción de muestras sintéticas. El segundo parámetro controla la aleatorización del algoritmo SMOTE en función de una semilla generadora de números aleatorios. Los valores elegidos para ambos parámetros han sido *k_neighbors* = 5 y *random_state* = 2 ya que estos valores han sido los utilizados en la publicación (Wei *et al.*, 2022).

```
from imblearn.over_sampling import SMOTE  
  
[...]
```

```
oversample=SMOTE(random_state=2,k_neighbors=5)
```

```
X_train,y_train=oversample.fit_resample(X_train, y_train)
```

ALGORITMO DE CLASIFICACIÓN: RANDOM FOREST

El algoritmo de clasificación *Random Forest* resulta de una combinación de árboles de decisión idénticos e independientes entre sí distribuidos de la misma forma en el “bosque” en los que cada árbol emite un voto para la clase más popular en su entrada. En los últimos años, este método de clasificación se ha convertido en uno de los más populares debido a la introducción de significantes mejoras en cuando a precisión en la clasificación. El voto por mayoría de las predicciones generadas por los distintos árboles de decisión es lo que genera la predicción de clasificación final. Este procedimiento es lo que se denomina *Random Forest* y en la *Figura 12* se puede ver el esquema general de este método implementado (Wei *et al.*, 2022).

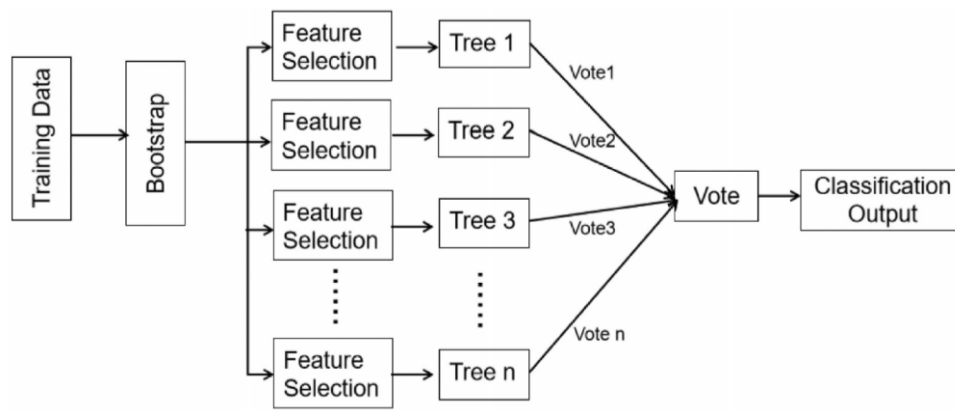


Figura 13: Estructura del clasificador *Random Forest*. *Random Forest* es un algoritmo de *machine learning* que combina varios árboles de decisión aleatorios y agrega sus predicciones promediando (Wei *et al.*, 2022)

En primer lugar, se va a hablar sobre la fase de *Bootstrap* que se lleva a cabo antes de cada árbol de decisión. *Bootstrapping* es un método estadístico que se basa en obtener varios conjuntos de entrenamiento al azar, no disjuntos con reemplazo de un único conjunto de datos de entrenamiento (*training data*). La probabilidad de escoger una muestra de una base de datos de tamaño N es $1/N$. Consecuentemente, después de N extracciones, cuando N es un valor muy alto, la probabilidad de que una muestra no sea seleccionada vendrá dada por la *ecuación 12*, cuyo comportamiento de dicha función se puede ver en la *Figura 13* en función del tamaño de N . Si esto ocurre, es decir, cuando hay datos que no se incluyen dentro del *Bootstrap*, estas muestras forman parte del grupo *out-of-bag-data* y son extremadamente útiles para la estimación de errores e importancia de características. Gracias a este método se consigue aumentar la estabilidad y precisión del clasificador así como disminuir la varianza (C. Zhang & Ma, 2012).

$$\left(1 - \frac{1}{N}\right)^N \approx e^{-1} \approx 0.368 \quad (12)$$

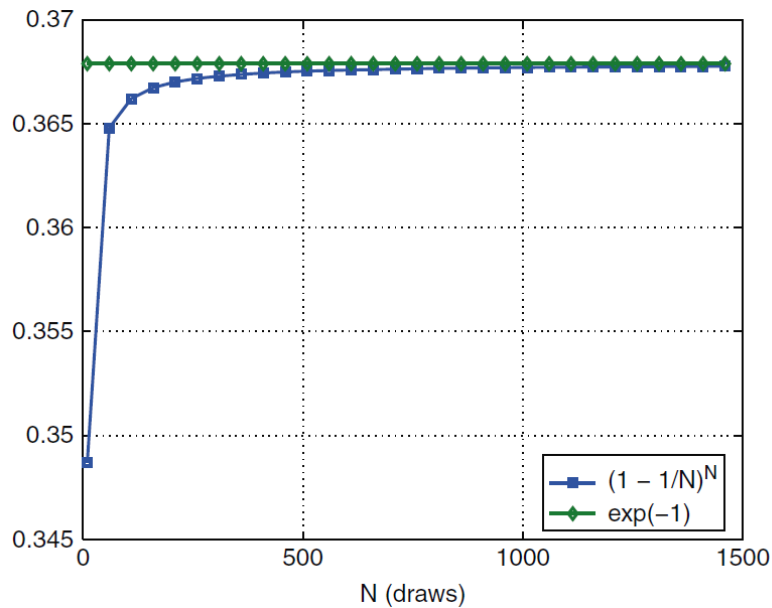


Figura 14: Procedimiento Bootstrap: probabilidad de no seleccionar una muestra del conjunto de entrenamiento tras N extracciones (C. Zhang & Ma, 2012)

En segundo lugar, se describe el algoritmo de *Random Forest*. Una de las principales ventajas computacionales que presenta este algoritmo frente al resto son:

- Rapidez en cuanto al tiempo invertido en el entrenamiento del sistema y predicción.
- Depende únicamente de tres parámetros.
- Puede ser implementado en paralelo.
- Permite implementar regresión y clasificación binaria o multiclase.
- Permite obtener medidas de importancia de características.

El objetivo final de este proceso consiste en dado un vector $X = (X_1, \dots, X_p)^T$, que en este caso correspondería a las características extraídas en las etapas anteriores, predecir su valor real Y asumiendo que tanto X como Y están relacionadas. Dicha predicción valdrá un 1 en caso de detectar un *spindle* y un 0 en el caso contrario. Para ello, el algoritmo trata de encontrar una función $f(X)$ que prediga Y . Se define, por tanto, la función de pérdidas $L(Y, f(X))$, la cual se trata de minimizar para hallar cuánto de cerca está $f(X)$ respecto a Y . Las típicas decisiones que se toman en base a esta función para un problema de clasificación binaria vienen dadas por la ecuación (13) (C. Zhang & Ma, 2012):

$$L(Y, f(X)) = \begin{cases} 0 & \text{si } Y = f(X) \\ 1 & \text{resto} \end{cases} \quad (13)$$

Los árboles que forman los bosques aleatorios están formados por nodos que presentan distintas funciones. En primer lugar, el nodo raíz es el origen del árbol y comprende todo el espacio del predictor. Cada nodo se va dividiendo en otros descendientes, hasta llegar a los nodos terminales u hoja, los cuales no se bifurcan en ninguno más. Normalmente, la

condición que se debe cumplir para elegir el nodo descendente al que debe continuar el dato viene determinada por un punto de separación o umbral, de tal manera que en función de si de cumple o no, se irá por uno de los lados. Para el caso en el que cada nodo se divide en 2 nodos sucesivos, si la característica es menor al umbral viajará por la izquierda y en caso contrario la decisión irá por la derecha (*Figura 14*). Una vez se ha producido la bifurcación, el proceso a seguir en los siguientes nodos será equivalente hasta llegar al nodo terminal.

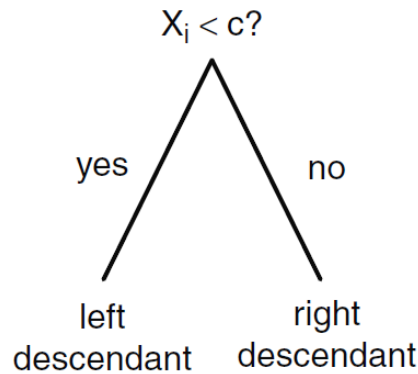


Figura 15: Esquema general de un nodo no terminal siendo la condición de separación "c" (C. Zhang & Ma, 2012)

Aunque el algoritmo de *Random Forest* tenga una buena reputación y funcione muy bien en prácticamente todos los casos, existen tres parámetros de configuración que pueden influir notablemente en la precisión de las predicciones (C. Zhang & Ma, 2012):

1. El número de árboles que conforman el algoritmo (*n-estimators*). Este valor se ha establecido en 100 de acuerdo con el artículo (Wei *et al.*, 2022)
2. El número mínimo de muestras para dividir un nodo interno (*min-samples-split*). Este valor se ha establecido en 120 por el mismo motivo al anterior.
3. El número mínimo de nodos terminales u hoja (*min-samples-leaf*). Este valor se ha establecido en 20 por el mismo motivo al anterior.

Todos estos pasos se han implementado en el proyecto gracias a la librería "sklearn.ensemble" de Python que contiene el método *RandomForestClassifier*. Gracias a este paquete se crea un nuevo modelo de clasificación *Random Forest*, posteriormente se entrena dicho modelo con el conjunto de entrenamiento (empleando el método *fit*) y, por último, se evalúa su capacidad predictiva empleando el conjunto de test (con el método *predict*) obteniendo así el valor de clasificación final (*classification output*). Resulta importante mencionar que todas las etapas que se pueden ver en el esquema de la *Figura 12* se implementan directamente con el uso de este clasificador.

```

from sklearn.ensemble import RandomForestClassifier

[...]

model=RandomForestClassifier(n_estimators=100,min_samples_split=120
,min_samples_leaf=20,bootstrap=True)

model.fit(X_train,y_train)

y_pred=model.predict(X=X_test)
  
```

MÉTRICAS DE EVALUACIÓN

En un problema de clasificación típico, existen métricas de evaluación que se emplean tanto en la etapa de entrenamiento como en la de test. El objetivo principal por el que resulta de gran utilidad dichas métricas se centra en conseguir optimizar el algoritmo de clasificación. Para este estudio, simplemente se han utilizado las herramientas para evaluar la efectividad del clasificador en la fase de testeo. En este estudio se distinguen principalmente tres grupos de métricas: discriminadores en función de un umbral, error cuadrático medio (MSE, *Mean Square Error*) y la importancia de características. A continuación, se lleva a cabo una breve explicación de cada una de ellas.

MÉTRICAS DE DISCRIMINACIÓN BASADAS EN UMBRAL

La evaluación de discriminación de la solución óptima dentro de un problema de clasificación binaria se basa en lo que se conoce como matriz de confusión. En esta tabla se representan en la columna la clase predicha por el clasificador y en la fila la clase real, de tal forma que se reflejan en ella los valores TP (*true positive*), TN (*true negative*), FP (*false positive*) y FN (*false negative*) (Tabla 3). Los valores TP y TN denotan el número de valores predichos positivos y negativos que concuerdan con los valores reales, mientras que FP y FN se emplean para identificar el número de falsos positivos y falsos negativos que se han dado tras la clasificación (Hossin & Sulaiman, 2015).

	Clase real positiva	Clase real negativa
Clase predictiva positiva	TP	FP
Clase predictiva negativa	FN	TN

Tabla 3: Matriz de confusión

En base a estos valores, se obtienen varias medidas de evaluación que permiten discriminar la solución más óptima. Entre todas ellas destacan (Hossin & Sulaiman, 2015):

- ❖ **Exactitud** (*accuracy*): Es la métrica de evaluación más utilizada en la práctica tanto para problemas de clasificación binaria como multiclase. Se basa en medir el porcentaje de predicciones correctas frente al total de las predicciones (*ecuación 14*). Una desventaja de este parámetro reside en que cuando hay desequilibrio entre los datos, suele ser favorable a la clase mayoritaria, proporcionando así un resultado poco convincente.

$$acc = \frac{TP + TN}{TP + FP + TN + FN} * 100 \quad (14)$$

- ❖ **Tasa de error** (*error rate*): Es la medida complementaria al *accuracy*. Mide, por tanto, el porcentaje de predicciones falsas frente al total de las predicciones (*ecuación 15*). Tanto la tasa de error como la exactitud constituyen unas de las métricas más comunes a la hora de encontrar una solución óptima al problema. Presentan la

ventaja de que son medidas computacionalmente fáciles y rápidas y por tanto muy útiles para problemas multiclase.

$$err = \frac{FP + FN}{TP + FP + TN + FN} * 100 \quad (15)$$

- ❖ **Sensibilidad** (*sensitivity*): Este parámetro permite conocer la proporción de casos positivos que fueron correctamente clasificados (*ecuación 16*). Para un modelo perfecto, este valor debería ser 1 o el 100%.

$$sn = \frac{TP}{TP + FN} * 100 \quad (16)$$

- ❖ **Especificidad** (*specificity*): Se emplea para medir la fracción de casos negativos correctamente clasificados (*ecuación 17*).

$$sp = \frac{TN}{TN + FP} * 100 \quad (17)$$

- ❖ **Precisión o valor predictivo positivo** (*precision*): La precisión se utiliza para evaluar el número correcto de casos positivos obtenidos frente a todos los casos positivos, reales o falsos que se han clasificado (*ecuación 18*).

$$p = \frac{TP}{TP + FP} * 100 \quad (18)$$

El cálculo de todos estos parámetros se ha llevado a cabo gracias a la herramienta Python tras la evaluación del algoritmo con el conjunto de test y validación. La librería “sklearn.metrics” proporciona un método *confusion_matrix* que calcula los valores TP, TN, FP y FN. A partir de ellos, aplicando las fórmulas anteriores se han podido obtener las distintas métricas de evaluación.

```
tn, fp, fn, tp=metrics.confusion_matrix(y_test, y_pred).ravel()
```

ERROR CUADRÁTICO MEDIO (MSE)

El error cuadrático medio se emplea en problemas de clasificación supervisados y mide la diferencia entre las soluciones predichas y las deseadas o reales (*ecuación 19*). Valores pequeños para el MSE son preferibles ya que inducen un comportamiento bueno para el clasificador. Una limitación que presenta el MSE, al igual que el *accuracy*, es que su valor se invalida para el caso de desequilibrio entre las clases de datos ya que se pondrá un mayor peso en la clase mayoritaria. La definición del error cuadrático medio se proporciona en la ecuación 18, donde P_i corresponde al valor predicho y A_i al valor real de la instancia i , donde N es el número total de instancias (Hossin & Sulaiman, 2015).

$$MSE = \frac{1}{N} \sum_{i=1}^N (P_i - A_i)^2 \quad (19)$$

En este caso se ha calculado la raíz del error cuadrático medio (RMSE) ya que éste se puede interpretar como la desviación estándar de la varianza y, además, tiene la ventaja de trabajar en las mismas unidades que la variable predicha. Cuando más pequeño es su valor, mejor ajuste tendrá el modelo implementado. La implementación de esta medida se ha realizado de forma similar a las anteriores, mediante la librería “sklearn.metrics” gracias al método *mean_squared_error*.

```
rmse=metrics.mean_squared_error(y_true=y_test,y_pred=y_pred,squared=False)
```

IMPORTANCIA DE CARACTERÍSTICAS

Medidas respecto a la importancia de características del predictor son importantes a la hora de realizar una selección de características e interpretar correctamente el algoritmo entrenado. Es posible que muchas de las variables que se están utilizando para la clasificación no aporten información relevante para el problema de predicción. En ese caso resulta de gran utilidad obtener la importancia de características directamente del algoritmo implementado (C. Zhang & Ma, 2012).

Los algoritmos *Random Forest* utilizan un procedimiento específico, aunque bastante intuitivo a la hora de obtener un parámetro de importancia de características. Supóngase que se tiene una variable “k”. El valor de esta variable se va permutando aleatoriamente, manteniendo fijos el resto de características. La importancia, por lo tanto, se estima cuando el MSE o error cuadrático medio varía al introducir cambios en esa variable dentro del grupo de test. Teniendo esto en cuenta, la variable “k” se eliminará del modelo siempre que se obtengan resultados similares y, por lo tanto, carezca de importancia (C. Zhang & Ma, 2012).

Para obtener esta nueva medida, se ha utilizado la librería “sklearn.ensemble” de Python, la cual presenta un método de cálculo de importancia de estas variables basándose en la permutación tras emplear un algoritmo de *Random Forest*.

```
sorted_idx=model.feature_importances_.argsort()
plt.barh(df['Name'][sorted_idx],model.feature_importances_[sorted_idx])
```

CAPÍTULO 4: RESULTADOS

Este capítulo está orientado a exponer los resultados obtenidos tras la implementación del método desarrollado en el capítulo anterior. En concreto, se presentan los resultados obtenidos tras la evaluación de tres modelos distintos en función del número de características con las que se ha entrenado y la manera en la que se ha trabajado con la señal.

En primer lugar, se expone el análisis de un espectrograma de la señal original con la que se ha trabajado. Esto ha permitido identificar de manera cualitativa los *spindles* con las marcas reales puestas por expertos y así verificar que tanto las señales como las marcas se correspondían para todos los sujetos de la base de datos.

Una vez analizado el espectrograma, se diferencian dos procedimientos diferentes. El primero de ellos trata de elaborar un modelo y de evaluarlo de manera similar al artículo duplicado con la idea de realizar una comparación en el capítulo de discusión acerca de las diferencias obtenidas. El otro procedimiento se basa en recortar la señal original y quedarnos únicamente con las fases N2 y N3 del sueño para mejorar los resultados, ya que son las que contienen *spindles* del sueño. Para este último método se han evaluado dos modelos distintos. Uno entrenado con las 14 características obtenidas tras su selección en el artículo (Wei et al., 2022) y otro entrenado con todas las características estimadas.

Además, por cada procedimiento se han representado *boxplots* de las características que se han calculado, lo que permite identificar su distribución y estudiar su comportamiento en función de los valores que toma si se trata o no de un *spindle*.

Por último, se elabora una gráfica en la que se representa la importancia de las características obtenida tras el entrenamiento del modelo en las fases N2-N3 con todas las variables extraídas de la señal y un ejemplo de predicción realizado para este último modelo.

ANÁLISIS DEL ESPECTROGRAMA

En primer lugar, para verificar que tanto las señales como las marcas etiquetadas por expertos se correspondían a *spindles* del sueño, se han realizado varios espectrogramas de 30 segundos de duración sobre cada uno de los sujetos para estudiar su contenido frecuencial y su distribución de la energía en función de tiempo. Como ya se ha comentado en anteriores capítulos, los *spindles* corresponden a oscilaciones que se dan

mayoritariamente en la fase N2 del sueño, aunque también en la fase N3, a frecuencias entre los 8-15 Hz en niños. En la *Figura 15* se presenta un ejemplo en que se visualiza el espectrograma del sujeto COG001 en un segmento de 30 segundos. Además, se grafican las marcas de *spindles* que se corresponden en ese tiempo, así como la señal correspondiente en el dominio temporal.

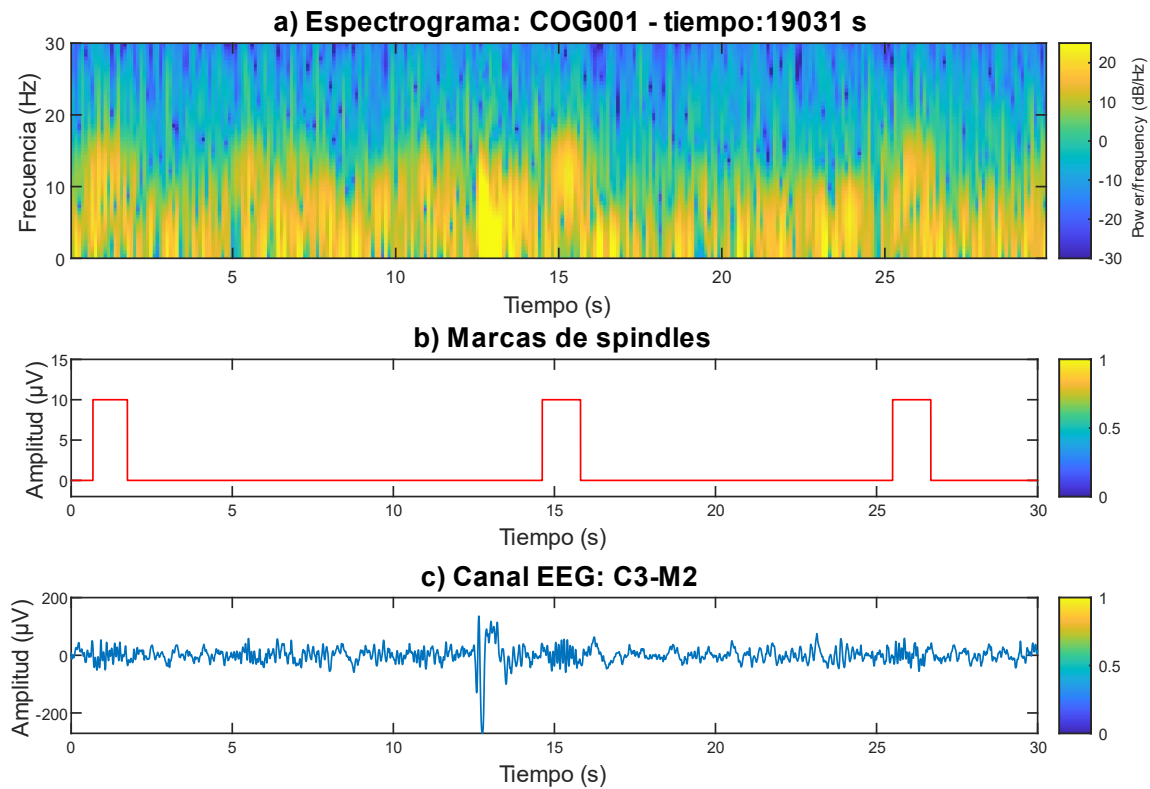


Figura 16: Representación de 30 segundos del sujeto COG001 a partir del segundo 19031 de grabación: a) representación del espectrograma, b) representación de las marcas de *spindles* etiquetadas por un experto y c) canal C3-M2 del EEG.

Analizando varias imágenes similares a esta probando con distintos sujetos, se puede observar que efectivamente las marcas de *spindles* se corresponden en el espectrograma a frecuencias comprendidas entre 8-15 Hz, además de que la señal EEG durante esas marcas parece pertenecer a un *spindle*. Se puede inferir, por tanto, que esta imagen se trata de un buen ejemplo para poder comprobar de forma visual que las señales y las marcas adquiridas con correctas y que se corresponden con *spindles* del sueño.

RESULTADOS CON EL MODELO ORIGINAL

El primer experimento que se realizó fue obtener los resultados de forma similar al modelo original propuesto por el artículo (Wei *et al.*, 2022). Para ello, se ha utilizado el grupo de entrenamiento, formado por los primeros cinco sujetos de la base de datos, con el objetivo de entrenar al sistema. Las características estimadas en este caso han sido las 14 que

previamente fueron seleccionadas por el autor cuyo procedimiento ha sido explicado en el capítulo anterior.

A lo largo de este apartado se muestran *boxplots* de dichas características para estudiar su comportamiento en presencia o no de *spindles* del sueño. A continuación, se recogen los resultados obtenidos al evaluar el algoritmo con el conjunto de test.

BOXPLOTS DE CARACTERÍSTICAS

Tras definir en el capítulo anterior lo que consiste un *boxplot*, se ha representado por cada característica normalizada del conjunto de entrenamiento, una gráfica de este tipo para poder visualizar cómo se distribuyen los valores en función de si ese segmento pertenece a un *spindle* o no. Para ello se han separado las representaciones en dos distintas, diferenciando por un lado las características extraídas en el tiempo y por otro las extraídas en frecuencia (*Figuras 17 y 18*).

CARACTERÍSTICAS EN EL TIEMPO

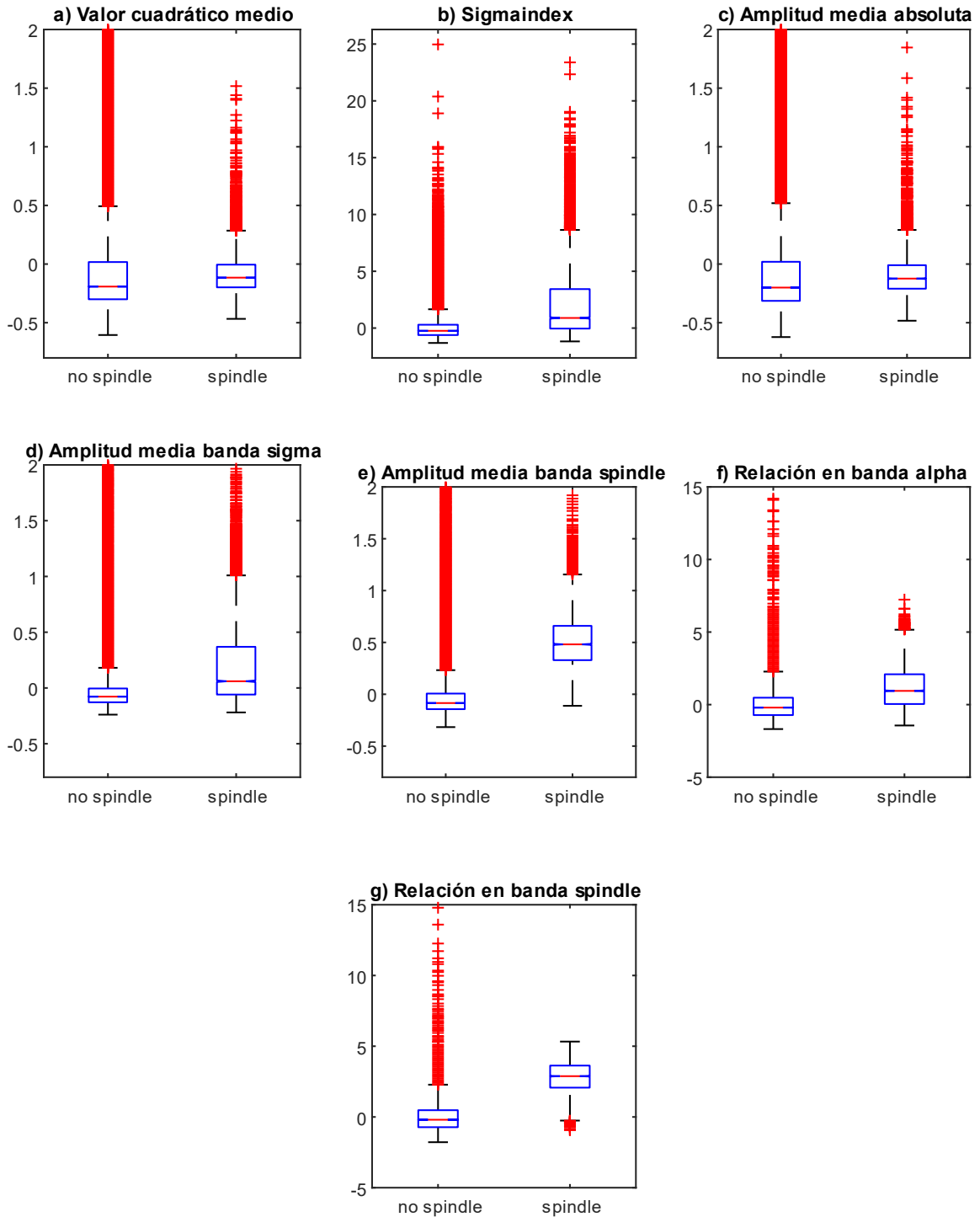


Figura 17: *Boxplots* de características normalizadas en el tiempo pertenecientes al conjunto de entrenamiento para el modelo original: a) valor cuadrático medio, b) sigmaindex, c) amplitud media absoluta, d) amplitud media banda sigma, e) amplitud media banda spindle, f) relación en banda alpha y g) relación en banda spindle.

CARACTERÍSTICAS EN FRECUENCIA

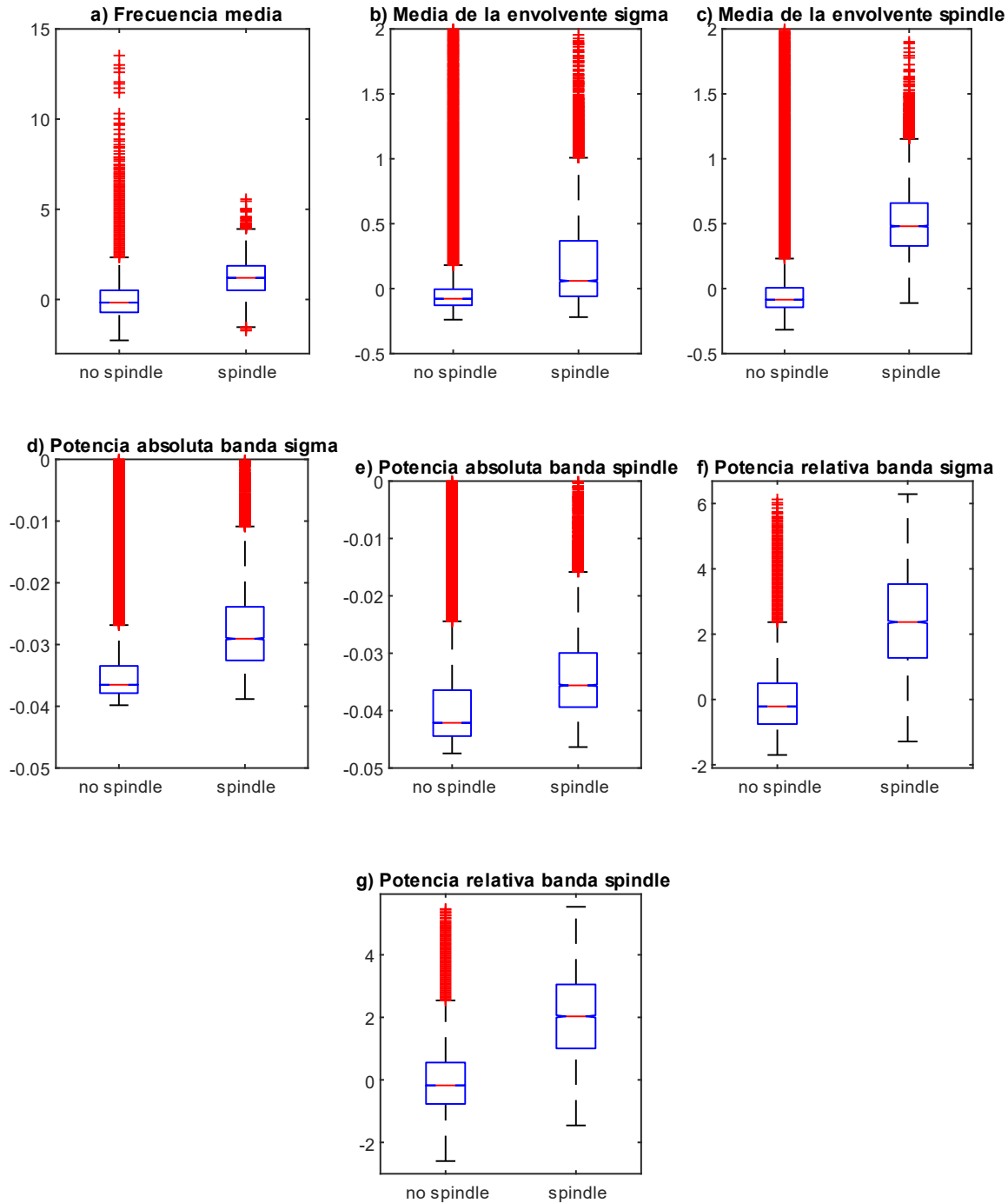


Figura 18: *Boxplots* de las características normalizadas en frecuencia pertenecientes al conjunto de entrenamiento para el modelo original: a) frecuencia media, b) valor medio de la envolvente en la banda sigma, c) valor medio de la envolvente en la banda spindle, d) potencia absoluta en la banda sigma, e) potencia absoluta en la banda spindle, f) potencia relativa en la banda sigma y g) potencia relativa en la banda spindle.

RESULTADOS CON EL CONJUNTO DE TEST

Tras finalizar la fase de entrenamiento, se pasa a evaluar el modelo con el conjunto de test, formado por los sujetos 6, 7, 8 y 9 de la base de datos. Se ha obtenido, por ello la matriz de confusión representada en la *Tabla 4*. En base a estos valores se ha podido estimar los cálculos de exactitud, tasa de error, sensibilidad, especificidad y precisión recogidos en la *Tabla 5* y representados sus valores en porcentaje.

	Spindles reales	No spindles reales
Predicción de spindles	2458	21981
Predicción de no spindles	99	609058

Tabla 4: Matriz de confusión obtenida tras evaluar el modelo original con el conjunto de test.

Métricas de discriminación basadas en umbral	Valor en %
Exactitud (<i>accuracy</i>)	96.51
Tasa de error (<i>error rate</i>)	3.48
Sensibilidad (<i>sensitivity</i>)	96.13
Especificidad (<i>specificity</i>)	96.52
Precisión o valor predictivo positivo (<i>precision</i>)	10.10

Tabla 5: Métricas de discriminación basadas en umbral obtenidas a partir de la matriz de confusión del conjunto de test.

Asimismo, se ha calculado la raíz del error cuadrático medio (RMSE) y el coeficiente de correlación de Matthew (MCC) para medir la calidad del clasificador. Cuanto más pequeño sea el valor del RMSE, mayor ajuste presentará el modelo. El valor del MCC, por su parte, debe estar comprendido entre -1 y +1, siendo +1 el valor que representa una predicción perfecta, el 0 una predicción aleatoria y el -1 una predicción inversa. Los resultados obtenidos se presentan en la *Tabla 6*.

Métricas auxiliares	Valor
Error cuadrático medio (RMSE)	0.18
Coficiente de correlación de Matthew (MCC)	0.31

Tabla 6: Error cuadrático medio y coeficiente de correlación de Matthew para el modelo original evaluado con el conjunto de test.

RESULTADOS CON FASES N2 Y N3

El siguiente paso que se ha dado en este TFG y con el objetivo de implementar mejoras en los resultados obtenidos, ha sido aplicar un algoritmo que separe de manera automática la señal inicial C3-M2 del EEG en sus diferentes fases del sueño de tal manera que dicha señal quede reducida a las fases N2 y N3 eliminando el resto. Así se consigue disminuir el número de etapas en las que se divide la señal y centrar la clasificación únicamente en aquellas fases en las que los *spindles* del sueño toman protagonismo. El objetivo principal de ello fue evitar falsos positivos y conseguir una mayor precisión en la clasificación.

El algoritmo utilizado está basado en técnicas de *deep learning* y ha sido desarrollado por otro TFG anterior (Calvo Merino, 2020). En este trabajo se obtuvieron precisiones del 87%

para el modelo original en el cual se utilizó una base de datos con registros de EEG de adultos. Estos resultados son muy buenos y por ello se puede considerar un método fiable para localizar las fases 2 y 3 del sueño para los EEG de este Trabajo Fin de Grado.

Para esta nueva metodología se han elaborado dos modelos diferentes. El primero de ellos se ha entrenado con un *training set* formado únicamente por las 14 características seleccionadas por el autor del artículo (Wei *et al.*, 2022), mientras que el segundo modelo se ha entrenado con las 29 características en total extraídas de cada época con el objetivo de obtener un gráfico que indique la importancia de características y ver las diferencias en cuanto a la clasificación.

A lo largo de este apartado se recogen los *boxplots* de todas las características en total estimadas para la nueva señal y se evalúa cada modelo tanto con el grupo de test formado por 4 sujeto obteniendo así las métricas de evaluación gracias a las cuales se podrá llevar a cabo una comparación entre los distintos clasificadores y con ello poder establecer ciertas conclusiones.

BOXPLOTS DE CARACTERÍSTICAS

En primer lugar y de mismo modo que se hizo para el modelo original, se han representado en varios diagramas de cajas y bigotes las distintas características normalizadas del conjunto de entrenamiento para poder estudiar su comportamiento dependiendo de si se trata o no de un spindle. Estos *boxplots* se han separado en varias categorías. Por un lado, se representan los *boxplots* pertenecientes a características extraídas en el tiempo (*Figuras 19 y 20*), en las que se separan por un lado las 7 características pertenecientes a la selección (*Figura 19*) y por otro, el resto de ellas (*Figura 20*). Lo mismo ocurre con las características frecuenciales. Por un lado, se representan las 7 características correspondientes a la selección (*Figura 21*) y, por otro lado, el resto de ellas (*Figura 22*). Esta clasificación se ha realizado para poder comparar mejor dichas gráficas con el modelo original para el caso de selección de 14 características.

CARACTERÍSTICAS EN EL TIEMPO

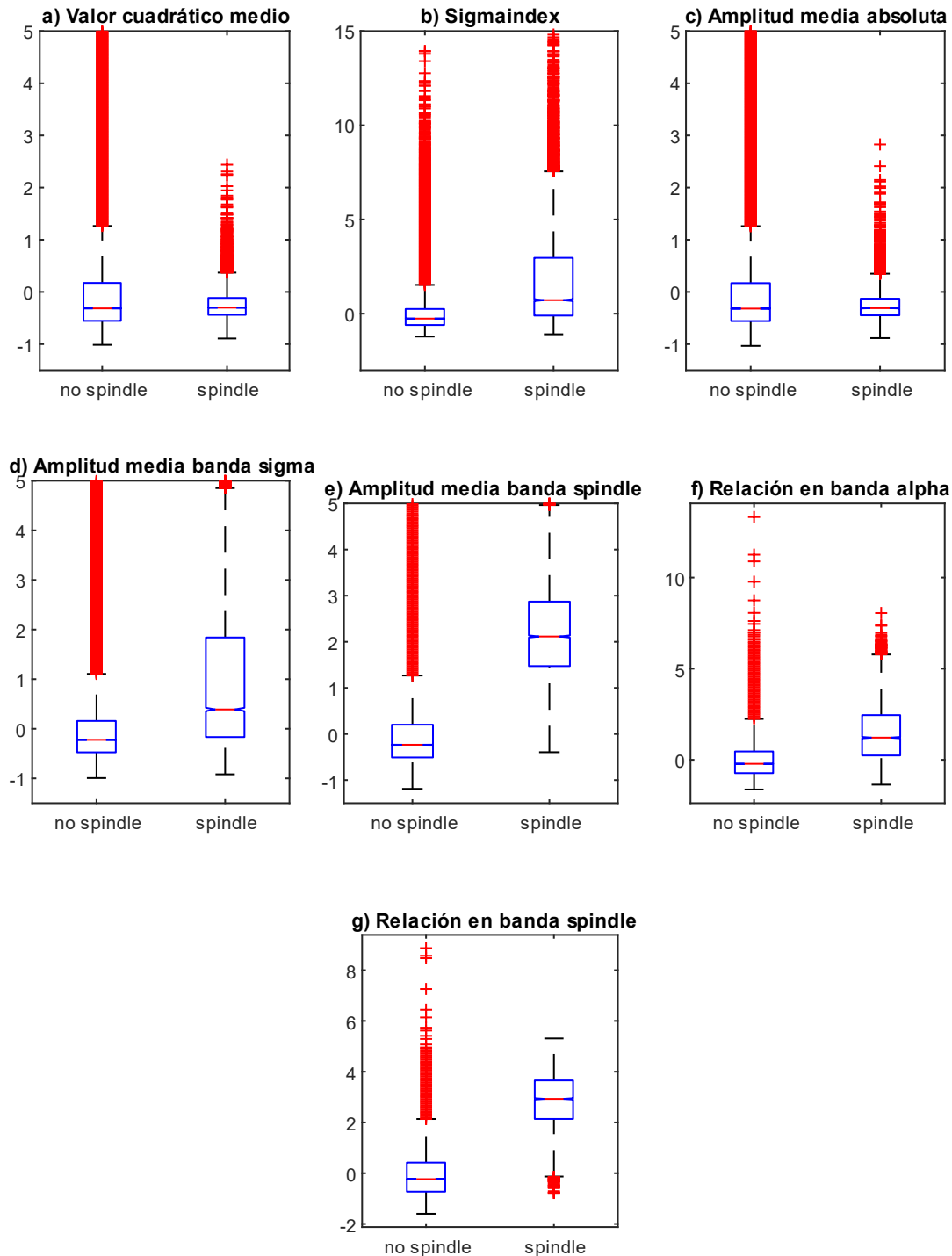


Figura 19: Boxplots de características normalizadas en el tiempo pertenecientes al conjunto de entrenamiento tras la selección para el modelo N2-N3: a) valor cuadrático medio, b) sigmaindex, c) amplitud media absoluta, d) amplitud media absoluta en banda sigma, e) amplitud media absoluta en banda spindle, f) relación en banda alpha y g) relación en banda spindle.

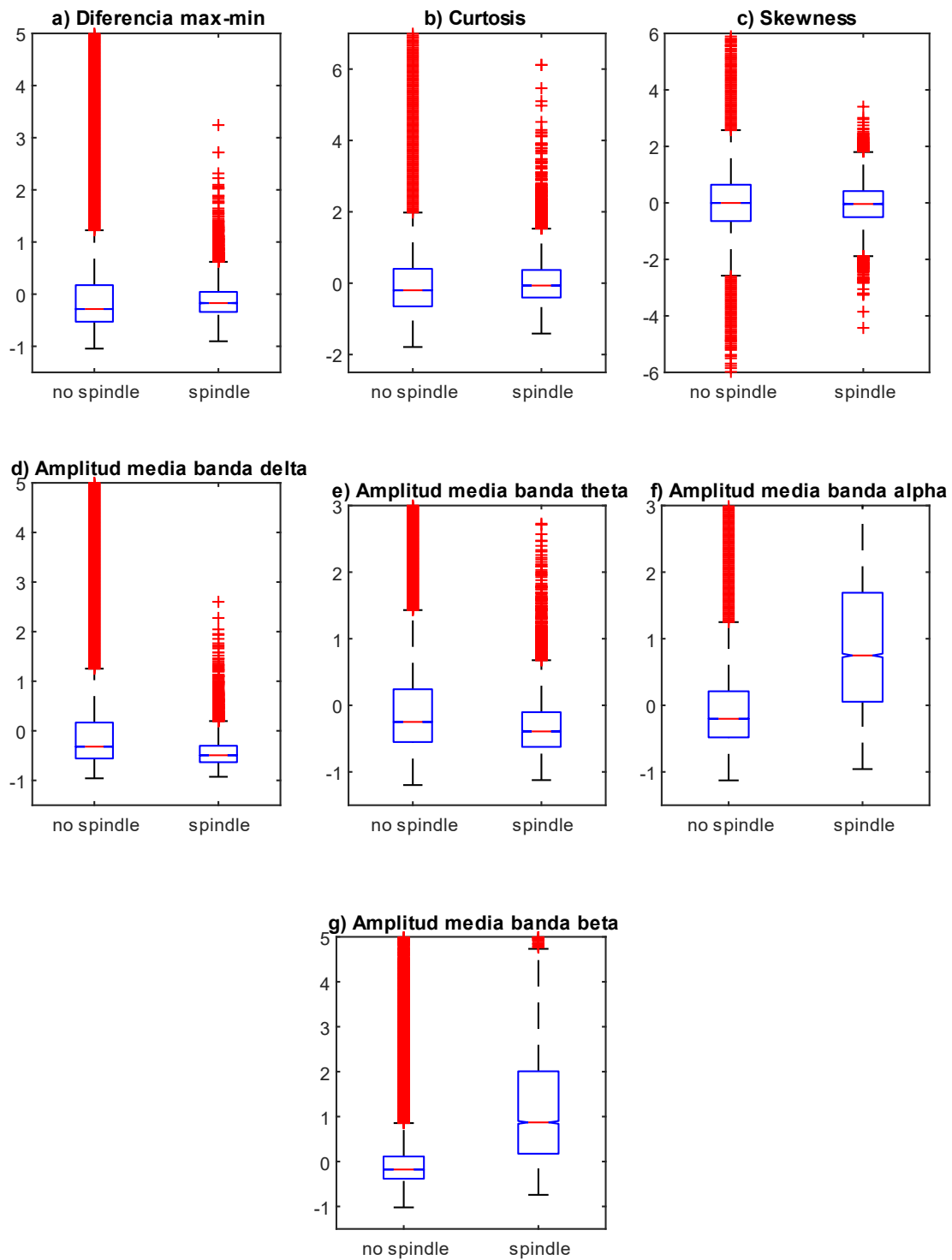


Figura 20: *Boxplots* de características adicionales normalizadas en el tiempo pertenientes al conjunto de entrenamiento para el modelo N2-N3: a) diferencia entre amplitud máxima y mínima, b) curtosis, c) skewness d) amplitud media absoluta en la banda delta, e) amplitud media absoluta en la banda theta, f) amplitud media absoluta en la banda alpha y g) amplitud media absoluta en la banda beta.

CARACTERÍSTICAS EN FRECUENCIA

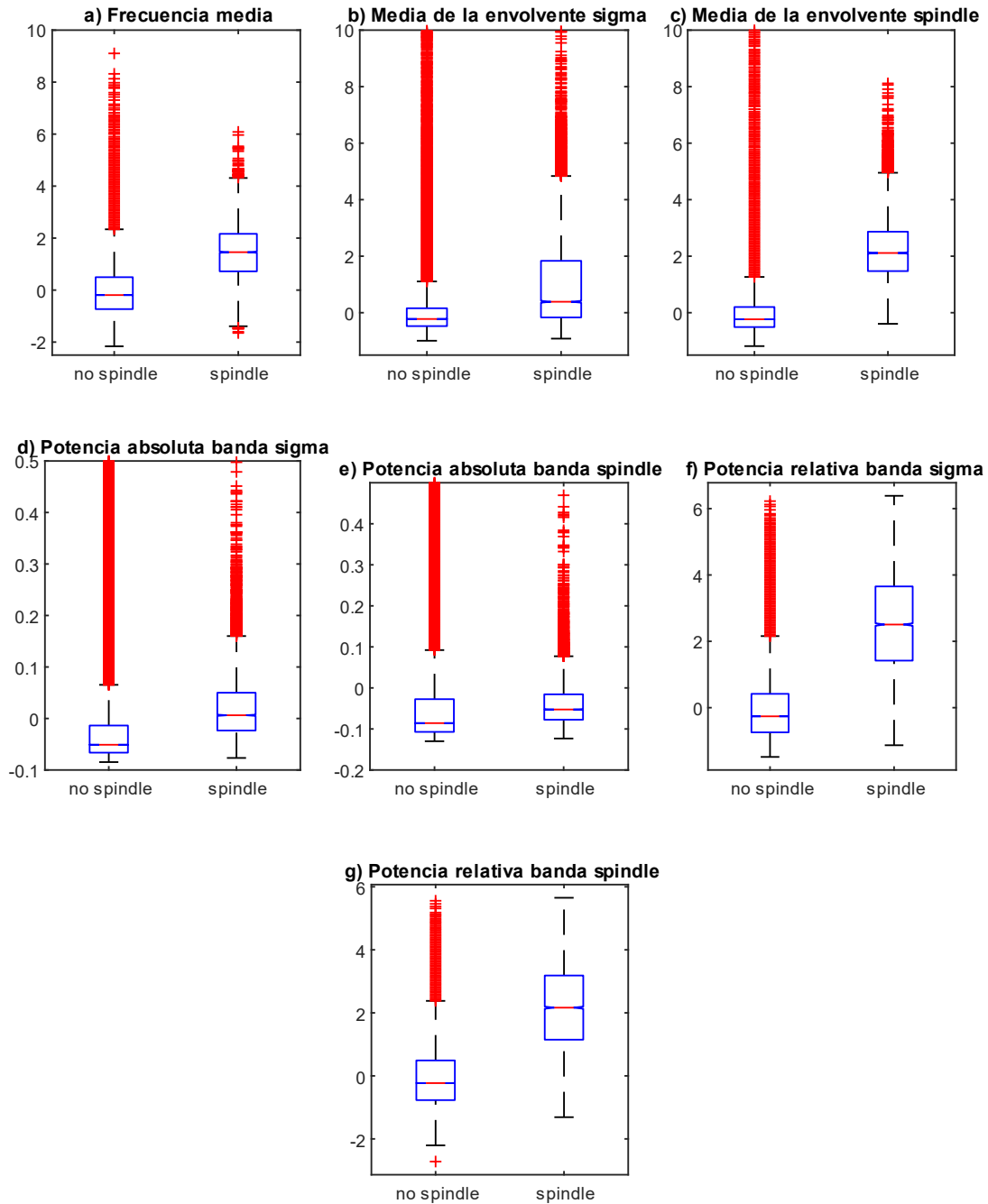


Figura 21: Boxplots de las características normalizadas en frecuencia del conjunto de entrenamiento tras la selección para el modelo N2-N3: a) frecuencia media, b) valor medio de la envolvente en la banda sigma, c) valor medio de la envolvente en la banda spindle, d) potencia absoluta en la banda sigma, e) potencia absoluta en la banda spindle, f) potencia relativa en la banda sigma y g) potencia relativa en la banda spindle.

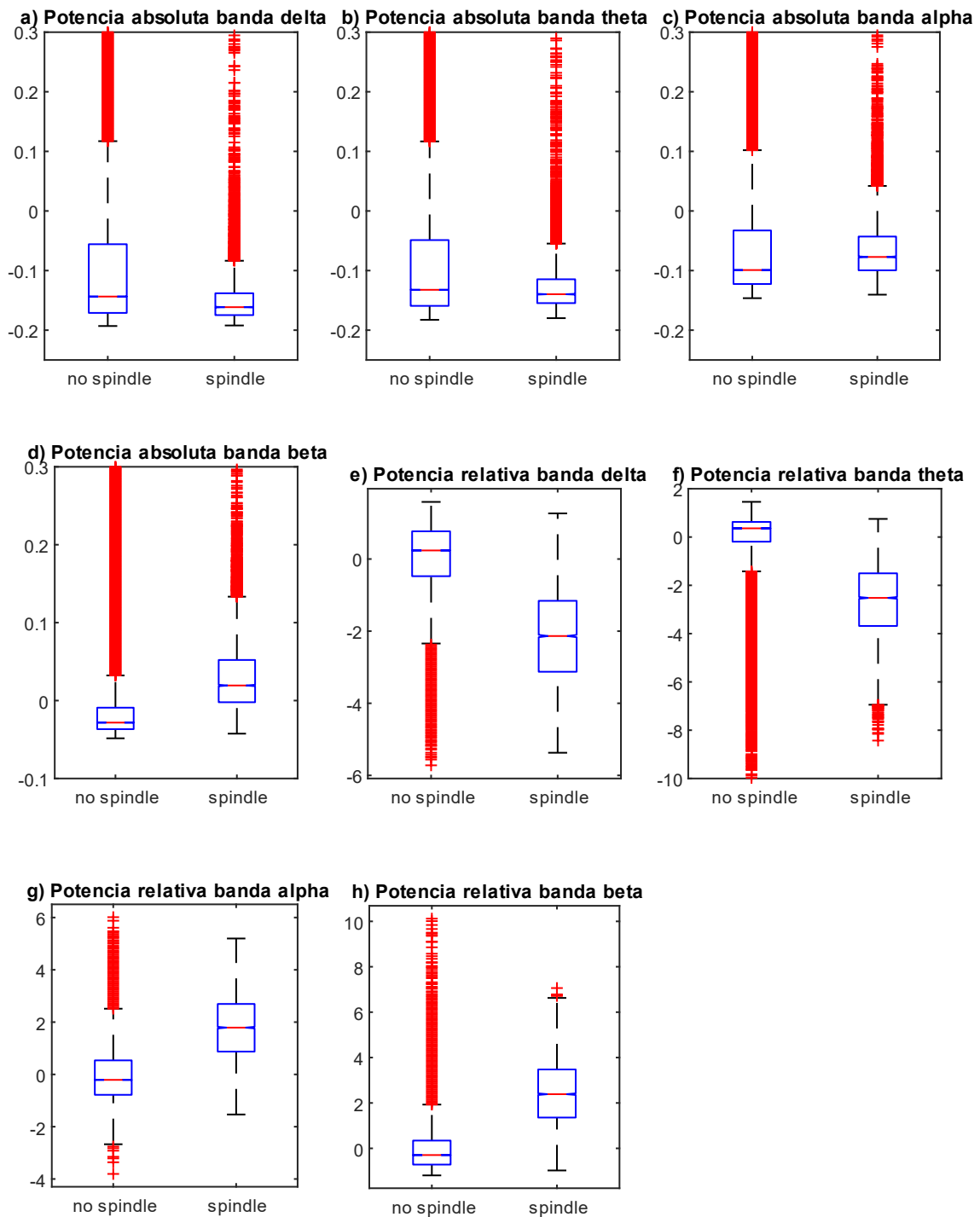


Figura 22: *Boxplots* de características adicionales normalizadas en frecuencia pertenecientes al conjunto de entrenamiento para el modelo N2-N3: a) potencia absoluta en la banda delta, b) potencia absoluta en la banda theta, c) potencia absoluta en la banda alpha, d) potencia absoluta en la banda beta, e) potencia relativa en la banda delta, f) potencia relativa en la banda theta, g) potencia relativa en la banda alpha y h) potencia relativa en la banda beta.

RESULTADOS CON EL CONJUNTO DE TEST

A continuación, se recogen los resultados tras evaluar los modelos con el conjunto de test. Teniendo en cuenta que se han entrenado dos clasificadores diferentes para la nueva señal C3 en la que únicamente está formada por las fases N2 y N3, se han representado en tablas diferentes por un lado la matriz de confusión perteneciente al modelo entrenado únicamente con las 14 características seleccionadas (*Tabla 7*) y, por otro lado, la matriz de confusión del modelo entrenado con todas las características en total extraídas de la señal (*Tabla 8*). Finalmente, las métricas de evaluación asociadas a dichos parámetros se reflejan en una tabla única en la que se pueden visualizar mejor los resultados (*Tabla 9*).

	Spindles reales	No spindles reales
Predicción de spindles	2376	18671
Predicción de no spindles	103	393156

Tabla 7: Matriz de confusión obtenida tras evaluar el modelo N2-N3 con el conjunto de test para las 14 características seleccionadas

	Spindles reales	No spindles reales
Predicción de spindles	2374	16529
Predicción de no spindles	105	395298

Tabla 8: Matriz de confusión obtenida tras evaluar el modelo N2-N3 con el conjunto de test para todas las características extraídas (sin selección de características).

Métricas de discriminación basadas en umbral	Valor en % con selección de características	Valor en % sin selección de características
Exactitud (<i>accuracy</i>)	95.47	95.99
Tasa de error (<i>error rate</i>)	4.53	4.01
Sensibilidad (<i>sensitivity</i>)	95.85	95.76
Especificidad (<i>specificity</i>)	95.47	95.99
Precisión o valor predictivo positivo (<i>precision</i>)	11.29	12.56

Tabla 9: Métricas de discriminación basadas en umbral obtenidas a partir de las matrices de confusión de la evaluación de los modelos N2-N3 con los distintos conjuntos de test.

Por último, se muestran los valores obtenidos en cuanto a la raíz del error cuadrático medio (RMSE) y al coeficiente de correlación de Matthew en la *Tabla 10*.

Métricas auxiliares	Valor para el modelo con selección de características	Valor para el modelo sin selección de características
Error cuadrático medio (RMSE)	0.18	0.17
Coeficiente de correlación de Matthew (MCC)	0.32	0.34

Tabla 10: Error cuadrático medio y coeficiente de correlación de Matthew para ambos modelos N2-N3 evaluados con el conjunto de test.

IMPORTANCIA DE CARACTERÍSTICAS

Para el modelo entrenado con todas las características extraídas, se han calculado sus importancias con la finalidad de poder comparar los modelos y sacar conclusiones sobre posibles mejoras y motivos por los que se obtienen resultados diferentes.

La gráfica proporcionada en la *Figura 23*, representa la importancia de características ordenadas de mayor a menor, de tal forma que se puede ver cómo la media de la envolvente en la banda spindle corresponde a la más significativa, mientras que el *skewness* a la característica menos importante o que menos información aporta a la clasificación.

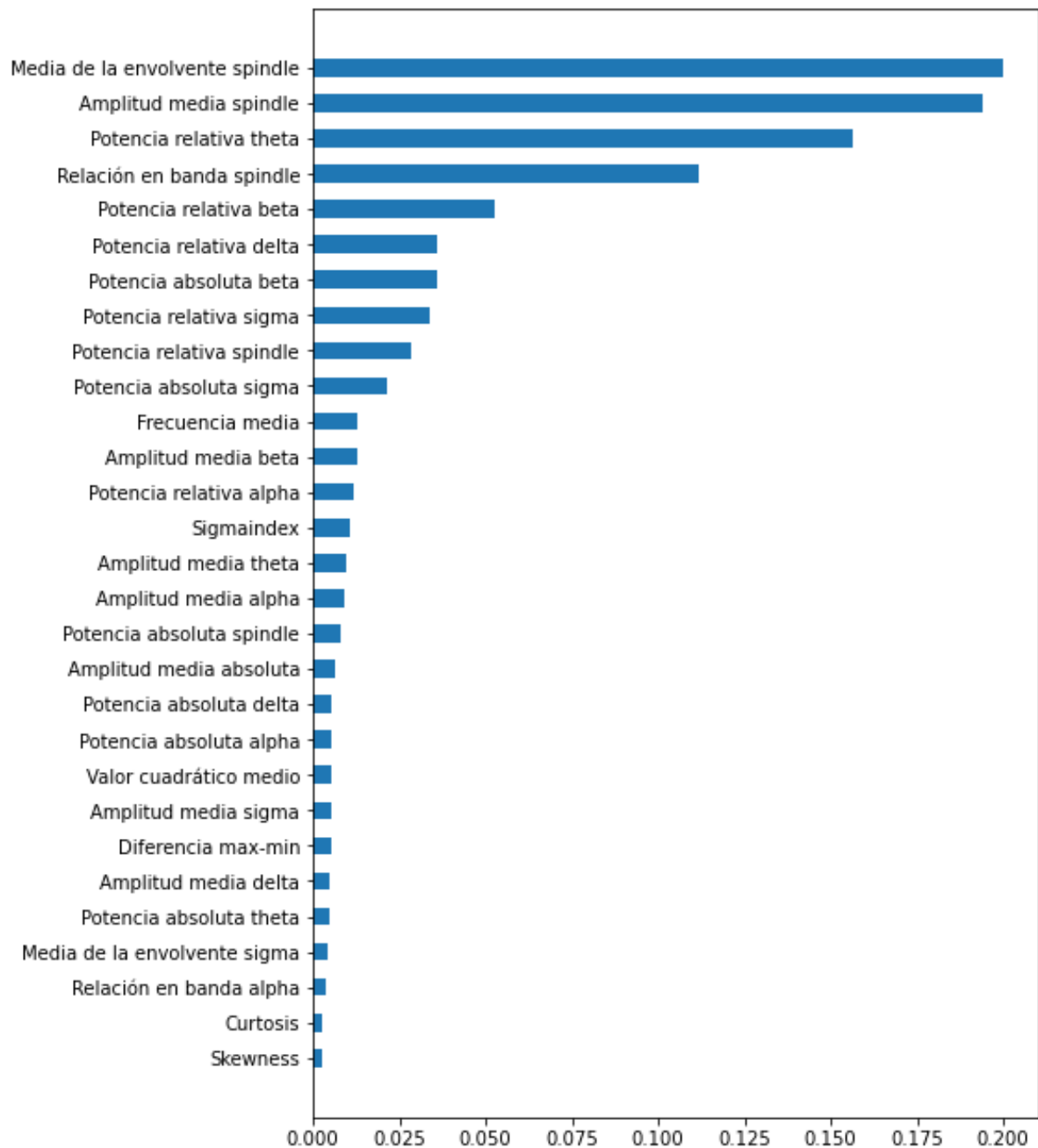


Figura 23: Importancia de características obtenida a partir del modelo N2-N3 entrenado con el total de características estimadas.

Por otro lado, en la *Figura 24* se puede ver la importancia de características seleccionadas obtenida en el artículo original. Los números que aparecen en el eje de ordenadas se corresponden a las siguientes características:

1. Amplitud media absoluta.
2. Valor cuadrático medio.
3. Frecuencia media.
4. Sigma index.
5. Relación en la banda alpha.
6. Relación en la banda spindle.
7. Amplitud media absoluta en la banda sigma.
8. Valor medio de la envolvente en la banda sigma.
9. Potencia relativa en la banda sigma.
10. Potencia absoluta en la banda sigma.
11. Amplitud media absoluta en la banda spindle
12. Valor medio de la envolvente en la banda spindle.
13. Potencia relativa en la banda spindle.
14. Potencia absoluta en la banda spindle.

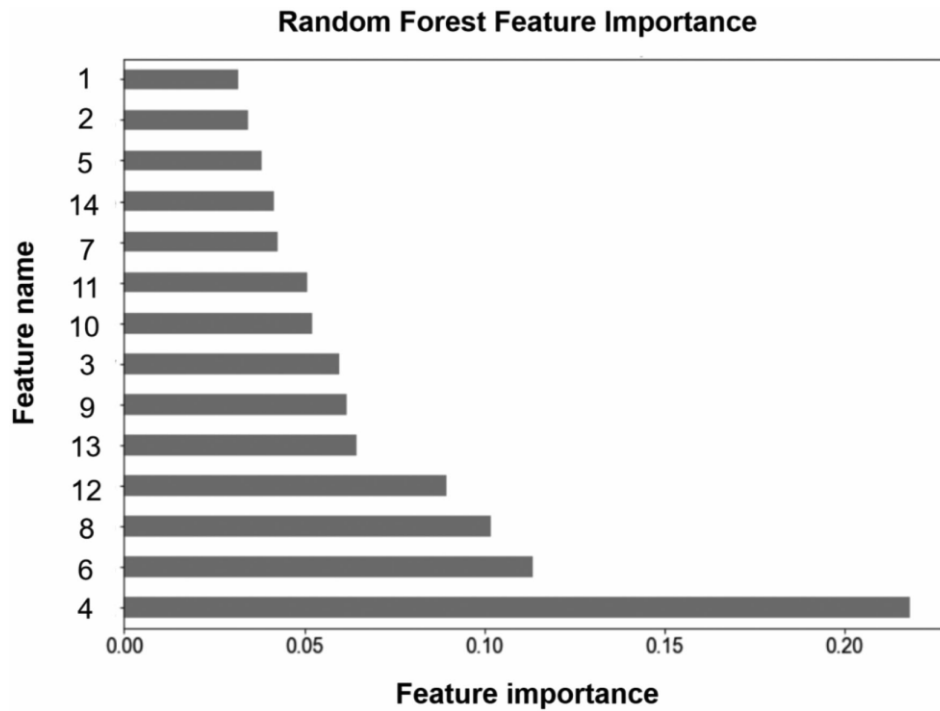


Figura 24: Importancia de las características seleccionadas obtenidas en el artículo original (Wei et al., 2022).

Se pueden ver varias diferencias entre las Figuras 23 y 24 entre las que destacan la característica más y menos importante. Para el artículo original estas corresponden al *sigma index* y amplitud media absoluta, respectivamente, mientras que para este TFG las características son el valor medio de la envolvente en la banda spindle y el *skewness*. Estas diferencias serán comentadas con mayor profundidad en el siguiente capítulo de discusión.

EJEMPLO DE PREDICCIÓN

Por último, se pretende reflejar un ejemplo sobre las marcas predichas y las reales, para ver que el modelo funciona correctamente detectando todos los *spindles* excepto por algún falso positivo. En esta representación se ha modificado la altura de ambas marcas para poder visualizarlo mejor (*Figura 25*).

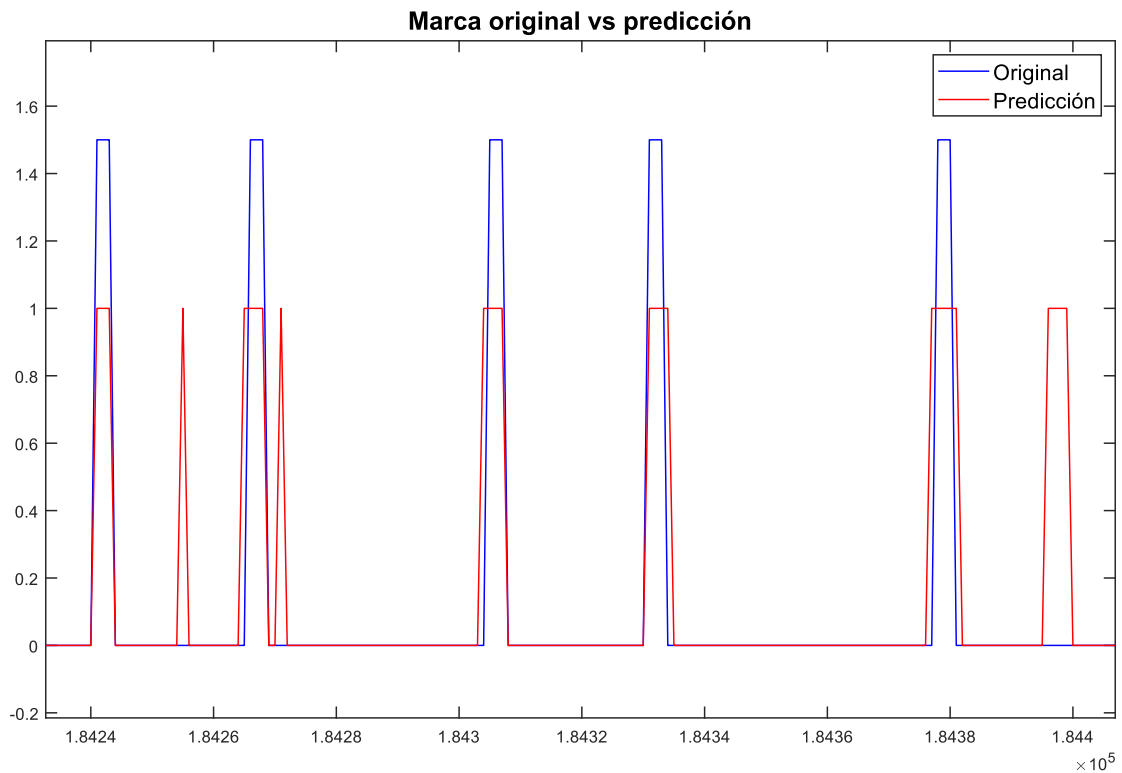


Figura 25: Gráfica en la que se representan las marcas originales en azul y en rojo las predicciones. Esta predicción es la correspondiente a la evaluación del modelo N2-N3 para todas las características con el conjunto de validación.

CAPÍTULO 5: DISCUSIÓN

En este estudio se presentan varios métodos de detección de *spindles* basados en una arquitectura de *Random Forest* empleando para ello varios registros EEG de niños con patologías de apnea del sueño, demostrando que es posible conseguir un clasificador que presente buenas prestaciones al respecto. La forma de proceder en todo momento ha estado guiada por la publicación de Wei et al. (2022), gracias a la cual ha servido de motivación a la hora de replicar el estudio.

El objetivo principal de toda esta detección se centra en ayudar a expertos a la hora de realizar el marcaje de *spindles*, ya que es un trabajo tedioso que implica una gran inversión de tiempo. Gracias a ello y conociendo la densidad de *spindles* que presenta una persona, se pueden diagnosticar o anticipar enfermedades relacionadas con el proceso cognitivo para poder tratarlo.

En el capítulo anterior se trató de exponer los resultados obtenidos tras la evaluación de varios modelos, sin embargo, es en este capítulo donde se pretenden analizar e interpretar además de comparar unos modelos con otros. Asimismo, se exponen los resultados del artículo a replicar y una comparación donde se exponen las diferencias que hay con este modelo. Para finalizar, se exponen algunas limitaciones y problemas con sus soluciones que se han dado para el desarrollo de este trabajo.

DISCUSIÓN DEL MODELO ORIGINAL

En esta discusión se pretende comentar los resultados obtenidos con el modelo original. Se comenzará hablando a cerca de las características extraídas y a continuación de los resultados obtenidos tras la evaluación del modelo.

En primer lugar, visualizando el comportamiento de dichas características tanto en el tiempo y en frecuencia mediante las representaciones de *boxplots* (Figuras 17 y 18), se pueden apreciar diferencias entre los valores obtenidos en presencia o no de *spindles*. Llama la atención que la mayoría de *outliers* contenidos en dichas representaciones pertenecen a la clase de no *spindle*, mientras que la clase *spindle* parece presentar unos diagramas de cajas más estables en torno a un valor. Además, se puede ver que hay características que parecen presentar un mismo comportamiento y, por lo tanto, nos hace pensar que no aportarán mucha información relevante a la hora de realizar la clasificación. Grupos de características que parecen presentar la misma distribución podrían ser:

- Amplitud media absoluta, amplitud media banda sigma y amplitud media banda spindle.
- Relación en banda alpha y en banda spindle.
- Valor medio de la envolvente en la banda sigma y spindle.
- Potencias absolutas en las bandas sigma y spindle
- Potencias relativas en las bandas sigma y spindle.

Los parecidos entre las características extraídas en las bandas spindle y sigma pueden tener sentido puesto que la banda sigma comprende las frecuencias entre 12.5-15 Hz, pertenecientes a la banda spindle que se ha utilizado en este estudio para niños entre 8-15 Hz. Lo mismo ocurre con las similitudes encontradas para las variables calculadas en las bandas alpha y spindle, puesto que la banda alpha contiene frecuencias entre 8-12 Hz presentes también en la banda spindle. A pesar de haber obtenido estos resultados con las representaciones, se ha decidido no eliminar aquellas características que presentan un mismo comportamiento debido al objetivo de imitar el artículo que se ha ido siguiendo paso a paso.

En cuanto al conjunto de test, los resultados obtenidos respecto a la matriz de confusión se pueden considerar muy elevados con una exactitud del 96.51%, una sensibilidad de 96.13% y una especificidad el 96.52%. Esto nos quiere decir que, en términos de clasificación, prácticamente se detectan todos los *spindles* que se contiene la señal. El total de estas oscilaciones marcadas para en el conjunto de test son 2557, de los cuales 2458 se detectan correctamente. El principal problema que presenta este modelo es que, a pesar de una especificidad muy elevada, el desbalanceo entre clases presenta un número de falsos positivos muy elevado en relación con el número de verdaderos positivos ($FP = 21981$), lo cual hace que la precisión o valor predictivo positivo obtenido sea bajo, del 10.10%. El motivo principal de ello reside en que, al igual que en los propios registros de sueño, hay pocos ejemplos de *spindles* en el conjunto de test en comparación con los no *spindles*. Para paliar este desajuste en los datos se utilizó la técnica de SMOTE, sin embargo, con esta técnica, se utilizó un $k = 5$, como en el artículo original, siendo insuficiente para compensar el desbalanceo en nuestra base de datos. Por ello, sería deseable en el futuro probar con diferentes valores de este parámetro para intentar optimizar los resultados.

DISCUSIÓN DE LOS MODELOS ESPECÍFICOS

Como se ha comentado anteriormente, se elabora un nuevo modelo con el objetivo de mejorar la precisión de los anteriores resultados. Para ello se recorta la señal del EEG eliminando todas las fases del sueño excepto N2 y N3, ya que son las que contienen *spindles* del sueño. Aplicando esta técnica, se pretende disminuir considerablemente el número de falsos positivos, así como el error cometido.

Es importante destacar que, para este nuevo método, se han elaborado dos modelos diferentes atendiendo al número de características con las que se ha entrenado. Por un lado, el primer modelo es entrenado con las 14 características seleccionadas en el artículo y, por otro, el segundo modelo se entrena con las 29 características calculadas en total sin aplicar

ningún tipo de método de selección. Ambos modelos han sido evaluados con el conjunto de test formado por el EEG de 4 sujetos.

En primer lugar, se van a comparar los resultados del modelo entrenado con 14 características con el original. En cuanto a los *boxplots* de las características, presentan una distribución similar a las características del modelo original con la salvedad de que para este nuevo caso parece disminuir el número de *outliers*, lo cual implica que estas sean más fiables y de mayor calidad a la hora de distinguir entre ambas clases.

Por otro lado, observando las matrices de confusión, se consigue disminuir considerablemente el número de falsos positivos, se pasa de 21981 a 16529. Las métricas obtenidas tras la evaluación con el conjunto de test son similares a las que se obtienen en el artículo original. Se consigue aumentar la precisión del 10.10% al 11.29% a pesar de que disminuye la exactitud y la especificidad y aumenta el error cometido.

En vista de que, a pesar de mejorar, aunque insuficientemente, los valores de precisión y de falsos positivos, se elaboró un nuevo modelo que intentara perfeccionar la clasificación aún más evaluado con todas las características calculadas. Se consideró que se podía dar el caso de que alguna de ellas contuviera información relevante para el rango de edad considerado en este proyecto, que es totalmente distinto al del artículo original y que no se ha tenido en cuenta para los anteriores modelos.

Atendiendo a los *boxplots* de las nuevas características extraídas en el dominio del tiempo, se puede ver en la *Figura 20* cómo las amplitudes medias en las distintas bandas pueden ser irrelevantes debido a su comportamiento similar a las amplitudes medias previamente calculadas. En cuanto al *skewness* y la *curtosis* parecen comportarse de forma similar, tanto si se trata de un *spindle* como si no a pesar de las diferencias en cuanto al número de *outliers*. Sin embargo, las nuevas características frecuenciales (*Figura 22*) parecen contener información adicional principalmente en las potencias relativas de las distintas bandas lo cual es un indicador de una posible mejoría en los nuevos resultados. La importancia de las características se comentará más adelante.

Estudiando los nuevos resultados obtenidos con la matriz de confusión se puede ver que el número de falsos positivos se reduce considerablemente para ambos conjuntos de evaluación en comparación con los anteriores modelos. Este hecho hace que aumenten los TN y se consigue disminuir no solo la tasa de error, sino también el error cuadrático medio. La precisión obtenida es mejor, aumentando hasta el 12.56%. En comparación con el mismo modelo entrenado con 14 características, se consigue aumentar tanto la exactitud como la sensibilidad y especificidad para ambos conjuntos de test así como disminuir la tasa de error y el RMSE, por lo que se puede considerar una buena decisión la elaboración de este nuevo modelo.

El principal motivo de obtener estos resultados reside en la incorporación de características que añaden información a la hora de entrenar el modelo y que previamente no se había tenido en cuenta. Principalmente, debido a la potencia relativa en la banda theta junto con el resto de potencias relativas y absolutas se consigue alcanzar el objetivo principal de mejorar la calidad de clasificador.

COMPARATIVA DE RESULTADOS

En este apartado se pretenden comparar los modelos generados en este TFG con el modelo implementado en el artículo que se quiere replicar. Para ello, y antes de comenzar con la comparativa, se presentan brevemente los resultados alcanzados por la publicación (Wei et al., 2022):

TEST	Spindles reales	No spindles reales
Predicción de spindles	3979	284
Predicción de no spindles	439	4736

Tabla 11: Matriz de confusión para el conjunto de test del artículo original (Wei et al., 2022).

VALIDACIÓN	Spindles reales	No spindles reales
Predicción de spindles	4050	254
Predicción de no spindles	590	4982

Tabla 12: Matriz de confusión para el conjunto de validación del artículo original (Wei et al., 2022).

Métricas de discriminación basadas en umbral	Valor en % con el conjunto de test	Valor en % con el conjunto de validación
Sensibilidad (<i>sensitivity</i>)	93.3	94.1
Especificidad (<i>specificity</i>)	91.5	89.4
Precisión o valor predictivo positivo (<i>precision</i>)	90.1	87.3

Tabla 13: Métricas de discriminación basadas en umbral obtenidas a partir de la matriz de confusión para el conjunto de test y validación del artículo original (Wei et al., 2022).

Métricas auxiliares	Valor para el conjunto de test	Valor para el conjunto de validación
Coeficiente de correlación de Matthew (MCC)	0.85	0.83

Tabla 14: Coeficiente de correlación de Matthew para ambos conjuntos de test y validación obtenidos en el modelo del artículo original (Wei et al., 2022).

En primer lugar, el estudio realizado por Wei et al. (2022) presenta unos valores de sensibilidad y especificidad muy altos (Tabla 13), aunque son más altos los obtenidos en este TFG para el conjunto de test (Tabla 5). Sin embargo, llama la atención la diferencia que se obtiene para la precisión (valor predictivo positivo) en el artículo a replicar, siendo muy elevada en torno al 90%. Este resultado no se consigue en ningún caso para ninguno de los modelos desarrollados en este TFG, siendo el mayor valor obtenido un 12.56%. El motivo de esta disimilitud se refleja principalmente en que el artículo balancea artificialmente el número de segmentos con *spindle* y sin *spindle* en el grupo de test, sin embargo, en este trabajo simplemente se balancea la clase *spindle* en el grupo de entrenamiento haciendo que las métricas, aunque más bajas, sean más realistas.

En cuanto al coeficiente de correlación de Matthew, en la Tabla 14 se puede ver cómo este valor se aproxima mucho más a +1 que los valores obtenidos en el estudio, lo que quiere decir que el clasificador implementado no es tan perfecto como lo es el desarrollado por la publicación. Teniendo en cuenta la ecuación 11 vista en el capítulo 3, se puede deducir que el motivo principal por el que este coeficiente es menor en esta investigación reside en el hecho de tener un elevado número de falsos positivos en comparación con el artículo (Wei et al., 2022).

Los principales motivos por los que se presentan estas diferencias en los resultados del artículo y los del TFG son los siguientes:

1. La base de datos con la que se trabaja en el artículo es mucho mayor (141 sujetos) que la que se ha conseguido para este estudio (9 sujetos).
2. En el artículo se trabaja con EEG de bebés recién nacidos (4 meses), mientras que en este TFG se trabaja con niños de entre 6 y 9 años, los cuales presentan diferencias intrínsecas a la hora de analizar los *spindles*.
3. Los resultados presentados en las *Tablas 11-14*, son los que se obtienen tras aplicar, después del balanceo, un nuevo criterio propio en función del número y duración de *spindles*, mientras que, en los resultados obtenidos en este TFG, dicho criterio no se ha tenido en cuenta. Aplicando esta nueva técnica en el artículo se consigue eliminar un gran número de falsos positivos, lo que les permite obtener una buena precisión.
4. En el artículo se trabaja con las señales F4-C4 y F3-C3, mientras que en este estudio únicamente se trabaja con la señal C3--M2 del EEG.

Además, se ha comprobado que las características importantes de los modelos desarrollados en este TFG son distintas a las obtenidas por el artículo original (*Figuras 23 y 24*). El motivo de ello puede residir en que las características extraídas por el artículo como *sigmaindex* o la potencia relativa en la banda sigma son específicas de neonatos, mientras que características como el valor de la envolvente o la amplitud media absoluta en la banda *spindle* son más específicas para el rango de edad en la que se ha centrado (6-9 años). Otro posible razonamiento puede ser que las características sean más importantes dependiendo del electrodo que se haya utilizado ya que es distinto al del artículo original. Incluso podría darse el caso que tanto la edad como el uso de canales distintos hayan influido en la diferencia de la importancia de las características.

Otro motivo que podría haber ayudado significativamente a obtener predicciones más precisas sería haber implementado una selección de características propias, en vez de utilizar directamente la selección del artículo, puesto que se puede ver en la *Figura 23* que algunas de ellas no se corresponden con las más importantes. Hay diferencias apreciables en cuanto a las características más y menos importantes para ambos modelos. Además, otra buena mejora hubiera sido probar con otros valores en la variable *k* del SMOTE para intentar optimizar los resultados.

A pesar de ello, en cuanto a resultados, se puede decir que el objetivo principal de toda esta investigación centrado en la detección automática de *spindles* del sueño aplicando *machine learning* se ha conseguido de manera satisfactoria (*Figura 25*) y que podría ser utilizado en el futuro como ayuda a expertos en su trabajo.

LIMITACIONES

Antes de finalizar este capítulo se detallan varias limitaciones que han estado presentes a lo largo del desarrollo del TFG.

La primera de las limitaciones, y la cual se puede considerar más importante, ha sido el hecho de trabajar con una base de datos en la que únicamente formaban parte de ella 9 sujetos. Esto implica que el alcance de los resultados obtenidos está limitado a un número reducido de sujetos. No obstante, la similitud entre los resultados presentados en este estudio con aquellos del estudio original, con una base de datos amplia de neonatos, sugiere que la utilidad observada en esta propuesta de detección automática de *spindles* puede tener continuidad en poblaciones más amplias y de diferentes características. Sin embargo, esta tendencia positiva deberá corroborarse en estudios futuros.

Otra de las limitaciones encontradas es el desbalanceo entre épocas con *spindles* y sin *spindles*, ocurriendo estas últimas en un número muy superior. Para tratar de minimizar esta limitación, se ha utilizado la técnica SMOTE configurada de acuerdo con el estudio original. No obstante, el problema no se ha logrado reducir suficientemente y, como consecuencia, los clasificadores entrenados generan un gran número de falsos positivos en relación al número de verdaderos positivos.

La tercera de las limitaciones, relacionada con el tipo de sujetos con el que se ha trabajado, es que la mayoría de ellos son pacientes con sospecha de apnea del sueño. La apnea del sueño conlleva a una reducción la densidad de *spindles* a lo largo del sueño en las fases N2 y N3 y, como consecuencia, puede inferir en el número de ejemplos pertenecientes a esa clase lo que puede originar problemas a la hora de la clasificación.

Otra limitación puede estar relacionada con el algoritmo de *machine learning* que se ha utilizado para llevar a cabo la clasificación de *spindles*. Aunque Random Forest es una técnica que ha mostrado rendimientos muy elevados en múltiples problemas de bioingeniería, la utilización de otras técnicas, como por ejemplo aquellas derivadas del *deep learning*, podría generar resultados diferentes. Por otro lado, la motivación de replicar con exactitud el algoritmo desarrollado por Wei *et al.* (2022) ha provocado que los hiperparámetros de ajuste que describen las características del *Random Forest*, así como los utilizados para la técnica del SMOTE hayan sido los mismos en vez de buscar unos propios atendiendo a propiedades específicas del estudio. Para ello, hubiera sido necesario utilizar un tercer grupo de validación, el cual no se ha utilizado en este estudio.

La última limitación podría estar relacionada con el hecho de que el modelo formado únicamente por las fases N2 y N3 del sueño está limitado por el algoritmo de *deep learning* desarrollado en el TFG de Eva Calvo (Calvo Merino, 2020). A pesar de que los resultados alcanzados por este algoritmo sean buenos (precisiones del 87%), presenta imperfecciones en cuanto a la asignación de las fases del sueño en señales EEG. Por otro lado, a raíz del alto rendimiento de dicho algoritmo, se puede interpretar que no ha habido una influencia excesiva en la detección en las fases N2-N3 y, por tanto, no comprometen las conclusiones obtenidas en este trabajo.

PROBLEMAS Y SOLUCIONES

Para finalizar, se pretende exponer el principal problema que se ha producido durante toda la implementación del clasificador relacionado con la base de datos inicial, así como los pasos seguidos hasta conseguir llegar a la solución final.

El problema principal con el que se ha topado a lo largo del desarrollo de este TFG y el cual ha entorpecido el avance de esta metodología ha estado relacionado con el hecho de que las marcas de *spindles* inicialmente no coincidían con las señales a trabajar. Por lo tanto, los resultados de clasificación iniciales ofrecían rendimientos muy bajos ya que no tenían ningún tipo de relación con los *spindles* que se trataba de detectar. A continuación, se enumeran varias sugerencias propuestas para intentar corregir este error y, finalmente, se recoge la solución final gracias a la cual este problema se pudo erradicar.

1. Se trató de modificar la señal realizando un re-referenciado promedio para ver si el problema estaba en la influencia que había entre unos canales y otros. La forma de proceder fue escoger todas las señales disponibles (C3, C4, F3, F4, O1 y O2) y con todo ello hacer un promedio para trabajar con una señal final formada por la resta del canal C3 con dicho promedio. Sin embargo, se observó que esta propuesta no mejoraba la situación inicial.
2. El siguiente procedimiento fue trabajar con la señal obtenida tras restar el canal C3 del mastoide M2 al cual estaba referenciado para eliminar posibles interferencias entre ellos. Sin embargo, este método tampoco introdujo ninguna mejora.
3. Finalmente, se optó por visitar el Hospital Universitario Río Hortega de Valladolid, proveedor de las bases de datos, para descartar que el problema estuviera en las bases de datos proporcionadas. Gracias a esta visita se pudo ver que efectivamente el problema residía en dichas señales y se cambió la forma de exportación de ellas pudiendo finalmente implementar el método correctamente con la señal C3-M2.

Todo ello ha servido como experiencia para el futuro a la hora de trabajar con cualquier tipo de señal. Verificar en todo momento que los datos iniciales no están corruptos es una buena práctica que ahorra tiempo a la hora de descartar problemas cuando se está desarrollando cualquier tipo de investigación.

CAPÍTULO 6: CONCLUSIONES Y LÍNEAS FUTURAS

Durante este TFG se ha trabajado con señales EEG en niños con el objetivo de elaborar un algoritmo capaz de detectar *spindles* del sueño. Después de explicar la metodología utilizada, los resultados obtenidos y la discusión sobre ellos, se finaliza este trabajo exponiendo las contribuciones realizadas al campo de investigación, las conclusiones finales de todo el proyecto y posibles líneas futuras que se pueden seguir para continuar el estudio.

CONTRIBUCIONES AL CAMPO DE INVESTIGACIÓN

Al finalizar este estudio, se deben reflejar las contribuciones que dicho trabajo aporta al campo de investigación, las cuales se enumeran a continuación:

- En primer lugar, el desarrollo de un algoritmo capaz de detectar *spindles* del sueño en señales procedentes del EEG en niños empleando técnicas de *machine learning* ha aportado una nueva metodología a la hora de trabajar con estas señales, que incorpora novedades como la de trabajar con segmentos de 0.5 segundos con solapamiento, la cual no se había explorado hasta el momento y que parece obtener resultados adecuados.
- Por otro lado, debido al problema expuesto en el capítulo anterior, se han podido corregir las señales corruptas con las que se ha estado trabajando hasta el momento, las cuales han podido ocasionar errores en proyectos de este tipo. Además, se han aplicado técnicas de referenciado para tratar de paliar esta cuestión, que pueden ser de gran utilidad como solución a otro tipo de dificultades.
- Se ha comprobado que las características importantes para los modelos desarrollados en este TFG son diferentes a los del artículo original.
- También se ha utilizado para este estudio únicamente el canal C3 del EEG, mientras que en el artículo original se utilizan a mayores el C4, F3 y F4.

- Por último, el desarrollo de ese TFG demuestra la posibilidad de desarrollar algoritmos clasificadores que detecten de manera automática *spindles* del sueño en niños entre 6 y 9 años. De esta forma se facilita la labor de su marcaje a expertos y el poder diagnosticar, en caso de detectar irregularidades en la densidad de *spindles* a lo largo del sueño, enfermedades relacionadas con el proceso cognitivo del paciente más rápidamente.

CONCLUSIONES

A continuación, se exponen las diferentes conclusiones extraídas en base al trabajo desarrollado a lo largo de este TFG:

- La elaboración de técnicas para la detección automática de *spindles* del sueño puede conseguir resultados de clasificación elevados en comparación a los obtenidos tras un proceso manual por expertos. De esta forma se agiliza el proceso y se pueden reducir sus costes en su procedimiento.
- Al trabajar con señales procedentes del EEG, se ha podido verificar que esta señal contiene información relevante y adecuada para caracterizar a este tipo de oscilaciones. Además, se puede afirmar que trabajar con un solo canal C3 es suficiente para llevar a cabo este tipo de tareas, aunque es cierto que la probabilidad de conseguir otro tipo de características con otros canales podría haber sido una buena idea para tratar de mejorar aún más los resultados.
- Como conclusión para el modelo original, presenta un alto rendimiento a la hora de clasificar o detectar *spindles* del sueño, ya que detecta prácticamente todos los marcados por profesionales. Sin embargo, no garantiza que todos los *spindles* predichos se correspondan a ellos debido al elevado número de falsos positivos. A pesar de ello, este método puede ayudar en gran medida a los expertos a detectar *spindles* con sólo fijarse en las zonas en las que el algoritmo los ha marcado, sin necesidad de revisar la señal completa, acelerando y facilitando su trabajo. Por otra parte, debido al problema del balanceado de datos y con la finalidad de intentar mejorar los resultados disminuyendo el número de falsos positivos para mejorar la precisión, se sugiere modificar la señal original trabajando únicamente con las fases N2 y N3 del sueño.
- Por otro lado, acortar la señal original para elaborar nuevos modelos centrándose únicamente las fases N2 y N3 del sueño ha permitido mejorar los resultados disminuyendo el número de falsos positivos de 21981 a 16529 y aumentar la precisión de 10.10% a 12.56%, tal y como se esperaba inicialmente, aunque la mejora no sea excesivamente mayor. Cabe destacar que de los dos modelos entrenados únicamente con las fases N2 y N3, el mejor de ellos ha sido el que

mantiene todas las características interpretando por ello que en el resto de los casos se está perdiendo información de los *spindles*.

- Respecto a los modelos N2-N3, se puede concluir que estos nuevos modelos específicos consiguen mejorar los resultados esperados, aunque la mejora no sea excesivamente mayor. Los falsos positivos, aunque disminuyen de 21981 a 16529, siguen siendo muy altos y el motivo de ello reside en el desequilibrio que hay entre las clases que se pretenden diferenciar. Cabe destacar que el modelo que mejores resultados obtiene de manera general es el último, ya que se consigue la mayor precisión del 12.56% y muy bajo error cometido 4.01%.
- Gracias al análisis espectral realizado, los *boxplots* de las características extraídas y la representación de su importancia, obtenida para el último modelo desarrollado, se ha podido observar el interés que presenta la banda spindle, entre 8 y 15 Hz en niños, a la hora de detectar *spindles* del sueño. Sin embargo, la extracción de características fuera de esa banda como la alpha, delta, beta, sigma, parecen no aportar demasiada información a la hora de realizar la clasificación. Además, hay que destacar que el análisis en frecuencia para estimar las potencias absolutas y relativas de la señal resulta imprescindible en el desarrollo de este trabajo.

Se puede concluir que, a lo largo de este TFG se ha desarrollado un algoritmo eficaz de detección de *spindles* del sueño basado en técnicas de *machine learning* en niños entre 6 y 9 años, empleando para ello un único canal C3 procedente del EEG de cada sujeto, referenciado al mastoide M2. Los resultados en ámbito general han sido elevados, aunque con las limitaciones descritas en apartados anteriores.

LÍNEAS FUTURAS

Con el objetivo de continuar este proyecto en líneas futuras, conviene resaltar algunos aspectos de este TFG que podrían estudiarse con mayor detalle y poder conseguir así un modelo con mejores prestaciones. Principalmente, dichos aspectos se encuentran relacionados con las limitaciones expuestas en el capítulo anterior.

Uno de los primeros pasos a realizar sería continuar la investigación realizando una búsqueda de más EEG pertenecientes a niños del mismo rango de edades para aumentar en la mayor medida posible la base de datos con la que se trabaja. Además, sería interesante no solo realizar el estudio con niños en presencia de AOS, sino ampliar el número de sujetos con pacientes que presenten otro tipo de enfermedades con el objetivo de llevar a cabo una investigación más exhaustiva acerca del comportamiento que presentan los *spindles* en función de ello. Por otro lado, también sería una buena idea variar el rango de edades y el sexo de los pacientes con los que se realiza el estudio, aunque es cierto que hay un mayor número de investigaciones realizadas para este campo en adultos que en niños. En definitiva, todo lo relacionado con aumentar la base de datos con otro tipo de

pacientes con distintas enfermedades o incluso edades aportaría otro tipo de información al método de clasificación. Siguiendo con la idea de aumentar el tamaño de la base de datos, sería novedoso implementar algoritmos basados en *deep learning* debido a que son tecnologías más recientes y avanzadas orientadas también al ámbito de la clasificación como el que se propone en (Albawi et al., 2018).

Otra línea de investigación innovadora sería ampliar el modelo desarrollado a un multclasificador para que detecte no solo *spindles* del sueño, sino también otro tipo de señales contenidas en el EEG como son *ripples* o *k-complex*, las cuales también están relacionadas con el desarrollo cognitivo. Esto permitiría llevar a cabo un estudio mucho más amplio en el que se podría ayudar a detectar otro tipo de trastornos, que no se pueden diagnosticar únicamente observando *spindles* del sueño.

En relación con el preprocesado de la señal, sería interesante probar a modificar el tipo de referenciado y verificar si este cambio ha sido significativo para los resultados o no. Asimismo, realizar pruebas con otros canales del EEG e incluso introducir algunos nuevos para el estudio sería una alternativa bastante original y prometedora en cuanto a los resultados esperados.

Por otra parte, a la vista de que el mayor problema que presenta el modelo desarrollado es el alto número de falsos positivos, podría ser interesante aplicar dentro del algoritmo algún tipo de arquitectura que diferencie las etapas del sueño N2 y N3 del resto, sin necesidad de invertir tiempo en modificar la señal original cada vez que se desee estimar la densidad de *spindles* para un determinado sujeto. Además, cambiar la técnica SMOTE empleada para el balanceo de datos por otra de las que se han ido explicando a lo largo de este TFG sería una manera interesante de comparar cuál método es el más eficaz.

Para finalizar, vista la importancia que tiene para las personas el estudio de los *spindles* del sueño a la hora de diagnosticar posibles enfermedades, sería interesante desarrollar algún programa o aplicación, en base a los distintos modelos entrenados o similares, que permita realizar una prueba a cada sujeto desde cualquier dispositivo electrónico en el momento que se desee. Esto permitirá, entre otras cosas, concienciar a la población a cerca de lo importante que es tener un buen descanso y diagnosticar enfermedades como la apnea del sueño a tiempo para poder ser tratados lo antes posible y mejorar su calidad de vida.

REFERENCIAS

- Acir, N., & Güzeliş, C. (2004). Automatic recognition of sleep spindles in EEG by using artificial neural networks. *Expert Systems with Applications*, 27(3), 451–458. <https://doi.org/10.1016/j.eswa.2004.05.007>
- Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis*, 22(5), 717–727. [https://doi.org/10.1016/S0731-7085\(99\)00272-1](https://doi.org/10.1016/S0731-7085(99)00272-1)
- Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2018). Understanding of a convolutional neural network. *Proceedings of 2017 International Conference on Engineering and Technology, ICET 2017, 2018-Janua*, 1–6. <https://doi.org/10.1109/ICEngTechnol.2017.8308186>
- Author, K., Groeneveld, R. A., & Meeden, G. (1984). Measuring Skewness Institute of Statisticians Measuring Skewness and Kurtosis. *Source: Journal of the Royal Statistical Society. Series D (The Statistician)*, 33(4), 391–399.
- Batta, M. (2020). Machine Learning Algorithms - A Review. *International Journal of Science and Research (IJ, 9(1)*, 381–386. <https://doi.org/10.21275/ART20203995>
- Boughorbel, S., Jarray, F., & El-Anbari, M. (2017). Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. *PLoS ONE*, 12(6), 1–17. <https://doi.org/10.1371/journal.pone.0177678>
- Calvo Merino, E. (2020). *Análisis de electroencefalogramas para la detección automática de las fases del sueño* (p. 99).
- Carmona, E. (2016). *Abstract Support Vector Machine 1 Introducción*. November, 1–27. https://www.researchgate.net/publication/263817587_Tutorial_sobre_Maquinas_de_Vectores_Soporte_SVM
- Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, 2(01), 20–28. <https://doi.org/10.38094/jastt20165>
- Chiappa, K. H. (1997). *Evoked potentials in clinical medicine* (3rd ed). Lippincott-Raven. <https://books.google.es/books?hl=es&lr=&id=h5mOjnTimT8C&oi=fnd&pg=PP11&dq=evoked+potentials&ots=uNoCVw0eOt&sig=t446oo33SjyCFpTTN9LVVIUB6vw#v=onepage&q=evoked+potentials&f=false>
- Druzgalski, C., Vega, J. E. M., & Zeljkovic, V. (2016). *Determinación del Tamaño Óptimo de Modelos HMM-GMM para Clasificación de las Señales Bioacústicas*. 37(1), 63–79.
- Duman, F., Erdamar, A., Eroglu, O., Telatar, Z., & Yetkin, S. (2009). Efficient sleep spindle detection algorithm with decision tree. *Expert Systems with Applications*, 36(6), 9980–9985. <https://doi.org/10.1016/j.eswa.2009.01.061>
- Eagan, B., Misfeldt, M., & Siebert-Evenstone, A. (2019). Advances in Quantitative Ethnography. In *Advances in Quantitative Ethnography* (Vol. 1). <https://doi.org/10.1007/978-3-030-33232-7>
- Ferreira, P., Le, D. C., & Zincir-Heywood, N. (2019). Exploring Feature Normalization and Temporal Information for Machine Learning Based Insider Threat Detection. *15th International Conference on Network and Service Management, CNSM 2019, Cnsm*. <https://doi.org/10.23919/CNSM46954.2019.9012708>
- Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters*, 27(4), 294–300. <https://doi.org/10.1016/j.patrec.2005.08.011>
- Gomez-Pilar, J., Gutiérrez-Tobal, G. C., Poza, J., Fogel, S., Doyon, J., Northoff, G., & Hornero,

- R. (2021). Spectral and temporal characterization of sleep spindles - methodological implications. *Journal of Neural Engineering*, 18(3). <https://doi.org/10.1088/1741-2552/abe8ad>
- Güneş, S., Dursun, M., Polat, K., & Yosunkaya, Ş. (2011). Sleep spindles recognition system based on time and frequency domain features. *Expert Systems with Applications*, 38(3), 2455–2461. <https://doi.org/10.1016/j.eswa.2010.08.034>
- Henry Olivi, R. (2013). Apnea del sueño: cuadro clínico y estudio diagnóstico. *Revista Médica Clínica Las Condes*, 24(3), 359–373. [https://doi.org/10.1016/s0716-8640\(13\)70173-1](https://doi.org/10.1016/s0716-8640(13)70173-1)
- Hermida L, R. C. (2016). El síndrome de apnea del sueño en los niños. *Revista Odontopediatría Latinoamericana*, 6(2), 1–19.
- Herrera, F. (2004). *Preprocesamiento de Datos en Minería de Datos*. 6–7.
- Herwig, U., Satrapi, P., & Schönfeldt-Lecuona, C. (2003). Using the International 10-20 EEG System for Positioning of Transcranial Magnetic Stimulation. *Brain Topography*, 16(2), 95–99. <https://doi.org/10.1023/B:BRAT.0000006333.93597.9d>
- Hossin, M., & Sulaiman, M. . (2015). a Review on Evaluation Metrics for Data. *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, 5(2), 1–11.
- Jorge Dagnino, S. (2014). Correlación. *Revista Chilena de Anestesia*, 43(2), 150–153.
- Khalid, S., Khalil, T., & Nasreen, S. (2014). A survey of feature selection and feature extraction techniques in machine learning. *Proceedings of 2014 Science and Information Conference, SAI 2014*, 372–378. <https://doi.org/10.1109/SAI.2014.6918213>
- Kinoshita, T., Fujiwara, K., Kano, M., Ogawa, K., Sumi, Y., Matsuo, M., & Kadotani, H. (2020). Sleep Spindle Detection using RUSBoost and Synchrosqueezed Wavelet Transform. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(2), 390–398. <https://doi.org/10.1109/TNSRE.2020.2964597>
- Kovács, B., Tinya, F., Németh, C., & Ódor, P. (2020). Unfolding the effects of different forestry treatments on microclimate in oak forests: results of a 4-yr experiment. *Ecological Applications*, 30(2), 321–357. <https://doi.org/10.1002/eap.2043>
- Kulkarni, P. M., Xiao, Z., Robinson, E. J., Jami, A. S., Zhang, J., Zhou, H., Henin, S. E., Liu, A. A., Osorio, R. S., Wang, J., & Chen, Z. (2019). A deep learning approach for real-time detection of sleep spindles. *Journal of Neural Engineering*, 16(3). <https://doi.org/10.1088/1741-2552/ab0933>
- Manoach, D. S., Pan, J. Q., Purcell, S. M., & Stickgold, R. (2016). Reduced Sleep Spindles in Schizophrenia: A Treatable Endophenotype That Links Risk Genes to Impaired Cognition? *Biological Psychiatry*, 80(8), 599–608. <https://doi.org/10.1016/j.biopsych.2015.10.003>
- Martinek, R., Ladrova, M., Sidikova, M., Jaros, R., Behbehani, K., Kahankova, R., & Kawala-Sterniuk, A. (2021). Advanced bioelectrical signal processing methods: Past, present and future approach—Part II: Brain signals. *Sensors*, 21(19), 1–32. <https://doi.org/10.3390/s21196343>
- Mohammadi, H., Aarabi, A., Rezaei, M., Khazaie, H., & Brand, S. (2021). Sleep Spindle Characteristics in Obstructive Sleep Apnea Syndrome (OSAS). *Frontiers in Neurology*, 12(February), 1–14. <https://doi.org/10.3389/fneur.2021.598632>
- Palacios, L. (2002). Breve historia de la electroencefalografía. *Acta Neurológica Colombiana*, 2(2), 104–107.
- Peralta, R. G., González-andino, S., & Gómez-gonzález, C. M. (2004). *grave_REN_2004*. 39(8), 748–756.
- Purcell, S. M., Manoach, D. S., Demanuele, C., Cade, B. E., Mariani, S., Cox, R., Panagiotaropoulou, G., Saxena, R., Pan, J. Q., Smoller, J. W., Redline, S., & Stickgold, R. (2017). Characterizing sleep spindles in 11,630 individuals from the National Sleep

- Research Resource. *Nature Communications*, 8(May), 1–16. <https://doi.org/10.1038/ncomms15930>
- Sadikoglu, F., Kavalcioglu, C., & Dagman, B. (2017). Electromyogram (EMG) signal detection, classification of EMG signals and diagnosis of neuropathy muscle disease. *Procedia Computer Science*, 120, 422–429. <https://doi.org/10.1016/j.procs.2017.11.259>
- Schimicek, P., Zeitlhofer, J., Anderer, P., & Saletu, B. (1994). Automatic Sleep-Spindle Detection Procedure: Aspects of Reliability and Validity. *Clinical EEG and Neuroscience*, 25(1), 26–29. <https://doi.org/10.1177/155005949402500108>
- Sörnmo, L., & Laguna, P. (2005). *Bioelectrical signal processing in cardiac and neurological applications* (N. Stjernan (ed.); Elsevier A). https://books.google.es/books?hl=es&lr=&id=RQv7tFFXYlC&oi=fnd&pg=PP5&dq=bioelectrical+signal+processing&ots=hGBr8i9QxH&sig=SLUiiNg_FJat8rugnftjkJ3eq8#v=onepage&q=bioelectrical+signal+processing&f=false
- Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications*, 134, 93–101. <https://doi.org/10.1016/j.eswa.2019.05.028>
- Uong, E. C., Jeffe, D. B., Gozal, D., Arens, R., Holbrook, C. R., Palmer, J., Cleveland, C., & Schotland, H. M. (2005). Development of a measure of knowledge and attitudes about obstructive sleep apnea in children (OSAKA-KIDS). *Archives of Pediatrics and Adolescent Medicine*, 159(2), 181–186. <https://doi.org/10.1001/archpedi.159.2.181>
- Ventouras, E. M., Monoyiou, E. A., Ktonas, P. Y., Paparrigopoulos, T., Dikeos, D. G., Uzunoglu, N. K., & Soldatos, C. R. (2005). Sleep spindle detection using artificial neural networks trained with filtered time-domain EEG: A feasibility study. *Computer Methods and Programs in Biomedicine*, 78(3), 191–207. <https://doi.org/10.1016/j.cmpb.2005.02.006>
- Wang, L. (2005). *Support Vector Machines: Theory and Applications* (Springer). <https://books.google.es/books?hl=es&lr=&id=uTzMPJjVjsMC&oi=fnd&pg=PA1&dq=svm+support+vector+machine&ots=GFEKczZFnd&sig=mfD2KfgL8whoQztPaibhLUEXXnE#v=onepage&q=svm+support+vector+machine&f=false>
- Warby, S. C., Wendt, S. L., Welinder, P., Munk, E. G. S., Carrillo, O., Sorensen, H. B. D., Jennum, P., Peppard, P. E., Perona, P., & Mignot, E. (2014). Sleep-spindle detection: Crowdsourcing and evaluating performance of experts, non-experts and automated methods. *Nature Methods*, 11(4), 385–392. <https://doi.org/10.1038/nmeth.2855>
- Wei, L., Ventura, S., Mathieson, S., Boylan, G., Lowery, M., & Mooney, C. (2022). Spindle-AI: Sleep Spindle Number and Duration Estimation in Infant EEG. *IEEE Transactions on Biomedical Engineering*, 69(1), 465–474. <https://doi.org/10.1109/TBME.2021.3097815>
- Yetton, B. D., Niknazar, M., Duggan, K. A., McDevitt, E. A., Whitehurst, L. N., Sattari, N., & Mednick, S. C. (2016). Automatic detection of rapid eye movements (REMs): A machine learning approach. *Journal of Neuroscience Methods*, 259, 72–82. <https://doi.org/10.1016/j.jneumeth.2015.11.015>
- Zhang, C., & Ma, Y. (2012). Ensemble Machine Learning. In *Ensemble Machine Learning*. <https://doi.org/10.1007/978-1-4419-9326-7>
- Zhang, G. Q., Cui, L., Mueller, R., Tao, S., Kim, M., Rueschman, M., Mariani, S., Mobley, D., & Redline, S. (2018). The National Sleep Research Resource: Towards a sleep data commons. *Journal of the American Medical Informatics Association*, 25(10), 1351–1358. <https://doi.org/10.1093/jamia/ocy064>