

Programa de Doctorado en Neurociencias

Detección de TDAH en niños mediante oscilaciones fNIRS prefrontales de muy baja frecuencia durante una tarea mental rítmica

Tesis Doctoral presentada por

SERGIO ORTUÑO MIRÓ

Director/a de Tesis:

Dr. D. Joaquín Ibáñez Ballesteros

Instituto de Neurociencias

Universidad Miguel Hernández de Elche

- 2024 -

La presente Tesis Doctoral, titulada "Detección de TDAH en niños mediante oscilaciones fNIRS prefrontales de muy baja frecuencia durante una tarea mental rítmica" se presenta bajo la modalidad de tesis por compendio de las siguientes publicaciones:

Ortuño-Miró, S., Molina-Rodríguez, S., Belmonte, C. y Ibañez-Ballesteros, J. (2023). Identifying ADHD boys by very-low frequency prefrontal fNIRS fluctuations during a rhythmic mental arithmetic task. *Journal of Neural Engineering*, 20(3), 036018. DOI 10.1088/1741-2552/acad2b





El Prof. Dr. D. Joaquin Ibañez Ballesteros, director de la tesis doctoral titulada "Detección de TDAH en niños mediante oscilaciones fNIRS prefrontales de muy baja frecuencia durante una tarea mental rítmica"

INFORMA:

Que D. Sergio Ortuño Miró ha realizado bajo mi supervisión el trabajo titulado "Detección de TDAH en niños mediante oscilaciones fNIRS prefrontales de muy baja frecuencia durante una tarea mental rítmica" conforme a los términos y condiciones definidos en su Plan de Investigación y de acuerdo con el Código de Buenas Prácticas de la Universidad Miguel Hernández de Elche, cumpliendo los objetivos previstos de forma satisfactoria para su defensa pública como tesis doctoral.

Lo que firmo para los efectos oportunos, en San Joan d'Alacant a 4 de Junio de 2024

Director de la tesis

Prof. Dr. D. Joaquin Ibañez Ballesteros



La Dra. Dña. María Cruz Morenilla Palao, Coordinadora del Programa de Doctorado en NEUROCIENCIAS

INFORMA:

Que D. Sergio Ortuño Miró ha realizado bajo la supervisión de nuestro Programa de Doctorado el trabajo titulado "Detección de TDAH en niños mediante oscilaciones fNIRS prefrontales de muy baja frecuencia durante una tarea mental rítmica." conforme a los términos y condiciones definidos en su Plan de Investigación y de acuerdo al Código de Buenas Prácticas de la Universidad Miguel Hernández de Elche, cumpliendo los objetivos previstos de forma satisfactoria para su defensa pública como tesis doctoral.

Lo que firmo para los efectos oportunos, en San Juan de Alicante a 4 de Junio de 2024

Dra. Dña. María Cruz Morenilla Palao

Coordinadora del Programa de Doctorado en NEUROCIENCIAS

AGRADECIMIENTOS

Quiero agradecer al profesor Dr. Joaquín Ibáñez su generosidad por haberme brindado la oportunidad para realizar esta tesis y permitirme participar en el gran trabajo que se desarrolla en el Departamento de Fisiología de la Universidad Miguel Hernández. Aprovecho estas líneas para expresar mi admiración por su infatigable capacidad para el trabajo y por su dominio de distintas materias y cuestiones, que hacen de él un científico difícil de imitar. Para mí ha sido un auténtico lujo tenerlo de director y recibir su docencia, que además de lo académico incluye valores humanos.

Igualmente quiero agradecer al Dr. Sergio Molina su colaboración en el trabajo realizado en el laboratorio, así como sus consejos para elaborar esta tesis. Ha sido un privilegio ver su crecimiento como científico y profesor, todo un ejemplo a imitar.

Mi especial gratitud para los padres y niños que han participado en los experimentos de forma altruista, y para todos aquellos que han colaborado, ya sean estudiantes o profesionales de distintas materias, con el trabajo que se desarrolla en el Departamento de Fisiología.

También quiero agradecer el apoyo técnico e instrumental prestado desinteresadamente por la compañía Newmanbrain, SL, fabricante del dispositivo fNIRS.

No me olvido de la mujer de mi vida, Marián, que algún consejo me ha dado, y cuya existencia es fundamental en mi andadura vital.

ABREVIATURAS	8
LISTA DE FIGURAS	
LISTA DE TABLAS	14
RESUMEN	15
ABSTRACT	17
1. INTRODUCCIÓN	19
1.1. Concepto de TDAH	19
1.1.1. El TDAH en el DSM-5	20
1.1.2. El TDAH en la CIE-11	23
1.2. Epidemiología	24
1.3. Evaluación y diagnóstico del TDAH	27
1.3.1. Procedimientos diagnósticos: consenso clínico	27
1.3.2. Controversia sobre la evaluación del TDAH	29
1.3.3. Otras herramientas diagnósticas	30
1.4. Pregunta de investigación	31
1.5. La espectroscopía funcional de infrarrojo cercano (fNIRS)	31
1.5.1. Principios operacionales fNIRS	32
1.5.2. La corteza prefrontal	35
1.5.3. Estudios NIRS en población TDAH	
1.6. Componentes oscilatorios hemodinámicos y función cerebral	
1.7. Hipótesis de trabajo	
1.8. Objetivos	
2. MÉTODO	41
2.1. Población y muestra	41
2.2. Procedimiento	41
2.3. Dispositivo fNIRS	43
2.4. Distribución de la energía espectral durante la tarea	45
2.5. Descomposición tiempo-frecuencia	46
2.6. Detección de la sincronización tiempo-escala	49
2.7. Subbandas de frecuencia relevantes y extracción de características	51
2.8. Algoritmos de clasificación y selección de características	54
3. RESULTADOS	58
3.1. Análisis PSD	

ÍNDICE

	3.2. Mapas de tiempo-frecuencia ISS	58
	3.3. Subbandas de frecuencia y shapelets referencia	60
	3.4. Rendimiento de los clasificadores	63
4.	DISCUSIÓN	68
	4.1. Inducción de oscilaciones por medio de la aritmética mental	68
	4.2. Selección de características y rendimiento de clasificadores	69
	4.3. Acuerdo en la VC	71
	4.4. Contribución de SS y CS a la clasificación	72
	4.4. Precisión, sensibilidad y especificidad	74
	4.5. Limitaciones y trabajos futuros	74
5.	CONCLUSIONES	75
6.	REFERENCIAS	76
7.	ANEXO	96



ABREVIATURAS

ADL: Análisis discriminante lineal
APA: Asociación Americana de Psicología
CPF: Corteza prefrontal
CIE: Clasificación internacional de las enfermedades
CS: Señal neural limpia
COI: Cono de influencia
CWT : Continuous wavelet transform
DSM: Manual diagnóstico y estadístico de los trastornos mentales
DT: Desarrollo típico
EEG: electroencefalografía
FFT: Fast Fourier transform
fNIRS: espectroscopia funcional de infrarrojo cercano
fMRI: resonancia magnética funcional
HbO: oxihemoglobina
HbR: desoxihemoglobina
IC: intervalo de confianza
ISS : inter-subject synchronization
IA: Inteligencia Artificial
MRI: resonancia magnética estructural
ML: Machine learning
NB: Naïve Bayes
OMS: Organización Mundial de la Salud
PSD: densidad espectral de potencia
ROI: Región de interés
RL: Regresión logística
SDD: distancia fuente-detector
SEM: Error estándar de la media
SFFS: Sequential forward floating selection

SS: Señal superficial

- **SVM**: Support vector machine
- TFA: Análisis tiempo-frecuencia

TDAH: Trastorno por déficit de atención e hiperactividad

VC: Validación cruzada



LISTA DE FIGURAS

- Figura 1. Representación del espectro electromagnético por longitud de onda.
- Figura 2. Representación esquemática de un montaje NIRS típico en el que se observa la luz cercana al infrarrojo (línea discontinua) viajando desde el emisor (E) hasta el detector (D).
- Figura 3. Representación esquemática de un montaje NIRS de "múltiples distancias" de medida entre los optodos" en el que se observa la luz cercana al infrarrojo (línea discontinua) viajando desde el emisor (E) hasta el detector (D) tanto en el canal corto como en el canal largo.
- Figura 4. Representación esquemática de la corteza frontopolar en el encéfalo humano.
- Figura 5. Esquema del procedimiento experimental.
- Figura 6. Posicionamiento del dispositivo NIRS. Los cuadrados amarrillos representan los emisores de luz, y los círculos blancos los detectores. Las líneas negras muestran los 16 los canales cortos y las líneas verdes los 12 canales largos. Los rectángulos formados por línea discontinua representan las tres regiones de interés (ROI) exploradas en este trabajo.
- Figura 7. Transformada wavelet continua (CWT) de la señal profunda de HbO desde la mitad del retorno de la inversión para un participante representativo. Izquierda: Escalograma de los datos extendidos simétricamente con el cono de influencia trazado en magenta. Las líneas discontinuas blancas contienen la señal original y las líneas negras continuas delimitan el período de la tarea. El rectángulo rojo representa la región del escalograma de la que se extrajeron los coeficientes para el análisis. Derecha: Esquemas de la matriz de coeficientes extraída que contiene 41 escalas y 3600 puntos temporales.
- Figura 8. Esquemas del procedimiento para el cálculo de un mapa representativo de sincronización entre sujetos (ISS). Comenzando con las matrices de coeficientes de cada individuo dentro de un grupo (izquierda), se reúnen datos de valores complejos para cada intervalo de frecuencia (centro) y la ISS se calcula momento a momento para obtener una representación del poder de sincronización dentro del grupo en cada frecuencia a lo largo del tiempo (mapa de color derecho). El gráfico rojo de más a la derecha muestra los máximos y mínimos alcanzados

por la ISS a lo largo de las frecuencias, que se utilizan para delimitar las subbandas a analizar (líneas discontinuas blancas horizontales).

- Figura 9. Ejemplo esquemático del procedimiento para extraer *shapelets* de referencia para el grupo DT y contrastarlos con el grupo TDAH. Las series de tiempo por participante dentro de una subbanda de frecuencia específica se obtienen a partir de su CWT inverso de banda limitada y se promedian solo para que el grupo DT obtenga el *shapelet*. Las distancias euclidianas al *shapelet* se calculan para las series de tiempo de ambos grupos y luego se contrastan las varianzas dentro y entre grupos para obtener el valor *F-stat*.
- Figura 10. PSD normalizados promedio para cada ROI (derecha, media e izquierda), cromóforo y tipo de señal en el rango de frecuencia de 0 a 80 mHz. Las curvas de colores gruesos se refieren al grupo DT (HbO en rojo y HbR en azul) mientras que las negras gruesas corresponden al grupo TDAH. Las líneas finas representan el IC del 95 % de la media. Rectángulos magenta delimitan los rangos de frecuencia que muestran diferencias significativas entre grupos. El panel SS muestra los PSD para señales poco profundas y CS para señales limpias.
- Figura 11. Mapas de color de sincronización entre sujetos (ISS) para las señales superficiales (SS) en los diferentes ROIs. Las gráficas de la parte superior corresponden a las señales de HbO para los grupos TD y TDAH, mientas que las de la parte inferior muestran las de HbR. A la izquierda de cada mapa se muestran un gráfico con los valores máximos de ISS en cada una de las bandas de frecuencia. Las líneas verticales negras delimitan el intervalo de la tarea. Las líneas discontinuas blancas delimitan las diferentes subbandas (etiquetas con letras A D) obtenidas en base a los valores máximos de ISS en cada una de las regiones de interés.
- Figura 12. Mapas de color de sincronización entre sujetos (ISS) para las señales limpias (CS) en los diferentes ROIs. Las gráficas de la parte superior corresponden a las señales de HbO para los grupos TD y TDAH, mientas que las de la parte inferior muestran las de HbR. A la izquierda de cada mapa se muestran un gráfico con los valores máximos de ISS en cada una de las bandas de frecuencia. Las líneas verticales negras delimitan el intervalo de la tarea. Las líneas discontinuas blancas delimitan las diferentes subbandas (etiquetas con

letras A - D) obtenidas en base a los valores máximos de ISS en cada una de las regiones de interés.

- Figura 13. Patrones promedio de sincronización entre sujetos (ISS) en todas las frecuencias. Las gráficas de la fila superior corresponden a los datos de SS y las de la fila inferior a CS. Los trazos rojos y azules se representan la HbO y HbR del grupo DT, mientras que los trazos negros muestran los datos del grupo TDAH. Cada subbanda se identifica con los rectángulos etiquetados horizontalmente, cuyos límites están definidos por las líneas discontinuas verticales.
- Figura 14: Shapelets referencia de señales superficiales (SS) para cada subbanda (A-D) dentro de cada ROI. Las gráficas de la parte superior muestran los patrones de HbO (trazos rojos) para los grupos DT y TDAH, mientras que las gráficas de la parte inferior muestran las señales de HbR (trazos azules). Las líneas finas representan el error estándar de la media de cada una de la shapelets. Las líneas discontinuas muestran el curso temporal de la sincronización entre sujetos (ISS) en cada subbanda (escala en el eje derecho); en color magenta se muestran los correspondientes a shapelets con un valor p <.01. Las shapelets seleccionadas por el wrapper son identificadas con cuadrados de colores. Los rectángulos de color gris indican cada uno de los periodos de 15 segundos de cálculo mental durante la tarea.
- Figura 15: Shapelets referencia de señales limpias (CS) para cada subbanda (A-D) dentro de cada ROI. Las gráficas de la parte superior muestran los patrones de HbO (trazos rojos) para los grupos DT y TDAH, mientras que las gráficas de la parte inferior muestran las señales de HbR (trazos azules). Las líneas finas representan el error estándar de la media de cada una de la shapelets. Las líneas discontinuas muestran el curso temporal de la sincronización entre sujetos (ISS) en cada subbanda (escala en el eje derecho); en color magenta se muestran los correspondientes a shapelets con un valor p <.01. Las shapelets seleccionadas por el wrapper son identificadas con cuadrados de colores. Los rectángulos de color gris indican cada uno de los periodos de 15 segundos de cálculo mental durante la tarea.
- Figura 16. Rendimiento de los diferentes modelos de clasificación empleando el subconjunto de características seleccionadas por el *wrapper* para SS, CS y SS + CS. Los recuadros coloreados a la derecha identifican las características elegidas

y su orden de entrada. Cada característica específica tiene un color diferente y su etiqueta es la combinación "Grupo-Banda-Cromóforo-ROI", siendo TD = grupo TD, AD = TDAH, L = izquierda, M = media y R = derecha. Nota: LOO = VC dejando uno fuera; 5-FO = 5 VC de 5 iteraciones; SVM = máquina de soporte vector; LR = regresión logística; ADL = análisis discriminante lineal; NB = naïve Bayes Gaussiano.



LISTA DE TABLAS

- Tabla 1: Límites de frecuencia de cada subbanda para cada tipo de señal, cromóforo y grupo.
- Tabla 2. Rendimiento alcanzado por los diferentes modelos de clasificación entrenados con el subconjunto de características seleccionadas mediante el filtro basado en el estadístico *F* para la SS, CS y SS+CS.
- Tabla 3. Tabla 2. Rendimiento alcanzado por los diferentes modelos de clasificación entrenados con el subconjunto de mejores características seleccionadas por el *wrapper* para SS + CS. La significación estadística se indica mediante valores p y los IC del 95 % se representan entre corchetes.



RESUMEN

El trastorno por déficit de atención e hiperactividad (TDAH) está reconocido como un trastorno del neurodesarrollo de alta prevalencia en niños en edad escolar, que a menudo persiste en la adolescencia y la edad adulta, y que suele solaparse con otras comorbilidades psiquiátricas. En la práctica diaria, el diagnóstico se basa en signos y síntomas clínicos que requieren una evaluación detallada por parte de profesionales expertos mediante entrevistas con los padres/cuidadores y, si procede, con el propio paciente. La evaluación clínica profunda requiere experiencia y consume bastante tiempo, lo que, dada la escasez de profesionales entrenados, conduce a frecuentes retrasos en el diagnóstico, a pasar por alto algunos casos o incluso a decisiones clínicas erróneas. Desde el punto de vista del desarrollo, el diagnóstico precoz es importante para realizar intervenciones farmacológicas y psicosociales más eficaces. Entre otros retos, encontrar biomarcadores objetivos de TDAH sería esencial para apoyar la toma de decisiones clínicas, aliviando la carga de trabajo que suponen los diagnósticos y el seguimiento de los tratamientos. Con este fin, diferentes modalidades tradicionales de neuroimagen combinadas con técnicas de machine-learning (ML) y deep-learning (DL) han sido propuestas para la evaluación objetiva del TDAH. A pesar de su potencial, la variabilidad en los resultados, escasa disponibilidad y los costes asociados suponen una barrera importante para el uso regular de estas herramientas diagnósticas.

Recientemente, la espectroscopia funcional de infrarrojo cercano (fNIRS) ha emergido como prometedora técnica de neuroimagen para examinar la función cerebral de forma no invasiva y segura, pudiendo aplicarse fácilmente en diferentes entornos clínicos gracias a su portabilidad y coste asumible. En comparación con otras tecnologías, pocos estudios han usado fNIRS para discriminar entre niños con TDAH y niños controles con desarrollo típico (DT) a nivel individual. El objetivo principal de este trabajo ha sido desarrollar un marco metodológico para identificar patrones oscilatorios fNIRS característicos de la condición TDAH o DT.

En el presente estudio se ha utilizado un novedoso dispositivo fNIRS de onda continua y registro multidistancia, denominado *Theia*, para monitorizar la actividad hemodinámica superficial y profunda en 15 niños TDAH y 15 DT durante la ejecución de una tarea de aritmética mental rítmica.

Utilizando métodos de descomposición espectral tiempo-frecuencia se calcularon medidas de sincronización para localizar patrones oscilatorios, específicos de frecuencia, máximamente representativos del grupo TDAH o DT. Se usaron las distancias euclídeas a dichos patrones como características para generar cuatro modelos predictivos lineales comúnmente utilizados en ML para clasificación binaria supervisada: *support vector machine* (SVM), regresión logística (RL), análisis discriminante lineal (ADL) y *naïve* Bayes (NB). Para elegir solo el subconjunto de características más relevantes se combinó el filtrado previo de características por su estadístico F con un algoritmo de selección secuencial flotante (*Sequential forward floating selection wrapper*). El rendimiento de los modelos predictivos se evaluó mediante validación cruzada (VC) con 5 particiones (*K-5 folds*) y VC dejando uno fuera (*Leave-one-out*). La significación estadística se estimó mediante procedimientos de remuestreo no paramétricos.

Las métricas de precisión, sensibilidad y especificidad obtenidas con los modelos RL y ADL fueron cercanas al 100 % (p < 0,001) para ambos esquemas de VC cuando se entrenaron con solo tres características clave derivadas de componentes oscilatorios superficiales y profundos de muy baja frecuencia.

Proporcionamos evidencia preliminar de que las oscilaciones fNIRS de muy baja frecuencia, inducidas/moduladas por una tarea mental rítmica, diferencian con precisión a los niños con TDAH de los controles TD. Puesto que los resultados predictivos obtenidos superan a otros estudios similares, el enfoque aquí propuesto resulta prometedor y técnicamente viable para la identificación de biomarcadores funcionales lo suficientemente confiables e interpretables como para informar la práctica clínica diaria.

ABSTRACT

Attention deficit hyperactivity disorder (ADHD) is recognized as a neurodevelopmental disorder with a high prevalence in school-aged children, often persisting into adolescence and adulthood, and often overlapping with other psychiatric comorbidities. In daily practice, diagnosis is based on clinical signs and symptoms that require detailed assessment by expert professionals through interviews with parents/caregivers and, if appropriate, with the patient him/herself. In-depth clinical evaluation requires expertise and is quite time-consuming, which, given the shortage of trained professionals, leads to frequent delays in diagnosis, overlooking some cases or even erroneous clinical decisions. From a developmental point of view, early diagnosis is important for more effective pharmacological and psychosocial interventions. Among other challenges, finding objective biomarkers of ADHD would be essential to support clinical decision making, alleviating the burden of diagnostics and treatment monitoring. To this end, different traditional neuroimaging modalities combined with machinelearning (ML) and deep-learning (DL) techniques have been proposed for the objective assessment of ADHD. Despite their potential, the variability of results, low availability and associated costs are a major barrier to the regular use of these diagnostic tools.

Recently, "functional near-infrared spectroscopy" (fNIRS) has emerged as a promising neuroimaging technique to examine brain function in a noninvasive and safe manner, and can be easily applied in different clinical settings due to its portability and affordable cost. Compared to other technologies, few studies have used fNIRS to discriminate between children with ADHD and typically developing (TD) controls at the individual level. The main objective of this work has been to develop a methodological framework to identify fNIRS oscillatory patterns characteristic of the ADHD or TD condition.

In the present study, a novel continuous wave, multi-distance recording fNIRS device, named Theia, was used to monitor surface and deep hemodynamic activity in 15 ADHD and 15 TD children during the performance of a rhythmic mental arithmetic task.

Using time-frequency spectral decomposition methods, synchronization measures were calculated to locate frequency-specific oscillatory patterns maximally representative of the ADHD or TD group. Euclidean distances to these patterns were used as features to generate four linear predictive models commonly used in ML for supervised binary classification: support vector machine (SVM), logistic regression (LR), linear discriminant analysis (LDA) and naïve Bayes (NB). To choose only the most relevant subset of features, prior filtering of features by their F-statistic was combined with a sequential forward floating selection wrapper algorithm. The performance of the predictive models was evaluated by cross-validation (CV) with 5 partitions (K-5 folds) and leave-one-out. Statistical significance was estimated by nonparametric resampling procedures.

The accuracy, sensitivity and specificity metrics obtained with the RL and ADL models were close to 100 % (p < 0.001) for both CV schemes when trained with only three key features derived from very low frequency shallow and deep oscillatory components.

We provide preliminary evidence that very low frequency fNIRS oscillations induced/modulated by a rhythmic mental task accurately differentiate ADHD children from TD controls. Since the predictive results obtained outperform other similar studies, the approach proposed here holds promising and technically feasible for the identification of functional biomarkers sufficiently reliable and interpretable to inform daily clinical practice.



1. INTRODUCCIÓN

El Trastorno por déficit de atención e hiperactividad/impulsividad (TDAH) es un síndrome heterogéneo, de carácter politético, en tanto en cuanto su expresión depende de distintas variables latentes que se pueden combinar de diferente forma e intensidad. La dificultad para su diagnóstico y la falta de pruebas objetivas podría estar provocando la inclusión de menores en esta categoría sin cumplir estrictamente los criterios del cuadro clínico. En el caso que nos ocupa, un diagnóstico certero cobra una relevancia de especial calibre por cuanto se derivan importantes consecuencias, tanto a nivel individual como psicosocial. En primer término, porque estos niños son sometidos a terapia farmacológica y, en segundo lugar, por sus problemas para el aprendizaje y adaptación social. Y es que el TDAH es una disfunción cognitivo-comportamental que repercute en la calidad de vida de estos niños, especialmente en su esfera psicológica y social. Comparado con los de desarrollo típico (DT), los niños TDAH muestran peor autonomía, desregulación emocional y estados de ánimo negativos, mayores dificultades en el aprendizaje y fracaso escolar, además de peor aceptación social (López-Villalobos et al., 2018). Igualmente, sus relaciones socioafectivas se encuentran afectadas, puesto que sus conductas disruptivas e intrusivas provocan un alto porcentaje de rechazo entre sus iguales. En el ámbito familiar, también aparecen consecuencias negativas al generarse un ambiente estresante y sentimientos de falta de competencia en los padres (Castellar et al., 2006; Presentación-Herrero et al., 2006).

1.1. Concepto de TDAH

La agitación motora, la impulsividad y los problemas de atención como categoría clínica ha sido objeto de interés por la medicina desde antaño con autores como Alexander Crichton (1763-1856) y George F. Still (1868-1941) (Ramos Platón, 1987; Taylor, 2011).

El antecedente directo del TDAH como entidad nosológica apareció en la segunda edición del *Manual Diagnóstico y Estadístico de los trastornos mentales* (DSM, acrónimo del original inglés *Diagnostic and Statistical Manual of Mental Disorders*) (1968) con la etiqueta reacción hipercinética de la infancia, introduciéndose oficialmente en la taxonomía con la denominación trastorno por déficit de atención/hiperactividad en el DSM-III (1980). Sin embargo, la conformación semiológica de dicho trastorno ha seguido experimentando cambios a lo largo de las ediciones de las distintas clasificaciones médicas. Actualmente existen dos sistemas de clasificación de los

trastornos mentales que tienen una relevancia capital en el diagnóstico de TDAH. Por un lado, el DSM editado por la Asociación Americana de Psicología (APA) y la *Clasificación Internacional de las Enfermedades* (CIE) establecida por la Organización Mundial de Salud (OMS). A continuación, revisaremos los diferentes criterios diagnósticos en cada uno de estos sistemas de clasificación.

1.1.1. El TDAH en el DSM-5

En la quinta edición del DSM el apartado *Trastornos de inicio en la infancia, niñez o adolescencia* pasa a denominarse *Trastornos del neurodesarrollo*, con lo que se está reconociendo una base orgánica de etiología neuronal en estos trastornos, es decir, problemas en la madurez del sistema nervioso.

Los criterios diagnósticos para el TDAH se presentan de la siguiente forma (APA, 2014):

- A. Patrón persistente de inatención y/o hiperactividad-impulsividad que interfiere con el funcionamiento o desarrollo, que se caracteriza por (1) y/o (2):
 - Inatención: Seis (o más) de los siguientes síntomas se han mantenido durante al menos 6 meses en un grado que no concuerda con el nivel de desarrollo y que afecta directamente las actividades sociales y académicas/laborales:

Nota: Los síntomas no son sólo una manifestación del comportamiento de oposición, desafío, hostilidad o fracaso en la comprensión de tareas o instrucciones. Para adolescentes mayores y adultos (a partir de los 17 años), se requiere un mínimo de cinco síntomas.

- a. Con frecuencia falla en prestar la debida atención a detalles o por descuido se comenten errores en las tareas escolares, en el trabajo o durante otras actividades (p. ej., se pasan por alto o se pierden detalles, el trabajo no se lleva a cabo con precisión).
- b. Con frecuencia tiene dificultades para mantener la atención en tareas o actividades recreativas (p. ej., tiene dificultad para mantener la atención en clases, conversaciones o la lectura prolongada).
- c. Con frecuencia parece no escucha cuando se le habla directamente (p. ej., parece tener la mente en otras cosas, incluso en ausencia de cualquier distracción aparente).
- d. Con frecuencia no sigue las instrucciones y no termina las tareas escolares, los quehaceres o los deberes laborales (p. ej., inicia tareas pero se distrae rápidamente y se evade con facilidad).
- e. Con frecuencia tiene dificultad para organizar tareas y actividades (p. ej., dificultad para gestionar tareas secuenciales, dificultad para poner los materiales y pertenencias en orden, descuido y desorganización en el trabajo, mala gestión del tiempo; no cumple los plazos).

- f. Con frecuencia evita, le disgusta o se muestra poco entusiasta en iniciar tareas que requieren un esfuerzo mental sostenido (p. ej., tareas escolares o quehaceres domésticos; en adolescentes mayores y adultos, preparación de informes, completar formularios, revisar artículos largos).
- g. Con frecuencia pierde cosas necesarias para tareas o actividades (p. ej., materiales escolares, lápices, libros, instrumentos, billetero, llaves, papeles del trabajo, gafas, móvil).
- h. Con frecuencia se distrae con facilidad por estímulos externos (para adolescentes mayores y adultos, puede incluir pensamientos no relacionados).
- Con frecuencia olvida las actividades cotidianas (p. ej., hacer las tareas, hacer las diligencias; en adolescentes mayores y adultos, devolver las llamadas, pagar las facturas, acudir a citas).
- Hiperactividad e impulsividad: Seis (o más) de los siguientes síntomas se han mantenido durante, al menos, 6 meses en un grado que no concuerda con el nivel de desarrollo y que afecta directamente a las actividades sociales y académicas/laborales:

Nota: Los síntomas no son sólo una manifestación del comportamiento de oposición, desafío, hostilidad o fracaso para comprender tareas o instrucciones. Para adolescentes mayores y adultos (a partir de 17 años de edad), se requiere un mínimo de cinco síntomas.

- a. Con frecuencia juguetea con o golpea las manos o los pies o se retuerce en el asiento.
- b. Con frecuencia se levanta en situaciones en que se espera que permanezca sentado (p. ej., se levanta en la clase, en la oficina o en otro lugar de trabajo, o en otras situaciones que requieren mantenerse en su lugar).
- c. Con frecuencia corretea o trepa en situaciones en las que no resulta apropiado. (Nota: En adolescentes o adultos, puede limitarse a estar inquieto.)
- d. Con frecuencia es incapaz de jugar o de ocuparse tranquilamente en actividades recreativas.
- e. Con frecuencia está "ocupado", "actuando como si lo impulsara un motor" (p. ej., es incapaz de estar o se siente incómodo estando quieto durante un tiempo prolongado, como en restaurantes, reuniones; los otros pueden pensar que está intranquilo o que le resulta difícil seguirlos).
- f. Con frecuencia habla excesivamente.
- g. Con frecuencia responde inesperadamente o antes de que se haya concluido una pregunta (p. ej., termina las frases de otros, no respeta el turno de conversación).
- h. Con frecuencia le es difícil esperar su turno (p. ej., mientras espera en la cola).
- i. Con frecuencia interrumpe o se inmiscuye con otros (p. ej., se mete en las conversaciones, juegos o actividades, puede empezar a utilizar las cosas de otras

personas sin esperar o recibir permiso; en adolescentes y adultos, puede inmiscuirse o adelantarse a lo que hacen otros).

- B. Algunos síntomas de inatención o hiperactivo-impulsivos estaban presentes antes de los 12 años.
- C. Varios síntomas de inatención o hiperactivo-impulsivos están presentes en dos o más contextos (p. ej., en casa, en la escuela o en el trabajo, con los amigos o parientes, en otras actividades).
- D. Existen pruebas claras de que los síntomas interfieren con el funcionamiento social, académico o laboral, o reducen la calidad de los mismos.
- E. Los síntomas no se producen exclusivamente durante el curso de la esquizofrenia o de otro trastorno psicótico y no se explican mejor por otro trastorno mental (p. ej., trastorno del estado de ánimo, trastorno de ansiedad, trastorno disociativo, trastorno de la personalidad, intoxicación o abstinencia de sustancias).

Una novedad importante de esta edición es que se suprime el término subtipo y se establece, según los resultados de los ítems, una clasificación por la presentación de los síntomas, que puede ser: presentación combinada, presentación predominante con falta de atención y presentación predominante hiperactiva/impulsiva.

Otra novedad es la introducción del diagnóstico para mayores de 16 años, en cuyo caso se reduce de seis a cinco síntomas —tanto en inatención como en hiperactividad— para el diagnóstico. También se cambia la edad de inicio de los síntomas pasando de los siete años a los doce. Por otra parte, se permite el diagnóstico de TDAH y autismo al mismo tiempo. Se recalca la necesidad de que la presentación de síntomas se produzca en más de un contexto y de que interfiera en el funcionamiento social, académico o social.

En esta edición se reconoce la problemática en los procesos cognitivos de base relacionados con la falta de atención, así como la existencia de ciertos marcadores biológicos (aumento de ondas lentas, menor volumen cerebral total) que pueden diferenciar a la población TDAH, sin embargo, son descartados como herramientas diagnósticas.

1.1.2. El TDAH en la CIE-11

La undécima revisión de la CIE de la OMS reserva el capítulo 6 para los *Trastornos mentales, del comportamiento y del neurodesarrollo*, y dentro del apartado de los *Trastornos del neurodesarrollo* se cataloga el TDAH con los siguientes requisitos para el diagnóstico de la siguiente forma *ad litteram* (OMS, 2019):

Características esenciales (obligatorias):

Un patrón persistente (p. ej., al menos 6 meses) de síntomas de falta de atención y/o una combinación de síntomas de hiperactividad e impulsividad que está fuera de los límites de la variación normal esperada para la edad y el nivel de desarrollo intelectual. Los síntomas varían según la edad cronológica y la gravedad del trastorno.

Inatención

- Varios síntomas de falta de atención que son persistentes y lo suficientemente graves como para tener un impacto negativo directo en el funcionamiento académico, laboral o social. Los síntomas suelen ser de los siguientes grupos:
 - Dificultad para mantener la atención en tareas que no proporcionan un alto nivel de estimulación o recompensa o que requieren un esfuerzo mental sostenido; falta de atención al detalle; cometer errores por descuido en las tareas escolares o laborales; no completar las tareas.
 - Se distrae fácilmente con estímulos extraños o pensamientos no relacionados con la tarea en cuestión; a menudo parece no escuchar cuando se le habla directamente; frecuentemente parece estar soñando despierto o tener la mente en otra parte.
 - Pierde cosas; es olvidadizo en las actividades diarias; tiene dificultad para recordar completar las próximas tareas o actividades diarias; dificultad para planificar, administrar y organizar el trabajo escolar, las tareas y otras actividades.

Nota: La falta de atención puede no ser evidente cuando el individuo se dedica actividades que proporcionan estimulación intensa y recompensas frecuentes.

Hiperactividad impulsividad

- Varios síntomas de hiperactividad/impulsividad que son persistentes y lo suficientemente graves como para tener un impacto negativo directo en el funcionamiento académico, laboral o social. Estos tienden a ser más evidentes en situaciones estructuradas que requieren autocontrol conductual. Los síntomas suelen ser de los siguientes grupos:
 - Actividad motora excesiva; deja el asiento cuando se espera que se quede quieto; a menudo corre; tiene dificultad para sentarse quieto sin moverse (niños más pequeños); sensación de inquietud física, sensación de incomodidad al estar quieto o sentado (adolescentes y adultos).
 - o Dificultad para participar en actividades en silencio; habla demasiado.
 - Suelta respuestas en la escuela, comentarios en el trabajo; dificultad para esperar su turno en conversaciones, juegos o actividades; interrumpe o se entromete en otras conversaciones o juegos.

- Una tendencia a actuar en respuesta a estímulos inmediatos sin deliberación o consideración de riesgos y consecuencias (p. ej., involucrarse en comportamientos con potencial de daño físico, decisiones impulsivas, conducción imprudente)
- Evidencia de síntomas significativos de falta de atención y/o hiperactividadimpulsividad antes de los 12 años, aunque algunas personas pueden buscar atención médica por primera vez más tarde en la adolescencia o en la edad adulta, a menudo cuando las demandas superan la capacidad del individuo para compensar las limitaciones.
- Las manifestaciones de falta de atención y/o hiperactividad-impulsividad deben ser evidentes en múltiples situaciones o entornos (p. ej., el hogar, la escuela, el trabajo, con amigos o familiares), pero es probable que varíen según la estructura y las demandas del entorno.
- Los síntomas no se explican mejor por otro trastorno mental (p. ej., un trastorno relacionado con la ansiedad o el miedo, un trastorno neurocognitivo como el delirio).
- Los síntomas no se deben a los efectos de una sustancia (p. ej., cocaína) o medicamento (p. ej., broncodilatadores, medicamentos de reemplazo de la tiroides) en el sistema nervioso central, incluidos los efectos de abstinencia, y no se deben a una enfermedad del sistema nervioso.

Como se puede observar, no se fija un número determinado de ítems para el diagnóstico, ya que el trastorno se puede manifestar de forma variable tal y como se advierte en la descripción. Así, se enfatiza la importancia del impacto de los síntomas en el funcionamiento de la persona y no solo la mera presencia, lo que permite al profesional sanitario establecer un diagnóstico por el valor clínico. Es decir, se pueden tener más síntomas y no presentar un cuadro clínico, mientras que se pueden manifestar menos síntomas con una gravedad disfuncional.

En esta edición, se ha hecho un esfuerzo por armonizar la categorización del TDAH con el DSM-5. Comparten la estructura bidimensional del trastorno y las distintas formas bajo el término "presentación", debiendo manifestarse en más de un ambiente y repercutir en la esfera psicosocial del sujeto. Si bien, hay alguna diferencia como tener en cuenta las conductas de riesgo dentro de la impulsividad.

1.2. Epidemiología

Cuantificar con exactitud el número de personas con TDAH en una población determinada es un aspecto crucial para la planificación de los recursos sanitarios. Dicha estimación permite determinar los costes asociados a este trastorno facilitando la gestión eficiente de los recursos económicos. Además, el conocimiento de la variación de los datos epidemiológicos entre países y su evolución a lo largo del tiempo proporciona información valiosa sobre la validez actual del diagnóstico de TDAH (Posner, et al., 2020). A continuación, revisaremos algunos de los trabajos más relevantes sobre la epidemiologia del TDAH tanto a nivel internacional como nacional.

En el contexto internacional, encontramos el trabajo de Skounti et al. (2007) quienes realizaron una revisión de 39 artículos que incluían indicadores epidemiológicos del TDAH en diferentes países. Estos autores señalaron la existencia de una prevalencia heterogénea y variable en función del país considerado. La prevalencia más baja fue la hallada en Inglaterra que se sitúa alrededor del 2,2%, mientras que la mayor se encontró en Alemania con un 17,8%. Algunas variables que podrían explicar estas diferencias son: las diferencias en el informante, factores culturales, el instrumento utilizado para cuantificar los síntomas, la utilización de los criterios DSM o CIE, o la variación de dichos criterios a lo largo del tiempo.

Por otro lado, Polancyk et al. (2007) realizaron una revisión de trabajos epidemiológicos de diversos países publicados entre 1978 y 2005. Después de someter los datos a un análisis meta-regresivo encontraron que los criterios diagnósticos utilizados, el evaluador, el origen geográfico y la utilización del criterio de funcionalidad para establecer el diagnóstico fueron los principales factores responsables de la heterogeneidad en la prevalencia. Los autores destacan que los estudios cuya metodología y criterio diagnóstico son semejantes informan de similares prevalencias (por ejemplo, Rusia y Gran Bretaña, 1,3% y 1,4%, respectivamente). Sin embargo, cuando el criterio diagnóstico difiere (presencia o no de disfuncionalidad) aparecen diferencias ostensibles con variaciones entre 3,7% y el 8,9% en un mismo país.

Otro trabajo interesante es el realizado por Döpfner et al. (2008), donde se compara la variación diagnostica utilizando DSM-IV o CIE-10. Para una misma muestra la prevalencia con los criterios DSM fue del 5% frente al 1% de la CIE-10. Además, estos autores ajustaron las prevalencias en función del cumplimiento de los otros criterios establecidos, y no exclusivamente con los síntomas como se ha procedido generalmente. Así, vemos que para el DSM se reduce la prevalencia al 4,1% si se añade el deterioro significativo en el funcionamiento. Cuando se incluye la presencia en sintomatología en dos o más ambientes la prevalencia desciende a un 3,9%, mientras que si además se utiliza al criterio de mostrar síntomas antes de los 7 años se reduce al 2,5%. Por último, cuando se aplica el criterio de una duración superior a seis meses obtenemos una prevalencia del 2,2%. La misma tendencia se observa cuando se usa la CIE-10, reduciéndose la prevalencia hasta el 0,6% si se incorporan la totalidad de los criterios diagnósticos.

En el 2012, Willcutt (2012) realizó un metaanálisis calculando las diferentes prevalencias en función de si el paciente presentaba 6 ítems de los establecidos en el DSM-IV o si mostraba la totalidad de criterios diagnósticos. Mediante esta metodología encontró que la estabilidad temporal para un mismo observador era adecuada pero que la concordancia entre evaluadores era moderada o baja, llegando a diferir en el diagnóstico del subtipo de TDAH en un mismo participante.

Centrándonos en las diferencias según el criterio diagnóstico, Guardiola et al. (2000) compararon la prevalencia utilizando el rendimiento neuropsicológico con los criterios del DSM-IV cuantificados por profesores. Las diferencias en la prevalencia entre ambas aproximaciones fueron considerables: 18% según el DSM-IV, en comparación con un 3,5% cuando se aplicaba también herramientas de evaluación neuropsicológica. Es interesante recalcar que no existieron diferencias en función del sexo utilizando los criterios neuropsicológicos mientras que la variable sexo sí que fue relevante cuando se emplearon los criterios DSM-IV (25,1% vs 11%). Los autores achacan estos resultados a que los criterios propuestos por la APA son útiles exclusivamente como herramientas de cribado debido a que su valor predictivo es bajo.

Por lo que respecta a los datos epidemiológicos del TDAH en España, el metaanálisis realizado por Catalá-López et al. (2012) encontró una prevalencia corregida estimada del 5,3%. En el estudio realizado por Navarro-Pardo et al., (2012) en población clínica infanto-juvenil se obtuvo una prevalencia del 10,9%, siendo la tercera categoría diagnóstica de mayor prevalencia, detrás de los trastornos de conducta y de los trastornos de ansiedad. Si nos centramos en los datos por edades, destaca la gran diferencia en el grupo de edad entre 6-11 años, donde se dispara la prevalencia a un 19,57%, en comparación con el resto de los grupos etarios: menores de 5 años (5,56%), 12-15 años (8,43%) y mayores de 15 años (5,88%). En función del género, los varones superan ampliamente a las mujeres, con un 14,6% frente a un 4,5%, respectivamente. Además, diversos trabajos muestran que la prevalencia varia ostensiblemente en función de la comunidad autónoma. Por ejemplo, la prevalencia en Castilla y León se sitúa alrededor del 6,66% (Rodríguez Molinero et al., 2009), mientras que en la comunidad Valenciana encontramos prevalencias del 8% (Andrés et al., 1999). Asi mismo encontramos una prevalencia del 4,57 y del 4,9% en la isla de Mallorca y en la Comunidad Autónoma de Canarias, respectivamente (Jiménez et al., 2012; Cardo y Servera, 2005)

En conjunto, tanto en España como en otros países encontramos una gran variabilidad de los datos epidemiológicos. Las diferencias entre criterios diagnósticos, fuentes de información, empleo de pruebas psicométricas y/o evaluación neuropsicológica para la cuantificación de los síntomas son factores metodológicos que explican la inconsistencia

de los datos. En realidad, esta situación refleja que el diagnóstico y evaluación del TDAH es un proceso complejo que constituye un reto en la actualidad.

1.3. Evaluación y diagnóstico del TDAH

1.3.1. Procedimientos diagnósticos: consenso clínico

En la práctica diaria, el diagnóstico de TDAH se basa en signos y síntomas clínicos que requieren una evaluación detallada por parte de profesionales expertos mediante diferentes métodos involucrando a los padres, profesores y, si procede, con el propio paciente. El procedimiento para el diagnóstico del TDAH incluye la elaboración de una historia clínica mediante una entrevista, las escalas de valoración psicométricas, la observación conductual directa, pruebas neuropsicológicas, e incluso el análisis comparativo de distintos fármacos. Si bien, en la práctica clínica habitual la evaluación se reduce a la historia clínica y a la aplicación de escalas psicométricas (Gualtieri y Johnson, 2005). Obtener información fiable por medio de estos procedimientos requieren de mucha experiencia clínica y grandes conocimientos de psicopatología infantil (Crippa et al., 2017).

La entrevista para la elaboración de una historia clínica se basa en la obtención de información sobre los signos y síntomas que el paciente muestra en diferentes contextos. Por este motivo es habitual que la entrevista incluya al niño, a los padres y, en lo posible, a otras personas relevantes como los profesores. La parte más relevante en este sentido es determinar la presencia del TDAH a través del diagnóstico clínico, que en estos casos está fundamentado, principalmente, en los criterios establecidos por los diferentes sistemas diagnósticos (Quinlan, 2003) prestando especial atención a la afectación emocional, familiar, académica y social. Valorar la sintomatología mediante la entrevista es complejo y requiere de mucha experiencia ya que tanto la CIE como el DSM establecen criterios diagnósticos conductuales sin concretar qué medios deben ser empleados para determinar su cumplimiento. Con la idea de objetivar el proceso de recogida de datos y la cuantificación de los síntomas suelen utilizarse las escalas psicométricas.

Las escalas psicométricas están compuestas por una serie de afirmaciones que representan el constructo de TDAH y que el interlocutor debe de valorar en función de su experiencia subjetiva. De esta forma se puede cuantificar numéricamente la sintomatología del TDAH. La utilización de un punto de corte nos permite, además, determinar si las puntuaciones en una determinada escala son compatibles con el diagnóstico clínico de TDAH. Entre las escalas psicométricas más importantes para la evaluación clínica del TDAH podemos resaltar:

-Las Escalas de Conners (Conners et al. 2011) son probablemente las pruebas psicométricas más utilizadas para evaluar el TDAH. Las escalas tienen diferentes versiones (padres, profesores y autoinforme) y proporcionan un índice específico TDAH y de psicopatología general, así como diferentes subescalas que aportan información sobre el funcionamiento ejecutivo, las relaciones sociales y la conducta agresiva. La primera edición fue adaptada para población española por Farré y Narbona (1989).

-La Escala de evaluación del trastorno por déficit de atención/hiperactividad se encuentra compuesta por 20 ítems y distribuidos en tres subescalas: hiperactividad (5 ítems), déficit de atención (5 ítems) y trastornos de conducta (10 ítems). Esta prueba psicométrica se elaboró para población española a partir de la Escala Conners Escolar porque presentaba mejores cualidades psicométricas (Farré-Riba y Narbona, 1997).

-La Escala *Attention–Deficit/Hyperactivity Disorder Rating Scales IV* constituye la prueba psicométrica por antonomasia para la valoración del TDAH debido a que se basa en los 18 criterios diagnósticos de DSM-IV. Tiene una versión para padres y otra para profesores. Servera y Cardo (2007) normalizaron dicha escala en población española. Encontraron diferencias con la población americana, debido principalmente a causas socioculturales.

Con independencia del método utilizado para recabar la información, lo cierto es que la evaluación clínica del TDAH y el establecimiento de un diagnóstico constituye un reto. El consenso clínico actual apoya la validez del diagnóstico cuando la evaluación se realiza por profesionales especializados (Faraone et al., 2021; Adesman, 2001). Por tanto, un diagnóstico fiable y válido requiere de gran pericia profesional y consume bastante tiempo, lo que, dada la escasez de profesionales entrenados, conduce a frecuentes retrasos en el diagnóstico y/o al sobrediagnóstico. Las altas prevalencias del TDAH de algunos estudios y el aumento de estas mismas en los últimos años hasta cifras casi epidémicas, hace pensar a algunos autores en el sobrediagnóstico de dicho trastorno (Narbona, 2001; García Peñas & Domínguez Carral, 2012; Gimeno Morales y Galbe Sánchez-Ventura, 2015). Efectivamente, Kazda et al. (2021) han comprobado que existe un

sobrediagnóstico en el TDAH, debido principalmente a la adición de casos con síntomas más leves. Téngase en cuenta que un sobrediagnóstico tiene unas consecuencias nocivas en todos los aspectos: Físicas, por el tratamiento farmacológico; Psicológicas, por el estigma que provoca la etiqueta de la enfermedad; Sociales, por el gasto sanitario innecesario (García Peñas y Domínguez Carral, 2012). Asimismo, el retraso en el diagnostico puede ser igualmente nocivo y llevar a decisiones clínicas erróneas. Desde el punto de vista del desarrollo, el diagnóstico precoz es vital para realizar intervenciones farmacológicas y/o psicosociales eficaces evitando las consecuencias a largo plazo de este trastorno.

1.3.2. Controversia sobre la evaluación del TDAH

Aunque la utilización de las pruebas psicométricas supone un avance en la cuantificación de la sintomatología del TDAH lo cierto es que actualmente su uso es controvertido por varios motivos: 1) La estructura factorial de las pruebas psicométricas no parece ser estable y varía en función de las poblaciones de estudio planteado dudas sobre la validez de constructo (García Garrido et al., 2014; Casanova et al., 2003). 2) Al proporcionar una puntuación que cuantifica la sintomatología del TDAH pueden generar una falsa sensación de objetividad en el profesional que las utiliza. 3) Existen variaciones en las puntuaciones referidas a la sintomatología en función de que persona responda al cuestionario (padre, madre, profesor o paciente), lo que pone de manifiesto la subjetividad de este tipo de pruebas (Gualtieri y Johnson, 2005; Abikoff et al., 1993). 4) La fiabilidad y la validez de este método de obtención de información depende en gran medida del grado de adiestramiento del informante para evaluar con precisión los síntomas del TDAH. Las investigaciones previas han mostrado que la experiencia de los padres en la cuantificación de los síntomas de inatención es especialmente problemática (Muñoz y García, 2008). 5) A menudo se advierte que las escalas de calificación de conductas que valoran el TDAH deberían de ser utilizadas como un elemento de cribado, pero en la práctica diaria se emplean como herramienta diagnóstica y no se tienen en cuenta otros elementos de juicio que sobre la disfunción del paciente (Fernández-Jaén et al., 2016). 6) El establecimiento de un punto de corte a partir del cual las conductas se consideran patológicas es también un elemento problemático. El punto de corte se establece de forma arbitraria en función de la intensidad de una constelación de síntomas que no tienen por qué ser patológicos per se (Reid y Maag, 1994). 7) Las pruebas de calificación de conductas no tienen en cuenta la afectación en la vida del individuo en términos emocionales, familiares, académicos o sociales (Fernández-Jaén et al., 2016). 8) Existen grandes diferencias en el poder predictivo de los ítems que conforman las escalas (Cardo et al., 2007; Cardo et al., 2009), por lo que sería necesario debatir si se debiera abandonar el modelo politético del TDAH por uno jerárquico. 9) Existe una concordancia limitada entre las diferentes escalas de calificación. Por ejemplo, Cardo y Serera (2007) realizaron un estudio utilizando cuatro instrumentos diferentes para evaluar la sintomatología TDAH (según criterios diagnósticos DSM-IV y CIE-10, y según las escalas *ADHD Rating Scale IV* y *SNAP-IV*, ambas basadas en la DSM-IV) encontrando que únicamente un 0,8% de los participantes era identificado como TDAH en las cuatro pruebas al mismo tiempo.

En virtud de lo expuesto hasta el momento, se puede considerar que la utilización de las pruebas psicométricas como herramientas diagnosticas puede ser problemático. Por tanto, se hace necesario buscar y desarrollar pruebas objetivas complementarias que permitan la identificación precisa de este tipo de población infanto-juvenil, para lo cual los marcadores biológicos se presentan como un procedimiento ineludible mediante la aplicación de instrumentos tecnológicos como la neuroimagen.

1.3.3. Otras herramientas diagnósticas

Numerosos estudios han intentado evaluar el TDAH mediante diferentes herramientas diagnósticas objetivas, la mayoría utilizando imágenes de resonancia magnética funcional (fMRI) o estructural (MRI) y electroencefalografía (EEG), junto a otras modalidades (magnetoencefalografía, electrocardiografía, actigrafía, etc.) utilizadas con menos frecuencia; para una revisión, véase Pereira-Sánchez y Castellanos (2021) y Loh et al., (2022). En el marco de las iniciativas del "Consorcio ADHD-200" (2012) se han realizado notables esfuerzos en el campo de la MRI y la fMRI. A pesar de los avances significativos en la comprensión de las anomalías relacionadas con la maduración y la función cerebral, los hallazgos de neuroimagen en la investigación del TDAH aún no pueden utilizarse para apoyar la práctica clínica debido a diferentes problemas (Posner et al., 2020; Pereira-Sánchez y Castellanos, 2021).

Una alternativa prometedora, de creciente interés, es la combinación de las citadas herramientas con técnicas de inteligencia artificial (IA), aprendizaje automático (*machine-learning*) y profundo (*deep-learning*), para mejorar la evaluación objetiva del TDAH. Del mismo modo, aunque prometedores, los estudios dedicados a la predicción del TDAH en un único sujeto mediante técnicas de IA han arrojado resultados muy variables (Loh et al., 2022; Arbabshirani et al., 2017), lo que plantea ciertas dudas metodológicas (Pulini et al., 2019). En la práctica real, incluso si se dispusiera de indicadores de neuroimagen útiles para apoyar las decisiones clínicas, la disponibilidad de estas herramientas de diagnóstico y sus costes asociados representarían una barrera importante para su uso regular, aumentando aún más la carga sobre los recursos sanitarios.

Una herramienta alternativa para evaluar el TDAH que merece la pena explorar es la espectroscopia funcional del infrarrojo cercano (fNIRS), caracterizada por ser no invasiva, portátil, rentable y aplicable en entornos más amigables/ecológicos (para una revisión, véase Mauri et al., 2018). fNIRS ha demostrado su utilidad en la monitorización de los cambios hemodinámicos funcionales asociados a la activación cerebral cortical (para una revisión, véase Pinti et al., 2020). En comparación con otras modalidades de neuroimagen, se han realizado pocos estudios de fNIRS para discriminar a los niños con TDAH de los controles sin TDAH, algunos de ellos intentando mejorar la clasificación combinando diferentes modalidades (por ejemplo, EEG + fNIRS) como en Güven et al., (2020). Aún menos estudios se han centrado en enfoques unimodales utilizando "exclusivamente" datos fNIRS (Monden et al., 2015; Crippa et al., 2017; Gu et al., 2017).

1.4. Pregunta de investigación

¿Es factible utilizar biomarcadores funcionales basados en neuroimagen óptica junto con herramientas computacionales de apoyo para diferenciar/clasificar niños con TDAH de aquellos con desarrollo típico?

1.5. La espectroscopia funcional de infrarrojo cercano (fNIRS)

Existen dos hitos relevantes en el siglo XX que constituyen el inicio de la espectroscopia de luz cercana al infrarrojo (NIRS por sus siglas en inglés) (véase la revisión histórica de Ferrari y Quaresima (2012)). El primero de ellos corresponde al fisiólogo inglés Glenn Allen Millikan quien desarrolló un dispositivo denominado oxímetro para medir la saturación de oxígeno sanguínea en los pilotos de la Royal Air Force. En segundo lugar, Frans Jöbsis, en 1977, demostró que era posible detectar mediante transiluminación infrarroja los cambios en la concentración de hemoglobina cortical en gatos anestesiados. Desde entonces y gracias al desarrollo tecnológico, NIRS ha sido empleado en diversas disciplinas tanto básicas como aplicadas (estudio de procesos cognitivos como memoria, lenguaje y atención, monitorización en accidentes

cerebrovasculares en fase subaguda, demencias, epilepsia, desórdenes del estado de ánimo, trastornos de ansiedad, etc.).

1.5.1. Principios operacionales fNIRS

La espectroscopia funcional de infrarrojo cercano (fNIRS) es una técnica de neuroimagen óptica no invasiva que mide cambios relativos en la concentración de oxihemoglobina (HbO) y desoxihemoglobina (HbR) cerebral por medio de luz en el espectro del infrarrojo cercano (750-1400 nm) (Figura 1) (Scholkmann *et al.*, 2014; Obrig, 2014).



Figura 1. Representación del espectro electromagnético por longitud de onda.

El fundamento biológico comúnmente aceptado de la tecnología fNIRS es el acoplamiento neurovascular (o hiperemia funcional), que describe el estrecho vínculo temporal y regional entre la actividad neuronal y las respuestas del flujo sanguíneo cerebral. El incremento en la actividad neuronal induce un aumento del flujo sanguíneo para satisfacer la demanda metabólica. Dicho incremento en el flujo sanguíneo se traduce en cambios en la concentración de HbO y HbR que son detectados por la tecnología NIRS. Aunque los mecanismos celulares no se conocen con precisión, el acoplamiento neurovascular está regulado por mecanismos de señalización entre los componentes de la denominada unidad neurovascular (Schaeffer y Iadecola, 2021) y es la base no solo de la tecnología fNIRS sino también de la resonancia magnética funcional (fMRI), la tomografía por emisión de positrones (PET) y la imagen óptica difusa (DOI) (Cauli y Hamel, 2010; Hosford y Gourine, 2019).

Entre todas las técnicas de neuroimagen, fNIRS destaca por su bajo coste, seguridad y facilidad de aplicación en entornos clínicos debido a su portabilidad y comodidad (Scholkmann et al., 2014). Sin embargo, la utilización de dicha técnica requiere considerar una importante limitación: las señales fNIRS son el resultado de la suma de varias fuentes entre las que se encuentra la actividad cerebral pero también la interferencia hemodinámica sistémica (cardiacas, respiratorias, presión arterial, actividad vasomotora, ruido instrumental, etc) y la interferencia superficial del cuero cabelludo (Tachtsidis y Scholkmann, 2016). Por lo tanto, es incorrecto asumir que la tecnología fNIRS solo registra cambios vinculados a la actividad cerebral funcional (Minati et al., 2011).

En un montaje NIRS típico la luz cercana al infrarrojo viaja desde el emisor hasta el detector atravesando las diferentes capas del cuero cabelludo hasta alcanzar la corteza cerebral para luego traspasar de nuevo estos tejidos en su camino hasta el detector (Figura 2) (Gagnon et al., 2012). La actividad hemodinámica de estos tejidos no cerebrales (piel, musculo y cráneo) modifica la luz cercana al infrarrojo lo que puede generar no pocas confusiones en la interpretación de las señales NIRS (Tachtsidis y Scholkmann, 2016). De hecho, la tecnología fNIRS es del orden de 20 veces más sensible a la hemodinámica del cuero cabelludo que a la propia corteza cerebral (Haeussinger et al.,2011).



Figura 2. Representación esquemática de un montaje NIRS típico en el que se observa la luz cercana al infrarrojo (línea discontinua) viajando desde el emisor (E) hasta el detector (D).

Por lo tanto, es patente que un desafío crucial en el campo de fNIRS es dilucidar de una manera fiable qué cambios hemodinámicos están vinculados al acoplamiento neurovascular y cuales son debido a los componentes confundentes. Para resolver este problema, una solución particularmente efectiva es el uso de "multiples distancias" de medida entre los optodos (emisores y detectores de luz) (Yücel et al., 2017). Debido a que la distancia fuente-detector (SDD por sus siglas en inglés) es el principal factor que determina la penetración de la luz en el cuero cabelludo, podemos diferenciar entre canales cortos (SS: 5-15 mm) y canales largos (SDD: 25-40 mm) en función de si registran la actividad hemodinámica superficial (piel del cuero cabelludo, músculo y cráneo) o si exploran tanto la actividad cortical como la superficial, respectivamente (Brigadoi y Cooper, 2015). La idea subyacente a este enfoque es utilizar la señal de superficie registrada por canales cortos como un regresor para estimar la supuesta señal cortical (Gagnon et al., 2012; Saager, 2008). La señal registrada por los canales cortos puede ser utilizada como regresor para eliminar la actividad superficial (no cerebral) de las señales más profundas de los canales largos, aislando así la putativa respuesta cerebral. Curiosamente, la mayoría de los instrumentos fNIRS no ofrecen esta posibilidad de registro "multi-distancia". Sin embargo, el instrumento NIRS (Theia) que se ha utilizado en este trabajo incluye una particular geometría, que le permite disponer de múltiples distancias inter-optodo para este fin.



Figura 3. Representación esquemática de un montaje NIRS de ""multiples distancias" de medida entre los optodos" en el que se observa la luz cercana al infrarrojo (línea discontinua) viajando desde el emisor (E) hasta el detector (D) tanto en el canal corto como en el canal largo.

Además, lejos de considerar las "multiples distancias" de medida entre los optodos como un método para hacer frente a la contaminación, la señal de superficie podría ser un valioso indicador de la actividad del sistema nervioso autónomo. De hecho, los vasos de la frente se encuentran inervados por el sistema nervioso simpático y parasimpático, por lo que la actividad vasomotora registrada por los canales cortos podría constituir un buen indicador de la actividad autonómica (Molina-Rodríguez et al., 2022). Un reciente metaanálisis sugiere que el TDAH está caracterizado por una desregulación autonómica manifestada por una pobre adaptación a las demandas de la tarea (Bellato et al., 2020).

Por todo ello, consideramos que la señal de superficial puede ser un indicador útil para clasificar a los participantes con TDAH y TD.

1.5.2. La corteza prefrontal

La corteza prefrontal se localiza en la parte más anterior y caudal de los lóbulos frontales extendiéndose desde la corteza premotora hasta el polo frontal. Debido a que es una región cortical de gran tamaño y con elevada heterogeneidad en términos funcionales, clásicamente se han diferenciado varias divisiones: región dorsomedial, dorsolateral, ventrolateral, ventromedial orbitofrontal, cingulado anterior y frontopolar (Carlén, 2017). La mayoría de las funciones cognitivas de orden superior (memoria de trabajo, la inhibición, flexibilidad cognitiva, gestión atencional) muestran un sustrato anatómico vinculado parcial o totalmente con las diferentes regiones de la corteza prefrontal. Por lo tanto, no es de extrañar que exista una gran cantidad de trabajos que relacionan en funcionamiento de las regiones prefrontales con la sintomatología del TDAH (Mauri et al., 2018; Shaw, 2007; Halperin y Schulz, 2006). El fNIRS Theia se encuentra especialmente diseñado para medir la actividad hemodinámica de la región más caudal de la corteza prefrontal: la región frontopolar. La corteza frontopolar se ubica por tanto en el área más anterior del polo frontal, extendiéndose en un eje rostro-caudal a través de las superficies mediales, orbitales y laterales. La corteza frontopolar se encuentra implicada en multitud de procesos cognitivos como exploración de diferentes alternativas, memoria prospectiva, razonamiento relacional y calculo (Burgess et al., 2013).



Figura 4. Representación esquemática de la corteza frontopolar en el encéfalo humano.

1.5.3. Estudios NIRS en población TDAH

Los estudios sobre fNIRS y TDAH pueden diferenciarse en dos grandes grupos. En primer lugar, encontramos los trabajos que tienen como objetivo analizar las diferencias funcionales entre TDAH y sujetos controles desde una perspectiva grupal (Mauri et al., 2018). La mayor parte de estos estudios señalan que durante la ejecución de tareas cognitivas los niños con desarrollo típico muestran incrementos más acusados en HbO en comparación con los niños TDAH. Este hecho sugiere que la sintomatología TDAH podría estar vinculada con la hipoactivación de algunas regiones de la corteza prefrontal (Arai et al., 2016; Inoue et al., 2012; Mauri et al., 2018; Negoro et al., 2010; Yasumura et al., 2014). Aunque este tipo de trabajos incrementa el conocimiento científico sobre el sustrato neuronal del TDAH, no proporcionan información que ayude a los profesionales en el diagnóstico (Gu et al., 2018). De hecho, este tipo de investigaciones ofrecen datos promediados de tendencias grupales que pueden o no darse en los individuos que conforman cada una de las clases, por lo que son escasamente aplicables en los entornos clínicos.

El segundo conjunto de investigaciones se caracteriza por emplear métodos de clasificación supervisada que son capaces de determinar hasta qué punto las características de la señal NIRS permiten distinguir/identificar a los participantes con TDAH de forma individual (Dolu et al., 2019; Gu et al., 2018; Güven et al., 2020; Monden et al., 2015). Este tipo de enfoque ofrece la posibilidad de utilizar las características hemodinámicas como biomarcador para calcular la probabilidad de que un sujeto desconocido pertenezca al grupo TDAH, lo que tiene una aplicabilidad clínica considerable (Crippa et al., 2017). Los trabajos previos apuntan a que la combinación de fNIRS y los métodos de aprendizaje automático pueden ser útiles para diferenciar entre participantes con TDAH y TD. Sin embargo, la literatura muestra variabilidad en el éxito de clasificación logrado, así como en las características de la señal NIRS y el tipo de paradigma utilizado. Por ejemplo, Monden et al (2015) informaron de una precisión de clasificación del 85% con una sensibilidad del 90% mediante el análisis de curvas ROC obtenidas a partir de datos de activación prefrontal derecha de oxihemoglobina durante una tarea "Go/No-Go". Utilizando una tarea "N-Back" y una máquina de vectores de soporte (SVM) Crippa et al., (2017) evaluaron la capacidad de clasificación de ambos cromóforos, encontrando que la HbR (78% de precisión, 72% de sensibilidad y 82% de especificidad) mostró un mayor poder de discriminación que la HbO (57% de precisión,
48% de sensibilidad y 67% de especificidad). Por último y también empleando una tarea N-back y un SVM encontramos el trabajo de *Gu et al* (2018) en el que alcanzaron un 86% de precisión con datos de oxihemoglobina medidos en el córtex prefrontal y temporal. Hay que destacar que en ninguna de estas investigaciones no se aplicaron medidas de control de la interferencia hemodinámica superficial y/o sistémica, lo cual es especialmente importante cuando se explora la corteza prefrontal a través de la frente (Takahashi et al., 2011; Haeussinger et al., 2014). A pesar de este conocido inconveniente de fNIRS, los algoritmos de clasificación pueden lograr un rendimiento aceptable aprendiendo algún tipo de representación de características a partir de los datos no corregidos, pero la incertidumbre sobre la naturaleza y el origen de las características dificulta la interpretación de los modelos predictivos.

En conclusión, queda claro que la tecnología NIRS ha sido utilizada con éxito variable para encontrar marcadores biológicos objetivables que faciliten el diagnostico de TDAH. Si bien los trabajos anteriormente citados han encontrado características funcionales que pueden definir a los niños con TDAH, hasta la fecha no existen marcadores biológicos consensuados que faciliten el diagnóstico.

1.6. Componentes oscilatorios hemodinámicos y función cerebral

Las señales fNIRS son una mezcla de componentes hemodinámicos que oscilan a diferentes escalas temporales asociadas a procesos globales sistémicos (fluctuaciones en la frecuencia cardiaca, respiración y presión arterial), regulación local del flujo sanguíneo (actividad vasomotora) y a la activación neuronal (acoplamiento neurovascular) (Tachtsidis & Scholkmann, 2016). Reflejan la contribución de múltiples procesos fisiológicos regulatorios para asegurar una adecuada oxigenación cerebral (actividad del sistema nervioso autónomo, repuesta vascular miogénica y neurogénica, liberación endotelial de sustancias vasoactivas, concentración de CO2 en sangre, etc.) (Andersen et al., 2018). Estudios previos sobre dinámica cardiovascular han identificado oscilaciones características en varias bandas de frecuencia: 0.005-0.0095 Hz (endotelial no relacionada con NO), 0.0095-0.021 Hz (endotelial relacionada con NO), 0.021-0.052 Hz (neurogénica), 0.052-0.145 Hz (miogénica), 0.145-0.6 Hz (respiratoria) y 0.6-2 Hz (cardíaca). También se ha propuesto la existencia de un vínculo entre los sistemas cardiovascular y neuronal en ciertas oscilaciones, actuando como una unidad funcional (Stefanovska, 2007).

En un estudio anterior en el que participaron adultos jóvenes sanos, nuestro equipo de investigación demostró que una tarea de aritmética mental rítmica inducía con éxito fluctuaciones hemodinámicas cíclicas acopladas a la frecuencia de la tarea (33 mHz), y que los patrones oscilatorios eran consistentes en todos los individuos tanto en las señales fNIRS superficiales como profundas registradas en la región frontopolar (Molina-Rodríguez et al., 2022). El análisis espectral también mostró actividad oscilatoria a frecuencias más bajas (<33 mHz) observada en reposo y durante la tarea mental, y con un pico prominente alrededor de 5-10 mHz.

En el campo de investigación en TDAH, estudios de fMRI en reposo han reportado que los pacientes con TDAH muestran diferencias significativas en la banda de baja frecuencia (10-80 mHz) en múltiples regiones cerebrales (Yu-Feng et al., 2007; Tang et al., 2018; Zhang et al., 2015), con una contribución separable de subbandas de frecuencia específicas, incluidas las frecuencias muy bajas (0-10 mHz) (Yu et al., 2016). Estas diferencias se han relacionado con anomalías en el funcionamiento de las redes de saliencia, atención y modo por defecto (*default-mode network*), pero las inconsistencias observadas en muchos estudios apuntan a una gran heterogeneidad en la actividad cerebral espontánea en el TDAH (Wang et al., 2017).

En conjunto, las pruebas sugieren que algunos atributos de la actividad cerebral son sensibles a bandas de frecuencia específicas y que, entre otros trastornos, podrían presentar características distintivas en el TDAH.

1.7. Hipótesis de trabajo

En virtud de sus principios operacionales, la técnica fNIRS es especialmente adecuada para el examen del TDAH por ser relativamente insensible a los artefactos de movimiento y no requerir restricciones de movimiento significativas (*i.e.* muy adecuada para pacientes con inquietud motora o con poca cooperación). La actividad prefrontal, de especial interés en el TDAH, es fácilmente registrable a través de la frente de forma estable y sin causar particular incomodidad al paciente. Gracias a su tolerancia, es factible monitorizar mediante fNIRS la respuesta hemodinámica funcional de niños TDAH y DT ante una tarea cognitiva.

Entre otras tareas cognitivas, la aritmética mental implica genéricamente el área frontopolar al sustentar los procesos atencionales y de memoria de trabajo necesarios para lograr un buen rendimiento matemático (Arsalidou y Taylor, 2011). Sería, por tanto, una tarea adecuada para estimular la actividad neuronal y su respuesta hemodinámica asociada en esta región cerebral.

Si la tarea aritmética está organizada de forma rítmica, alternando cíclicamente fases de esfuerzo mental con fases de pausa de la misma duración, se podrían inducir oscilaciones igualmente rítmicas en las señales fNIRS y asociadas a la frecuencia de la tarea. La inducción de tal ritmicidad, a una frecuencia fundamental conocida, facilitaría la identificación de patrones oscilatorios mediante técnicas de análisis espectral.

Con base en estas premisas, partimos de la hipótesis de que una tarea mental rítmica induciría una respuesta funcional susceptible de ser capturada mediante fNIRS en la región frontopolar. Dicha respuesta sería separable en diferentes componentes oscilatorios de origen cerebral y extracerebral, siendo distintivos en frecuencias específicas como potenciales biomarcadores de la condición TDAH.

1.8. Objetivos

El objetivo general del presente trabajo es desarrollar un marco metodológico para la identificación de biomarcadores funcionales, basados en propiedades espectrales de las señales fNIRS, aplicables en herramientas computacionales de apoyo a la evaluación objetiva del TDAH.

Con este propósito, se establecieron los siguientes objetivos específicos:

1. Evaluar la capacidad de una tarea rítmica de aritmética mental para inducir oscilaciones fNIRS en niños TDAH y DT.

- Examinar las señales fNIRS originadas en capas tisulares superficiales y profundas, y aislar eficientemente la señal de respuesta cortical de otros componentes hemodinámicos no cerebrales.
- Separar los diferentes componentes oscilatorios presentes en las señales fNIRS mediante técnicas de análisis espectral tiempo-frecuencia.
- 4. Identificar las frecuencias más relevantes para TDAH y DT mediante análisis de sincronización de las bandas espectrales.
- 5. Extraer los patrones oscilatorios comunes asociados a las bandas espectrales más relevantes.
- 6. Extraer características numéricas con base en métricas de similitud entre las señales individuales y los patrones oscilatorios comunes de TDAH y DT.
- 7. Seleccionar el subconjunto de características con mayor capacidad discriminativa entre TDAH y DT.
- Identificar los patrones oscilatorios máximamente representativos de la condición TDAH y DT como potenciales biomarcadores funcionales.
- 9. Generar modelos predictivos usando diferentes algoritmos lineales para clasificación binaria supervisada.
- Evaluar y comparar la capacidad de los modelos predictivos para diferenciar entre TDAH y DT a nivel individual.

2. MÉTODO

2.1. Población y muestra

La muestra inicial estuvo compuesta por 16 niños clínicamente diagnosticados de TDAH (criterios DSM-V de la Asociación Americana de Psiquiatría) con presentación combinada (es decir, inatentos y con hiperactividad/impulsividad) y 17 controles con desarrollo típico. Todos los participantes fueron caucásicos, diestros, con visión normal o corregida, sin otras comorbilidades médicas o psiquiátricas y su idioma materno era el castellano. El estudio fue aprobado por el Comité de Ética de la Universidad Miguel Hernández. Con anterioridad a su participación obtuvimos el consentimiento informado de ambos padres/tutores y los participantes no recibieron compensación económica. Los niños con el diagnostico de TDAH fueron reclutados entre los usuarios de una asociación de apoyo local al TDAH mientras que los niños con desarrollo típico fueron reclutados en los colegios de la zona. Dado que todos los participantes con TDAH estaban en tratamiento con metilfenidato decidimos provechar los "descansos" en la administración, pautados por prescripción médica, para realizar las evaluaciones. El periodo de "descanso" mínimo fue de cuatro días en todos los casos lo que es suficiente para eliminar los efectos del metilfenidato. Un participante del grupo TDAH y dos del grupo control fueron eliminados por problemas en la calidad de sus señales NIRS (ver sección 2.4.). La muestra final incluyó a 15 participantes con TDAH (edad media 11,9, DE 1,4, rango de 10 a 13 años) y 15 controles con desarrollo típico (edad media de 11,6, DE 1,1, rango de 10 a 13 años).

2.2. Procedimiento

Los registros se realizaron en una habitación confortable con temperatura y luminosidad controlada. Los participantes se sentaron cómodamente en un sillón ergonómico situado a 80 cm de una pantalla en la que visionaron una versión ligeramente modificada de una tarea de aritmética mental rítmica descrita en un trabajo anterior de nuestro grupo de investigación (Molina-Rodríguez et al., 2022). El paradigma estaba compuesto por 10 bloques consecutivos de 30 s, cada uno de ellos se componía de una fase de cálculo aritmético de 15 s seguida por 15 s de reposo (Figura 5). El periodo de la tarea impone una frecuencia de 0,033 Hz a la que a partir de ahora denominaremos como "frecuencia de la tarea". En la fase de cálculo se solicitó a los participantes que sumaran mentalmente y de forma iterativa una cifra de un dígito (entre 3 y 9) a una cifra de 3

dígitos (entre 100 y 199), ambas elegidas al azar, lo más rápida y precisamente posible durante 15 s. Por ejemplo, si en la pantalla se muestra "112 + 5", las iteraciones serían: 112 + 5 = 117; 117 + 5 = 122; 122 + 5 = 127, etc. Transcurridos los 15s en la pantalla aparece "¿resultado?" y el participante debe verbalizar el valor final alcanzado, por ejemplo "127". Utilizar un diseño experimental que minimiza el habla durante la tarea reduce los cambios en la respiración que puede afectar a la hemodinámica cerebral (Scholkmann et al., 2013). Tras la comunicación del resultado el examinador anota el valor final que, implícitamente, refleja también el número de iteraciones mentales realizadas y su exactitud. Utilizamos estos indicadores para valorar el rendimiento de los participantes en la tarea aritmética. Con el objetivo de verificar la adecuada comprensión de la tarea y la disminución de las respuestas de estrés, los participantes realizaron dos ensayos previos al registro experimental. Se instruyó a los participantes para que mantuvieran la vista en la pantalla, lo más relajados posible, y que realizaran los cálculos aritméticos mentalmente sin hablar ni gesticular. El protocolo fue estructurado en tres fases consecutivas e ininterrumpidas de registro: (i) 300 s de línea de base en reposo; (ii) 300 s de tarea mental y (iii) 300 s de recuperación en reposo. Esquemáticamente, la secuencia del procedimiento fue la siguiente (Figura 5): (1) registro basal en reposo durante 3 minutos (participante relajado, con los ojos abiertos y sin realizar ninguna acción) mientras se muestra una pantalla en negro; (2) se muestra una cruz de fijación durante 2 segundos que avisa del inicio de la fase de cálculo; (3) se muestran los números a sumar iterativamente durante 15 s; (4) tras 15 s, se muestra "¿resultado?" durante 5 s y el participante informa verbalmente la suma final que ha obtenido; (5) se muestra pantalla en negro indicando al participante que se relaje; (6) Los pasos 2 a 5 se repiten durante los 10 ciclos de cálculo-recuperación; (7) Tras los 10 ciclos de cálculo-pausa se procede a realizar un registro final en reposo durante 3 minutos.



Figura 5. Esquema del procedimiento experimental.

2.3. Dispositivo fNIRS

En el presente trabajo se ha utilizado un instrumento NIRS multicanal y de onda continua (Tehia[®], Newmanbrain, S.L., Elche, España), validado en un trabajo previo de nuestro grupo de investigación (Molina-Rodríguez et al., 2022). El dispositivo transfiere datos de forma inalámbrica (a través de Bluetooth) a una frecuencia de muestreo de 10 Hz. El sistema NIRS se encuentra equipado con cuatro emisores de luz con dos fuentes LED (740 y 850 nm) y diez detectores integrados en una plataforma rectangular de 80x20 mm. La combinación de pares de optodos a diferentes distancias proporciona 16 canales cortos (14 mm) y 12 canales largos (32 mm). El dispositivo corrige la luz ambiental y dispone de un acelerómetro de 3 ejes para la detección de artefactos de movimientos. El sistema NIRS se colocó en el frente centrado en AFpz, de acuerdo con el sistema internacional 10-5 (Oostenveld y Praamstra, 2001; Jurcak et al., 2007), para registrar la región frontopolar de la corteza prefrontal (área 10 de Brodmann).

Con el objetivo de identificar los registros con interferencias no fisiológicas, señales saturadas o con una baja relacion señal-ruido se verificaron los datos ópticos sin procesar eliminando las señales que mostraron valores extremos (< 5 % o > 95 % del rango dinámico del dispositivo) o con un coeficiente de variación excesivo (> 7,5%) (Orihuela-Espina et al., 2010; Zimeo Morais et al., 2018). Además, mediante la inspección visual de las señales NIRS y de acelerometría identificamos y eliminamos aquellos registros con artefactos de movimiento. Como se mencionó anteriormente, tres participantes fueron excluidos debido al incumplimiento de algunos de estos criterios de calidad.

Se calcularon los cambios de concentración relativos de HbO y HbR mediante de la ley de Lambert-Beer modificada¹ (Delpy et al., 1988; Kocsis et al., 2006), utilizando funciones del programa Homer2 NIRS (Huppert et al., 2009) basado en MATLAB (Versión R2021b, Mathworks, Natick, MA, EE. UU.). Se empleó un factor *differential path lenght* de 6 para ambas longitudes de onda. A continuación, las señales de HbO y HbR se filtraron digitalmente con un filtro de paso-bajo Butterworth de quinto orden y fase cero, con un corte en 0,08 Hz (MATLAB Signal Processing Toolbox); no se aplicó filtrado paso-alto. De esta forma, se eliminaron los componentes de presión arterial, respiratorios, cardíacos, ondas de Mayer e instrumentales de alta frecuencia, preservando la banda de muy baja frecuencia (Reddy et al., 2021). Como resultado, para cada cromóforo se obtuvieron 16 series temporales de los canales cortos (SS por sus siglas en ingles) y 12 de los canales largos (LS por sus siglas en ingles). SS contiene únicamente componentes de capas superficiales (no cerebrales), mientras que el registro LS, incluye tanto componentes superficiales como profundos (Tachtsidis y Scholkmann, 2016).



Figura 6. Posicionamiento del dispositivo NIRS. Los cuadrados amarrillos representan los emisores de luz, y los círculos blancos los detectores. Las líneas negras muestran los 16 los canales cortos y las líneas verdes los 12 canales largos. Los rectángulos formados por línea discontinua representan las tres regiones de interés (ROI) exploradas en este trabajo.

Debido a que DS está contaminada por oscilaciones de origen no cerebral, fue necesario aislar la respuesta cortical relacionada con la tarea de la interferencia sistémica y de las capas superficiales del cuero cabelludo (Kirilina et al., 2012; Scholkmann et al., 2014; Takahashi et al., 2011). El dispositivo NIRS utilizado en este trabajo se encuentra equipado con tecnología de "múltiples distancias" y para cada canal largo existen tres canales cortos cercanos que pueden utilizarse como referencias para eliminar la contaminación (Brigadoi y Cooper, 2015; Gagnon et al., 2012; Yücel et al., 2017). Tal y como sugiere Gagnon et al., (2014) aplicamos un enfoque basado en el "doble canal corto" en la que cada canal largo fue regresado por la suma de los dos canales cortos más cercanos al emisor y al receptor. La regresión lineal fue calculada mediante de la función "robusfit" de MATLAB la cual es más robusta a los valores extremos. Tras la regresión, se obtuvo una "señal limpia" (CS) que fundamentalmente contiene la verdadera señal de origen cortical.

Sería posible que la posición de los optodos no fuese la misma en todos los participantes, debido a las diferencias individuales en la forma y tamaño de la cabeza, por lo que la interpretación de canales aislados podría ser engañosa (Tak, et al., 2016). Por ello, con el objetivo de mejorar la consistencia espacial (Plichta et al., 2006; Schecklmann et al., 2008), para cada participante promediamos las SS y CS de cada una de las tres regiones de interés (ROI) mostradas en la Figura 6. Debido a que cada ROI se obtuvo promediando la misma cantidad de optodos vecinos, las señales resultantes tienen una relación señal-ruido comparable. Finalmente, para operar en la misma escala, estandarizamos las señales en puntuaciones Z. Todo el procesamiento subsecuente fue realizado con estas señales.

2.4. Distribución de la energía espectral durante la tarea

Las tareas cognitivas pueden afectar los componentes de la señal fNIRS en múltiples frecuencias, debido a la contribución interrelacionada de factores hemodinámicos asociados a la actividad sistémica, a las capas de tejido superficial y a la actividad neuronal (Nambu et al., 2017; Näsi et al., 2013; Zimeo Morais et al., 2017). Desde esta perspectiva general, en una primera comparación a nivel de grupo se identificaron los componentes oscilatorios más relevantes durante la tarea cognitiva. Con este fin, se realizó un análisis de densidad espectral de potencia (Power Spectral Density, PSD) usando el periodograma promediado de Welch (Welch, 1967). Para el análisis se utilizó el segmento de señal comprendido entre 30 s antes del inicio de la tarea y 30 s después del final de la tarea, lo que supone un intervalo temporal de 30+300+30 = 360segundos (3600 muestras a una frecuencia de muestreo = 10 Hz). Los 30 segundos adicionales se incluyeron para tener en cuenta posibles efectos de anticipación o persistencia. Para mejorar la resolución en frecuencia, extendimos los datos a la siguiente potencia de dos (4096 muestras) mediante reflexión simétrica. A continuación, se calcularon los PSD mediante la función "pwelch" de MATLAB con una FFT² de 4096 muestras, ventana Hanning de 2000 muestras y superposición = 80 % para suavizar el espectro y reducir la varianza del ruido (Ilvedson, 1998). Para permitir comparaciones, los PSD se normalizaron a valores porcentuales relativos mediante el cálculo de la relación de potencia de cada intervalo de frecuencia con respecto a la potencia total de

² FFT: *Fast Fourier Transform* (Transformada rápida de Fourier)

todo el espectro (Aarabi y Huppert, 2016). Este procedimiento se aplicó a los datos de SS y CS de cada participante.

Para valorar diferencias significativas entre los PSD de los grupos TDAH y TD, se aplicaron pruebas t de dos muestras a cada frecuencia definida en el PSD. El estadístico t observado se corrigió para comparaciones múltiples mediante un método no paramétrico basado en *clusters* de frecuencias (Maris y Oostenveld, 2007). Se generaron 2000 permutaciones mezclando aleatoriamente los datos entre clases, y para cada permutación se calcularon los valores t de los *clusters* observados. Luego, se calculó el valor p a partir de la proporción de permutaciones que mostraron un valor t mayor que el observado. El umbral de significancia de los *clusters* de frecuencia se estableció para un alfa crítico = 0.01. Finalmente, los PSD se promediaron entre los participantes de cada clase para obtener el PSD normalizado promedio de HbO y HbR para cada tipo de señal y ROI. El intervalo de confianza (IC) del 95 % para la media en cada intervalo de frecuencia se calculó mediante 2000 remuestreos por *bootstrapping*.

2.5. Descomposición tiempo-frecuencia

Aunque puede proporcionar información valiosa, el análisis de las diferencias espectrales a nivel de grupo no permite discriminar la condición TDAH o DT a nivel individual. Por tanto, una cuestión crucial del presente trabajo fue la identificación de los componentes de frecuencia con capacidad potencial para diferenciar entre las dos condiciones. Para ello, necesitábamos un método adecuado para localizar las oscilaciones relacionadas con la tarea en diferentes escalas temporales (i.e. bandas de frecuencia), apropiado para el análisis de señales no estacionarias y capaz de proporcionar alguna medida de similitud para definir la pertenencia a una clase (TDAH o DT).

Las técnicas convencionales de análisis espectral pueden resolver cómo se distribuye la energía de las señales en diferentes frecuencias, pero revela poco sobre cómo cambia esta distribución a lo largo del tiempo. Las técnicas de análisis tiempo-frecuencia (*time-frequency analysis*, TFA) son una mejor opción para descubrir patrones oscilatorios en determinadas frecuencias, sobre todo cuando varían sustancialmente con el tiempo y/o hay múltiples componentes presentes en la señal (Cohen, 1995; Gröchenig et al., 2001). Además, permiten separar los componentes de magnitud y fase asociados a la señal, lo que resulta muy útil para detectar sincronizaciones transitorias (Morales y Bowers, 2022; Glerean et al., 2012). Sin embargo, la selección de la técnica TFA óptima depende en gran medida del conocimiento previo sobre las características de la señal que se pretende

analizar (Wacker y Witte, 2013). Puesto que no disponíamos de información previa sobre los componentes oscilatorios más útiles a efectos de clasificación, el método convencional de filtrado pasa banda seguido de transformada de Hilbert no era fiable debido a la incertidumbre sobre las bandas de paso a aplicar; aparte de otras limitaciones inherentes (Munia y Aviyente, 2019). También evitamos los métodos adaptativos que dependen en gran medida de un ajuste específico de parámetros para lograr resultados significativos, como la descomposición empírica o variacional de modos (Liu et al., 2018; Kijewski-Correa y Kareem, 2006). Finalmente, decidimos usar un enfoque impulsado por datos (*data-driven*), basado en una transformada *wavelet* continua (*continuous wavelet transform*, CWT) compleja y en la detección de sincronización tiempo-escala.

La CWT proporciona una representación tiempo-frecuencia (o tiempo-escala) de las características de una señal a partir de la dilatación y traslación de una función wavelet madre; su teoría y descripción matemática pueden consultarse en Cohen y Kovacevic (1996) y Hramov et al., 2015. La CWT puede verse como un banco de filtros complejo, definido automáticamente por las escalas de la wavelet (Sadowsky, 1996) y del que puede extraerse información de magnitud y fase en cada una de las escalas (subbandas de frecuencia) del análisis; evitando los inconvenientes de utilizar múltiples filtros personalizados (Munia y Aviyente, 2019). Para calcular la CWT utilizamos las wavelets Morse generalizadas, una superfamilia flexible de *wavelets* exactamente analíticas y especialmente útiles para analizar señales con amplitud y frecuencia variables en el tiempo, es decir, señales moduladas (Olhede y Walden, 2002; Lilly y Olhede, 2010). Dado que las wavelets Morse pueden ajustarse para abarcar muchas otras wavelets analíticas de uso común, proporcionan un marco unificado como punto de referencia. Así, por ejemplo, fijando el parámetro de simetría $\gamma = 3$ se define la familia "Airy" *wavelets* que puede sustituir con éxito a la popular *wavelet* Morlet (Lilly y Olhede, 2012). Además de otras aplicaciones, las wavelets de Morse se han propuesto en diversos estudios biomédicos, como la locomoción humana (Suwansawang et al., 2017), la electrocardiografía y la electromiografía (Wachowiak et al., 2018; Wiklendt et al., 2020), el acoplamiento de oscilaciones neuronales (Nakhnikian et al., 2016; Munia y Aviyente, 2019), la clasificación de datos electroencefalográficos (Huggins et al., 2021; Ajith et al., 2021) y la supresión de artefactos fNIRS (Perpetuini et al., 2021)

Para mantener el procedimiento lo más simple y reproducible posible, calculamos la CWT aplicando los parámetros de Morse predeterminados recomendados en MATLAB Wavelet Toolbox (simetría = 3, producto tiempo-ancho de banda = 60), y que definen una wavelet perfectamente simétrica. Para lograr una representación tiempo-frecuencia lo más precisa posible, tuvimos en cuenta los efectos de borde producidos cuando las wavelets se extienden fuera de los límites de la señal. Dado que tales efectos dependen de la escala wavelet, se tuvo en cuenta el denominado "cono de influencia" (COI) para evitar las regiones de CWT con posibles coeficientes inexactos³. Para reducir los efectos de borde, y antes de la CWT, extendimos la longitud de las señales mediante reflexión simétrica (i.e. en ambos extremos). Como el análisis PSD preliminar mostró que la mayor parte de la potencia espectral estaba en la banda 0 - 50 mHz (Fig. X), la CWT se centró en ese rango de frecuencia aplicándose a las señales extendidas simétricamente con el parámetro de discretización de escala "voces por octava" = 10. Con esta configuración, y con base en el paso de banda mínimo y máximo permitido, se obtuvieron coeficientes CWT para 45 escalas (frecuencias aproximadas entre 2,4 y 50 mHz). A pesar de la extensión previa de la señal, observamos que las frecuencias inferiores a 3 mHz estaban fuera de los límites del COI (según lo estimado por la función "cwt" de MATLAB) y eran, por tanto, más propensas a sufrir efectos de borde. Por esta razón, los análisis posteriores se limitaron al rango de frecuencias 3 - 50 mHz y, consecuentemente, el número de escalas utilizables se redujo a 41 (Figura 7).



Figura 7. Transformada wavelet continua (CWT) de la señal profunda de HbO desde la mitad del retorno de la inversión para un participante representativo. Izquierda: Escalograma de los datos extendidos simétricamente con el cono de influencia trazado en magenta. Las líneas discontinuas blancas contienen la señal original y las líneas negras continuas delimitan el período de la tarea. El rectángulo rojo representa la región del escalograma de la que se extrajeron los coeficientes para el análisis. Derecha: Esquemas de la matriz de coeficientes extraída que contiene 41 escalas y 3600 puntos temporales.

³ en el caso de las *wavelets* Morse, el concepto equivalente a COI es "huella" wavelet (Lilly, 2017).

La matriz completa de coeficientes CWT se reservó para el cálculo posterior de la CWT inversa (es decir, incluyendo los segmentos extendidos), aunque para el siguiente paso de análisis solo se utilizó una porción central de la matriz correspondiente a la señal original sin extender.

2.6. Detección de la sincronización tiempo-escala

Bajo la premisa de que la tarea cognitiva induciría fluctuaciones fNIRS diferenciadas para los grupos DT y TDAH, nos propusimos identificar las subbandas de frecuencia que mostrasen sincronización significativa a nivel de grupo. Dado que la sincronización grupal puede aparecer transitoriamente y no de forma constante, realizamos un detallado análisis momento a momento para poder capturar oscilaciones comunes que evolucionan dinámicamente a lo largo del tiempo. Debe mencionarse que esta estrategia se inspira en la lógica subyacente del denominado "análisis de correlación inter-sujeto" (*inter-subject correlation*, ISC) (Nastase et al., 2019; Hasson et al., 2004). El análisis ISC es un método, basado en datos, para detectar actividad cerebral sincronizada en un grupo de participantes expuestos a un estímulo común. Si el estímulo induce respuestas consistentes en los participantes la correlación de sus señales será significativa, en caso contrario la correlación será baja. Aunque conceptualmente similar, en este estudio no se usaron medidas de correlación sino la información de magnitud y fase obtenida de los coeficientes complejos de la CWT.

En estudios de fMRI, se han validado medidas como la "sincronización de fase inter-sujeto" (Bolt et al., 2018) y la "consistencia de fase por pares" (Kauppi et al., 2014; Berens, 2009) para capturar sincronizaciones de fase instantánea entre sujetos. La primera es similar a la "longitud del vector medio circular" en la literatura de estadística circular (Delorme y Makeig, 2004), también conocida en estudios de EEG relacionados con eventos como "coherencia de fase" (Cohen, 2014), "agrupación de fase" (Tallon-Baudry et al., 1996) o "factor de bloqueo de fase" (Bastos y Schoffelen, 2016). Sin embargo, estas medidas se basan únicamente en la dispersión de los ángulos de fase, ignorando la magnitud. Por tanto, estas medidas asumen que las señales de baja amplitud afectan a la medición igual que las que tienen una amplitud significativa. Sin embargo, este enfoque puede no ser apropiado para fNIRS, donde los cambios de amplitud están relacionados con la magnitud de la respuesta hemodinámica. Dado que a medida que aumenta la amplitud mejora la relación señal-ruido, es razonable argumentar que las observaciones con amplitudes más altas pueden contribuir a una estimación más realista de la sincronización de fase (Richter et al., 2015). Bajo este supuesto, decidimos utilizar una medida estrechamente relacionada llamada "coherencia lineal inter-ensayo", que combina la magnitud y la fase en el paso de normalización (Cohen, 2014). Dado que aquí comparábamos las observaciones de los sujetos, denominamos la medida como "sincronización entre sujetos" (*inter-subject synchronization*, ISS), omitiendo "lineal" para simplificar y formulada de manera similar como:

$$ISS(f,t) = \frac{\sum_{k=1}^{n} F_k(f,t)}{\sqrt{n \sum_{k=1}^{n} |F_k(f,t)|^2}},$$

donde $F_k(f, t)$ es el estimado espectral de la observación k en la frecuencia f y tiempo t, y el módulo \parallel representa la norma compleja. ISS también adopta valores entre 0 (ausencia de sincronización) y 1 (sincronización perfecta). El ISS se calculó momento a momento para cada escala a partir de las matrices de coeficientes CWT de cada grupo (*i.e.* 15 observaciones de participantes por escala). En este punto, limitamos el análisis al intervalo de tiempo entre -30 s y +30 s alrededor de la tarea. Este procedimiento se aplicó independientemente a los datos SS y CS de cada grupo para cada cromóforo y ROI, obteniendo una representación de ISS en el plano tiempo-frecuencia que, con fines ilustrativos, puede visualizarse como un mapa de colores (Figura 8). A continuación, elegimos el máximo ISS observado en cada escala y que representa la máxima sincronización grupal alcanzada en esa frecuencia específica (Figura 8). En este trabajo sólo se consideraron los máximos de ISS, independientemente del punto temporal en el que se alcanzaron. Tampoco se estimaron umbrales de significación en este paso, confiando en el poder discriminativo inherente a la selección de características y clasificación.

En cada mapa ISS se identificaron varios picos máximos en determinadas subbandas de frecuencias, reflejando la presencia de componentes oscilatorios con cierta sincronización grupal. Puesto que la suposición era que estos componentes podrían distinguir entre el grupo DT y TDAH, ahora debíamos localizar las bandas que presentasen mayor poder discriminativo. Para simplificar el procedimiento, se promediaron los datos de ISS de las tres ROIs para obtener la media de los máximos de ISS por frecuencia. Así, el análisis se redujo a un único patrón ISS común por cromóforo y tipo de señal para cada uno de los dos grupos, es decir, 2 grupos (DT, TDAH) x 2 cromóforos (HbO, HbR) x 2 tipos de señal (SS, CS) = 8 patrones ISS. Por último, para reducir el ruido los patrones ISS se suavizaron mediante una media móvil utilizando una

ventana deslizante de longitud igual a la mitad de las voces por octava (es decir, 10/2 = 5). Estos patrones se examinaron en los pasos siguientes para identificar los componentes de frecuencia más relevantes.



Figura 8. Esquemas del procedimiento para el cálculo de un mapa representativo de sincronización entre sujetos (ISS). Comenzando con las matrices de coeficientes de cada individuo dentro de un grupo (izquierda), se reúnen datos de valores complejos para cada intervalo de frecuencia (centro) y la ISS se calcula momento a momento para obtener una representación del poder de sincronización dentro del grupo en cada frecuencia a lo largo del tiempo (mapa de color derecho). El gráfico rojo de más a la derecha muestra los máximos y mínimos alcanzados por la ISS a lo largo de las frecuencias, que se utilizan para delimitar las subbandas a analizar (líneas discontinuas blancas horizontales).

2.7. Subbandas de frecuencia relevantes y extracción de características

Al igual que otras métricas de sincronización, ISS es una medida compuesta que no existe por sí sola a nivel de un solo sujeto, sino que representa una estadística conjunta de la sincronización del grupo. Por tanto, desentrañar la contribución a ISS de cada individuo no es una cuestión sencilla (Fulcher, 2018). Sin embargo, los picos de ISS sugieren que algunos componentes de frecuencia muestran cursos temporales similares entre individuos, al menos dentro de ciertos intervalos de tiempo. En otras palabras, existen patrones secuenciales comunes al grupo que pueden proporcionar información distintiva para definir la pertenencia a una clase. Este concepto entra dentro del campo interdisciplinar y muy estudiado de la clasificación de series temporales, que engloba una variedad de técnicas para identificar aquellas propiedades (características) que tienen suficiente poder discriminatorio para distinguir entre diferentes clases (para una revisión, véase Ye y Keogh, (2011)). En el contexto del presente trabajo, una técnica adecuada podría ser la basada en el concepto shapelet, que aborda el problema de la clasificación descubriendo secuencias primitivas de series temporales (shapelets) que se utilizan para cuantificar la (di)similitud entre clases (Hills et al., 2014; Richard, et al., 2020). Las shapelets proporcionan información directamente interpretable sobre patrones (formas) que son importantes para comprender cómo difieren las clases de datos, una propiedad deseable para los sistemas de apoyo a la toma de decisiones clínicas (Hearst et al., 1998).

Aquí aplicamos los fundamentos de la técnica *shapelets*, pero en lugar de buscar subsecuencias similares e independientes de la fase (es decir, las subsecuencias pueden estar situadas en cualquier lugar de la serie) realizamos el análisis dentro de un intervalo de tiempo fijo, teniendo todas las subsecuencias la misma longitud. No aplicamos ninguna traslación de subsecuencias en el tiempo, lo que supone que la similitud de las series temporales también depende de la fase (es decir, de una alineación temporal coherente). Por lo tanto, en lugar de patrones locales, capturamos patrones globales presentes en todo un intervalo de tiempo. Con este enfoque, el término "*shapelet*" podría no ser del todo apropiado, aunque lo usaremos por razones prácticas ya que se basa en principios comparables y es fácil de conceptualizar. No obstante, conviene aclarar que en el presente estudio el término '*shapelet*' hace referencia a un patrón máximamente representativo de una clase en un marco temporal específico.

Para extraer las series temporales en las que identificar *shapelets* representativos nos basamos en los patrones ISS medios, ya que sugieren las frecuencias que probablemente contengan oscilaciones sincronizadas. Calculando la CWT inversa dentro de la subbanda específica definida por los límites de un pico ISS, podemos reconstruir dichos componentes oscilatorios en el dominio temporal. Para reducir los efectos de borde, la CWT inversa se calculó a partir de la matriz de coeficientes completa obtenida en un paso anterior (sección 2.6). A continuación, las series temporales resultantes se truncaron en el intervalo comprendido entre -30 s y +30 s alrededor de la tarea. Tras aplicar este procedimiento a las CWT de todos los individuos pertenecientes a un grupo, se dispone de un conjunto de series temporales (n = 15) para encontrar un *shapelet* de referencia para ese grupo en una subbanda concreta.

Dado que todas las series temporales tienen la misma longitud y están dentro del mismo marco temporal, se puede obtener una *shapelet* de referencia simplemente promediando. El razonamiento es que, si las series temporales comparten un patrón común, su media debería representar suficientemente bien al grupo. Para cuantificar la similitud con la *shapelet* de referencia, entre otras posibilidades, se puede usar una medida tan sencilla como la distancia euclidiana:

$$D(S,T) = \sqrt{\sum_{i=1}^{n} (S_i - T_i)^2}$$

donde *S* designa la *shapelet* referencia y *T* una serie temporal, ambas de longitud *n*. *S* y *T* deben normalizarse para que tengan media 0 y varianza 1, lo que garantiza operar en la misma escala. La estandarización también nos permite relacionar la distancia euclidiana con el coeficiente de correlación ($r = 1 - D^2/2n$).

Para evaluar la capacidad de la *shapelet* referencia para discriminar entre grupos, se contrastaron sus distancias a un grupo con las del otro. Por ejemplo, una *shapelet* representativa del grupo DT debería tener distancias menores con los miembros de este grupo que con los del grupo TDAH, y viceversa. Entre otras medidas, el estadístico *F* del análisis de la varianza puede utilizarse para evaluar el poder discriminativo de la *shapelet* (Richard et al., 2020). Este estadístico indica la relación entre varianza intergrupo e intragrupo:

$$F = \frac{\sum_{i=1}^{C} \frac{(\bar{D}_{i} - \bar{D})^{2}}{C - 1}}{\sum_{i=1}^{C} \sum_{d_{j} \in D_{i}} \frac{(d_{j} - \bar{D}_{i})^{2}}{n - C}}$$

donde *C* es el número de clases (o grupos; en nuestro caso = 2), \overline{D} es la media global de todas las distancias, \overline{D}_i es la media de las distancias de la clase *i*, y *n* es el número de series temporales. Cuanto mejor sea la *shapelet*, mayor será el valor *F*, porque la diferencia intergrupo aumenta mientras que disminuye la intragrupo. El valor *p* correspondiente puede calcularse de la distribución *F*.

Basándonos en los patrones ISS medios, identificamos cuatro subbandas candidatas en cada una de ellas que se etiquetaron como A, B, C y D en orden decreciente de frecuencia. Cada subbanda contiene un pico (máximo local) flanqueado por dos valles (mínimos locales) delimitando los límites de frecuencia. Para cada patrón ISS, cada uno perteneciente a un grupo (DT o TDAH), realizamos el siguiente procedimiento para cada subbanda (Figura 9):

(1) Calcular la CWT inversa dentro de la subbanda a partir de las matrices de coeficientes extendidas de ambos grupos para obtener las series temporales correspondientes (n = 15 + 15 = 30).

(2) Truncar las series temporales en el intervalo de tiempo de interés.

(3) Generar la *shapelet* de referencia promediando sólo las series temporales del grupo objetivo (n = 15); también puede calcularse una medida de variabilidad como el error estándar de la media (SEM).

(4) Calcular las distancias a la *shapelet* para las series temporales de ambos grupos.

(5) Calcular F y el valor p a partir de las 15 + 15 = 30 distancias.

Dado que cada patrón ISS y sus subbandas son comunes a las tres ROIs, este procedimiento se aplicó por separado a los datos de cada ROI pero utilizando las mismas subbandas. Así, para cada patrón ISS se obtiene una matriz de 30×12 distancias, donde cada fila contiene las medidas de distancia de un individuo y las columnas corresponden a las 4 subbandas x 3 ROIs. Al haber ocho patrones ISS se obtuvo una matriz final de tamaño 30×96 (15 filas para DT y 15 para TDAH; 48 columnas para SS y 48 para CS). Cada columna tenía asociado su correspondiente valor F y *p*, indicando lo bien que su *shapelet* diferenciaba los grupos.

En síntesis, empleamos el principio *shapelet* para transformar las observaciones de datos a diferentes escalas temporales en un simple espacio de características basadas en distancias euclídeas (único tipo de características en el presente trabajo).



Figura 9. Ejemplo esquemático del procedimiento para extraer *shapelets* de referencia para el grupo DT y contrastarlos con el grupo TDAH. Las series de tiempo por participante dentro de una subbanda de frecuencia específica se obtienen a partir de su CWT inverso de banda limitada y se promedian solo para que el grupo DT obtenga el *shapelet*. Las distancias euclidianas al *shapelet* se calculan para las series de tiempo de ambos grupos y luego se contrastan las varianzas dentro y entre grupos para obtener el valor *F*-stat.

2.8. Algoritmos de clasificación y selección de características

En el presente estudio, la diferenciación entre la condición TDAH y DT se trató como un problema de clasificación binaria supervisada. La capacidad discriminativa de las características extraídas mediante el procedimiento propuesto se evaluó mediante cuatro algoritmos de aprendizaje automático: máquina de soporte vector (SVM) (Hearst et al., 1998), regresión logística (RL) (Hosmer et al., 2013), análisis discriminante lineal (ADL) (Fisher, 1936) y *naïve* Bayes Gaussiano (NB) (Kuncheva, 2006). Estos algoritmos son bien conocidos, intrínsecamente interpretables, eficientes desde el punto de vista computacional y pueden funcionar con tamaños de muestra relativamente pequeños. Con diferentes parametrizaciones, SVM es muy utilizado en estudios de neuroimagen de trastornos cerebrales, siendo RL y ADL otras opciones habituales (Arbabshirani et al., 2017); también en el campo de investigación del TDAH (Loh et al., 2022). Aunque menos popular, se incluyó NB por su facilidad de aplicación y buen rendimiento en una variedad de aplicaciones (Hand y Yu, 2001).

Aunque se basan en reglas diferentes, los cuatro algoritmos pertenecen a la categoría de clasificadores lineales, es decir, para hacer predicciones los clasificadores intentan estimar la línea que mejor separa las observaciones de las dos clases mediante una combinación lineal de las variables explicativas (Pereira et al., 2009; Hastie et al., 2009). Así, es posible saber en qué medida influye cada característica en la predicción, lo que mejora enormemente su interpretabilidad en comparación con algoritmos más complejos (p. ej. redes neuronales). Utilizamos las implementaciones de MATLAB de los clasificadores (*Machine Learning Toolbox*), usando la configuración predeterminada para mayor simplicidad y reproducibilidad; no se aplicó optimización de hiperparámetros.

Además de la respuesta funcional putativa, las señales fNIRS también contienen componentes comunes de origen sistémico y de actividad local impredecible. Por lo tanto, es muy probable que, entre las extraídas, ciertas características fuesen redundantes y/o irrelevantes, lo que podría degradar el rendimiento del clasificador por problemas de sobreajuste y ruido. Abordamos el problema reduciendo el número de características para obtener un subconjunto más pequeño y relevante, reduciendo la complejidad y mejorando al mismo tiempo la precisión de la clasificación y la capacidad de generalización (Somol et al., 2010). Con este fin se aplicó un enfoque híbrido de dos pasos: (1) selección preliminar de características por simple filtrado con base en las propiedades intrínsecas de cada característica; (2) un algoritmo de selección secuencial flotante (*Sequential forward floating selection wrapper*) que usa el propio clasificador como evaluador para escoger las características más relevantes (Amann et al., 2022).

1) Filtrado

A pesar de la amplia oferta de métodos de filtrado (Jović et al., 2015), y tras probar algunos de los más populares (p. ej. *relief and Minimum Redundancy-Maximum Relevance*), nos decantamos por una opción sencilla y directa basada en el estadístico *F* del análisis de la varianza. Simplemente seleccionamos las características que mostraban valores p <.01, asumiendo que era muy improbable que sus *shapelets* generadoras separaran los grupos por azar. Con este filtrado, se redujo efectivamente el número de características de 48 a 5 para SS y de 48 a 10 para CS. Los algoritmos de clasificación se aplicaron a las características filtradas de SS y CS por separado, y también en conjunto (SS + CS).

Para valorar el rendimiento predictivo de los clasificadores se utilizaron dos esquemas de validación cruzada (VC):

- a) VC de 5 iteraciones (*K-5 cross-validación*), con estratificación para asegurar un 50% de observaciones en cada clase y 20 repeticiones aleatorias Monte-Carlo (5 x 20 = 100 reparticiones de los datos) para reducir la varianza del error [103]. Puesto que el número de observaciones es 30 (15 TDAH y 15 TD), con el esquema de 5 iteraciones se asignan aleatoriamente 24 observaciones para entrenamiento (80%) y 6 para test (20%) en cada una de las 100 reparticiones usadas para la validación.
- b) VC dejando uno fuera (*Leave-one-out cross-validation*), donde cada observación, de las 30 disponibles, se usa una sola vez para el test y las 29 restantes para entrenamiento (30 reparticiones en total).

Dado que sólo tratamos con dos clases (TDAH y TD) y las observaciones estaban bien equilibradas (50% en cada clase), las métricas de precisión, especificidad y sensibilidad pueden utilizarse para evaluar el rendimiento (Luque et al., 2019; Hossin y Sulaiman, 2015). Dichas métricas se calcularon como el promedio obtenido de las matrices de confusión correspondientes a cada partición de VC. En este punto, nos centramos en la precisión (una métrica comúnmente utilizada en la práctica) para evaluar el rendimiento de la clasificación. Con este fin, calculamos el umbral teórico de significación estadística para la precisión basándonos en la distribución binomial acumulativa para p < 0.001, dos clases (probabilidad = 0,5) y un tamaño de muestra = 30 (Combrisson y Jerbi, 2015).

2) Wrapper

Partiendo del subconjunto de características previamente filtradas se afinó la selección mediante una función *wrapper*, personalizada y programada en MATLAB, que implementa un algoritmo de selección flotante secuencial hacia delante (*Sequential forward floating selection*, SFFS) (Somol et al., 2010; Pudil et al., 1994).

SFFS utiliza al propio clasificador como evaluador, comenzando con una selección vacía y añadiendo secuencialmente solo una característica cada vez para crear subconjuntos candidatos que son evaluados por VC. Después, entre todas las evaluadas, la mejor característica se añade a la selección. Cuando el conjunto seleccionado es >2, un paso hacia atrás intenta optimizarlo eliminando una o más características si fuese necesario. Este procedimiento se repite hasta que la adición de nuevas características ya no mejora la clasificación. El *wrapper* se aplicó independientemente a cada uno de los cuatro clasificadores con ambos esquemas de VC. Cabe destacar que el orden de entrada de las características finalmente seleccionadas nos permite conocer su importancia relativa. También merece la pena mencionar que nuestro *wrapper* puede clasificar características por múltiples métricas al mismo tiempo, y que en este trabajo utilizamos dos métricas concretas para seleccional (o más) características mejoran por igual la especificidad, se selecciona la que tenga mejor precisión.

Una vez seleccionado el mejor subconjunto de características para cada clasificador, estimamos la significación estadística del rendimiento observado mediante un procedimiento no paramétrico de permutación de etiquetas que no asume ninguna propiedad estadística particular de los datos (Combrisson y Jerbi, 2015; Bzdok, 2017; Golland y Fischle, 2003). Generamos 5000 remuestreos, cada uno permutando aleatoriamente las etiquetas de las dos clases y bajo la hipótesis nula de que las características no definen la pertenencia a una clase. Para cada remuestreo se evaluó el rendimiento de la clasificación utilizando el mismo esquema de VC que para los datos reales. Las métricas de rendimiento observadas se compararon con la distribución nula correspondiente para estimar un valor p. Además, se estimó el intervalo de confianza (IC) al 95% mediante el método de percentiles con corrección de sesgo para cada métrica; usando bootstrapping (con reemplazo) sobre 2000 remuestreos, manteniendo la misma proporción de clases (50%) y al menos tres observaciones distintas en cada clase (Fu et al., 2005). Finalmente, comprobamos si el rendimiento del wrapper podría estar sesgado debido al uso de características filtradas, es decir, el efecto "peeking" (Kriegeskorte et al., 2009; Pulini et al., 2019). Para ello, aplicamos el wrapper utilizando el conjunto original de características (sin filtrar) y comparamos los resultados de rendimiento.

3. RESULTADOS

3.1. Análisis PSD

La Figura 10 muestra la PSD promedio de las señales fNIRS para el grupo DT y TDAH en cada ROI (derecho, medio, izquierdo). Se observa un patrón común con un pico alrededor de 33 mHz correspondiente a la frecuencia de la tarea y un pico más prominente alrededor de 4 mHz. Otras oscilaciones, con menor magnitud, estaban presentes sobre los 17, 24 y 67 mHz. El test de permutación solo mostró diferencias significativas entre los grupos a 33 mHz para CS-HbO en el ROI derecho. En todos los casos, la mayor parte de la energía espectral se localizó en el rango de frecuencia 0 a 50 mHz.



Figura 10. PSD normalizados promedio para cada ROI (derecha, media e izquierda), cromóforo y tipo de señal en el rango de frecuencia de 0 a 80 mHz. Las curvas de colores gruesos se refieren al grupo DT (HbO en rojo y HbR en azul) mientras que las negras gruesas corresponden al grupo TDAH. Las líneas finas representan el IC del 95 % de la media. Rectángulos magenta delimitan los rangos de frecuencia que muestran diferencias significativas entre grupos. El panel SS muestra los PSD para señales poco profundas y CS para señales limpias.

3.2. Mapas de tiempo-frecuencia ISS

La Figura 11 muestra las variaciones ISS de la SS en el plano tiempo-frecuencia obtenidas a partir de la CWT para ambos grupos; dentro del rango 3 a 50 mHz.



Figura 11. Mapas de color de sincronización entre sujetos (ISS) para las señales superficiales (SS) en los diferentes ROIs. Las gráficas de la parte superior corresponden a las señales de HbO para los grupos TD y TDAH, mientas que las de la parte inferior muestran las de HbR. A la izquierda de cada mapa se muestran un gráfico con los valores máximos de ISS en cada una de las bandas de frecuencia. Las líneas verticales negras delimitan el intervalo de la tarea. Las líneas discontinuas blancas delimitan las diferentes subbandas (etiquetas con letras A - D) obtenidas con base en los valores máximos de ISS en cada una de las regiones de interés.

De manera similar, en la Figura 12 se muestran los mapas de tiempo-frecuencia de ISS correspondiente a la CS. A primera vista se pueden observar zonas de sincronización bien definidas en determinadas subbandas de frecuencias, algunas similares en ambos grupos y otras claramente diferentes. A continuación, destacaremos algunas de ellas a modo de ejemplo. En lo referente a SS, se puede observar una intensa sincronización en todos los ROIs alrededor de 33 mHz para la HbO del grupo DT, y en los ROI derecho y medio a 7 mHz y por debajo de 4 mHz. El grupo de TDAH presenta un ISS aún más intenso alrededor de 4 mHZ en todas las regiones de interés, pero mucho menos evidente a 7 o 33 mHz. Respecto a la CS, aunque en menor medida, el grupo DT también sincroniza a 33 mHz mientras que el grupo TDAH lo hace de una forma más débil y difusa. Además, el grupo DT parece sincronizarse con mayor intensidad durante la primera parte de la tarea a 7 mHz (ROI medio e izquierdo), mientras que el grupo TDAH se sincroniza cerca del final alrededor de 17 mHz. En conclusión, el análisis

de ISS revela la existencia de una pluralidad de subbandas que podrían aportar información sobre las similitudes y diferencias entre ambos grupos.



Figura 12. Mapas de color de sincronización entre sujetos (ISS) para las señales limpias (CS) en los diferentes ROIs. Las gráficas de la parte superior corresponden a las señales de HbO para los grupos TD y TDAH, mientas que las de la parte inferior muestran las de HbR. A la izquierda de cada mapa se muestran un gráfico con los valores máximos de ISS en cada una de las bandas de frecuencia. Las líneas verticales negras delimitan el intervalo de la tarea. Las líneas discontinuas blancas delimitan las diferentes subbandas (etiquetas con letras A - D) obtenidas en base a los valores máximos de ISS en cada una de las regiones de interés.

3.3. Subbandas de frecuencia y shapelets referencia

La tabla 1 muestra la sincronización en las diferentes bandas (etiquetadas como A, B, C y D, donde A corresponde a la frecuencia de la tarea) estimada a partir de los valores máximos de ISS promediando los diferentes ROIs. La Figura 13 ilustra el proceso de delimitación de estas bandas mediante la localización de las regiones con ISS mínimos adyacentes a un valor máximo. Se pueden observar máximos (picos) de ISS en las subbandas A y C para la HbO del grupo DT tanto en SS como en CS, mientras que el grupo TDAH muestra picos considerables en D para SS y en B para CS. Respecto a HbR, se observan picos claros en A y D en todos los casos. Debe tenerse en cuenta que dentro de una misma subbanda algunos picos se encuentran claramente desplazados en frecuencia dependiendo del grupo (por ejemplo, subbanda C para CS-HbO). Por lo tanto, resulta evidente que los máximos de ISS revelan diferencias en determinadas frecuencias.

Las diferentes subbandas también se representan mediante líneas discontinuas blancas en los mapas de ISS de las Figuras 11 y 12.

			Subbanda de frecuencia (mHz)				
Tipo de señal	Cromóforo	Grupo	А	В	С	D	
	НЬО	TD	50 - 21.8 21.8 - 50 - 23.3 23.3 - 50 - 23.0 23.0 - 50 - 23.0 23.0 -	21.8 - 11.7	11.7 - 4.7	4.7 – 3.0	
55	Odi i	TDAH	50 - 23.3	23.3 - 14.3	14.3 - 6.7	6.7 – 3.0	
00		TD	50 - 23.0	23.0 - 14.3	14.3 - 5.4	5.4 – 3.0	
	TIDIN	TDAH 50 - 23.0		23.0 - 14.3	14.3 - 7.7	7.7 – 3.0	
	НЬО	TD	50 - 21.8	21.8 - 11.7	11.7 - 5.4	5.4 – 3.0	
CS	Odi i	TDAH	50 - 20.3	20.3 - 7.7	7.7 - 4.7	4.7 – 3.0	
00		TD	50 - 21.8	21.8 - 11.7	11.7 - 6.7	6.7 – 3.0	
	ПЛК	TDAH	50 - 16.5	16.5 - 8.2	8.2 - 5.8	5.8 – 3.0	

Tabla 1. Límites de frecuencia de cada subbanda para cada tipo de señal, cromóforo y grupo.



Figura 13. Patrones promedio de sincronización entre sujetos (ISS) en todas las frecuencias. La graficas de la fila superior corresponde a los datos de SS y la fila inferior a CS. Los trazos rojos y azules se representan la HbO y HbR del grupo DT, mientras que los trazos negros muestran los datos del grupo TDAH. Cada subbanda se identifica con los rectángulos etiquetados horizontalmente, cuyos límites están definidos por las líneas discontinuas verticales.

La Figura 14 muestra las *shapelets* referencia obtenidas en casa subbanda para la SS, mientras que la Figura 15 muestra los correspondientes a la CS. Se observa una rica variedad de patrones, algunos de los cuales son similares entre los grupos y otros son claramente diferentes. Asi, por ejemplo, el grupo DT exhibe fluctuaciones rítmicas en la subbanda A de la SS-HbO, que son muy consistentes entre los participantes como refleja

la elevada ISS (trazo discontinuo); por el contrario, el grupo TDAH exhibe una mayor variabilidad entre sujetos. Otro ejemplo es el perteneciente a CS-HbR-D, donde los participantes de DT muestran un patrón consistente de incremento-disminución de hemoglobina mientras que dicho patrón no aparece en el grupo TDAH. También se puede observar que en CS-HbO-B los participantes TDAH se sincronizan hacia el final de la tarea, mientras que grupo DT lo hace visiblemente antes. En conclusión, parece que ciertas *shapelets* representan bien la respuesta prototípica de su grupo mientras que no se ajustan al otro.



Figura 14: *Shapelets* referencia de señales superficiales (SS) para cada subbanda (A-D) dentro de cada ROI. Las gráficas de la parte superior muestran los patrones de HbO (trazos rojos) para los grupos DT y TDAH, mientras que las gráficas de la parte inferior muestran las señales de HbR (trazos azules). Las líneas finas representan el error estándar de la media de cada una de la *shapelets*. Las líneas discontinuas muestran el curso temporal de la sincronización entre sujetos (ISS) en cada subbanda (escala en el eje derecho); en color magenta se muestran los correspondientes a *shapelets* con un valor p < .01. Las *shapelets* seleccionadas por el *wrapper* son identificadas con cuadrados de colores. Los rectángulos de color gris indican cada uno de los periodos de 15 segundos de cálculo mental durante la tarea.



Figura 15: *Shapelets* referencia de señales limpias (CS) para cada subbanda (A-D) dentro de cada ROI. Las gráficas de la parte superior muestran los patrones de HbO (trazos rojos) para los grupos DT y TDAH, mientras que las gráficas de la parte inferior muestran las señales de HbR (trazos azules). Las líneas finas representan el error estándar de la media de cada una de la *shapelets*. Las líneas discontinuas muestran el curso temporal de la sincronización entre sujetos (ISS) en cada subbanda (escala en el eje derecho); en color magenta se muestran los correspondientes a *shapelets* con un valor p <.01. Las *shapelets* seleccionadas por el *wrapper* son identificadas con cuadrados de colores. Los rectángulos de color gris indican cada uno de los periodos de 15 segundos de cálculo mental durante la tarea.

3.4. Rendimiento de los clasificadores

En la Tabla 2 se muestra el rendimiento de los diferentes clasificadores entrenados con las características seleccionadas mediante el filtrado basado en el estadístico *F*. Específicamente, cinco *shapelets* pertenecientes a la SS y diez de la CS cumplieron el criterio de filtrado (valores p <.01) (Señales magentas en las Figuras 14 y 15). En general el rendimiento de todos los clasificadores mejoró cuando se combinaron características provenientes de la SS y de la CS. Ambos esquemas de VC alcanzaron un valor de precisión $\geq 76.7\%$ lo que corresponde a un valor $p < 10^{-3}$ de acuerdo con la distribución teórica de la binomial acumulativa para dos clases y un tamaño muestral de 30 participantes. Respecto a la especificidad, ADL fue el clasificador que mostró valores más altos (88,7% y 93,3%, respectivamente, para VC de 5 iteraciones y dejando uno fuera). Cuando ambas señales fueron utilizadas de forma independiente el rendimiento de los clasificadores disminuyó, especialmente para la SS. Los clasificadores entrenados con

CS mostraron precisiones superiores al umbral mientras que solo RL y NB lo hicieron para la SS. Estos datos sugieren que la combinación de SS y CS contiene información útil para discriminar las clases con independencia del modelo de clasificación.

55-05.										
		SV	VM	R	RL		ADL		NB	
Tipo señal	P. metric	5FO	LOO	5-FO	LOO	5-FO	LOO	5-FO	LOO	
SS	Precisión	73,3	70	80	80	74,5	73,3	80,5	80	
5	Sensibilidad	78,7	73,3	85,7	86,7	80,7	80	88,7	86,7	
características	Especificidad	78,7	73,3	85,7	86,7	80,7	80	88,7	86,7	
CS	Precisión	81,2	80	83,5	80	78,7	80	85,5	86,7	
10	Sensibilidad	82,3	80	82	80	81,7	80	84,7	86,7	
características	Especificidad	82,3	80	82	80	81,7	80	84,7	86,7	
CS y SS	Precisión	88,7	86,7	88,2	90	85,7	83,3	85,7	86,7	
15	Sensibilidad	91,3	93,3	91,7	93,3	82,7	73,3	84	86,7	
características	Especificidad	86	80	84.7	86.7	88.7	93.3	87.3	86.7	

Tabla 2. Rendimiento alcanzado por los diferentes modelos de clasificación entrenados con el subconjunto de características seleccionadas mediante el filtro basado en el estadístico F para la SS, CS y SS+CS.

Nota: LOO = VC dejando uno fuera; 5-FO = 5 VC de 5 iteraciones; SVM = máquina de soporte vector; LR = regresión logística; ADL = análisis discriminante lineal; NB = *naïve* Bayes Gaussiano

La Figura 16 demuestra como el uso del *wrapper* mejoró ostensiblemente la selección de características. Todos los modelos de clasificación fueron muy parsimoniosos y en ninguno de los casos se emplearon más de tres características. Tras la incorporación del *wrapper*, las características provenientes de la CS mostraron mejor rendimiento en comparación con las de SS. Respecto al orden en la selección de características, en el caso de SS se seleccionó en primer lugar la perteneciente al grupo DT, subbanda A, HbO, ROI izquierdo o derecho dependiendo del clasificador (recuadro magenta 'TD-A-HbO-L' o el rojo 'TD -A-HbO-R' en la Figura 16). En lo referente a la CS, todos los clasificadores coinciden en las dos primeras características 'AD-B-HbO-R' y 'TD-D-HbR-L' (cuadrados verdes y amarillos, respectivamente, en la Figura 16). Nuevamente los mejores rendimientos se alcanzaron con la combinación de SS y CS. De hecho, RL y ADL mostraron una puntuación superior al 99% en las tres métricas para ambos esquemas de VC, lo que representa un rendimiento muy elevado. NB obtuvo métricas más bajas, entre el 91,3% al 93,3%, mientras que SVM fue el clasificador con peor rendimiento.



Figura 16. Rendimiento de los diferentes modelos de clasificación empleando el subconjunto de características seleccionadas por el *wrapper* para SS, CS y SS + CS. Los recuadros coloreados a la derecha identifican las características elegidas y su orden de entrada. Cada característica específica tiene un color diferente y su etiqueta es la combinación "Grupo-Banda-Cromóforo-ROI", siendo TD = grupo TD, AD =

TDAH, L = izquierda, M = media y R = derecha. Nota: LOO = VC dejando uno fuera; 5-FO = 5 VC de 5 iteraciones; SVM = máquina de soporte vector; LR = regresión logística; ADL = análisis discriminante lineal; NB = naïve Bayes Gaussiano.

Respecto a SS + CS, el wrapper seleccionó las mismas características para RL y ADL en ambas VC. La primera característica en introducirse fue 'AD-B-HbO-R', que también fue la primera en la CS. En segundo lugar, se añadió TD-A-HbO-L', que fue la primera en seleccionarse cuando solo se consideró la SS. Finalmente, se incluyó 'TD-D-HbR-M' de CS. Al observar el historial de interacciones del wrapper, comprobamos que la característica 'AD-B-HbO-R' logra, por si sola, alrededor del 80% de precisión, sensibilidad y especificidad, lo cual no es sorprendente ya que tiene el valor de estadístico F más elevado (19,6, p <0,0001). La adición de 'TD-A-HbO-L' mejoró las métricas hasta un 90% y con 'TD-D-HbR-M' se alcanzaron valores próximos al 100%. Por tanto, la característica con mayor poder de discriminación proviene de la CS-HbO del grupo TDAH, específicamente de la banda B en la que precisamente podemos observar un pico de ISS muy prominente (Figura 15). El grupo de DT aportó la siguiente característica proveniente de la SS en la frecuencia de la tarea y de HbO (advierta el pico de ISS en la subbanda A). Por último, la clasificación fue optimizada por un componente CS-HbR muy lento del grupo DT (ver el pico en D). Las shapelets que generaron estas características pueden ser identificadas mediante los cuadrados verde, rojo y azul, respectivamente (Figuras 14 y 15). Cabe destacar que NB también compartió la primera característica ('AD-B-HbO-R'), mientras que la segunda y la tercera provinieron del mismo grupo, subbanda y cromóforo, pero de un ROI adyacente.

En la Tabla 3 se muestran las métricas obtenidas mediante la selección de características de la SS + CS mediante el *wrapper*. En todos los casos (excepto para la sensibilidad de SVM en la VC de 5 iteraciones) los análisis de permutación indicaron valores p < .001. Los modelos de clasificación RL y ADL fueron los que mostraron las mejores métricas y los IC más estrechos. En comparación NB obtuvo un rendimiento más bajo, aunque significativo, mientras que SVM fue el modelo que peor desempeño mostro.

Cabe destacar que cuando se utilizó el conglomerado completo de características (48 SS+ 48 CS) en el entrenamiento de la RL y ADL se obtuvieron exactamente el mismo subconjunto de características para ambos esquemas de VC y, por lo tanto, las mismas métricas y estadísticas. En consecuencia, queda claro que la preselección de características no indujo métricas artificialmente optimistas, aunque si redujo el coste computacional.

Tabla 3. Tabla 2. Rendimiento alcanzado por los diferentes modelos de clasificación entrenados con el subconjunto de mejores características seleccionadas por el wrapper para SS + CS. La significación estadística se indica mediante valores p y los IC del 95 % se representan entre corchetes.

		SV	/M	RL		ADL		NB	
Señal	P. metric	5-FO	LOO	5-FO	LOO	5-FO	LOO	5-FO	LOO
CS & SS	Precisión	$82.5 \\ p < .001 \\ [68.2, 95.7]$	93.3 <i>p</i> < .001 [80.0, 100]	99.3 p < .001 [95.8, 100]	$100 \ p < .001 \ [96.7, 100]$	99.5 p < .001 [96.7, 100]	$100 \ p < .001 \ [96.7, 100]$	92.2 p < .001 [80.3, 98.0]	93.3 p < .001 [83.3, 100]
	Sensibilidad	$72.3 \\ p = .046 \\ [56.0, \\ 73.3]$	86.7 p < .001 [66.7, 100]	99.7 p < .001 [93.3, 100]	$100 \ p < .001 \ [93.3, 100]$	99.7 p < .001 [92.0, 100]	100 <i>p</i> < .001 [93.3, 100]	91.3 p < .001 [73.3, 99.7]	93.3 p < .001 [80.0, 100]
	Especificidad	92.7 p < .001 [78.2, 100]	$100 \ p < .001 \ [86.7, 100]$	99.0 p < .001 [91.8, 100]	$100 \ p < .001 \ [93.3, 100]$	99.3 p < .001 [90.4, 100]	100 p < .001 [93.3, 100]	93.0 <i>p</i> < .001 [80, 100]	93.3 p < .001 [80.0, 100]
	Características	2	3	3	3	3	3	3	3
Nota: LOO = VC dejando uno fuera; 5-FO = 5 VC de 5 iteraciones; SVM = máquina de soporte vector; LR = regresión logística; ADL = análisis discriminante lineal; NB = naïve Bayes Gaussiano									

67

4. DISCUSIÓN

El objetivo general del presente trabajo era desarrollar un marco metodológico para la identificación de biomarcadores funcionales, basados en propiedades espectrales de las señales fNIRS, aplicables en herramientas computacionales de apoyo a la evaluación objetiva del TDAH. Los resultados mostrados en apartados anteriores nos permiten afirmar que dicho objetivo se ha alcanzado. En este sentido, hemos mostrado que, mediante una adecuada selección de características, los clasificadores obtuvieron métricas predictivas realmente elevadas. Además, nos gustaría recalcar que las características utilizadas en la construcción de los modelos fueron obtenidas mediante un enfoque unimodal centrado únicamente en datos NIRS. También es interesante señalar que las características utilizadas se encuentran vinculadas a patrones oscilatorios que son visualmente reconocibles lo que incrementa su interpretabilidad. Por último, hemos empleado únicamente algoritmos lineales supervisados sin optimizar sus hiperparámetros y que, además, son fácilmente interpretables. Por lo tanto, queda claro que no hemos pretendido incrementar el rendimiento de los modelos a expensas de combinar diferentes tipos de características o de parametrizar los clasificadores, sino que hemos explorado la viabilidad de un marco metodológico para la identificación de biomarcadores funcionales, de una manera lo suficientemente accesible e interpretable como para ser usados en la práctica clínica.

4.1. Inducción de oscilaciones por medio de la aritmética mental

En relación con el primer objetivo específico, descubrimos que la tarea aritmética indujo con éxito cambios cíclicos en las señales SS y CS, en forma de picos espectrales altamente característicos centrados en la frecuencia de la tarea (es decir, 33 mHz). Estos picos no parecen discernir entre los grupos TDAH y DT ya que solo aparecieron diferencias significativas para CS-HbO en el ROI derecho. La capacidad de esta tarea rítmica de aritmética mental para inducir oscilaciones NIRS ya había sido testeada en un trabajo con jóvenes universitarios (Molina-Rodríguez et al., 2022). Sin embargo, aquí por primera vez mostramos que este hecho es extensible a poblaciones infantiles y con diagnóstico de TDAH.

4.2. Selección de características y rendimiento de clasificadores.

En cuanto a los objetivos específicos cinco, seis, siete y ocho podemos afirmar que hemos corroborado que la selección eficiente de características mejora y estabiliza el rendimiento de los diferentes modelos de predicción, siendo particularmente importante cuando la cantidad de características excede al número de observaciones como suele ocurrir con tamaños muestrales reducidos (Jović et al., 2015). La selección de características tiene como objetivo descartar predictores irrelevantes/redundantes, lo que reduce el sobreajuste, mejora la capacidad predictiva, y conduce a modelos de clasificación más simples, interpretables y eficaces en terminos computacionales. Dado que encontrar el subconjunto ideal de características mediante una búsqueda exhaustiva entre todas las posibles suele ser poco práctico, los métodos se selección subóptimos se utilizan habitualmente como soluciones alternativas. Numerosos estudios han propuesto diferentes estrategias para la selección de características; sin embargo, no se ha encontrado ninguna aproximación que funcione de manera óptima en todos los escenarios (Somol et al., 2010; Remeseiro y Bolon-Canedo, 2019). En el presente trabajo hemos empleado un enfoque combinado basado en la obtención de un primer conjunto de características candidatas mediante una prueba estadística univariada para luego aplicar un wrapper para encontrar el mejor subconjunto utilizando el propio algoritmo de clasificación como evaluador. Dado que nuestro tamaño muestral era relativamente pequeño y que los métodos de selección de características más complejos podían ser propensos al sobreajuste, nos decantamos por una aproximación simple y totalmente independiente del método de aprendizaje (Vabalas et al., 2019). En el primer paso, utilizamos el estadístico F para seleccionar aquellas características con valores p < .01. A pesar de su simplicidad, esta aproximación fue lo suficientemente buena como para lograr precisiones estadísticamente significativas ($p < 10^{-3}$ de acuerdo con la distribución teórica de la binomial acumulativa) para todos los clasificadores y esquemas de VC cuando se utilizaron la CS y SS + CS. De hecho, utilizando SS + CS encontramos precisiones entre el 83,3% y el 90% que son comparables y en algunos casos mejores a los valores obtenidos en trabajos anteriores (Monden et al., 2015; Crippa et al., 2017; Gu et al., 2018). Por lo tanto, las características seleccionadas mediante este método parecen ser intrínsecamente distintivas y muestran un funcionamiento adecuado con independencia del modelo de clasificación y esquema de VC utilizado.

A pesar de estos resultados prometedores, era posible que algunas características irrelevantes estuvieran agregando innecesariamente cierta complejidad e impidieran alcanzar un rendimiento óptimo. Para abordar esta posibilidad aplicamos el método *wrapper* para afinar en mayor medida la selección de características. Específicamente, empleamos un algoritmo SFFS que permite evitar el "efecto de anidamiento" que ocurre cuando una característica previamente seleccionada ya no puede ser eliminada (Somol et al., 2010; Liu y Yu, 2005; Amann et al., 2022; Jović et al., 2015; Kim, 2009; Luque et al., 2019; Hossin y Sulaiman, 2015; Combrisson y Jerbi, 2015; Pudil et al., 1994). De hecho, comprobamos como nuestro algoritmo SFFS realizó con éxito "pasos hacia atrás" para encontrar mejores soluciones. Sorprendentemente, la utilización del método *wrapper* condujo a modelos muy sencillos formados por no más de 3 características (relación características-muestra de 3/30) en los que es poco probable que se esté produciendo un sobreajuste (Vabalas et al., 2019).

En relación con los objetivos específicos 9 y 10, encontramos que el rendimiento fue particularmente elevado cuando RL y ADL fueron entrenados con el mismo subconjunto de características provenientes de SS + CS, logrando puntuaciones alrededor del 100% en todas las métricas y esquemas de VC. NB obtuvo métricas ligeramente más bajas, aunque superiores al 91%. Además, este clasificador mostró acuerdo en la elección de la primera característica con RL y ADL mientras que las otras dos seleccionadas mostraron propiedades similares a las elegidas por RL y ADL (Figura 16). En conjunto, estos resultados sugieren que las características escogidas son muy estables entre los diferentes algoritmos de clasificación y esquemas de VC. Este hecho es especialmente observable en el caso de CS, en el que el consenso entre clasificadores es aún más amplio (Figura 16). Por todo ello, parece plausible considerar que el rendimiento observado no se debió a un efecto de sobre-selección ("*peeking*") (Raudys, 2006), sino más bien a la verdadera naturaleza distintiva de las características seleccionadas.

Sorprendentemente, SVM obtuvo el peor desempeño cuando fue entrenado con la SS + CS, especialmente en el esquema de VC de 5 iteraciones donde mostró las puntuaciones más bajas y los IC más amplios. En cambio, el resto de los clasificadores no se vieron afectados por la VC, alcanzando puntuaciones e IC similares con ambos esquemas (Tabla 3). Como ya hemos mencionado, uno de nuestros propósitos era evitar cualquier ajuste de hiper-parámetros y, por lo tanto, descartamos el uso de SVM en el presente estudio debido a sus métricas de clasificación.

Como se puede observar en la Figura 16 y en la Tabla 3, RL y ADL obtuvieron el mejor rendimiento cuando se entrenaron con las características de SS+CS. Ambos clasificadores coincidieron en la misma solución propuesta por el wrapper, que era exactamente la misma que para las características preseleccionadas mediante el filtrado utilizando el estadístico F y para el conjunto completo. Además, tanto RL como ADL lograron puntuaciones comparables en todas las métricas (>99 % y con p < 0,001utilizando el remuestreo para definir la hipótesis nula). Además, el rango de los ICs al 95% fue igualmente estrecho, aunque probablemente sesgado debido a que las puntuaciones fueron cercanas al 100%. Se sabe que la VC tiende a generar IC más estrechos a medida que la precisión se acerca al 100%, pero estos intervalos se vuelven más amplios y asimétricos a medida que disminuye el tamaño de la muestra, lo que puede subestimar los errores de predicción, especialmente en el esquema "dejando uno fuera" (Varoquaux, 2018). A pesar de trabajar con una muestra de 30 participantes, encontramos que el límite inferior del IC de la precisión nunca fue <95,8%, estuvo a menos del 5% de la media en todos los casos, con una desviación por debajo del 15% máximo esperado para una clasificación binaria con nuestro tamaño muestral (Varoquaux, 2018). Por lo tanto, los resultados sugieren que las características seleccionadas cumplen con los supuestos estadísticos requeridos por la RL y ADL para hacer que las observaciones sean separables en dos clases. El resto de la discusión se centra en RL y ADL y en sus tres características compartidas seleccionadas por el wrapper.

4.3. Acuerdo en la VC

Los esquemas de VC "dejando uno fuera" y de "5 iteraciones" son los adoptados con mayor frecuencia en los estudios de TDAH (Loh et al., 2022). Algunos autores han sugerido que "dejando uno fuera" puede ser útil en un escenario diagnóstico (Saeb et al., 2017; Tougui et al., 2021), mientras que otros recomiendan el uso de K particiones o bien divisiones aleatorias repetidas para estimaciones más estables (Hastie et al., 2009; Varoquaux et al., 2017). En cualquier caso, se sabe que la VC puede verse comprometida por los tamaños muestrales reducidos, particularmente si el número de predictores es elevado, lo que puede producir sobrestimaciones de la precisión en un grado variable y dependiendo de las particularidades del estudio (Arbabshirani et al., 2017; Flint et al., 2021). Por ello, esperábamos que los diferentes esquemas VC mostrasen diferencias en el rendimiento, en el que "dejando uno fuera" obtuviera IC más amplios y métricas más optimistas (Varoquaux, 2018). Sin embargo, encontramos que ambos esquemas mostraron resultados muy similares para RL y ADL (Tabla 3), lo que sugiere que las características seleccionadas son predictoras lo suficientemente buenos para obtener métricas estables e independientes del esquema de VC.

4.4. Contribución de SS y CS a la clasificación

En relación con los objetivos específicos número dos, tres y cuatro debemos comentar en primer lugar que este trabajo no se basó en medidas de activación/desactivación como es habitual en los estudios fNIRS, sino más bien en el estudio de la información distintiva contenida en los diferentes componentes de frecuencia. Entre otras, esta es una de las razones por las que preferimos denominar a la CS de esta forma, sin mayores suposiciones sobre su verdadera naturaleza. Por otro lado, no consideramos a la SS únicamente como un componente contaminador a descartar, sino que creemos que puede contener información representativa de cada una de las clases debido a la estrecha interacción entre el funcionamiento cognitivo y autonómico (Forte et al., 2019a; Forte et al., 2019b; Wang et al., 2016). De hecho, cuando combinamos SS y CS se alcanzó una precisión cercana al 100%, lo que supera a otros estudios de clasificación que utilizan datos fNIRS unimodales.

Sorprendentemente, la primera característica seleccionada provino de la CS del grupo TDAH en forma de un patrón consistente de HbO en la banda B (7,7–20,3 mHz) del ROI derecho (Figura 14 y 15). Específicamente, dicho patrón se sincroniza cada vez más entre los participantes a medida que la tarea avanza y alcanza su punto máximo cerca del final, mientras que el grupo de DT alcanza su punto máximo de sincronización más rápidamente y disminuye hacia el final (observamos evoluciones similares en los tres ROIs). Por lo tanto, parece que esta actividad hemodinámica es una característica distintiva del grupo TDAH a esta frecuencia. Los estudios de resonancia magnética funcional en estado de reposo (rs-fMRI) han informado de anomalías en la actividad espontanea local de diversas regiones cerebrales, detectadas en la banda de frecuencia típica de fMRI de 10 a 80 mHz (Tang et al., 2018), y también en otras subbandas más estrechas como 0-10, 10-27 y 27-73 mHz, (Yu et al., 2016). Sin embargo, no existe en la literatura una conclusión consistente al respecto debido a la elevada diversidad en la presentación del TDAH. Sin embargo, junto con los hallazgos de los estudios de conectividad rs-fMRI (Wang y Li, 2015; Castellanos y Proal, 2012), la evidencia actual sugiere que el análisis de oscilaciones de baja frecuencia podría proporcionar información valiosa sobre la disfunción de redes cerebrales observadas en el TDAH (para una revisión
ver: Posner et al., 2020 o Castellanos y Aoki, 2016). En este sentido, se puede extraer pocas conclusiones consensuadas del gran número de resultados heterogéneos presentado en los estudios mencionados. Sin embargo, la mayoría de ellos coinciden en señalar el papel clave del funcionamiento anómalo de la CPF en las redes por defecto y de atención/saliencia observadas en el TDAH, probablemente relacionadas con un retraso en la maduración cortical (Shaw, 2007; Halperin y Schulz, 2006). Dado que nuestro dispositivo NIRS explora parte de la CPF, existe la posibilidad de que el patrón distintivo de TDAH se deba a una desactivación retrasada/interferida de la red por defecto durante la tarea, reflejándose como un componente de frecuencia especifico. En consonancia con este hallazgo, Salmi et al. (2020) encontraron una mayor sincronización de la CPF medial en los pacientes con TDAH asociada con la impulsividad durante una tarea de atención naturalística.

Se podría aplicar una discusión similar a la tercera característica que también surgió de la CS, pero del cromóforo HbR en la banda D de ultrabaja frecuencia del grupo de DT (también visibles en todos los ROIs). Específicamente, observamos un patrón de incremento de HbR al inicio de la tarea con un retorno lento a niveles más bajos y un nuevo incremento al final (Figura 15). Es difícil de dilucidar únicamente con los datos fNIRS si el efecto inicial se debe al consumo sostenido de oxígeno, a la desactivación o a la interrelación de ambos. Curiosamente, se ha informado que las señales BOLD ultralentas (0-10 mHz) parecen discriminar el TDAH del DT (Yu et al., 2016). Si bien es cierto que se requiere de más investigaciones, no se debe subestimar la importancia de considerar tanto la HbO como la HbR.

La segunda característica seleccionada pertenece a la HbO-SS en la frecuencia de la tarea y contribuyó significativamente a mejorar la clasificación (Figura 16). Obviamente, este patrón no es de origen cerebral, sino que refleja una actividad hemodinámica sistémica (incluso local) coordinada y registrada en los tejidos superficiales de la frente. Específicamente, esta característica SS pertenece al grupo DT que mostraba mayor sincronización entre participantes que el grupo TDAH (Figura 14). Esta falta de sincronía probablemente refleja la incapacidad de los participantes con TDAH para coordinar los recursos autonómicos con las demandas cognitivas (Bellato et al., 2020). Este hallazgo está en línea con otros trabajos que utilizan imágenes térmicas (Coben y Myers, 2009) o medidas de entropía del electrocardiograma (Koh et al., 2022) para diferenciar el TDAH con una precisión prometedora.

4.4. Precisión, sensibilidad y especificidad

La aproximación utilizando el wrapper favoreció en primer lugar la especificidad y después la precisión. La razón detrás de esta decisión es que, al primar la especificidad, los clasificadores son menos propensos a los falsos positivos y, por tanto, más fiables a la hora de predecir el TDAH. Etiquetar a un niño como TDAH cuando no lo es puede conducir a la aplicación de tratamientos farmacológicos innecesarios, al estigma social y a otras consecuencias iatrogénicas. Por el contrario, un falso negativo puede descartarse con relativa facilidad mediante una evaluación complementaria y/o un seguimiento en el tiempo. Sin embargo, dado que la especificidad mostró una puntuación tan elevada y al hecho de que la muestra estaba perfectamente equilibrada, la precisión (y la sensibilidad) aumentaron en consecuencia, lo que produjo muy poca diferencia entre las distintas métricas. Aunque de escasa utilidad en este trabajo, consideramos que definir el orden de importancia de las diferentes métricas puede conducir a una mejor selección de datos relevantes a la hora de considerar los objetivos y particularidades de cada ensayo clínico. Por otro lado, aunque no tuvimos en cuenta medidas predictivas, como la prevalencia, creemos que las métricas utilizadas en este trabajo son suficientes como prueba de concepto, ya que son útiles en la evaluación de pruebas de detección/diagnóstico en la investigación clínica (Trevethan, 2017; Eusebi, 2013).

4.5. Limitaciones y trabajos futuros

Este trabajo no tiene como objetivo validar una prueba diagnóstica para TDAH, sino que pretende proponer un marco de trabajo, a modo de prueba de concepto, para encontrar biomarcadores funcionales que ayuden a desarrollar dicha prueba diagnóstica. Así pues, se hace necesario abordar algunas cuestiones para enmarcar correctamente los hallazgos de esta investigación. En primer lugar, a pesar del buen rendimiento de los modelos de clasificación, el tamaño limitado de la muestra podría haber llevado a una sobreestimación de las métricas. Sin embargo, debido a que nuestra muestra es muy homogénea en términos de edad, sexo y presentación de TDAH, el tamaño de muestra efectivo es comparable o incluso mayor al de investigaciones previas. La homogeneidad de nuestra muestra podría considerase una amenaza a la generalización de nuestros resultados a la población general, sin embargo, consideramos que es más bien una fortaleza. La condición TDAH es tan heterogénea que tratar de encontrar una solución "todo en uno" puede ser objetivo difícil de lograr. Creemos que encontrar biomarcadores funcionales para diferentes casos de usos (por ejemplos, niños que nunca han utilizado metilfenidato, niñas o adolescentes) podría ser una aproximación más productiva y prometedora. Obviamente, el marco de trabajo propuesto aquí puede ser aplicado fácilmente a otros casos de uso, proyecto que actualmente tenemos en marcha. También consideramos relevante investigar en el futuro si este enfoque puede beneficiarse de las "medidas muti-distancia extendidas" con correcciones de "*path length*" parcial (Song et al., 2020). En segundo lugar, consideramos que otra limitación de nuestro trabajo es la ausencia de un conjunto de datos para la validación de los modelos, lo que podría generar dudas sobre la generalización de nuestros resultados. Por lo tanto, a pesar de nuestros resultados prometedores queremos enfatizar que para probar la utilizad del marco de trabajo actual como apoyo a las decisiones clínicas, sería necesario investigaciones con mayores tamaños muestrales y una validación externa adecuada.

5. CONCLUSIONES

- 1. Una tarea de aritmética mental rítmica tiene la capacidad de inducir oscilaciones fNIRS tanto en chicos TDAH como DT.
- 2. Las técnicas de análisis espectral tiempo-frecuencia permiten separar los diferentes componentes oscilatorios presentes en las señales fNIRS.
- 3. El análisis de sincronización propuesto permite identificar las bandas de frecuencia más relevantes para diferenciar entre TDAH y DT.
- 4. Las oscilaciones fNIRS de muy baja frecuencia, inducidas/moduladas por una tarea mental rítmica, discriminan fiablemente las condiciones TDAH y DT.
- 5. La utilización del método *wrapper* ha demostrado ser útil para identificar el mejor subconjunto de características.
- 6. El método *wrapper* generó modelos predictivos muy sencillos formados por no más de 3 características.
- 7. Una selección eficiente de características mejora y estabiliza el rendimiento de los diferentes modelos predictivos.
- 8. Los clasificadores RL y ADL lograron puntuaciones de precisión, sensibilidad y especificidad de casi el 100 %, superando a otros estudios fNIRS similares.
- 9. SS no es un mero componente contaminador a descartar, sino que contiene información representativa de cada una de las condiciones, TDAH y DT.
- 10. Combinando SS y CS se alcanzó una precisión cercana al 100%, lo que supera a otros estudios de clasificación que utilizan solo señales fNIRS profundas.
- 11. Las características escogidas son muy estables entre los diferentes algoritmos de clasificación y esquemas de VC.

6. REFERENCIAS

- Aarabi, A. and Huppert, T. J. (2016). Characterization of the relative contributions from systemic physiological noise to whole-brain resting-state functional near-infrared spectroscopy data using single-channel independent component analysis. *Neurophotonics 3*, 025004.
- Abikoff, H., Courtney, M., Pelham Jr, W. E. y Koplewicz, H. S. (1993). Teachers' ratings of disruptive behaviors: The influence of halo effects. *Journal of abnormal child psychology*, 21(5), 519-533.
- Adesman, A. R. (2001). The diagnosis and management of attention-deficit/hyperactivity disorder in pediatric patients. *Primary care companion to the Journal of clinical psychiatry*, *3*(2), 66.
- ADHD-200 consortium. (2012). The ADHD-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. *Frontiers in systems neuroscience*, *6*, 62.
- Ajith, K., Menaka, R., y Kumar, S. S. (2021). *EEG based mental state analysis*. *In Journal* of *Physics: Conference Series*. IOP Publishing.
- Amann, J., Vetter, D., Blomberg, S. N., Christensen, H. C., Coffee, M., Gerke, S., ... y Z-Inspection initiative. (2022). To explain or not to explain?—Artificial intelligence explainability in clinical decision support systems. *PLOS Digital Health*, 1(2), e0000016.
- American Psychiatric Association (1968). DSM-II. Diagnostic and statistical manual of mental disorders, second edition. Washington DC: APA.
- American Psychiatric Association (1980). DSM-III. Manual diagnóstico y estadístico de los trastornos mentales. Texto revisado. Barcelona: Masson.
- American Psychiatric Association (2014). DSM-5. Manual diagnóstico y estadístico de los trastornos mentales. Madrid: Editorial Médica Panamericana.
- Andersen, A. V., Simonsen, S. A., Schytz, H. W., y Iversen, H. K. (2018). Assessing lowfrequency oscillations in cerebrovascular diseases and related conditions with

near-infrared spectroscopy: a plausible method for evaluating cerebral autoregulation? Neurophotonics, 5(03), 1. https://doi.org/10.1117/1.nph.5.3.030901

- Andrés, M. A., Catala, M. A. y Gómez-Beneyto, M. (1999). Prevalence, comorbidity, risk factors and service utilisation of disruptive behaviour disorders in a community sample of children in Valencia (Spain). *Social psychiatry and psychiatric epidemiology*, 34(4), 175-179.
- Arai, S., Okamoto, Y., Fujioka, T., Inohara, K., Ishitobi, M., Matsumura, Y., ... y Kosaka, H. (2016). Altered frontal pole development affects self-generated spatial working memory in ADHD. *Brain and development*, 38(5), 471-480.
- Arbabshirani, M. R., Plis, S., Sui, J., & Calhoun, V. D. (2017). Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage*, 145, 137-165.
- Arsalidou, M., y Taylor, M. J. (2011). Is 2+2=4? Meta-analyses of brain areas needed for numbers and calculations. *NeuroImage*, 54(3), 2382–2393. https://doi.org/10.1016/j.neuroimage.2010.10.009
- Bastos, A. M., y Schoffelen, J. M. (2016). A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Frontiers in systems neuroscience*, 9, 175.
- Bellato, A., Arora, I., Hollis, C. y Groom, M. J. (2020). Is autonomic nervous system function atypical in attention deficit hyperactivity disorder (ADHD)? A systematic review of the evidence. *Neuroscience and biobehavioral reviews*, 108, 182–206.
- Berens, P. (2009). CircStat: a MATLAB toolbox for circular statistics. *Journal of* statistical software, 31, 1-21.
- Bolt, T., Nomi, J. S., Vij, S. G., Chang, C., y Uddin, L. Q. (2018). Inter-subject phase synchronization for exploratory analysis of task-fMRI. *NeuroImage*, 176, 477-488.

- Brigadoi, S. y Cooper, R. J. (2015). How short is short? Optimum source-detector distance for short-separation channels in functional near-infrared spectroscopy. *Neurophotonics*, 2(2), 025005.
- Burgess, P. W. y Wu, H-C. (2013). Rostral prefrontal cortex (Brodmann area 10): metacognition in the brain. (2013). En D. T. Stuss & R. T. Knight (Eds.), *Principles of Frontal Lobe Function* (pp. 524-534). Oxford University Press.
- Bzdok, D. (2017). Classical statistics and statistical learning in imaging neuroscience. *Frontiers in neuroscience*, *11*, 543.
- Cardo, E., Bustillo, M., Riutort, L., Bernad, M. M., Meisel, V., García-Banda, G. y Servera, M. (2009). ¿Cuál es la combinación de síntomas según padres y maestros más fiable para el diagnóstico de TDAH? *Anales de pediatría*, 71(2), 141-147.
- Cardo, E., Bustillo, M. y Servera, M. (2007). Valor predictivo de los criterios del DSM-IV en el diagnóstico del trastorno por déficit de atención/hiperactividad y sus diferencias culturales. *Revista de Neurología*, 44(2), 19-22.
- Cardo, E. y Servera-Barceló, M. (2005). Prevalencia del trastorno de déficit de atención e hiperactividad. *Revista de Neurología*, 40(1), 11-15.
- Cardo, E. y Servera, M. (2007). ADHD Rating Scale-IV en una muestra escolar española: datos normativos y consistencia interna para maestros, padres y madres. Revista de Neurología, 45, 393-9.
- Carlén, M. (2017). What constitutes the prefrontal cortex? Science, 358(6362), 478–482.
- Casanova, J. A. A., Campos, J. A. A., Alecha, M. A. I. y Cebollero, M. P. (2003). Estructura factorial de la Escala de Conners para profesores en muestras comunitaria y clínica. *Revista de psicología general y aplicada: Revista de la Federación Española de Asociaciones de Psicología*, 56(2), 173-184.
- Castellanos, F. X., y Proal, E. (2012). Large-scale brain systems in ADHD: beyond the prefrontal-striatal model. *Trends in cognitive sciences*, *16*(1), 17-26.
- Castellanos, F. X. y Aoki, Y. (2016). Intrinsic functional connectivity in attentiondeficit/hyperactivity disorder: a science in development. *Biological psychiatry: cognitive neuroscience and neuroimaging*, 1(3), 253-261.

- Castellar, R. G., Herrero, M. J. P., Hierro, R. S. y Casas, A. M. (2006). Estado sociométrico de los niños con trastorno por déficit de atención con hiperactividad subtipo combinado. *Revista de Neurología*, 42(2), 13-17.
- Catalá-López, F., Peiró, S., Ridao, M., Sanfélix-Gimeno, G., Gènova-Maleras, R. y Catalá, M. A. (2012). Prevalence of attention deficit hyperactivity disorder among children and adolescents in Spain: a systematic review and meta-analysis of epidemiological studies. *BMC Psychiatry*, 12(1), 1-13.
- Cauli, B. y Hamel, E. (2010). Revisiting the role of neurons in neurovascular coupling. *Frontiers in neuroenergetics*, *2*, 9.
- Coben, R., y Myers, T. E. (2009). Sensitivity and specificity of long wave infrared imaging for attention-deficit/hyperactivity disorder. *Journal of Attention Disorders*, 13(1), 56-65.
- Cohen, A., y Kovacevic, J. (1996). Wavelets: The mathematical background. *Proceedings of the IEEE, 84*(4), 514-522.
- Cohen, L. (1995). Time-frequency analysis. New Jersey: Prentice hall.
- Cohen, M. X. (2014). *Analyzing neural time series data: theory and practice*. MIT press.
- Conners, C. K., Pitkanen, J. y Rzepa, S. R. (2011). Conners 3rd Edition (Conners 3; Conners 2008). En Encyclopedia of Clinical Neuropsychology, 675-678. New York: Springer.
- Combrisson, E., y Jerbi, K. (2015). Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of neuroscience methods*, 250, 126-136.
- Crippa, A., Salvatore, C., Molteni, E., Mauri, M., Salandi, A., Trabattoni, S., Agostini, C., Molteni, M., Nobile, M. y Castiglioni, I. (2017). The utility of a computerized algorithm based on a multi-domain profile of measures for the diagnosis of attention deficit/hyperactivity disorder. *Frontiers in psychiatry*, *8*, 189.
- Delorme, A., y Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1), 9-21.

- Delpy, D. T., Cope, M., van der Zee, P., Arridge, S., Wray, S. y Wyatt, J. S. (1988). Estimation of optical pathlength through tissue from direct time of flight measurement. *Physics in Medicine & Biology*, 33(12), 1433.
- Dolu, N., Altınkaynak, M., Güven, A., Özmen, S., Demirci, E., İzzetoğlu, M., y Pektaş, F. (2019). Effects of methylphenidate treatment in children with ADHD: a multimodal EEG/fNIRS approach. *Psychiatry and Clinical Psychopharmacology*, 29(3), 285-292.
- Döpfner, M., Breuer, D., Wille, N., Erhart, M., Ravens-Sieberer, U., y Bella Study Group.
 (2008). How often do children meet ICD-10/DSM-IV criteria of attention deficit-/hyperactivity disorder and hyperkinetic disorder? Parent-based prevalence rates in a national sample–results of the BELLA study. *European child & adolescent psychiatry*, 17, 59-70.
- Eusebi, P. (2013). Diagnostic accuracy measures. *Cerebrovascular Diseases, 36*(4), 267-272.
- Faraone, S. V., Banaschewski, T., Coghill, D., Zheng, Y., Biederman, J., Bellgrove, M. A., Newcorn, J.H., Gignac, M., Al Saud, N. M., Manor, I., Rohde, L. A., Yang, L., Cortese, S., Almagor, D., Stein, M. A., Albatti, T. H., Aljoudi, H. F., Alqahtani, M. M. J., Asherson, P., Atwoli, L., Bölte, S., Buitelaar, J. K., Crunelle, C. L., Daley, D., Dalsgaard, S., Döpfner, M., Espinet, S., Fitzgerald, M., Franke, B., Gerlach, M., Haavik, J., Hartman, C. A., Hartung, C. M., Hinshaw, S. P., Hoekstra, P. J., Hollis, C., Kollins, S. H., Sandra Kooij, J. J., Kuntsi, J., Larsson H., Li, T., Liu, J., Merzon, E., Mattingly, G., Mattos, P., McCarthy, S., Mikami, A. Y., Molina, B. S. G., Nigg, J. T., Purper-Ouakil, D., Omigbodun, O. O., Polanczyk, G. V., Pollak, Y., Poulton, A. S., Rajkumar, R. P., Reding, A., Reif, A., Rubia, K., Rucklidge, J., Romanos, M., Ramos-Quiroga, J. A., Schellekens, A., Scheres, A., Schoeman, R., Schweitzer, J. B., Shah, H., Solanto, M. V., Sonuga-Barke, E., Soutullo, C., Steinhausen, H. C., Swanson, J. M., Thapar, A., Tripp, G., van de Glind, G., van de Brink, W., Van der Oord, S., Venter, A., Vitiello, B., Walitza, S. y Wang, Y. (2021). The world federation of ADHD international consensus statement: 208 evidence-based conclusions about the disorder. Neuroscience & Biobehavioral Reviews, 128, 789-818.

- Farré, M. y Narbona, J, (1989). Índice de hiperquinesia y rendimiento escolar: validación del cuestionario de Conners en nuestro medio. Acta Pediátrica Española, 47, 103-109.
- Farré-Riba, A. y Narbona, J. (1997). Escalas de Conners en la evaluación del trastorno por déficit de atención con hiperactividad: nuevo estudio factorial con niños españoles. *Revista de Neurología*, 25(138), 200-204.
- Fernández-Jaén, A., Martín Fernández-Mayoralas, D., Fernández-Perrone, A. L., Calleja-Pérez, B., Albert, J., López-Martín, S. y López-Arribas, S. (2016). Disfunción en el trastorno por déficit de atención/hiperactividad: evaluación y respuesta al tratamiento. *Revista de Neurología*, 62(Supl. 1), 79-S84.
- Ferrari, M. y Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *Neuroimage*, 63(2), 921-935.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals* of eugenics, 7(2), 179-188.
- Flint, C., Cearns, M., Opel, N., Redlich, R., Mehler, D. M., Emden, D., ... y Hahn, T. (2021). Systematic misestimation of machine learning performance in neuroimaging studies of depression. *Neuropsychopharmacology*, 46(8), 1510-1517.
- Forte, G., De Pascalis, V., Favieri, F. y Casagrande, M. (2019a). Effects of blood pressure on cognitive performance: A systematic review. *Journal of clinical medicine*, 9(1), 34.
- Forte, G., Favieri, F., & Casagrande, M. (2019b). Heart rate variability and cognitive function: A systematic review. *Frontiers in neuroscience*, *13*, 710.
- Fu, W. J., Carroll, R. J., y Wang, S. (2005). Estimating misclassification error with small samples via bootstrap cross-validation. *Bioinformatics*, 21(9), 1979-1986.
- Fulcher, B.D. (2018). Feature-Based Time-Series Analysis. En G. Dong y H. Liu (Eds.), *Feature engineering for machine learning and data analytics*, (pp. 87-116). CRC Press.

- Gagnon, L., Cooper, R. J., Yücel, M. A., Perdue, K. L., Greve, D. N. y Boas, D. A. (2012). Short separation channel location impacts the performance of short channel regression in NIRS. *Neuroimage*, 59(3), 2518-2528.
- Gagnon, L., Yücel, M. A., Boas, D. A. y Cooper, R. J. (2014). Further improvement in reducing superficial contamination in NIRS using double short separation measurements. *Neuroimage*, 85, 127-135.
- García Garrido, J. V., Grau Rubio, C., & Carcés Ferrer, J. (2014). Cuestionarios TDAH para profesores: un análisis desde los criterios del DSM-IV-TR y DSM-V. *Revista Española de Orientación y Psicopedagogía*, 25(1), 62-77.
- García Peñas, J. J. y Domínguez Carral, J. (2012). ¿Existe un sobrediagnóstico del trastorno por déficit de atención e hiperactividad (TDAH)? Evidencias en Pediatría, 51 (8), 1-5.
- García-Pérez, M., y Magaz-Lago, A. (2006). Escala Magallanes de Detección de Déficit de Atención (EMA-DDA). COHS.
- Gimeno Morales, M. y Galbe Sánchez-Ventura, J. (2015). Trastorno por déficit de atención e hiperactividad: ¿incrementa su prevalencia o se promueve su diagnóstico? *Pediatría Atención Primaria*, 17(65), e95-e98.
- Glerean, E., Salmi, J., Lahnakoski, J. M., Jääskeläinen, I. P., y Sams, M. (2012). Functional magnetic resonance imaging phase synchronization as a measure of dynamic functional connectivity. *Brain connectivity*, 2(2), 91-101.
- Golland, P., y Fischl, B. (2003). Permutation tests for classification: towards statistical significance in image-based studies. *En Biennial international conference on information processing in medical imaging (pp. 330-341)*. Springer Berlin Heidelberg.
- Gröchenig, K. (2001). Foundations of Time-Frequency Analysis. Birkhäuser Boston.
- Gu, Y., Miao, S., Han, J., Zeng, K., Ouyang, G., Yang, J. y Li, X. (2017). Complexity analysis of fNIRS signals in ADHD children during working memory task. *Scientific reports*, 7(1), 1-10.

- Gu, Y., Miao, S., Han, J., Liang, Z., Ouyang, G., Yang, J., y Li, X. (2018). Identifying ADHD children using hemodynamic responses during a working memory task measured by functional near-infrared spectroscopy. *Journal of neural engineering*, 15(3), 035005.
- Gualtieri, C. T. y Johnson, L. G. (2005). ADHD: Is objective diagnosis possible? *Psychiatry*, 2(11), 44-53.
- Guardiola, A., Fuchs, F. D. y Rotta, N. T. (2000). Prevalence of attention-deficit hyperactivity disorders in students: comparison between DSM-IV and neuropsychological criteria. *Arquivos de Neuro-psiquiatria*, 58(2B), 401-407.
- Güven, A., Altınkaynak, M., Dolu, N., İzzetoğlu, M., Pektaş, F., Özmen, S., Demircis, E. y Batbat, T. (2020). Combining functional near-infrared spectroscopy and EEG measurements for the diagnosis of attention-deficit hyperactivity disorder. *Neural Computing and Applications*, 32, 8367-8380.
- Haeussinger, F. B., Dresler, T., Heinzel, S., Schecklmann, M., Fallgatter, A. J., y Ehlis,
 A. C. (2014). Reconstructing functional near-infrared spectroscopy (fNIRS) signals impaired by extra-cranial confounds: an easy-to-use filter method. *NeuroImage*, 95, 69-79.
- Haeussinger, F. B., Heinzel, S., Hahn, T., Schecklmann, M., Ehlis, A. C., y Fallgatter, A. J. (2011). Simulation of near-infrared light absorption considering individual head and prefrontal cortex anatomy: implications for optical neuroimaging. *PloS one*, *6*(10), e26377.
- Halperin, J. M. y Schulz, K. P. (2006). Revisiting the role of the prefrontal cortex in the pathophysiology of attention-deficit/hyperactivity disorder. *Psychological bulletin*, 132(4), 560.
- Hand, D. J., y Yu, K. (2001). Idiot's Bayes—not so stupid after all?. *International statistical review*, 69(3), 385-398.
- Hasson, U., Nir, Y., Levy, I., Fuhrmann, G. y Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, 303(5664), 1634-1640.

- Hastie, T., Tibshirani, R., Friedman, J. H., y Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). Springer.
- Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., y Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4), 18-28.
- Hills, J., Lines, J., Baranauskas, E., Mapp, J. y Bagnall, A. (2014). Classification of time series by shapelet transformation. *Data mining and knowledge discovery*, 28, 851-881.
- Hosford, P. S. y Gourine, A. V. (2019). What is the key mediator of the neurovascular coupling response? *Neuroscience & Biobehavioral Reviews*, *96*, 174-181.
- Hosmer Jr, D. W., Lemeshow, S., y Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons.
- Hossin, M., y Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International journal of data mining & knowledge management process*, 5(2), 1.
- Hramov, A. E., Koronovskii, A. A., Makarov, V. A., Pavlov, A. N. y Sitnikova, E. (2015). *Wavelets in Neuroscience*. Springer.
- Huggins, C. J., Escudero, J., Parra, M. A., Scally, B., Anghinah, R., De Araújo, A. V. L.,
 ... y Abasolo, D. (2021). Deep learning of resting-state electroencephalogram signals for three-class classification of Alzheimer's disease, mild cognitive impairment and healthy ageing. *Journal of Neural Engineering*, 18(4), 046087.
- Huppert, T. J., Diamond, S. G., Franceschini, M. A. y Boas, D. A. (2009). HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Applied optics*, 48(10), D280-D298.
- Ilvedson, C. R. (1998). *Transfer Function Estimates Using Time-Frequency Analysis*. Massachusetts Institute of Technology.
- Inoue, Y., Sakihara, K., Gunji, A., Ozawa, H., Kimiya, S., Shinoda, H., ... y Inagaki, M. (2012). Reduced prefrontal hemodynamic response in children with ADHD during the Go/NoGo task: a NIRS study. *Neuroreport*, 23(2), 55-60.

- Jiménez, J.E., Rodríguez, C., Camacho, J., Afonso, M. y Artiles, C. (2012). Estimación de la prevalencia del trastorno por déficit de atención con o sin hiperactividad (TDAH) en población escolar de la Comunidad Autónoma de Canarias. *European Journal of Education and Psychology*, 5(1), 13-26.
- Jović, A., Brkić, K. y Bogunović, N. (2015). A review of feature selection methods with applications. En 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 1200-1205.
- Jurcak, V., Tsuzuki, D. y Dan, I. (2007). 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. *Neuroimage*, *34*(4), 1600-1611.
- Kauppi, J. P., Pajula, J., y Tohka, J. (2014). A versatile software package for inter-subject correlation based analyses of fMRI. *Frontiers in neuroinformatics*, *8*, 2.
- Kazda, L., Bell, K., Thomas, R., McGeechan, K., Sims, R. y Barratt, A. (2021). Overdiagnosis of attention-deficit/hyperactivity disorder in children and adolescents: a systematic scoping review. JAMA network open, 4(4).
- Kijewski-Correa, T., y Kareem, A. (2006). Efficacy of Hilbert and wavelet transforms for time-frequency analysis. *Journal of engineering mechanics*, *132*(10), 1037-1049.
- Kim, J. H. (2009). Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Computational statistics & data analysis*, 53(11), 3735-3745.
- Kirilina, E., Jelzow, A., Heine, A., Niessing, M., Wabnitz, H., Brühl, R., Ittermann, B, Jacobs, A.M., Tachtsidis, I. (2012). The physiological origin of task-evoked systemic artefacts in functional near infrared spectroscopy. *Neuroimage*, 61(1), 70-81.
- Kocsis, L., Herman, P. y Eke, A. (2006). The modified Beer–Lambert law revisited. *Physics in Medicine & Biology*, 51(5), N91.
- Koh, J. E., Ooi, C. P., Lim-Ashworth, N. S., Vicnesh, J., Tor, H. T., Lih, O. S., ... y Fung,D. S. S. (2022). Automated classification of attention deficit hyperactivity

disorder and conduct disorder using entropy features with ECG signals. Computers in biology and medicine, 140, 105120.

- Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., y Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature neuroscience*, 12(5), 535-540.
- Kuncheva, L. I. (2006). On the optimality of Naïve Bayes with dependent binary features. *Pattern Recognition Letters*, 27(7), 830-837.
- Lilly, J. M., y Olhede, S. C. (2010). On the analytic wavelet transform. *IEEE transactions* on information theory, 56(8), 4135-4156.
- Lilly, J. M., y Olhede, S. C. (2012). Generalized Morse wavelets as a superfamily of analytic wavelets. *IEEE Transactions on Signal Processing*, 60(11), 6036-6041.
- Liu, T., Luo, Z., Huang, J., y Yan, S. (2018). A comparative study of four kinds of adaptive decomposition algorithms and their applications. *Sensors*, *18*(7), 2120.
- Liu, H., y Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on knowledge and data engineering*, 17(4), 491-502.
- Loh, H. W., Ooi, C. P., Barua, P. D., Palmer, E. E., Molinari, F. y Acharya, U. (2022). Automated detection of ADHD: Current trends and future perspective. *Computers in Biology and Medicine*, 105525.
- López-Villalobos, J. A., Garrido-Redondo M., Sacristán-Martín, A. M., Martínez-Rivera, M. T., López-Sánchez, M. V., Andrés-De Llano, J. M., Rodríguez-Molinero L. y Camina-Gutiérrez, A. B. (2018). Percepción de niños y adolescentes sobre la calidad de vida en casos de trastorno por déficit de atención/hiperactividad con y sin tratamiento farmacológico y en controles. *Revista de Neurología, 67*(06), 195-202.
- Luque, A., Carrasco, A., Martín, A., y de Las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, 91, 216-231.

- Nakhnikian, A., Ito, S., Dwiel, L. L., Grasse, L. M., Rebec, G. V., Lauridsen, L. N., y Beggs, J. M. (2016). A novel cross-frequency coupling detection method using the generalized Morse wavelets. *Journal of neuroscience methods*, 269, 61-73.
- Nambu, I., Ozawa, T., Sato, T., Aihara, T., Fujiwara, Y., Otaka, Y., Osu, R., Izawa, J. y Wada, Y. (2017). Transient increase in systemic interferences in the superficial layer and its influence on event-related motor tasks: a functional near-infrared spectroscopy study. *Journal of biomedical optics*, 22(3), 035008.
- Narbona, J. (2001). Alta prevalencia del TDAH: ¿niños trastornados o sociedad maltrecha. *Revista de Neurología*, 32(3), 229-231.
- Näsi, T., Mäki, H., Hiltunen, P., Heiskala, J., Nissilä, I., Kotilahti, K. y Ilmoniemi, R. J. (2013). Effect of task-related extracerebral circulation on diffuse optical tomography: experimental data and simulations on the forehead. *Biomedical optics express*, 4(3), 412-426.
- Nastase, S. A., Gazzola, V., Hasson, U. y Keysers, C. (2019). Measuring shared responses across subjects using intersubject correlation. *Social Cognitive and Affective Neuroscience*, 14(6), 667-685.
- Navarro-Pardo, E., Moral, J. C. M., Galán, A. S. y Beitia, M. D. S. (2012). Desarrollo infantil y adolescente: trastornos mentales más frecuentes en función de la edad y el género. *Psicothema*, 24(3), 377-383.
- Negoro, H., Sawada, M., Iida, J., Ota, T., Tanaka, S., y Kishimoto, T. (2010). Prefrontal dysfunction in attention-deficit/hyperactivity disorder as measured by nearinfrared spectroscopy. *Child Psychiatry & Human Development*, 41, 193-203.
- Maris, E. y Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEGdata. *Journal of neuroscience methods*, *164*(1), 177-190.
- Mauri, M., Nobile, M., Bellina, M., Crippa, A., & Brambilla, P. (2018). Light up ADHD:
 I. Cortical hemodynamic responses measured by functional Near Infrared
 Spectroscopy (fNIRS): Special Section on "Translational and Neuroscience
 Studies in Affective Disorders". *Journal of Affective Disorders*, 234, 358-364.

- Minati, L., Kress, I. U., Visani, E., Medford, N., y Critchley, H. D. (2011). Intra-and extra-cranial effects of transient blood pressure changes on brain near-infrared spectroscopy (NIRS) measurements. *Journal of neuroscience methods*, 197(2), 283-288.
- Molina-Rodríguez, S., Mirete-Fructuoso, M., Martínez, L. M. y Ibañez-Ballesteros, J. (2022). Frequency-domain analysis of fNIRS fluctuations induced by rhythmic mental arithmetic. *Psychophysiology*, 59(10), e14063.
- Monden, Y., Dan, I., Nagashima, M., Dan, H., Uga, M., Ikeda, T., Tsuzuki, D., Tsuzuki, D., Kyutoku, Y., Gunji, Y., Hirano, D., Taniguchi, T., Shimoizumi, H., Watanable, E. y Yamagata, T. (2015). Individual classification of ADHD children by right prefrontal hemodynamic responses during a go/no-go task as assessed by fNIRS. *NeuroImage: Clinical*, *9*, 1-12.
- Morales, S. y Bowers, M. E. (2022). Time-frequency analysis methods and their application in developmental EEG data. *Developmental Cognitive Neuroscience*, 54, 101067.
- Munia, T. T., y Aviyente, S. (2019). Time-frequency based phase-amplitude coupling measure for neuronal oscillations. *Scientific reports*, 9(1), 12441.
- Muñoz, J. A. L., y García, I. M. (2008). Perfil social de los subtipos del Trastorno por Déficit de Atención con Hiperactividad. *Apuntes de psicología*, *26*, 317-329.
- Muñoz, J. A. L. y García, I. M. (2010). Prevalencia del trastorno por déficit de atención en escolares: comparación entre criterios diagnósticos y criterios clínicos. *Psicología conductual*, 18(2), 365-384.
- Obrig, H. (2014). NIRS in clinical neurology—a 'promising'tool? *Neuroimage*, 85, 535-546.
- Olhede, S. C., y Walden, A. T. (2002). Generalized morse wavelets. *IEEE Transactions* on Signal Processing, 50(11), 2661-2670.
- Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for highresolution EEG and ERP measurements. *Clinical neurophysiology*, *112*(4), 713-719.

- Organización Mundial de la Salud (2019). *Clasificación Internacional de las Enfermedades, undécima revisión*. Ginebra: OMS. Disponible en <u>https://icd.who.int/browse11/l-m/es</u>
- Orihuela-Espina, F., Leff, D. R., James, D. R., Darzi, A. W., y Yang, G. Z. (2010). Quality control and assurance in functional near infrared spectroscopy (fNIRS) experimentation. *Physics in Medicine & Biology*, *55*(13), 3701.
- Pereira, F., Mitchell, T. y Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage*, 45(1), S199-S209.
- Pereira-Sánchez, V., y Castellanos, F. X. (2021). Neuroimaging in attentiondeficit/hyperactivity disorder. *Current opinion in psychiatry*, 34(2), 105.
- Perpetuini, D., Cardone, D., Filippini, C., Chiarelli, A. M., y Merla, A. (2021). A motion artifact correction procedure for fNIRS signals based on wavelet transform and infrared thermography video tracking. *Sensors*, 21(15), 5117.
- Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., & Burgess,
 P. W. (2020). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Annals of the New York Academy of Sciences*, 1464(1), 5-29.
- Plichta, M. M., Herrmann, M. J., Baehne, C. G., Ehlis, A. C., Richter, M. M., Pauli, P. y Fallgatter, A. J. (2006). Event-related functional near-infrared spectroscopy (fNIRS): are the measurements reliable? *Neuroimage*, 31(1), 116-124.
- Polanczyk, G., De Lima, M. S., Horta, B. L., Biederman, J., y Rohde, L. A. (2007). The worldwide prevalence of ADHD: a systematic review and metaregression analysis. *American journal of psychiatry*, 164(6), 942-948.
- Posner, J., Polanczyk G. V. y Sonuga-Barke, E. (2020). Attention-deficit hypereactivity disorder. *Lancet*, 395(10222), 450-62.
- Presentación-Herrero, M. D. J., García-Castellar, R., Miranda-Casas, A., Siegenthaler-Hierro, R. y Jara-Jiménez, P. (2006). Impacto familiar de los niños con trastorno por déficit de atención con hiperactividad subtipo combinado: efecto de los problemas de conducta asociados. *Revista de Neurología*, 42(3), 137-143.

- Pudil, P., Novovičová, J. y Kittler, J. (1994). Floating search methods in feature selection. *Pattern recognition letters*, 15(11), 1119-1125.
- Pulini, A. A., Kerr, W. T., Loo, S. K., y Lenartowicz, A. (2019). Classification accuracy of neuroimaging biomarkers in attention-deficit/hyperactivity disorder: effects of sample size and circular analysis. *Biological Psychiatry: Cognitive Neuroscience* and Neuroimaging, 4(2), 108-120.
- Quinlan D. M. (2003). Evaluación del trastorno por déficit de atención con hiperactividad y de los trastornos comórbidos. En T E Brown (Ed), *Trastornos por déficit de* atención y comorbilidades en niños, adolescentes y adultos, 455-507. Barcelona: Masson.
- Ramos Platón, M. J. (1987). Del síndrome de daño cerebral al trastorno de déficit de atención: revisión crítica. Revista de la Asociación Española de Neuropsiquiatría., 7(20), 040-071.
- Raudys, S. (2006). Feature over-selection. En D. Y. Yeung, J. T. Kwok, A. Fred, F. Roli,
 D. de Ridder (Eds.), *Structural, Syntactic, and Statistical Pattern Recognition*,
 622–631. Heidelberg: Springer.
- Reddy, P., Izzetoglu, M., Shewokis, P. A., Sangobowale, M., Diaz-Arrastia, R. y Izzetoglu, K. (2021). Evaluation of fNIRS signal components elicited by cognitive and hypercapnic stimuli. *Scientific Reports*, 11(1), 23457.
- Reid, R. y Maag, J. W. (1994). How many fidgets in a pretty much: A critique of behavior rating scales for identifying students with ADHD. *Journal of School Psychology*, 32(4), 339-354.
- Remeseiro, B. y Bolon-Canedo, V. (2019). A review of feature selection methods in medical applications. *Computers in biology and medicine*, 112, 103375.
- Richard, A., Mayag, B., Talbot, F., Tsoukias, A., y Meinard, Y. (2020). Transparency of classification systems for clinical decision support. En *Information Processing* and Management of Uncertainty in Knowledge-Based Systems: 18th International Conference, IPMU 2020, Lisbon, Portugal, June 15–19, 2020, Proceedings, Part III 18 (pp. 99-113). Springer International Publishing.

- Richter, C. G., Thompson, W. H., Bosman, C. A., & Fries, P. (2015). A jackknife approach to quantifying single-trial correlation between covariance-based metrics undefined on a single-trial basis. *Neuroimage*, 114, 57-70.
- Rodríguez Molinero, L., López Villalobos, J. A., Garrido Redondo, M., Sacristán Martín, A. M., Martínez Rivera, M. T. y Ruiz Sanz, F. (2009). Estudio psicométricoclínico de prevalencia y comorbilidad del trastorno por déficit de atención con hiperactividad en Castilla y León (España). *Pediatría Atención Primaria*, 11(42), 251-270.
- Saager, R., y Berger, A. (2008). Measurement of layer-like hemodynamic trends in scalp and cortex: implications for physiological baseline suppression in functional nearinfrared spectroscopy. *Journal of biomedical optics*, 13(3), 034017-034017.
- Sadowsky, J. (1996). Investigation of signal characteristics using the continuous wavelet transform. *Johns Hopkins Apl Technical Digest, 17*(3), 258-269.
- Saeb, S., Lonini, L., Jayaraman, A., Mohr, D. C. y Kording, K. P. (2017). The need to approximate the use-case in clinical machine learning. *Gigascience*, *6*(5).
- Salmi, J., Metwaly, M., Tohka, J., Alho, K., Leppämäki, S., Tani, P., ... y Laine, M. (2020). ADHD desynchronizes brain activity during watching a distracted multitalker conversation. *NeuroImage*, 216, 116352.
- Schaeffer, S., y Iadecola, C. (2021). Revisiting the neurovascular unit. *Nature neuroscience*, 24(9), 1198-1209. *https://doi.org/10.1038/s41593-021-00904-7*
- Schecklmann, M., Ehlis, A. C., Plichta, M. M. y Fallgatter, A. J. (2008). Functional nearinfrared spectroscopy: a long-term reliable tool for measuring brain activity during verbal fluency. *Neuroimage*, 43(1), 147-155.
- Scholkmann, F., Gerber, U., Wolf, M., y Wolf, U. (2013). End-tidal CO2: an important parameter for a correct interpretation in functional brain studies using speech tasks. *Neuroimage*, 66, 71-79.
- Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Pavia, J. M., Wolf, U. y Wolf, M. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *Neuroimage*, 85, 6-27.

- Servera, M. y Cardo, E. (2007). ADHD Rating Scale-IV en una muestra escolar española: datos normativos y consistencia interna para maestros, padres y madres. *Revista de Neurología*, 45(7), 393-9.
- Shaw, P., Eckstrand, K., Sharp, W., Blumenthal, J., Lerch, J. P., Greenstein, D., Clasen, L., Evans, A., Giedd, J. y Rapoport, J. L. (2007). Attention-deficit/hyperactivity disorder is characterized by a delay in cortical maturation. *Proceedings of the national academy of sciences*, 104(49), 19649-19654.
- Skounti, M., Philalithis, A. y Galanakis, E. (2007). Variations in prevalence of attention deficit hiperactivity disorder worldwide. *European Journal of Pediatrics*, 166, 117-123.
- Somol, P., Novovicová, J. y Pudil, P. (2010). Efficient feature subset selection and subset size optimization. En A. Herout (Ed.), *Pattern Recognition Recent Advances*, 1-24. Rijeka: Ed. InTech.
- Song, X., Chen, X., Chen, L., An, X., y Ming, D. (2020). Performance improvement for detecting brain function using fNIRS: a multi-distance probe configuration with PPL method. *Frontiers in Human Neuroscience*, 14, 569508.
- Stefanovska, A. (2007). Coupled Oscillatros: Complex But Not Complicated Cardiovascular and Brain Interactions. *IEEE Engineering in Medicine and Biology Magazine*, 26(6), 25–29. <u>https://doi.org/10.1109/EMB.2007.907088</u>
- Suwansawang, S. y Halliday, D. M. (2017). Time-frequency based coherence and phase locking value analysis of human locomotion data using generalized Morse wavelets BIOSIGNALS 2017. 10th Int. Conf. Bio-Inspired Syst. Signal Process, 4 34–41.
- Tachtsidis, I., y Scholkmann, F. (2016). False positives and false negatives in functional near-infrared spectroscopy: issues, challenges, and the way forward. *Neurophotonics*, 3(3), 031405. <u>https://doi.org/10.1117/1.NPh.3.3.031405</u>
- Tak, S., Uga, M., Flandin, G., Dan, I., y Penny, W. D. (2016). Sensor space group analysis for fNIRS data. *Journal of neuroscience methods*, 264, 103-112.

- Takahashi, T., Takikawa, Y., Kawagoe, R., Shibuya, S., Iwano, T. y Kitazawa, S. (2011). Influence of skin blood flow on near-infrared spectroscopy signals measured on the forehead during a verbal fluency task. *Neuroimage*, 57(3), 991-1002.
- Tallon-Baudry, C., Bertrand, O., Delpuech, C., y Pernier, J. (1996). Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *Journal* of Neuroscience, 16(13), 4240-4249.
- Tang, C., Wei, Y., Zhao, J., & Nie, J. (2018). Different developmental pattern of brain activities in ADHD: a study of resting-state fMRI. *Developmental neuroscience*, 40(3), 246-257.
- Taylor, E. (2011). Antecedents of ADHD: a historical account of diagnostic concepts. *ADHD Attention Deficit and Hyperactivity Disorders*, *3*(2), 69-75.
- Tougui, I., Jilbab, A. y El Mhamdi, J. (2021). Impact of the choice of cross-validation techniques on the results of machine learning-based diagnostic applications. *Healthcare informatics research*, 27(3), 189-199.
- Trevethan, R. (2017). Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. *Frontiers in public health*, *5*, 307.
- Vabalas, A., Gowen, E., Poliakoff, E., y Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PloS one, 14*(11), e0224365.
- Varoquaux, G. (2018). Cross-validation failure: small sample sizes lead to large error bars. *Neuroimage*, *180*, 68-77.
- Varoquaux, G., Raamana, P. R., Engemann, D. A., Hoyos-Idrobo, A., Schwartz, Y. y Thirion, B. (2017). Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. *NeuroImage*, 145, 166-179.
- Wachowiak, M. P., Wachowiak-Smolíková, R., Johnson, M. J., Hay, D. C., Power, K. E., y Williams-Bell, F. M. (2018). Quantitative feature analysis of continuous analytic wavelet transforms of electrocardiography and electromyography. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376*(2126), 20170250.

- Wacker, M. y Witte, H. (2013). Time-frequency techniques in biomedical signal analysis. *Methods of information in medicine*, 52(04), 279-296.
- Wang, J. B., Zheng, L. J., Cao, Q. J., Wang, Y. F., Sun, L., Zang, Y. F. y Zhang, H. (2017). Inconsistency in abnormal brain activity across cohorts of ADHD-200 in children with attention deficit hyperactivity disorder. *Frontiers in neuroscience*, 11, 320.
- Wang, X. H. y Li, L. (2015). Altered temporal features of intrinsic connectivity networks in boys with combined type of attention deficit hyperactivity disorder. *European Journal of Radiology*, 84(5), 947-954.
- Wang, X., Liu, B., Xie, L., Yu, X., Li, M. y Zhang, J. (2016). Cerebral and neural regulation of cardiovascular activity during mental stress. *Biomedical engineering online*, 15, 335-347.
- Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. IEEE *Transactions on audio and electroacoustics*, 15(2), 70-73.
- Wiklendt, L., Brookes, S. J., Costa, M., Travis, L., Spencer, N. J., & Dinning, P. G. (2020). A novel method for electrophysiological analysis of EMG signals using MesaClip. *Frontiers in Physiology*, 11, 484.
- Willcutt, E. G. (2012). The prevalence of DSM-IV attention-deficit/hyperactivity disorder: a meta-analytic review. *Neurotherapeutics*, 9(3), 490-499.
- Yasumura, A., Kokubo, N., Yamamoto, H., Yasumura, Y., Nakagawa, E., Kaga, M., ... y Inagaki, M. (2014). Neurobehavioral and hemodynamic evaluation of Stroop and reverse Stroop interference in children with attention-deficit/hyperactivity disorder. *Brain and Development*, 36(2), 97-106.
- Ye, L., y Keogh, E. (2011). Time series shapelets: a novel technique that allows accurate, interpretable and fast classification. *Data mining and knowledge discovery*, 22, 149-182.
- Yu, X., Yuan, B., Cao, Q., An, L., Wang, P., Vance, A., Silk, T.J., Zang, Y., Wang, Y. y Sun, L. (2016). Frequency-specific abnormalities in regional homogeneity among

children with attention deficit hyperactivity disorder: a resting-state fMRI study. *Science bulletin*, *61*, 682-692.

- Yu-Feng, Z., Yong, H., Chao-Zhe, Z., Qing-Jiu, C., Man-Qiu, S., Meng, L., ... y Yu-Feng,
 W. (2007). Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. *Brain and Development*, 29(2), 83-91.
- Yücel, M. A., Selb, J. J., Huppert, T. J., Franceschini, M. A. y Boas, D. A. (2017). Functional near infrared spectroscopy: enabling routine functional brain imaging. *Current opinion in biomedical engineering*, 4, 78-86.
- Zhang, J., Kendrick, K. M., Lu, G., y Feng, J. (2015). The fault lies on the other side: altered brain functional connectivity in psychiatric disorders is mainly caused by counterpart regions in the opposite hemisphere. *Cerebral Cortex*, 25(10), 3475-3486.
- Zimeo Morais, G. A., Scholkmann, F., Balardin, J. B., Furucho, R. A., de Paula, R. C. V., Biazoli, C. E. y Sato, J. R. (2018). Non-neuronal evoked and spontaneous hemodynamic changes in the anterior temporal region of the human head may lead to misinterpretations of functional near-infrared spectroscopy signals. *Neurophotonics*, 5(1), 011002.

7. ANEXO



PAPER • OPEN ACCESS

Identifying ADHD boys by very-low frequency prefrontal fNIRS fluctuations during a rhythmic mental arithmetic task

To cite this article: Sergio Ortuño-Miró et al 2023 J. Neural Eng. 20 036018

View the article online for updates and enhancements.

You may also like

- Identifying ADHD children using hemodynamic responses during a working memory task measured by functional nearinfrared spectroscopy
 Yue Gu, Shuo Miao, Junxia Han et al.
- Resting state dynamic functional connectivity in children with attention deficit/hyperactivity disorder Maliheh Ahmadi, Kamran Kazemi, Katarzyna Kuc et al.
- <u>Subtype classification of attention deficit</u> <u>hyperactivity disorder with hierarchical</u> <u>binary hypothesis testing framework</u> Yuan Gao, Huaqing Ni, Ying Chen et al.

REATH

OPS

Main talks

Early career sessions

Posters



Breath Biopsy Conference

Join the conference to explore the **latest challenges** and advances in **breath research**, you could even **present your latest work**!



Register now for free!

This content was downloaded from IP address 193.147.143.29 on 04/06/2024 at 09:13

CrossMark

OPEN ACCESS

13 December 2022

20 December 2022

ACCEPTED FOR PUBLICATION

Original content from this work may be used under the terms of the Creative Commons

Attribution 4.0 licence.

Any further distribution of this work must

maintain attribution to

citation and DOI.

the author(s) and the title of the work, journal

RECEIVED 28 September 2022

REVISED

PUBLISHED 23 May 2023

Journal of Neural Engineering

PAPER

Identifying ADHD boys by very-low frequency prefrontal fNIRS fluctuations during a rhythmic mental arithmetic task

Sergio Ortuño-Miró¹, Sergio Molina-Rodríguez², Carlos Belmonte² and Joaquín Ibañez-Ballesteros^{1,*} 💿

 $^1\,$ Department of physiology, Miguel Hernandez University, San Joan d'Alacant, Alicante, Spain

² Institute of Neurosciences, Spanish National Research Council- Miguel Hernandez University, San Joan d'Alacant, Alicante, Spain
 * Author to whom any correspondence should be addressed.

E-mail: charly.joa@umh.es

Keywords: attention-deficit/hyperactivity disorder, functional near-infrared spectroscopy, machine learning, classification, mental arithmetic, logistic regression, linear discriminant analysis

Abstract

Objective. Computer-aided diagnosis of attention-deficit/hyperactivity disorder (ADHD) aims to provide useful adjunctive indicators to support more accurate and cost-effective clinical decisions. Deep- and machine-learning (ML) techniques are increasingly used to identify neuroimaging-based features for objective assessment of ADHD. Despite promising results in diagnostic prediction, substantial barriers still hamper the translation of the research into daily clinic. Few studies have focused on functional near-infrared spectroscopy (fNIRS) data to discriminate ADHD condition at the individual level. This work aims to develop an fNIRS-based methodological approach for effective identification of ADHD boys via technically feasible and explainable methods. Approach. fNIRS signals recorded from superficial and deep tissue layers of the forehead were collected from 15 clinically referred ADHD boys (average age 11.9 years) and 15 non-ADHD controls during the execution of a rhythmic mental arithmetic task. Synchronization measures in the time-frequency plane were computed to find frequency-specific oscillatory patterns maximally representative of the ADHD or control group. Time series distance-based features were fed into four popular ML linear models (support vector machine, logistic regression (LR), discriminant analysis and naïve Bayes) for binary classification. A 'sequential forward floating selection' wrapper algorithm was adapted to pick out the most discriminative features. Classifiers performance was evaluated through five-fold and leave-one-out cross-validation (CV) and statistical significance by non-parametric resampling procedures. Main results. LR and linear discriminant analysis achieved accuracy, sensitivity and specificity scores of near 100% (p < .001) for both CV schemes when trained with only three key wrapper-selected features, arising from surface and deep oscillatory components of very low frequency. Significance. We provide preliminary evidence that very-low frequency fNIRS fluctuations induced/modulated by a rhythmic mental task accurately differentiate ADHD boys from non-ADHD controls, outperforming other similar studies. The proposed approach holds promise for finding functional biomarkers reliable and interpretable enough to inform clinical practice.

1. Introduction

Attention-deficit/hyperactivity disorder (ADHD) is recognized as a highly prevalent neurodevelopmental disorder in school-age children worldwide, often persisting into adolescence and adulthood, and frequently overlapped with other psychiatric comorbidities [1–3]. People with ADHD exhibit three core behavioral symptoms, inattention, hyperactivity and impulsivity, although each displayed to a varying degree [4]. ADHD is a highly heterogeneous impairing condition, extensively researched over many years, that is probably the best known childhood-onset disorder [5]. Based on the last 10–20 years of scientific evidence, updated information on ADHD was recently summarized by the 'World Federation of

BY

S Ortuño-Miró et al

ADHD' as an international consensus statement [6] and reviewed by Posner *et al* in [7], covering epidemiology, etiology, pathophysiology, diagnosis and treatment. It is currently accepted that ADHD is a complex, heterogeneous disorder, in which different expressions of impairment along with variable trajectories must be recognized in order to adopt personalized approaches that best target an individual. This is of crucial importance because, even despite serious distress/impairments, many patients lead rewarding and productive lives when properly managed.

Diagnosis of ADHD is event today based mainly on clinical signs and symptoms that require a detailed evaluation by an expert clinician through interviews with parents/caregivers and/or the patient himself, if applicable [8]. Noteworthy, diagnosis cannot be solely based on rating scales, neuropsychological test or brain imaging. Despite the criticisms that argue a risk of subjectivity, the current consensus supports the validity of the diagnostic criteria applied by welltrained professionals [6]. However, even for a specialist, clinical evaluation is quite time-consuming and requires several visits to be thoroughly performed [9]. Besides, the significant shortage of trained professionals also contributes to a frequent delay in diagnosis or even to overlook some cases. From a developmental perspective, an early diagnosis is very likely to be of value for more effective pharmacological and psychosocial interventions [10]. In this view, there is a need for objective biomarkers as useful adjunctive indicators to alleviate the workload of diagnoses and treatment follow-up.

Numerous studies have tried to assess ADHD through different objective diagnostic tools, most using functional (fMRI) or structural magnetic resonance imaging (MRI) and electroencephalography (EEG), with other modalities (magnetoencephalography (MEG), electrocardiography (EKG), etc) being deployed less frequently, and with an increasing use of artificial intelligence (AI) techniques (for reviews, see [11, 12]). Noticeable efforts in MRI and fMRI were made under the initiatives of the 'ADHD-200 Consortium' [13]. Despite significant advances in understanding abnormalities related to brain maturation and function, neuroimaging findings in ADHD research cannot yet be used to support clinical practice due to a variety of concerns [7, 11]. Likewise, though promising, studies devoted to single-subject prediction of ADHD via deep- and machine-learning (ML) methods have reported quite variable results [12, 14], raising certain methodological concerns [15]. In actual practice, even if useful neuroimaging indicators were available to support clinical decisions, the availability of these diagnostic tools and their associated costs would represent a major barrier to their regular use, further increasing the burden on healthcare resources.

An alternative tool to assess ADHD worth to explore is functional near-infrared spectroscopy (fNIRS), which is characterized by being noninvasive, wearable, cost-effective, and deployable in more friendly/ecological settings (for a review, see [16]). fNIRS has shown its usefulness in monitoring functional hemodynamic changes associated with cortical brain activation (for a review, see [17]). Compared to other neuroimaging modalities, few fNIRS studies have been conducted to discriminate children with ADHD from non-ADHD controls, some of them trying to improve classification by combining different modalities (e.g. EEG + fNIRS) as in [18]. Even fewer studies focused on single unimodal approaches by using 'exclusively' fNIRS data. For example, Monden et al reported a classification accuracy of 85% with a sensitivity of 90% by analyzing Receiver Operating Characteristic (ROC curves obtained from right prefrontal oxy-Hb activation data during a go/nogo task [19]. Using prefrontal cortex (PFC) activation measures during an N-back task, Crippa et al achieved mean accuracies of 78% with 72% sensitivity and 82% specificity when a support vector machine (SVM) classifier was trained on data from deoxy-Hb [20]. Also employing an N-back task and an SVM, Gu et al reached 86% of accuracy with oxy-Hb data measured in the prefrontal and temporal cortex [21]. It is worth noting that no correction for components of noncerebral origin was applied to the fNIRS signals in the aforementioned studies, which is especially important when scanning the PFC through the forehead [22, 23], since functional extra- and cerebral responses are interrelated processes that overlap in fNIRS recordings and with a greater confounding effect for oxy-Hb [24-26]. Notwithstanding this known drawback of fNIRS, classification algorithms can achieve appreciable performance by learning some type of feature representation from the uncorrected NIRS data, but uncertainty about the nature and origin of the features hampers the interpretability of predictive models. We also note that, in these studies, the features were based on some kind of measurement from the averaged fNIRS data across trials/epochs, a classic approach that, while often providing robust results, fails to uncover finer distinctive patterns embedded in the data.

In a previous study involving non-ADHD young adults, our research team showed that a rhythmic mental arithmetic task successfully induced cyclical hemodynamic fluctuations coupled to the task frequency (33 mHz), and that the oscillatory patterns were consistent across individuals both in superficial and deep fNIRS signals recorded in the frontopolar region [27]. Spectral analysis also showed oscillatory activity at lower frequencies (<33 mHz) seen at rest and during mental task, and with a prominent peak around 5–10 mHz. Resting-state fMRI studies have reported that ADHD patients show significant differences in the low-frequency oscillations (10–80 mHz) band across multiple brain regions [28–30], with separable contribution of specific frequency sub-bands

IOP Publishing

S Ortuño-Miró et al

including extra-low frequencies (0–10 mHz) [31]. These differences have been related to abnormalities in the salience, attentional and default-mode networks (DMNs) functioning, but the inconsistences observed across many studies point to a large heterogeneity in spontaneous brain activity in ADHD [32]. Despite this, evidence suggests that some characteristics of ADHD brain activity are sensitive to specific frequency bands.

We hypothesized that a rhythmic mental task might induce/modulate fNIRS-measurable oscillatory activity in particular sub-bands within the 0-80 mHz range, which might be distinctive enough to discriminate ADHD at the individual level. Thus, in the present study, we explored frequency-specific hemodynamic patterns during a cyclical mental arithmetic task administered to a homogeneous sample of 10- to 13-year-old boys with ADHD and non-ADHD controls. A data-driven approach, based on synchronization measures, guided us in selecting the relevant frequencies to extract discriminative features and assess the performance of four commonly used supervised ML methods. Analysis were performed on superficial and regression-corrected deep fNIRS signals recorded from the forehead through a recently introduced multi-distance, multi-channel device [27]. The present work aims to contribute to the objective assessment of ADHD through computer-aided affordable tools deployable in many clinical settings.

2. Method

In this study, we focused solely on fNIRS recordings to find distinctive hemodynamic information by following a data analysis pipeline consisting of: (i) signal preprocessing; (ii) delimitation of the frequency range of interest by spectral power analysis; (iii) identification of relevant sub-bands by time-frequency synchronization analysis; (iv) definition of grouprepresentative oscillatory patterns; (v) computation of features by distance measures; (vi) selection of discriminative features and (vii) classification of individuals using ML linear models.

2.1. Participants

Sixteen clinically referred boys meeting the American Psychiatric Association's DSM-V criteria for the combined ADHD presentation (i.e. inattention + hyperactivity/impulsivity) and 17 age-matched typically developing (TD) control boys were initially recruited. All participants were Caucasian, native Spanish speakers, right-handed, had normal or correctedto-normal vision and none had other major medical/psychiatric diagnoses. The study was approved by the Ethics Committee of the University Miguel Hernandez according to the Declaration of Helsinki. Written consent was obtained from the parents/guardians of all participants prior to enrollment, none of whom received any financial compensation. Boys diagnosed with ADHD were recruited from a local ADHD support association, while school psychologists referred the control group as TD boys with no history of mental disorder or medication use. Since all ADHD participants were on methylphenidate treatment, they were assessed during the 'drug holiday' period prescribed by the referring clinician, with a minimum four day washout in all cases. One boy from the ADHD group and two from the TD control group were excluded due to NIRS signal quality issues. Therefore, the study ultimately included 15 participants with ADHD (mean age 11.9, SD 1.4, range 10–13 years) and 15 TD controls (mean age 11.6, SD 1.1, range 10–13 years).

2.2. Experimental protocol

In a quiet room, participants sat in a comfortable chair and were asked to relax and keep their eyes on a computer screen 80 cm away. Visual cues and instructions were presented on the screen throughout the experiment. They underwent a slightly modified version of a rhythmic mental arithmetic task described in a previous work [27]. Briefly, the task comprised 10 consecutive 30 s trials, each starting with 15 s of mental calculation followed by a 15 s pause of relaxation. During mental math, participants were asked to iteratively add a small number (5-9) to a threedigit number (100-199) (both numbers randomly chosen), silently and as quickly and accurately as possible. The pause then begins by presenting the question 'Result?' for 5 s, prompting for the voicing of the final result reached, followed by a black screen indicating mental relaxation until a 2 s fixation cross announced the start of the next trial. The task lasted 300 s and was uninterruptedly preceded by 300 s of baseline recording in resting state and followed by another 300 s of recovery in relaxed state (figure 1). To check whether participants engaged with the task well enough, we assessed behavioral performance using two simple scores, total iterations and total exact results achieved throughout the entire task. Note that the 30 s period of the trials corresponds to a frequency of 0.033 Hz, which we will refer to as the task frequency throughout the text. We also note that this design minimizes speech during the task, thus avoiding significant changes in breathing that could affect cerebral hemodynamics [33].

2.3. fNIRS recordings and preprocessing

In this work, we used a newly developed NIRS device (Tehia, Newmanbrain, S.L., Elche, Spain), light and easy to use, which was recently introduced in [27]. In short, it is a multichannel continuous-wave NIRS device that has four emitters and ten detectors arranged in a rectangular patch of 80×20 mm, each emitter housing two LEDs at wavelengths 740 nm and 850 nm. Through its duty cycle, the device combines pairs of optodes at different separation distances to





Figure 2. Probe geometry and placement (yellow squares = emitters; white circles = detectors), providing 16 short-channels (black lines with numbers) and 12 long-channels (green lines with letters). White dashed rectangles depict the three regions-of-interests (ROIs) explored in this work.

provide multi-distance recordings through 16 shortchannels (14 mm) and 12 long-channels (32 mm). It also corrects for ambient light interference and records motion activity using a 3-axis accelerometer. Data is transferred via Bluetooth at a sample rate of 10 Hz. The NIRS probe was placed on the forehead centered on AFpz according to the international 10-5 system, mainly covering the frontopolar area (Brodmann area 10) of the PFC (figure 2).

To identify recordings suffering from poor signalto-noise ratio, saturation or unphysiological interferences, we checked the raw optical data to identify those that exhibited extreme values (<5% or >95% of the device's dynamic range) or an excessive coefficient of variation (>7.5%) [34, 35]. Furthermore, by visually inspecting sudden changes in signals aligned with sharp shifts in accelerometer data, we identified recordings degraded by motion artifacts. As previously mentioned, three participants were excluded due to signal quality issues.

Relative concentration changes in oxy- (HbO) and deoxy-hemoglobin (HbR) were computed via the modified Beer–Lambert law [36, 37], using functions of the Homer2 NIRS package [38] based in MATLAB (Version R2021b, Mathworks, Natick, MA, USA). A differential pathlength factor of 6 was used for both wavelengths. HbO and HbR data were then digitally low-pass filtered with a zero-phase, 5thorder Butterworth filter, cut-off 0.08 Hz (MATLAB Signal Processing Toolbox); no high-pass filtering was applied. Thus, we remove blood-pressure, respiratory, cardiac and high frequency instrumental components while preserving the band of very-lowfrequency oscillations [39]. As a result, for each chromophore we achieved 16 time-series from the shortchannels plus 12 from the long-channels that we call shallow- (SSs) and deep-signals (DSs), respectively. SS contains non-cerebral components recorded from superficial layers, whereas DS combines both shallowand deep-components [24].

Since DS is contaminated by confounding elements unrelated to functional brain activity, such as systemic hemodynamics and inhomogeneous blood flow changes in the superficial tissue layers [22, 26, 40], a cleanup step was necessary to highlight the task-induced cortical response. The NIRS device provides us with multi-distance measurements, where each DS has three candidate SS that can be used as spatially close references to remove contamination [41–43]. As suggested in [44], we applied a 'double SS' approach in which each DS was regressed on the sum of the two SSs recorded closest to the emitter and detector of the corresponding long-channel. We solved linear regression by using the MATLAB function 'robustfit', which uses an algorithm less sensitive to outliers than ordinary least-squares [45]. After regression, the raw residuals represent a clean (or corrected) signal (CS) that likely contains the actual neuronal signal but perhaps mixed with other unpredictable components not fully removed.

Due to variability in head shape and size [46], channel positions are not fully consistent across individuals, and thus interpretation of isolated channels can be misleading. To improve signal reliability by spatial clustering [47, 48], for every single participant we averaged across the SSs and CSs belonging to three regions of interests (ROIs), left, medial and right (figure 2). Therefore, the data for each ROI is now reduced to just two average signals which, for consistency, we will continue to denote as SS and CS. All further processing was done on these signals, which show a comparable signal-tonoise ratio in all three ROIs because they were obtained by averaging the same number of neighboring signals in each region. Finally, to operate on the same scale, all signals were standardized into z-scores.

2.4. Spectral power distribution during task

Cognitive tasks may affect fNIRS signal components at multiple frequencies, reflecting the interlinked contribution of multiple factors arising from systemic, superficial tissue layers and neuronal activity [34, 49, 50]. As a preliminary step, we aimed to identify the most relevant oscillatory components present during the math task, regardless of whether they were spontaneous or task-evoked. To this end, we conducted a power spectral density (PSD) analysis using the Welch's averaged periodogram method [51]. PSD was obtained from the signal segment between 30 s before the task onset and 30 s after the end of the task, which yields a data vector of 30 + 300 + 30 = 360 s in length (3600 samples at sampling rate = 10 Hz). Additional 30 s on both sides were included to account for potential anticipatory or persistence task-related effects. To improve frequency resolution, we extended the data to the next power-of-two (4096 samples) by symmetrical reflection. PSDs were then computed via the MAT-LAB 'pwelch' function with FFT length = 4096, Hanning window = 2000 samples and overlap = 80% to account for spectral smoothness and reduced noise variance [52]. To allow comparisons, the PSDs were normalized to relative percentage values by calculating the power ratio of each frequency bin to the total power of the entire spectrum [53]. This procedure was applied to the SS and CS data of each participant.

To assess significant PSD differences between TD and ADHD, we performed a two-sample *t*-tests along frequency bins. The observed *t*-statistics were corrected for multiple comparisons following the cluster-based nonparametric approach given in [54]. We computed Monte Carlo cluster tests over 2000 permutations of the same *t*-test by randomly shuffling the data between classes. Then, we estimated the permutation *p*-value from the proportion of random realizations showing a larger cluster-statistic than the observed one. We set a critical alpha-level = 0.01 to identify frequency bins significantly different. PSDs were finally averaged across the participants of each class to obtain the average normalized PSD of HbO and HbR for each signal type and ROI. The 95% confidence interval (CI) for the mean at each frequency bin was also calculated by bootstrapping over 2000 resamples.

2.5. Time-frequency decomposition

Identification of the frequency components with potential capacity to discriminate between the two classes of participants (i.e. TD and ADHD) was a crucial issue. To this end, we needed a suitable method for locating task-related oscillations on different time scales (i.e. frequency bands), appropriate for nonstationary signal analysis, and capable of providing some measure of similarity to define class membership. Conventional spectral analysis can resolve how the signals' power is distributed along frequencies, but reveals little about how frequency content changes over time or time-varying patterns. Timefrequency analysis (TFA) techniques are a better option for discovering oscillatory patterns at certain frequencies, particularly when these may vary substantially over time and/or multiple components are present in the signal [55, 56]. Furthermore, they allow for separating the magnitude and phase components associated with the signal, which is very useful for capturing transient oscillations alignments [57, 58]. As discussed in [59], the selection of an optimal TFA technique depends on the knowledge of the signal characteristic, which allows a better match with the properties of a particular analysis method. We had no prior information on the most useful oscillatory components for classification purposes, and thus the conventional bandpass filtering + Hilbert transform method was unreliable due to the likely misidentification of the passbands to apply; apart from other inherent limitations [60]. We also avoided signaladaptive methods that greatly rely on specific knowledge of their parameter settings to find meaningful results, such as empirical or variational mode decomposition [61, 62]. Hence, we decided to use a data-driven approach based on complex continuous wavelet transform (CWT) and time-scale synchronization detection.

CWT is a signal processing method that provides a time-frequency (or time-scale) representation of the characteristics of a signal on the basis of the dilation and translation of a mother wavelet function;





theory and mathematical description can be found elsewhere [63, 64]. CWT can be viewed as a bandpass filter with varying bandwidths automatically defined by the wavelet scale [65], which avoids the drawbacks of using custom filters [60]. To compute the CWT we made use of the generalized Morse wavelets, a flexible superfamily of exactly analytic wavelets particularly useful for analyzing signals with time-varying amplitude and frequency, i.e. modulated signals [66, 67]. Since Morse wavelets can be tuned to encompass many other analytic wavelets commonly used, they provide a unified framework as reference point. Thus, for example, setting the symmetry parameter $\gamma = 3$ defines a family members ('Airy' wavelets) that can successfully replace the popular Morlet wavelet as the default analytic wavelet for general-purpose use [68]. Besides other applications, Morse wavelets have been proposed in a variety of biomedical studies, such as human locomotion [69], electrocardiography and electromyography [70, 71], neural oscillations coupling [60, 72], electroencephalographic data classification [73, 74] and fNIRS artifactual denoising [75].

To keep the procedure as simple and reproducible as possible, we computed the CWT using the default Morse parameters recommended in the MATLAB Wavelet Toolbox (symmetry = 3, time-bandwidth product = 60), which yield a perfectly symmetric wavelet. To obtain the most accurate time-frequency representation of our time-series data, we accounted for edge effects produced when the wavelets extend ('see') outside the signal boundaries. Since such effects depend on the wavelet scale, the socalled 'cone of influence' (COI) must be considered to avoid possible inaccurate coefficient values. Of note, in the case of Morse wavelets the equivalent concept is 'wavelet footprint' [76]. In order to reduce edge effects, we extended the signals through symmetric reflection of the full signal length prior CWT.

As the preliminary PSD analysis (section 2.4) showed that most of the spectral power was within the 0-50 mHz band (figure 6), we focused the CWT on that frequency range. CWT was applied to the symmetrically extended signals with the scale discretization parameter voices-per-octave = 10, which after calculation of the minimum and maximum allowable bandpass results in 45 scales with approximate frequencies ranging from 2.4 to 50 mHz. Despite signal extension, we observed that frequencies below 3 mHz were slightly outside the COI boundaries (as estimated by the MATLAB 'cwt' function) and more prone to edge effects, so we exclude them from subsequent analysis (figure 3). Thus, the number of usable scales was limited to 41 (3–50 mHz), which in turn excludes any extremely slow trends that were not removed during signal preprocessing.

After CWT, we kept the full coefficient matrix for later computation of the inverse CWT (i.e. including the data corresponding to the extended segments) whereas for the next step we would use only a shorter portion of the matrix.

2.6. Time-scale synchronization detection

Under the hypothesis that the math task may induce differentiated fNIRS fluctuations for the TD and ADHD groups, we set out to identify the frequency sub-bands that showed higher group synchronizations during the task. Since group-wise synchronization may appear as transient peaks rather than constantly, we performed a time-point-by-time-point



Figure 4. Schematics of the procedure to compute a representative inter-subject synchronization (ISS) map. Starting with the coefficient matrices of each individual within a group (left), complex-valued data are put together for each frequency bin (middle) and the ISS is computed moment-to-moment to obtain a representation of the within-group synchronization strength at each frequency over time (right colormap). The rightmost red plot depicts the maxima and minima reached by the ISS along frequencies, which are used to delimit the sub-bands to analyze (horizontal white dashed lines).

analysis, which allows capturing common oscillatory patterns that evolve dynamically over time. It is worth noting that this strategy is inspired by the underlying logic of so-called inter-subject correlation analysis, a data-driven approach devoted to assessing consistent neural responses to stimuli across individuals [77, 78].

We measured instantaneous inter-subject synchronization (ISS) using the magnitude and phase information provided by the complex-valued CWT coefficients. In fMRI studies, measures as intersubject phase synchronization [58] and pairwise phase consistency [79, 80] have been validated for the assessment of voxel-wise instantaneous phase synchronization across subjects. The former is similar to the 'circular mean resultant length' in circular statistics literature [81], also known in EEG eventrelated studies as 'inter-trial phase coherence' [82], 'inter-trial phase clustering' [83], or 'phase-locking factor' [84]. However, these measures rely only on the uniformity of phase angles, ignoring the magnitude. Thus, when applied to fNIRS it means that lowamplitude signals affect the measurement the same as those with significant amplitude. Therefore, this approach may not be entirely appropriated for fNIRS data where amplitude changes are related to the magnitude of the hemodynamic response. Since as amplitude increases, the signal-to-noise ratio improves, it is reasonable to argue that observations with higher amplitudes can contribute to a more realistic estimate of phase synchronization [85]. Under this assumption, we decided to use a closely related measure called 'inter-trial linear coherence' that combines magnitude and phase in the normalization step [82]. Since the measurement here was across subject observations and not across trials, we call this ISS, omitting 'linear' for simplicity and similarly formulated as:

$$ISS(f,t) = \frac{\sum_{k=1}^{n} F_k(f,t)}{\sqrt{n \sum_{k=1}^{n} |F_k(f,t)|^2}},$$

where $F_k(f, t)$ is the spectral estimate of observation k at frequency f and time t, and the modulus || represents the complex norm. ISS also takes values between 0 (absence of sync) and 1 (perfect sync).

ISS was computed moment-to-moment for each scale from the CWTs coefficient matrices of each group (i.e. 15 participant observations per scale). As done in section 2.4, we limited the analysis only to the time-interval between -30 s and +30 s around the task. This procedure was applied independently to the SS and CS data of each group of boys for each chromophore and ROI, obtaining an ISS representation in the time-frequency plane that can also be visualized as a color map (figure 4). Next, we choose the maximum ISS observed along each scale, which represents the highest group synchronization achieved at each specific frequency (figure 4). Please note that in this work we only take into account the ISS maxima, regardless of the time point at which they are reached. We also did not consider calculating any significance threshold at this step, as we relied on the inherent discriminative power of the subsequent feature selection and classification procedures.

Based on the observed maxima, on each ISS map we identify several peaks within specific frequency sub-bands that contain oscillatory components showing some synchronization at the group level. Since we expected such components to provide information to differentiate between groups, it was necessary to locate the sub-bands exhibiting more discriminative power. To simplify the procedure, for each case we averaged the ISS data across the three ROIs to obtain the mean ISS maxima per frequency. Thus, we simplified the analysis to only one common ISS pattern by chromophore and signal type for each of the two groups, i.e. 2 groups (TD, ADHD) \times 2 chromophores (HbO, HbR) \times 2 signal types (SS, CS) = 8 ISS patterns. Finally, to reduce noise, the ISS patterns were smoothed by moving average using a sliding window of length half the voices-per-octave (i.e. 10/2 = 5). These patterns were examined in the following steps to identify the most relevant frequency components.

2.7. Determination of frequency sub-bands and feature extraction

Like the other aforementioned synchronization measures, ISS is a compound measure that does not exist on its own at a single-subject level but represents a summary statistic of group synchronization. Therefore, to disentangle the contribution to ISS of each individual is not a straightforward issue [86]. However, ISS peaks suggest that some frequency components show similar time courses across individuals, at least within certain time intervals. In other words, there are sequential patterns common to the group that can provide distinctive information to define class membership. This concept falls into the interdisciplinary and much-studied field of timeseries classification, which encompasses a variety of techniques for identifying those properties (features) that have sufficient discriminating power to distinguish between different classes of time series (for a review, see [87]). In the context of the present work, a well-suited technique could be the one based on the shapelet framework, which addresses the classification problem by discovering primitive time-series sequences (shapelets) that are used to quantify the (dis)similarity between classes of time series [88, 89]. Shapelets provide directly interpretable information about patterns (shapes) that are important for understanding how data classes differ, a desirable property for clinical decision support systems [90].

Here we applied the basics of the shapelets technique, but instead of looking for phase-independent subsequences similar in shape (i.e. subsequences may be located anywhere in the series) we performed the analysis within a fixed time-interval, all subsequences having the same length. We did not apply any subsequence translation over time, which implies that time-series similarity also depends on the phase (i.e. on a consistent time-alignment). Therefore, instead of local, we captured global patterns present over a whole time interval. Under this approach, the term 'shapelet' may not be appropriate; however, since it relies on comparable principles and is easy to conceptualize, we will keep it here but in terms of a pattern that is maximally representative of a class within a specific timeframe.

At this point, we need to extract the time-series to be used for identifying representative *shapelets*. The average ISS patterns suggest us the frequencies that are likely to contain synchronized oscillations. By computing the inverse CWT within the specific sub-band defined by the bounds of an ISS peak, we can reconstruct such band-limited components in the time-domain. To reduce edge-effects, the inverse CWT was computed from the extended coefficient matrix that we reserved in a previous step (see section 2.5). Then, the resulting time-series were truncated to the interval between -30 s and +30 s around the task. After applying this procedure to the CWT of all the individuals belonging to a group, a set of time-series (n = 15) is available to find a reference *shapelet* for that group in a particular sub-band.

Since all the time-series have the same length and are within the same timeframe, a suitable reference *shapelet* can be obtained simply by averaging. If the time-series share a common pattern, their average should represent the group well enough. To quantify similarity with the reference *shapelet*, among other possibilities, a simple measure as Euclidean distance can be computed:

$$D(S,T) = \sqrt{\sum_{i=1}^{n} (S_i - T_i)^2}$$

where *S* denotes the reference *shapelet* and *T* a timeseries, both of length *n*. Note that *S* and *T* should be standardized to have mean 0 and standard deviation 1, which ensures to operate on the same scale. Standardization also allows us to relate the Euclidean distance to the correlation coefficient ($r = 1 - D^2/2n$).

In order to assess the capability of the *shapelet* to discriminate between groups, we can contrast the distances obtained from one group with those of the other group. For example, a *shapelet* that is representative of the TD group should have smaller distances to members of this group than to members of the ADHD group, and vice versa. Among other quality measures, the *F*-statistic for analysis of variance can be used to assess the discriminative power of a *shapelet* [89]. This statistic indicates the ratio of the between-group variability to the within-group variability as:

$$F = \frac{\sum_{i=1}^{C} \frac{(\bar{D}_i - \bar{D})^2}{C - 1}}{\sum_{i=1}^{C} \sum_{d_i \in D_i} \frac{(d_i - \bar{D}_i)^2}{n - C}}$$

where *C* is the number of classes (or groups; in our case = 2), \overline{D} is the overall mean of all distances, \overline{D}_i is the mean of the distances within class *i*, and *n* is the number of time-series. The better the *shapelet* the greater the *F* value, because the difference between-groups increases while it decreases within a group. The corresponding *p*-value can be calculated from the *F*-distribution.

Based on the average ISS patterns, we identified four candidate sub-bands in each of them that were labeled *A*, *B*, *C* and *D* in decreasing order of frequency. Each sub-band contains a peak (localmaximum) flanked by two troughs (local-minima) that delimit the frequency boundaries. For each ISS pattern, each belonging to a target group (TD or ADHD), we performed the following procedure for each sub-band (figure 5): (i) compute the inverse CWT within the sub-band from the extended coefficient matrices of both groups to obtain the corresponding time-series (n = 15 + 15 = 30). (ii) Truncate time-series to the time-interval of interest. (iii) Generate the reference *shapelet* by averaging only the





time-series of the target group (n = 15), a variability measure such as the standard error of the mean can also be computed. (iv) Calculate distances to the shapelet for the time-series of both groups. (v) Calculate *F* and *p*-value from the 15 + 15 = 30 distances. Because each ISS pattern and its sub-bands are common to all three ROIs, this procedure was applied separately to each ROI data but using the same subbands. Thus, for each ISS pattern we obtain a matrix of 30×12 distances, where each row contains the distance measures of an individual and columns correspond to the 4 sub-bands \times 3 ROIs. Since there are eight ISS patterns, the final matrix was of size 30×96 , 48 columns for SS and 48 for CS, while 15 rows correspond to the TD group and 15 to ADHD. Please note that each column has an associated F-stat and p-value, indicating how well a particular shapelet differentiates the groups.

In summary, we employed the *shapelet* approach to transform data observations at different timescales into a simple feature space of Euclidean distances, which are the only feature type used in the present work.

2.8. Classification algorithms and feature selection

To assess the feasibility of the proposed procedure to differentiate between TD and ADHD, we tested it with four well-suited machine learning algorithms for supervised binary classification, namely linear SVM [91], logistic regression (LR) [92], linear discriminant analysis (LDA) [93] and Gaussian naïve Bayes (NB) [94]. We chose these algorithms because they are well known, inherently interpretable, computationally efficient, and can work with relatively small sample sizes. Under a variety of flavors (different kernel, regularization, etc), SVM is very frequently present in neuroimaging-based studies of brain disorders, with LDA and LR being the other most popular choices [14]. Noteworthy, a similar usage scenario occurs in the ADHD research field [12]. Although less commonly used, we included NB because its ease of application and good performance in a variety of applications despite the assumption of feature independence [95].

Although based on different models, discriminative (LR & SVM) vs generative (LDA & NB), all four are within the linear classifier category, i.e. to make predictions, the classifiers try to learn the line that best separates the points of the two classes [96, 97], which depends on a linear combination of the explanatory variables. Thus, it is possible to know to what extent each feature influences the prediction, which greatly improves its interpretability. In this sense, we avoided more complex classifiers such as those using artificial neural networks or non-linear kernels, wherein the relationship between features and prediction is less transparent. We used MATLAB's implementations of the classifiers (Machine Learning Toolbox) with default settings for simplicity and reproducibility. To minimize the risk of the algorithms overfitting the available data and losing generalizability, hyperparameter optimization was not applied [98]. As usually recommended, the features were standardized using the corresponding column mean and standard deviation.

In addition to the putative functional response, fNIRS signals also contain components originating from common systemic forces and unpredictable local activity. Therefore, it is very likely that our feature matrix also contains redundant and/or irrelevant data that can degrade classifier performance by cause of overfitting and noise issues. Model regularization can be applied to some algorithms to account for statistical overfitting, however that raises the problem of choosing a suitable technique (e.g. lasso) and finding appropriate regularization parameters. To avoid increasing the complexity of the models, we addressed the problem by reducing the feature space. Feature selection is a commonly used tool to obtain a smaller subset of the most relevant features, reducing complexity while improving classification accuracy and generalization capacity [99]. A reasonable hybrid approach is to first apply a filter method, before modeling, to select some features based only on their intrinsic properties. Then more sophisticated methods such as wrappers may be employed to find the best subset of features, using the classifier itself as evaluator [100]. Among others, a benefit of such selection is an easier explanation of the prediction because the models are simpler [101].

Despite the ample offer of filter methods [102], and after trying some of the popular ones (*relief*, *minimum redundancy-maximum relevance* and *chisquare*; results not shown), we settled on a fairly straight option based on the *F*-statistic. We simply selected the features that showed *p*-values < .01, assuming that their generating shapelets were very unlikely to separate the groups by chance. In this way, we significantly reduced the features from 48 to 5 for SS and from 48 to 10 for CS.

The classification methods were first applied to filter-selected features separately for SS and CS, and then for all of them together (SS + CS). To assess the predictive performance, we applied two crossvalidation (CV) techniques for comparison purposes, namely leave-one-out (LOO) and stratified 5-fold. In the first, data was partitioned into 30 folds where each observation was used once as a test set and the remaining ones formed the training set. In the latter, five partitions were randomly chosen, each with 24 observations as the training set and 6 as the test set; folds were repartitioned over 20 Monte-Carlo repetitions (5 \times 20 = 100 models) to reduce CV variance [103], while stratification ensured that sets had the same proportion of classes (50% in our case). We used 5- instead of 10-fold because with the latter the test set size = 3 would be too close to that of LOO = 1. Since we are dealing with only two classes and our datasets are well-balanced (i.e. equal proportion of both classes), accuracy, specificity and sensitivity can be used as metrics to assess performance [104, 105], as obtained from the corresponding confusion matrices and then averaged across folds. At this point, we focused on accuracy (a commonly used metric in practice) to test the statistically significant classification performance. Thus, we computed the theoretical above-chance accuracy threshold based on the binomial cumulative distribution at $p < 10^{-3}$ for 2-classes (probability = 0.5) and a sample size = 30[106].

Afterwards, we applied a *wrapper* method to fine-tune the feature selection. We used a custommade MATLAB wrapper function that implements a sequential forward floating selection (SFFS) algorithm [99, 107]. SFFS starts with an empty set and sequentially adds one feature at a time to create candidate subsets that are evaluated by CV. After that, the best feature is added to the set. When the size of the selected set is >2, a backward step tries to optimize it by removing one or more features. This procedure is repeated until there is no performance improvement. The two aforementioned CV techniques were also applied independently to each classifier. Noteworthy, the input order of the finally selected features allows us to know their relative importance. It is also worth mentioning that our wrapper can rank features by multiple metrics at the same time, and that in this work we used two criteria to select/remove features, specificity and then accuracy. Thus, if two (or more) features equally improve the specificity, the one with the best accuracy is selected.

Once the best subset of features was selected for each classifier, we estimated the statistical significance of the observed performance through a nonparametric label permutation procedure that does not assume any particular statistical property of the data [106, 108, 109]. We generated 5000 resamples, each of which randomly permuted the labels of the two classes; realizing the null hypothesis that features do not define class membership. For each resampled data, the classification performance was evaluated using the same CV scheme as for the real data. The observed performance metrics were ranked against the corresponding null-distribution to estimate a p-value. In addition, a 95% bias corrected percentile interval was estimated as CI for each metric by bootstrapping (with replacement) over 2000 resamples, with each realization keeping the same proportion of classes (50%) and at least three distinct observations in each class [110].

Finally, we checked whether the *wrapper* performance might have been biased due to the use of a prefiltered feature set that included all data in the selection process, i.e. the 'peeking' effect [15, 111]. To this end, we repeated the *wrapper* procedure but using the full feature set (i.e. 96 features), then comparing the performance outcomes.

3. Results

3.1. Behavioral performance

On average, TD participants achieved slightly higher scores for iterations ($M_{\text{TD}} = 36.20$; SD_{TD} = 16.48; $M_{\text{ADHD}} = 29.66$; SD_{ADHD} = 14.01) and precision of results ($M_{\text{TD}} = 7.26$; SD_{TD} = 1.79; $M_{\text{ADHD}} = 7.00$; SD_{ADHD} = 2.32) than the ADHD group. However, the unpaired *t*-test did not show significant differences in iterations (t(28) = 1.169; p = .252) or in precision (t(28) = 0.351; p = .728).



Figure 6. Average normalized power spectral density (PSD) for each ROI (right, medial and left), chromophore and signal type in the frequency range 0–80 mHz. Thick colored curves refer to TD group (HbO in red and HbR in blue) while thick black ones correspond to ADHD group. Thin lines depict the 95% CI of the mean. Magenta rectangles delimit the frequency ranges that show significant differences between groups. Panel SS shows the PSDs for shallow-signals and CS for clean-signals.



Figure 7. Inter-subject synchronization (ISS) color maps for shallow-signals (SS) within each ROI. Upper rows correspond to HbO for TD and ADHD groups, while lower rows refer to HbR. The small plots to the left of each map depict the ISS maxima across frequencies for each case. Vertical black lines delimit the task-interval. Horizontal white dashed lines delimit the common sub-bands obtained by averaging ISS maxima across ROIs. Labels (A)–(D) identify each of these sub-bands.

3.2. Spectral power analysis

Figure 6 shows the average PSD distribution of fNIRS signals for the TD and ADHD group within each ROI. A common pattern can be observed, with a peak around 33 mHz corresponding to the task frequency and a most prominent peak around 4 mHz. Smaller secondary peaks are also present around 17, 24 and 67 mHz, for example. The cluster-based permutation test only showed significant differences between groups at 33 mHz for CS-HbO in the right-ROI. In all cases, most of the spectral power can be allocated roughly within the 0–50 mHz frequency range.

3.3. Time-frequency ISS maps

Figure 7 shows the ISS representation in the time-frequency plane obtained from the complex CWTs of the SS data for both groups; within the range of 3-50 mHz.

Similarly, figure 8 depicts the ISS maps corresponding to the CS data. At first glance, well-defined synchronization zones can be seen within certain frequency sub-bands, some similar in both groups and others clearly differentiated. Here we highlight some of them as examples. Regarding SS, strong synchronization can be seen in all ROIs around 33 mHz for
and group.



 Table 1. Frequency bounds of each average sub-band A–D for shallow-signals (SS) and clean-signals (CS), and for each chromophore

			Frequency sub-band (mHz)				
Signal type	Chromophore	Group	A	В	С	D	
SS	uho	TD	50-21.8	21.8-11.7	11.7-4.7	4.7-3.0	
	HbO	ADHD	50-23.3	23.3-14.3	14.3-6.7	6.7-3.0	
	ԱԻՆ	TD	50-23.0	23.0-14.3	14.3-5.4	5.4-3.0	
	HUK	ADHD	50-23.0	23.0-14.3	14.3-7.7	7.7–3.0	
	UbO	TD	50-21.8	21.8-11.7	11.7–5.4	5.4-3.0	
CS	поо	ADHD	50-20.3	20.3-7.7	7.7-4.7	4.7-3.0	
	ULD	TD	50-21.8	21.8-11.7	11.7-6.7	6.7–3.0	
	TIUK	ADHD	50-16.5	16.5-8.2	8.2-5.8	5.8-3.0	

HbO of TD group, and in right- and mid-ROI at 7 mHz and below 4 mHz. Noteworthy, the ADHD group presents even stronger ISS around 4 mHz in all ROIs, but much less evident at 7 or 33 mHz. Regarding CS, albeit to a lesser extent, TD group also synchronizes at 33 mHz while TD group does so in a more diffuse and weak way. Furthermore, TD group seems to be more synchronized during the first part of the task at 7 mHz (mid- and left-ROI), whereas the ADHD group is synchronized in the last part around 17 mHz. Yet another remarkable sync is observed for the HbR of TD group at 4 mHz. Overall, ISS analysis reveals a plurality of sub-bands that can carry information about similarities and differences between groups.

3.4. Frequency sub-bands and reference shapelets

Table 1 shows the common synchronization subbands estimated from the average ISS maxima across ROIs, labeled *A*, *B*, *C* and *D* by decreasing frequency, with *A* corresponding to the task frequency. Figure 9 illustrates how these sub-bands were delineated by locating the ISS minima surrounding each peak. Higher peaks can be seen in sub-bands A and C for HbO of TD group in both SS and CS, while ADHD group shows notable peaks in D for SS and B for CS. Regarding HbR, it shows clear peaks in A and D in all cases. Note that within the same assigned sub-band, in some cases the peaks are clearly shifted in frequency depending on the group (e.g. C sub-band for CS-HbO). It is evident again that the ISS maxima also reveal differences at certain frequencies. These average sub-bands are also depicted in the ISS maps of figures 7 and 8 by white dashed lines.

Figure 10 shows the reference *shapelets* obtained in each sub-band for SS data and figure 11 those corresponding to CS data. A rich variety of patterns can be seen, some similar across groups and others clearly different. Thus, for example, TD group exhibits rhythmic fluctuations in the *A* sub-band of SS-HbO, which are very consistent across participants as reflected by the high ISS (dashed traces); in contrast, ADHD group shows greater inter-subject variability. Another example is visible in CS-HbR-D, were TD



Figure 9. Average inter-subject synchronization (ISS) patterns across frequencies. The top row corresponds to shallow-signals (SS) and the bottom row to clean-signals (CS). Red and blue traces relate respectively to HbO and HbR of the TD group, while ADHD data is always plotted in black. Horizontal labeled rectangles identify each sub-band, whose boundaries are defined by the minima marked by vertical dashed lines.





group shows a consistent pattern of increasing then decreasing, whereas ADHD group does not. It can also be seen that the ADHD group synchronizes CS-HbO-B towards the end of the task, while TD does so visibly earlier. Once again, certain *shapelets* seem to represent well the average response of their group, while they do not fit the other one.

3.5. Classification performance

Table 2 shows the performance achieved by classifiers trained with the features selected by filtering, i.e. those with an F *p*-value < .01. Five and ten shapelets, respectively for SS and CS, generated features that met the filter criteria, each identified in figures 10 and 11 by the magenta color of their ISS traces. Overall, performance improves for all classifiers when SS and CS features are combined. All of them reached accuracy values \geq 76.7% with both CV schemes, which is the above-chance threshold at $p < 10^{-3}$ according to the theoretical binomial cumulative distribution for 2-classes and a sample size = 30. Regarding to specificity, LDA showed the highest values (88.7% and 93.3%, respectively for five-fold and LOO). When SS and CS features were used



Figure 11. Clean-signals (CS) reference shapelets. The description of the drawn elements is the same as in figure 10.

Table 2. Performance scores achieved by each classification model trained with the subset of filter-selected features for shallow-signals (SS), clean-signals (CS) and SS + CS.

		SVM		LR		LDA		NB	
Signal type	P. metric	5-FO	LOO	5-FO	LOO	5-FO	LOO	5-FO	LOO
	Accuracy	73.3	70	80	80	74.5	73.3	80.5	80
SS 5 features	Sensitivity	78.7	73.3	85.7	86.7	80.7	80	88.7	86.7
	Specificity	78.7	73.3	85.7	86.7	80.7	80	88.7	86.7
	Accuracy	81.2	80	83.5	80	78.7	80	85.5	86.7
CS 10 features	Sensitivity	82.3	80	82	80	81.7	80	84.7	86.7
	Specificity	82.3	80	82	80	81.7	80	84.7	86.7
	Accuracy	88.7	86.7	88.2	90	85.7	83.3	85.7	86.7
CS & SS 15 features	Sensitivity	91.3	93.3	91.7	93.3	82.7	73.3	84	86.7
	Specificity	86	80	84.7	86.7	88.7	93.3	87.3	86.7

Note: LOO = leave-one-out; 5-FO = 5 folds; SVM = support vector machine; LR = logistic regression; LDA = linear discriminant analysis; <math>NB = na are Bayes.

independently, performance was worse, although all classifiers still showed above-threshold accuracies for CS, while only LR and NB did so for SS. These preliminary findings suggest that SS and CS together contain valuable information to discriminate groups independently of any classification model, CS being probably the more relevant.

Figure 12 demonstrates how performance was greatly improved by using the *wrapper* for feature selection. Note that, in all cases, the classification models were very parsimonious and no more than three features were used. When compared separately, *wrapper*-selected CS features perform better than SS overall. Regarding the input order, for SS the firstin feature always belongs to TD group, A sub-band, HbO, left- or right-ROI depending on the classifier ('TD-A-HbO-L' magenta box or 'TD-A-HbO-R' red box in figure 12). It should be noted, regarding CS, that all the classifiers agree on the first two features 'AD-B-HbO-R' and 'TD-D-HbR-L' (green and yellow boxes, respectively, in figure 12). Once again, the best results were obtained with SS + CS. In fact, LR and LDA scored over 99% on all three metrics for both CV schemes, which is really high performance. NB scored lower, from 91.3% to 93.3% overall while SVM was the weakest classifier when all metrics are considered.

When evaluating SS + CS, the *wrapper* selected the same features for LR and LDA in both CV cases. Noteworthy, the first-in feature was 'AD-B-HbO-R', which was also the first for CS. The second one was 'TD-A-HbO-L', the first for SS. Finally, 'TD-D-HbR-M' from CS completed the set. Looking at the *wrapper* history, we observed that 'AD-B-HbO-R' alone achieves about 80% of the accuracy, sensitivity and specificity, which is not surprising since it has the highest F (19.6, p-value < .0001). The addition of 'TD-A-HbO-L' improves the scores up to 90% and with 'TD-D-HbR-M' they approach 100%. Therefore, the most powerful feature comes from the CS-HbO data of ADHD group, specifically from



Figure 12. Performance scores achieved by each classification model using the subset of wrapper-selected features for shallow-signals (SS), clean-signals (CS) and SS + CS. Colored rectangles on the right identify the chosen features and their input order. Each specific feature has a different color and its label is the combination 'Group-Band-Chromophore-ROI', being TD = TD group, AD = ADHD, L = left, M = mid and R = right.

band B in which a prominent ISS peak can be seen (figure 9). TD group provides the next best feature in form of consistent HbO fluctuations at task frequency in the SS data (see the ISS peak in sub-band

A). Finally, a very-slow CS-HbR component of TD group optimizes the classification (see the peak in *D*). The *shapelets* that generated these features are respectively identified by green, red and blue boxes

in figures 10 and 11. Noteworthy, NB also shared the first feature while the second and third come from the same group, sub-band and chromophore but from an adjacent ROI.

Table 3 summarizes the results obtained by *wrappers* with SS + CS features. In all cases (except for SVM in sensitivity with five-fold) permutation testing indicated significance at p < .001. LR and LDA showed the highest scores in all metrics and also the narrowest CIs. Although highly significant, NB showed lower performance and larger ICs, whereas SVM performed the worst.

Noteworthy, when the full feature set (48 SS + 48 CS) was used to feed the *wrappers* for LR and LDA, we got exactly the same feature sub-set for both CV schemes and, hence, the same scores and statistics. Therefore, feature pre-selection did not lead to more optimistic solutions with inflated performance, albeit it did reduce computational cost.

4. Discussion

The present study aimed to assess the ability of a rhythmic mental math task to induce/modulate fNIRS oscillatory patterns (referred to here as shapelets) maximally representative of the ADHD or TD condition, and the feasibility of using them in automated classification of individuals. For this purpose, distinctive shapelets were first located on the basis of group synchronization strength at certain frequencies, and then simple measures of similarity were computed per-subject as features to train four popular machine learning algorithms suitable for binary classification. We found that with proper feature selection, classifiers achieved truly high predictive statistics when defining class membership from the individual oscillatory components. Noteworthy characteristics of the study are that it is based on a unimodal approach that focuses on data drawn exclusively from the fNIRS domain, uses Euclidean distance as the only type of feature to build the classification models, and features are linked to visually identifiable waveforms. In addition, we have limited ourselves to ready-to-use, inherently interpretable supervised linear algorithms (as available in MATLAB) without hyperparameter optimization. Therefore, we did not seek high performance at the expense of combining a variety of feature types and/or tuning classifiers, but rather explore the feasibility of proposing a methodological framework as a starting point for identifying hemodynamic biomarkers, in a way accessible and interpretable enough as to inform clinical practice.

4.1. Feature selection and classifiers performance

We corroborated that efficient feature selection greatly improves/stabilizes correct classification, being particularly important when the number of features exceeds the number of observations, as is often the case with small sample sizes [102]. Feature selection aims to discard irrelevant/redundant predictors, which improves predictive ability by reducing overfitting, leads to simpler classification models with less computational cost, and makes models easier to understand by knowing the most important variables. Since finding the optimal subset of features by exhaustively searching among all possible ones is often impractical, suboptimal selection methods are commonly used as workarounds, even with the awareness that they might not be optimal. Numerous studies have proposed different strategies to select features, however, no single method has been found that works best in all scenarios [99, 112]. Here, we employed a combined approach that consists of first obtaining a candidate set of features using a simple univariate statistical test (i.e. each feature is scored independently) to select the top ranked ones, and then applying a wrapper method to find the best subset among those features using the classification algorithm itself as evaluator. Given that the available data was relatively small and that complex search methods might be more prone to overfitting, we settled on a simpler strategy fully independent of any learning method [113]. Thus, for the first filtering step, we ranked the features by F-stat and then those with p < .01 were chosen. Despite its simplicity, the filter was good enough to achieve statistically significant accuracies (based on the binomial law at $p < 10^{-3}$) for all classifiers and CV schemes when CS and SS + CS features were used. In fact, with SS + CS the accuracies ranged between 83.3% and 90% depending on the case, which are comparable (some even better) to the previously reported values in [19–21]. Therefore, the features seemed to be intrinsically distinctive and worked well regardless of the different algorithmic architecture of the classifiers and CV partitioning.

Despite these promising results, some irrelevant features could potentially degrade performance and add unnecessary complexity. We addressed such a possibility by applying a wrapper to fine-tune feature selection. Specifically, we use an SFFS algorithm that is supposed to overcome simple sequential-selection by controlling the 'nesting effect' when, once selected, a feature cannot be dropped [99, 107]. In fact, we checked that our SFFS algorithm successfully conducted the backward steps to find better solutions. Remarkably, wrappers yielded very simple models with no more than 3 features, meaning a feature-tosample ratio of 3/30 which is very unlikely to lead to overfitting [114]. Performance was particularly high when LR and LDA were trained on SS + CS features, achieving scores of around 100% on all metrics and CVs, both classifiers pointing to the same feature subset. While showing slightly lower scores, albeit >91%, NB also agreed in the same first feature and the other two shared similar properties as LR and LDA (figure 12). These results suggest that selected features

Table 3. Performance scores achieved by each classification model using the best subset of wrapper-selected features obtained from shallow- (SS) plus clean-signals (CS). Statistical significance is indicated by *p*-values and performance 95% CIs are represented within square brackets.

	P. metric	SVM		LR		LDA		NB	
Signal		5-FO	LOO	5-FO	LOO	5-FO	LOO	5-FO	LOO
CS & SS	Accuracy	82.5 <i>p</i> < .001 [68.2, 95.7]	93.3 <i>p</i> < .001 [80.0, 100]	99.3 <i>p</i> < .001 [95.8, 100]	100 <i>p</i> < .001 [96.7, 100]	99.5 <i>p</i> < .001 [96.7, 100]	100 <i>p</i> < .001 [96.7, 100]	92.2 <i>p</i> < .001 [80.3, 98.0]	93.3 <i>p</i> < .001 [83.3, 100]
	Sensitivity	72.3 p = .046 [56.0, 73.3]	86.7 <i>p</i> < .001 [66.7, 100]	99.7 <i>p</i> < .001 [93.3, 100]	100 <i>p</i> < .001 [93.3, 100]	99.7 <i>p</i> < .001 [92.0, 100]	100 <i>p</i> < .001 [93.3, 100]	91.3 <i>p</i> < .001 [73.3, 99.7]	93.3 <i>p</i> < .001 [80.0, 100]
	Specificity	92.7 <i>p</i> < .001 [78.2, 100]	100 <i>p</i> < .001 [86.7, 100]	99.0 <i>p</i> < .001 [91.8, 100]	100 <i>p</i> < .001 [93.3, 100]	99.3 <i>p</i> < .001 [90.4, 100]	100 <i>p</i> < .001 [93.3, 100]	93.0 <i>p</i> < .001 [80, 100]	93.3 <i>p</i> < .001 [80.0, 100]
	Features	2	3	3	3	3	3	3	3

Note: LOO = leave-one-out; 5-FO = 5-fold; SVM = support vector machine; LR = logistic regression; LDA = linear discriminant analysis; NB = naïve Bayes.

are quite stable and survive different classification algorithms and CV schemes. Such stability can also be seen in the CS case, in which an even wider consensus is shared across classifiers (figure 12). Therefore, it appears that the observed performance was not due to an over-selection effect [115], but to the true distinctive nature of the features.

Surprisingly, SVM performed the worst in the SS + CS case, particularly with the five-fold CV where it achieved the lowest scores and also the widest CIs. In contrast, the other classifiers were not affected by the CV, reaching similar scores and CIs with both schemes (table 3). Most likely, SVM performance degraded due to lack of proper regularization, leading to insufficient penalty for misclassification. As mentioned, one of the goals was to avoid any hyperparameter tuning and therefore we preferred to rule out SVM in the context of the present study.

As can be seen from figure 12 and table 3, LR and LDA won in the classification task when trained with SS + CS features. They both agreed on the same wrapper solution, which was exactly the same for the features preselected by filtering and for the full set. Also, they achieved comparable scores (>99%) in all metrics (significant at p < .001, but now using permutation tests to define the null hypothesis). In addition, the 95% CIs were similarly narrow, although skewed due to scores being close to the 100% ceiling. CV is known to tend towards narrower confidence bounds as accuracy approaches 100% but increasingly wide and asymmetric as sample size decreases, which can lead to under-estimate prediction errors and specially with LOO CV [116]. Despite having only 30 samples, we found that the lower limit of the CI for accuracy was never <95.8% and was less than 5% away from the mean in all cases, a deviation below the overall 15% expected for a binary classification with this number of samples [116]. The results suggest that the few selected features meet the statistical assumptions required by LR and LDA to make observations highly separable into 2-classes. The rest of this discussion

focuses on LR and LDA and their three wrapperselected shared features.

4.2. CV agreement

As reported in [12], LOO and k-fold are the most commonly adopted CV methods in ADHD studies; plus hold-out which seems more appropriate for large datasets. Some researchers have suggested that LOO may be more useful in a diagnostic scenario [98, 117], whereas others recommend k-fold or repeated random splits for more stable estimates [97, 118]. In either case, it is known that CV is compromised by small sample sizes, particularly if there are many predictors, which tend to overestimate predictive accuracy to a variable degree depending on the particularities of the study [14, 119]. In light of this, we tested LOO and five-fold expecting differences in performance, with LOO showing more optimistic scores and larger confidence bounds [116]. However, we found that both gave very similar results with LR and LDA (table 3), suggesting that the three chosen features are good enough predictors to yield stable CV measures regardless of method.

4.3. Contribution of SS and CS to classification

The present work did not rely on activation/deactivation measures as usual in fNIRS studies, but rather on the distinctive information content of frequency components. Among others, this is one of the reasons why we cautiously prefer to consider CS as clean/ corrected DSs, without further assumptions about their true nature. On another hand, we did not consider SSs as nuisance components to be discarded, but as potential carriers of representative information due to the close interaction between cognitive and autonomic functioning [120–122]. In fact, we got accuracies close to 100% when SS and CS features were combined, which out-perform other classification studies using unimodal fNIRS data.

Remarkably, the first-selected feature came from CS of the ADHD group in form of a consistent HbO

pattern in the B band (7.7–20.3 mHz) of right-ROI (figures 11 and 12). Notably, this pattern increasingly synchronizes across participants as mental task progresses and peaks near the end, while TD group peaks faster and decreases towards the end (similar time-courses can be seen in all three ROIs). It seems that some kind of underlying hemodynamic activity works differently in ADHD at that frequency. Resting-state fMRI (rs-fMRI) studies have reported frequency-specific ADHD abnormalities in local spontaneous activity of multiple brain regions, detected not only in the conventional fMRI range of 10-80 mHz [29], but also in narrower sub-bands such as 0-10, 10-27 and 27-73 mHz among others [31]. However, probably due to the high heterogeneity of ADHD presentation, consistent conclusions have not yet been reached [32]. Nevertheless, together with the findings from rs-fMRI connectivity studies [123, 124], evidence suggests that low-frequency fluctuations analysis could provide valuable insights into the dysfunction of brain-networks that has been observed in many ADHD studies (for reviews, see [7, 125]). As can be deduced from the studies mentioned, few agreed conclusions can be drawn from the large number of heterogeneous reports. Nevertheless, reports frequently concur on the key role of the PFC in the altered relationships between DMN and attention/salience networks observed in ADHD, likely related to a delayed cortical maturation [126, 127]. Since our NIRS probe interrogates part of the PFC, there is a possibility that this distinctive ADHD pattern we found is due to delayed/interfered DMN deactivation during the task, which is reflected as a specific frequency component. In line with this finding, Salmi et al [128] found increased impulsivityassociated synchronization in the medial PFC of ADHD patients during a naturalistic attention task.

Similar discussion could be applied to the third feature that also arose from CS, but from an HbR *shapelet* in the ultra-low *D*-band of TD group (also visible in all ROIs). It can be seen a consistent pattern of increasing at task onset, slowly returning to lower levels and raising again at the end (figure 11), which is absent in the ADHD group. Whether the initial effect is due to sustained oxygen consumption, deactivation, or both being interrelated, is inherently difficult to elucidate with fNIRS data alone. Interestingly, it has been reported that ultra-slow BOLD signals (0-10 mHz) seem to better differentiate ADHD from TD [31]. While more research is needed, the importance of assessing both HbO and HbR should not be underestimated.

The second-selected feature arose from an SS *shapelet* of HbO in the task frequency *A*-band and contributed significantly to improving classification (figure 12). Obviously, this oscillatory pattern is not of cerebral origin but rather reflects a coordinated systemic (even local) hemodynamic activity recorded in

superficial tissue layers. Specifically, this SS feature pertains to TD boys who show greater within-group synchronization than the ADHD group, visible also in the other two ROIs (figure 10). This finding is in line with other studies using thermal imaging [129] or EKG entropy measures [130] to differentiate ADHD with promising accuracy. This lack of synchrony probably reflects the inability of ADHD boys to coordinate autonomic resources with cognitive demands [131].

4.4. Accuracy, sensitivity and specificity

We instructed the *wrappers* to favor specificity first and accuracy second. The rationale behind is that, by favoring specificity, classifiers are less prone to false positives and so more reliable when predicting ADHD. Marking a child as ADHD when it is not true can lead to risky drug treatments, social stigma and other annoyances. Conversely, without major drawbacks, a false negative can be easily ruled-out through a complementary evaluation and/or follow-up over time. However, since the specificity scored so high and the sample was exactly balanced, the accuracy (and sensitivity) rose correspondingly, with very little difference between the metrics. Although useless in the present work, we believe that defining the order of importance of the metrics could lead to a better selection of relevant data when considering the goals and particularities of clinical trials. On another hand, albeit not predictive measures (prevalence was not considered), the metrics used here suffice as a proofof-concept, as they are considered useful for evaluating screening/diagnostic tests in clinical research [132, 133].

4.5. Limitations and future work

The current study is not intended to validate an ADHD diagnostic tool for use in a clinical setting, but to propose a methodological framework as a proof-of-concept to find functional biomarkers to help develop such a tool. However, some issues need to be addressed in order to properly frame the current findings. First, despite the high scores achieved, the limited sample size could have led to overestimating the performance of the classification. However, since ours is a fairly pure sample of only boys within a narrow age range and focused on combined ADHD subtype, the effective sample size is comparable or greater than those of other classification studies. The homogeneity of our sample can be seen as a limitation that would prevent generalizing the findings to other groups, but we feel that it is rather a strength of the study. ADHD condition is so heterogeneous that trying to find an all-in-one solution can be a difficult goal to achieve. We believe that tailoring functional biomarkers for specific use-case samples (e.g. women, drug-naïve children, adolescents, etc) may be more productive and easily implementable as a decision

tree. Obviously, the approach proposed here can be easily applied to those other use-cases, work that we currently have under way. Whether this approach can benefit from extended multi-distance measurements with partial path length correction, as proposed in [134], will require future research.

Second, the unavailability of a separate dataset for validation may cast doubt on the generalizability of the models. Therefore, to test the usefulness of the current method in clinical decision support, further research with a larger sample size and external validation will be necessary.

5. Conclusion

Over the last decade, considerable efforts have been made in the development of AI-based models to aid clinical decisions in brain disorders such as ADHD, but with little impact on the healthcare workflow. Despite promising results in the field of neuroimaging, substantial problems remain that make it difficult to bridge the gap between research and daily clinic. Among other challenges, finding reliable biomarkers via technically feasible and explainable methods is essential to enable AI implementations as trustworthy clinical decision support systems.

To contribute to the objective assessment of ADHD, we provide a proof-of-concept that very-low frequency fNIRS fluctuations induced by a rhythmic mental task accurately differentiate ADHD boys (average age 11.9 years) from non-ADHD controls at an individual level. We propose a method, based on synchronization analysis in the time-frequency plane, to find distinctive oscillatory patterns from which to extract simple distance-based features following a shapelet-like approach. We also propose an SFFS wrapper algorithm combined with ML linear models as evaluators for efficient feature selection. The results showed that with only three key features, LR and LDA classifiers achieved accuracy, sensitivity and specificity scores of nearly 100%, outperforming other fNIRS studies. An advantage of this proposal is greater transparency of the classification outcomes, since LR and LDA models are linearly interpretable and key features are visualizable as physiologically significant hemodynamic patterns. We also suggest that predictive models can be improved by targeting specific samples of ADHD patients. Our observation of altered specific frequency components is compatible with previous studies pointing to ADHD-associated abnormalities in cortical maturation and neural networks connectivity, particularly those related to PFC functioning and autonomic-cognitive interaction. To our knowledge, the methodological approach presented here has not been previously tested in fNIRSbased ADHD studies. It could also be used to assess other types of brain disorders. Despite our promising results, we emphasize that further rigorous validation is required to confirm the ability of the method

to provide robust biomarkers as adjunct indicators applicable into the clinical practice.

Data availability statement

The data generated and/or analyzed during the current study are not publicly available for legal/ethical reasons but are available from the corresponding author on reasonable request.

Acknowledgments

The authors are immensely grateful to the participants and their families who selflessly participated in this research. We acknowledge the technical assistance of Mr Miguel Ibañez and Mr Alejandro Mendez.

Conflict of interest

Joaquín Ibañez-Ballesteros and Carlos Belmonte report that they are inventors of patents licensed to Newmanbrain, S L and co-founders and scientific advisors of Newmanbrain S L, the company responsible of manufacturing the NIRS device used in this research.

Ethical statement

All procedures performed in this study involving human participants were approved by the Ethical Committee of Universidad Miguel Hernandez and conform to the 1964 Helsinki declaration and its later amendments, as well as with local regulations. Written informed consent was obtained from parents and verbal assent from all the boys.

Funding

This work was conducted in absence of any funding.

CRediT authorship contribution statement

Sergio Ortuño-Miró: Conceptualization, Methodology, Formal analysis, Investigation, Writing— Original Draft. Sergio Molina-Rodríguez: Conceptualization, Methodology, Formal analysis, Investigation, Writing—Original Draft. Carlos Belmonte: Conceptualization, Writing—Review & Editing. Joaquín Ibañez-Ballesteros: Conceptualization, Methodology, Formal analysis, Software, Writing— Review & Editing, Supervision. All authors have reviewed and approved the manuscript.

ORCID iD

Joaquín Ibañez-Ballesteros bhttps://orcid.org/0000-0001-8606-4221

References

- [1] Biederman J 2005 Attention-deficit/hyperactivity disorder: a selective overview *Biol. Psychiatry* **57** 1215–20
- [2] Canals J, Morales-Hidalgo P, Jané M C and Domènech E 2018 ADHD prevalence in Spanish preschoolers: comorbidity, socio-demographic factors, and functional consequences J. Atten. Disord. 22 143–53
- [3] Thapar A and Cooper M 2016 Attention deficit hyperactivity disorder *Lancet* 387 1240–50
- [4] Gadow K D, Drabick D A G, Loney J, Sprafkin J, Salisbury H, Azizian A and Schwartz J 2004 Comparison of ADHD symptom subtypes as source-specific syndromes J. *Child Psychol. Psychiatry* 45 1135–49
- [5] Campbell S B 2000 Attention-deficit/hyperactivity disorder: a developmental view *Handbook of Developmental Psychopathology* (Boston, MA: Springer) pp 383–401
- [6] Faraone S V *et al* 2021 The world federation of ADHD international consensus statement: 208 evidence-based conclusions about the disorder *Neurosci. Biobehav. Rev.* 128 789–818
- [7] Posner J, Polanczyk G V and Sonuga-Barke E 2020 Attention-deficit hyperactivity disorder *Lancet* 395 450–62
- [8] American Psychiatric Association 2013 Diagnostic and Statistical Manual of Mental Disorders (Arlington, VA: American Psychiatric Association)
- [9] Adesman A R 2001 The diagnosis and management of attention-deficit/hyperactivity disorder in pediatric patients Prim. Care Companion J. Clin. Psychiatry 3 66–77
- [10] Halperin J M, Bédard A C V and Curchack-Lichtin J T 2012 Preventive interventions for ADHD: a neurodevelopmental perspective *Neurotherapeutics* 9 531–41
- [11] Pereira-Sanchez V and Castellanos F X 2021 Neuroimaging in attention-deficit/hyperactivity disorder *Curr. Opin. Psychiatry* 34 105–11
- [12] Loh H W, Ooi C P, Barua P D, Palmer E E, Molinari F and Acharya U R 2022 Automated detection of ADHD: current trends and future perspective *Comput. Biol. Med.* 146 105525
- [13] ADHD-200-Consortium 2012 The ADHD-200
 Consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience *Front. Syst. Neurosci.* 6 62
- [14] Arbabshirani M R, Plis S, Sui J and Calhoun V D 2017 Single subject prediction of brain disorders in neuroimaging: promises and pitfalls *Neuroimage* 145 137–65
- [15] Pulini A A, Kerr W T, Loo S K and Lenartowicz A 2019 Classification accuracy of neuroimaging biomarkers in attention-deficit/hyperactivity disorder: effects of sample size and circular analysis *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 4 108–20
- [16] Mauri M, Nobile M, Bellina M, Crippa A and Brambilla P 2018 Light up ADHD: I. Cortical hemodynamic responses measured by functional near infrared spectroscopy (fNIRS) *J. Affect. Disord.* 234 358–64
- [17] Pinti P, Tachtsidis I, Hamilton A, Hirsch J, Aichelburg C, Gilbert S and Burgess P W 2018 The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience Ann. New York Acad. Sci. 1464 1–25
- [18] Güven A, Altınkaynak M, Dolu N, İzzetoğlu M, Pektaş F, Özmen S, Demirci E and Batbat T 2020 Combining functional near-infrared spectroscopy and EEG measurements for the diagnosis of attention-deficit hyperactivity disorder *Neural Comput. Appl.* 32 8367–80
- [19] Monden Y et al 2015 Individual classification of ADHD children by right prefrontal hemodynamic responses during a go/no-go task as assessed by fNIRS *NeuroImage Clin.* 9 1–12
- [20] Crippa A, Salvatore C, Molteni E, Mauri M, Salandi A, Trabattoni S, Agostoni C, Molteni M, Nobile M and Castiglioni I 2017 The utility of a computerized algorithm based on a multi-domain profile of measures for the

diagnosis of attention deficit/hyperactivity disorder *Front. Psychiatry* **8** 1–10

- [21] Gu Y, Miao S, Han J, Liang Z, Ouyang G, Yang J and Li X 2018 Identifying ADHD children using hemodynamic responses during a working memory task measured by functional near-infrared spectroscopy *J. Neural Eng.* 15 035005
- [22] Takahashi T, Takikawa Y, Kawagoe R, Shibuya S, Iwano T and Kitazawa S 2011 Influence of skin blood flow on near-infrared spectroscopy signals measured on the forehead during a verbal fluency task *Neuroimage* 57 991–1002
- [23] Haeussinger F B, Dresler T, Heinzel S, Schecklmann M, Fallgatter A J and Ehlis A-C 2014 Reconstructing functional near-infrared spectroscopy (fNIRS) signals impaired by extra-cranial confounds: an easy-to-use filter method *Neuroimage* 95 69–79
- [24] Tachtsidis I and Scholkmann F 2016 False positives and false negatives in functional near-infrared spectroscopy: issues, challenges, and the way forward *Neurophotonics* 3 031405
- [25] Caldwell M, Scholkmann F, Wolf U, Wolf M, Elwell C and Tachtsidis I 2016 Modelling confounding effects from extracerebral contamination and systemic factors on functional near-infrared spectroscopy *Neuroimage* 143 91–105
- [26] Kirilina E, Jelzow A, Heine A, Niessing M, Wabnitz H, Brühl R, Ittermann B, Jacobs A M and Tachtsidis I 2012 The physiological origin of task-evoked systemic artefacts in functional near infrared spectroscopy *Neuroimage* 61 70–81
- [27] Molina-Rodríguez S, Mirete-Fructuoso M, Martínez L and Ibañez-Ballesteros J 2022 Frequency-domain analysis of fNIRS fluctuations induced by rhythmic mental arithmetic *Psychophysiology* 1 1–25
- [28] Yu-Feng W, Yong H, Chao-Zhe Z, Qing-Jiu C, Man-Qiu S, Meng L, Li-Xia T, Tian-Zi J and Yu-Feng W 2007 Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI *Brain Dev.* 29 83–91
- [29] Tang C, Wei Y, Zhao J and Nie J 2018 Different developmental pattern of brain activities in ADHD: a study of resting-state fMRI *Dev. Neurosci.* 40 246–57
- [30] Zhang J, Kendrick K M, Lu G and Feng J 2015 The fault lies on the other side: altered brain functional connectivity in psychiatric disorders is mainly caused by counterpart regions in the opposite hemisphere *Cereb. Cortex* 25 3475–86
- [31] Yu X, Yuan B, Cao Q, An L, Wang P, Vance A, Silk T J, Zang Y, Wang Y and Sun L 2016 Frequency-specific abnormalities in regional homogeneity among children with attention deficit hyperactivity disorder: a resting-state fMRI study *Sci. Bull.* 61 682–92
- [32] Wang J-B, Zheng L-J, Cao Q-J, Wang Y-F, Sun L, Zang Y-F and Zhang H 2017 Inconsistency in abnormal brain activity across cohorts of adhd-200 in children with attention deficit hyperactivity disorder *Front. Neurosci.* 11 1–10
- [33] Scholkmann F, Gerber U, Wolf M and Wolf U 2013 End-tidal CO₂: an important parameter for a correct interpretation in functional brain studies using speech tasks *Neuroimage* 66 71–79
- [34] Zimeo Morais G A, Scholkmann F, Balardin J B, Furucho R A, de Paula R C V, Biazoli C E and Sato J R 2017 Non-neuronal evoked and spontaneous hemodynamic changes in the anterior temporal region of the human head may lead to misinterpretations of functional near-infrared spectroscopy signals *Neurophotonics* 5 1
- [35] Orihuela-Espina F, Leff D R, James D R C, Darzi A W and Yang G Z 2010 Quality control and assurance in functional near infrared spectroscopy (fNIRS) experimentation *Phys. Med. Biol.* 55 3701–24
- [36] Kocsis L, Herman P and Eke A 2006 The modified Beer-Lambert law revisited Phys. Med. Biol. 51 N91
- [37] Delpy D T, Cope M, Van Der Zee P, Arridge S, Wray S and Wyatt J 1988 Estimation of optical pathlength through

tissue from direct time of flight measurement *Phys. Med. Biol.* **33** 1433–42

- [38] Huppert T J, Diamond S G, Franceschini M A and Boas D A 2009 HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain Appl. Opt. 48 0–33
- [39] Reddy P, Izzetoglu M, Shewokis P A, Sangobowale M, Diaz-Arrastia R and Izzetoglu K 2021 Evaluation of fNIRS signal components elicited by cognitive and hypercapnic stimuli *Sci. Rep.* 11 23457
- [40] Scholkmann F, Kleiser S, Metz A J, Zimmermann R, Mata Pavia J, Wolf U and Wolf M 2014 A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology *Neuroimage* 85 6–27
- [41] Brigadoi S and Cooper R J 2015 How short is short? Optimum source-detector distance for short-separation channels in functional near-infrared spectroscopy *Neurophotonics* 2 025005
- [42] Gagnon L, Cooper R J, Yücel M A, Perdue K L, Greve D N and Boas D A 2012 Short separation channel location impacts the performance of short channel regression in NIRS *Neuroimage* 59 2518–28
- [43] Yücel M A, Selb J J, Huppert T J, Franceschini M A and Boas D A 2017 Functional near infrared spectroscopy: enabling routine functional brain imaging *Curr. Opin. Biomed. Eng.* 4 78–86
- [44] Gagnon L, Yücel M A, Boas D A and Cooper R J 2014 Further improvement in reducing superficial contamination in NIRS using double short separation measurements *Neuroimage* 85 127–35
- [45] Holland P W and Welsch R E 1977 Robust regression using iteratively reweighted least-squares Commun. Stat. Theory Methods 6 813–27
- [46] Tak S, Uga M, Flandin G, Dan I and Penny W D 2016 Sensor space group analysis for fNIRS data J. Neurosci. Methods 264 103–12
- [47] Plichta M M, Herrmann M J, Baehne C G, Ehlis A-C, Richter M M, Pauli P and Fallgatter A J 2006 Event-related functional near-infrared spectroscopy (fNIRS): are the measurements reliable? *Neuroimage* 31 116–24
- [48] Schecklmann M, Ehlis A-C, Plichta M M and Fallgatter A J 2008 Functional near-infrared spectroscopy: a long-term reliable tool for measuring brain activity during verbal fluency *Neuroimage* 43 147–55
- [49] Näsi T, Mäki H, Hiltunen P, Heiskala J, Nissilä I, Kotilahti K and Ilmoniemi R J 2013 Effect of task-related extracerebral circulation on diffuse optical tomography: experimental data and simulations on the forehead *Biomed. Opt. Express* 4 412
- [50] Nambu I, Ozawa T, Sato T, Aihara T, Fujiwara Y, Otaka Y, Osu R, Izawa J and Wada Y 2017 Transient increase in systemic interferences in the superficial layer and its influence on event-related motor tasks: a functional near-infrared spectroscopy study J. Biomed. Opt. 22 035008
- [51] Welch P 1967 The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms *IEEE Trans. Audio Electroacoust.* 15 70–73
- [52] Ilvedson C R 1998 Transfer Function Estimates Using Time-Frequency Analysis (Cambridge: Massachusetts Institute of Technology)
- [53] Aarabi A and Huppert T J 2016 Characterization of the relative contributions from systemic physiological noise to whole-brain resting-state functional near-infrared spectroscopy data using single-channel independent component analysis *Neurophotonics* 3 025004
- [54] Maris E and Oostenveld R 2007 Nonparametric statistical testing of EEG- and MEG-data J. Neurosci. Methods 164 177–90
- [55] Cohen L 1995 *Time-frequency Analysis* (Englewood Cliffs, NJ: Prentice Hall)
- [56] Gröchenig K 2001 Foundations of Time-Frequency Analysis (Boston, MA: Birkhäuser Boston)

- [57] Morales S and Bowers M E 2022 Time-frequency analysis methods and their application in developmental EEG data *Dev. Cogn. Neurosci.* 54 101067
- [58] Glerean E, Salmi J, Lahnakoski J M, Jääskeläinen I P and Sams M 2012 Functional magnetic resonance imaging phase synchronization as a measure of dynamic functional connectivity *Brain Connect.* 2 91–101
- [59] Wacker M and Witte H 2013 Time-frequency techniques in biomedical signal analysis: a tutorial review of similarities and differences *Methods Inf. Med.* 52 279–96
- [60] Munia T T K and Aviyente S 2019 Time-frequency based phase-amplitude coupling measure for neuronal oscillations Sci. Rep. 9 1–15
- [61] Liu T, Luo Z, Huang J and Yan S 2018 A comparative study of four kinds of adaptive decomposition algorithms and their applications Sensors 18 1–51
- [62] Kijewski-Correa T and Kareem A 2006 Efficacy of Hilbert and wavelet transforms for time-frequency analysis J. Eng. Mech. 132 1037–49
- [63] Cohen A and Kovacevic J 1996 Wavelets: the mathematical background Proc. IEEE 84 514–22
- [64] Hramov A E, Koronovskii A A, Makarov V A, Pavlov A N and Sitnikova E 2015 Wavelets in Neuroscience (Berlin: Springer)
- [65] Sadowsky J 1996 Investigation of signal characteristics using the continuous wavelet transform *Johns Hopkins APL Tech. Dig.* 17 258–69
- [66] Olhede S C and Walden A T 2002 Generalized Morse wavelets IEEE Trans. Signal Process. 50 2661–70
- [67] Lilly J M and Olhede S C 2010 On the analytic wavelet transform *IEEE Trans. Inf. Theory* 56 4135–56
- [68] Lilly J M and Olhede S C 2012 Generalized Morse wavelets as a superfamily of analytic wavelets *IEEE Trans. Signal Process.* 60 6036–41
- [69] Suwansawang S and Halliday D M 2017 Time-frequency based coherence and phase locking value analysis of human locomotion data using generalized Morse wavelets *BIOSIGNALS 2017-10th Int. Conf. Bio-Inspired Syst. Signal Process* vol 4 pp 34–41
- [70] Wachowiak M P, Wachowiak-Smolíková R, Johnson M J, Hay D C, Power K E and Williams-Bell F M 2018 Quantitative feature analysis of continuous analytic wavelet transforms of electrocardiography and electromyography *Phil. Trans. R. Soc.* A **376** 20170250
- [71] Wiklendt L, Brookes S J H, Costa M, Travis L, Spencer N J and Dinning P G 2020 A novel method for electrophysiological analysis of EMG signals using mesaclip *Front. Physiol.* 11 1–10
- [72] Nakhnikian A, Ito S, Dwiel L L, Grasse L M, Rebec G V, Lauridsen L N and Beggs J M 2016 A novel cross-frequency coupling detection method using the generalized Morse wavelets J. Neurosci. Methods 269 61–73
- [73] Huggins C J, Escudero J, Parra M A, Scally B, Anghinah R, Vitória Lacerda De Araújo A, Basile L F and Abasolo D 2021 Deep learning of resting-state electroencephalogram signals for three-class classification of Alzheimer's disease, mild cognitive impairment and healthy ageing *J. Neural Eng.* 18 046087
- [74] Ajith K, Menaka R and Kumar S S 2021 EEG based mental state analysis J. Phys.: Conf. Ser. 1911 012014
- [75] Perpetuini D, Cardone D, Filippini C, Chiarelli A M and Merla A 2021 A motion artifact correction procedure for fNIRS signals based on wavelet transform and infrared thermography video tracking *Sensors* 21 5117
- [76] Lilly J M 2017 Element analysis: a wavelet-based method for analysing time-localized events in noisy time series *Proc. R. Soc.* A 473 20160776
- [77] Nastase S A, Gazzola V, Hasson U and Keysers C 2019 Measuring shared responses across subjects using intersubject correlation Soc. Cogn. Affect. Neurosci. 14 669–87

- [78] Hasson U, Nir Y, Levy I, Fuhrmann G and Malach R 2004 Intersubject synchronization of cortical activity during natural vision *Science* 303 1634–40
- [79] Bolt T, Nomi J S, Vij S G, Chang C and Uddin L Q 2018 Inter-subject phase synchronization for exploratory analysis of task-fMRI *Neuroimage* 176 477–88
- [80] Kauppi J-P, Pajula J and Tohka J 2014 A versatile software package for inter-subject correlation based analyses of fMRI *Front. Neuroinform.* 8 1–13
- [81] Berens P 2009 CircStat: a MATLAB toolbox for circular statistics J. Stat. Softw. 31 1–21
- [82] Delorme A and Makeig S 2004 EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis J. Neurosci. Methods 134 9–21
- [83] Cohen M X 2014 Analyzing Neural Time Series Data (Cambridge, MA: MIT Press)
- [84] Tallon-Baudry C, Bertrand O, Delpuech C and Pernier J 1996 Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human J. Neurosci. 16 4240–9
- [85] Bastos A M and Schoffelen J-M 2016 A tutorial review of functional connectivity analysis methods and their interpretational pitfalls *Front. Syst. Neurosci.* 9 1–23
- [86] Richter C G, Thompson W H, Bosman C A and Fries P 2015 A jackknife approach to quantifying single-trial correlation between covariance-based metrics undefined on a single-trial basis *Neuroimage* 114 57–70
- [87] Fulcher B D 2018 Feature-based time-series analysis Feature Engineering for Machine Learning and Data Analytics Dong G and Liu H (Boca Raton, FL: CRC press) pp 87–116
- [88] Ye L and Keogh E 2011 Time series shapelets: a novel technique that allows accurate, interpretable and fast classification *Data Min. Knowl. Discov.* 22 149–82
- [89] Hills J, Lines J, Baranauskas E, Mapp J and Bagnall A 2014 Classification of time series by shapelet transformation Data Min. Knowl. Discov. 28 851–81
- [90] Richard A, Mayag B, Talbot F, Tsoukias A and Meinard Y 2020 Transparency of Classification Systems for Clinical Decision Support Information Processing and Management of Uncertainty in Knowledge-Based Systems 1239 (London: Nature Publishing Group) pp 99–113
- [91] Hearst M A, Dumais S T, Osuna E, Platt J and Scholkopf B 1998 Support vector machines *IEEE Intell. Syst. Appl.* 13 18–28
- [92] Hosmer D W Jr, Lemeshow S and Sturdivant R X 2013 Applied Logistic Regression vol 398 (New York: Wiley)
- [93] Fisher R A 1936 The use of multiple measurements in taxonomic problems Ann. Eugen. 7 179–88
- [94] Kuncheva L I 2006 On the optimality of Naïve Bayes with dependent binary features Pattern Recognit. Lett. 27 830–7
- [95] Hand D J and Yu K 2001 Idiot's bayes: not so stupid after all? Int. Stat. Rev. 69 385
- [96] Pereira F, Mitchell T and Botvinick M 2009 Machine learning classifiers and fMRI: a tutorial overview *Neuroimage* 45 S199–209
- [97] Hastie T, Tibshirani R, Friedman J H and Friedman J H 2009 The Elements of Statistical Learning: Data Mining, Inference, and Prediction vol 2 (New York: Springer)
- [98] Saeb S, Lonini L, Jayaraman A, Mohr D C and Kording K P 2017 The need to approximate the use-case in clinical machine learning *Gigascience* 6 1–9
- [99] Somol P, Novovičová J and Pudil P 2010 Efficient feature subset selection and subset size optimization *Pattern Recognition Recent Advances* ed A Herout (Rijeka: Intech) (https://doi.org/10.5772/224)
- [100] Liu H and Yu L 2005 Toward integrating feature selection algorithms for classification and clustering *IEEE Trans. Knowl. Data Eng.* 17 491–502
- [101] Amann J et al 2022 To explain or not to explain?—artificial intelligence explainability in clinical decision support systems ed H H-S Lu PLoS Digit. Heal. 1 e0000016

- [102] Jovic A, Brkic K and Bogunovic N 2015 A review of feature selection methods with applications 2015 38th Int. Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) (IEEE) pp 1200–5
- [103] Kim J-H 2009 Estimating classification error rate: repeated cross-validation, repeated hold-out and bootstrap *Comput. Stat. Data Anal.* 53 3735–45
- [104] Luque A, Carrasco A, Martín A and de Las Heras A 2019 The impact of class imbalance in classification performance metrics based on the binary confusion matrix *Pattern Recognit.* 91 216–31
- [105] Hossin M and Sulaiman M N 2015 A review on evaluation metrics for data classification evaluations Int. J. Data Min. Knowl. Manage. Process 5 01–11
- [106] Combrisson E and Jerbi K 2015 Exceeding chance level by chance: the caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy J. Neurosci. Methods 250 126–36
- [107] Pudil P, Novovičová J and Kittler J 1994 Floating search methods in feature selection *Pattern Recognit. Lett.* 15 1119–25
- [108] Bzdok D 2017 Classical statistics and statistical learning in imaging neuroscience Front. Neurosci. 11 1–23
- [109] Golland P and Fischl B R 2003 Permutation tests for classification: towards statistical significance in image-based studies *Information Processing in Medical Imaging* vol 18 (Berlin: Springer) pp 330–41
- [110] Fu W J, Carroll R J and Wang S 2005 Estimating misclassification error with small samples via bootstrap cross-validation *Bioinformatics* 21 1979–86
- [111] Kriegeskorte N, Simmons W K, Bellgowan P S and Baker C I 2009 Circular analysis in systems neuroscience: the dangers of double dipping *Nat. Neurosci.* 12 535–40
- [112] Remeseiro B and Bolon-Canedo V 2019 A review of feature selection methods in medical applications *Comput. Biol. Med.* 112 25–29
- [113] Reunanen J 2003 Overfitting in making comparisons between variable selection methods J. Mach. Learn. Res. 3 1371–82
- [114] Vabalas A, Gowen E, Poliakoff E, Casson A J and Hernandez-Lemus E 2019 Machine learning algorithm validation with a limited sample size *PLoS One* **14** 1–20
- [115] Raudys S 2006 Feature over-selection Structural, Syntactic, and Statistical Pattern Recognition (Berlin: Springer) pp 622–31
- [116] Varoquaux G 2018 Cross-validation failure: small sample sizes lead to large error bars *Neuroimage* 180 68–77
- [117] Tougui I, Jilbab A and El Mhamdi J 2021 Impact of the choice of cross-validation techniques on the results of machine learning-based diagnostic applications *Healthc. Inform. Res.* 27 189–99
- [118] Varoquaux G, Raamana P R, Engemann D A, Hoyos-Idrobo A, Schwartz Y and Thirion B 2017 Assessing and tuning brain decoders: cross-validation, caveats, and guidelines *Neuroimage* 145 166–79
- [119] Flint C et al 2021 Systematic misestimation of machine learning performance in neuroimaging studies of depression Neuropsychopharmacology 46 1510–7
- [120] Forte G, De Pascalis V, Favieri F and Casagrande M 2019 Effects of blood pressure on cognitive performance: a systematic review J. Clin. Med. 9 34
- [121] Forte G, Favieri F and Casagrande M 2019 Heart rate variability and cognitive function: a systematic review *Front. Neurosci.* 13 1–11
- [122] Wang X, Liu B, Xie L, Yu X, Li M and Zhang J 2016 Cerebral and neural regulation of cardiovascular activity during mental stress *Biomed. Eng. Online* 15 335–47
- [123] Wang X-H and Li L 2015 Altered temporal features of intrinsic connectivity networks in boys with combined type of attention deficit hyperactivity disorder *Eur. J. Radiol.* 84 947–54

- [124] Castellanos F X and Proal E 2012 Large-scale brain systems in ADHD: beyond the prefrontal-striatal model *Trends Cogn. Sci.* 16 17–26
- [125] Castellanos F X and Aoki Y 2016 Intrinsic functional connectivity in attention-deficit/hyperactivity disorder: a science in development *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 1 253–61
- [126] Shaw P K, Eckstrand W, Sharp J, Blumenthal J P, Lerch D, Greenstein L, Clasen A and Evans J 2007 Attention-deficit/hyperactivity disorder is characterized by a delay in cortical maturation *Proc. Natl Acad. Sci.* 104 19649–54
- [127] Halperin J M and Schulz K P 2006 Revisiting the role of the prefrontal cortex in the pathophysiology of attention-deficit/hyperactivity disorder *Psychol. Bull.* 132 560–81
- [128] Salmi J, Metwaly M, Tohka J, Alho K, Leppämäki S, Tani P, Koski A, Vanderwal T and Laine M 2020 ADHD desynchronizes brain activity during watching a distracted multi-talker conversation *Neuroimage* 216 116352

- [129] Coben R and Myers T E 2009 Sensitivity and specificity of long wave infrared imaging for attention-deficit/ hyperactivity disorder J. Atten. Disord. 13 56–65
- [130] Koh J E W, Ooi C P, Lim-Ashworth N S, Vicnesh J, Tor H T, Lih O S, Tan R S, Acharya U R and Fung D S S 2022 Automated classification of attention deficit hyperactivity disorder and conduct disorder using entropy features with ECG signals *Comput. Biol. Med.* **140** 105120
- [131] Bellato A, Arora I, Hollis C and Groom M J 2020 Is autonomic nervous system function atypical in attention deficit hyperactivity disorder (ADHD)? A systematic review of the evidence *Neurosci. Biobehav. Rev.* 108 182–206
- [132] Trevethan R 2017 Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice *Front. Public Health* 5 1–7
- [133] Eusebi P 2013 Diagnostic accuracy measures *Cerebrovasc.* Dis. 36 267–72
- [134] Song X, Chen X, Chen L, An X and Ming D 2020
 Performance improvement for detecting brain function using fNIRS: a multi-distance probe configuration with PPL method *Front. Hum. Neurosci.* 14 1–11

Biblioteca