

Jornadas de Automática

Comparativa de Random Forest, XGBoost y SVM-RBF para la detección de TDAH

Adrián Jiménez-García^{a,*}, Raquel García-Ausín^{b,c,*}, Jesús Enrique Sierra-García^d, Bruno Baruque-Zanón^d

^a Escuela Politécnica Superior, Universidad de Burgos, Burgos, España

^b Departamento de Ciencias de la Educación, Universidad de Burgos, Burgos, España

^c Asociación ABUDAH-TDAH, Burgos, España.

^d Departamento de Digitalización, Universidad de Burgos, Burgos, España.

To cite this article: Jiménez-García, A., García-Ausín, R., Sierra-García, J.E., Baruque-Zanón, B. 2026. Comparativa de Random Forest, XGBoost y SVM-RBF para la detección de TDAH. Jornadas de Automática, 47. <https://doi.org/10.17979/ja-cea.2026.47.13843>

Resumen

El diagnóstico del Trastorno por Déficit de Atención e Hiperactividad (TDAH) se basa principalmente en evaluaciones subjetivas, como entrevistas clínicas y cuestionarios conductuales, lo que puede generar sesgos e inconsistencias diagnósticas. En este contexto, las señales electroencefalográficas (EEG) han surgido como una herramienta prometedora debido a las alteraciones observadas en sus patrones espectrales y dinámicas funcionales en personas con TDAH. Este trabajo explora la aplicación de técnicas de aprendizaje automático para la detección del TDAH a partir de señales EEG. En concreto, se implementaron y compararon tres modelos de clasificación: Random Forest, Máquinas de Vectores de Soporte con función de base radial (SVM-RBF) y XGBoost. Su rendimiento se evaluó mediante exactitud, precisión y sensibilidad para determinar su capacidad discriminativa. Los resultados evidencian el potencial de la inteligencia artificial como apoyo al diagnóstico clínico tradicional, aunque todavía resultan insuficientes para aplicaciones clínicas autónomas.

Palabras clave: TDAH, EEG, aprendizaje automático, Random Forest, SVM-RBF, XGBoost, clasificación automatizada, diagnóstico asistido por ordenador, neurociencia computacional, biomarcadores de EEG, detección precoz

Comparison of Random Forest, XGBoost, and RBF SVM for ADHD Detection

Abstract

The diagnosis of Attention-Deficit/Hyperactivity Disorder (ADHD) relies primarily on subjective assessments, such as clinical interviews and behavioral questionnaires, which may lead to diagnostic bias and inconsistencies. In this context, electroencephalographic (EEG) signals have emerged as a promising tool due to the alterations observed in their spectral patterns and functional dynamics in individuals with ADHD. This study explores the application of machine learning techniques for ADHD detection based on EEG signals. Specifically, three classification models were implemented and compared: Random Forest, Support Vector Machines with a Radial Basis Function kernel (SVM-RBF), and XGBoost. Their performance was evaluated using accuracy, precision, and sensitivity metrics to assess their discriminative capability. The results demonstrate the potential of artificial intelligence as a support tool for traditional clinical diagnosis, although they remain insufficient for fully autonomous clinical applications.

Keywords: ADHD, EEG, Machine Learning, Random Forest, SVM-RBF, XGBoost, Automated Classification, Computer-Aided Diagnosis, Computational Neuroscience, EEG Biomarkers, Early Detection.

1. Introducción

El trastorno por déficit de atención e hiperactividad (TDAH) es uno de los trastornos del neurodesarrollo más

frecuentes, con una prevalencia mundial del 7,2% [American Psychiatric Association, 2022; Koutsoklenis & Honkasilta, 2023]. Se caracteriza por síntomas persistentes de inatención, hiperactividad e impulsividad que afectan al funcionamiento

cognitivo, conductual y social [García-Ausin & Sierra-García, 2024; García-Ausin et al., 2026]. Pese a los avances en su comprensión clínica, el diagnóstico sigue basándose principalmente en entrevistas y cuestionarios conductuales, susceptibles de generar sesgos e inconsistencias diagnósticas [Lenartowicz & Loo, 2014].

En este contexto, la identificación de biomarcadores objetivos se ha convertido en una línea de investigación prioritaria. Entre las herramientas neurofisiológicas disponibles, la electroencefalografía (EEG) destaca por ser una técnica no invasiva, con alta resolución temporal y capacidad para registrar la actividad cerebral en tiempo real. Diversos estudios han evidenciado alteraciones en los patrones espectrales y funcionales del EEG en individuos con TDAH, sugiriendo su utilidad como fuente de características para el diagnóstico asistido por inteligencia artificial [Slater et al., 2022]. Paralelamente, el aprendizaje automático ha ampliado las posibilidades de análisis de señales biomédicas complejas. En particular, algoritmos basados en árboles de decisión, como Random Forest y XGBoost, han demostrado ser eficaces para modelar relaciones no lineales, manejar datos de alta dimensionalidad y mitigar el sobreajuste [Chen & Guestrin, 2016]. Estas características los convierten en opciones adecuadas para la clasificación automática de señales EEG, caracterizadas por una elevada variabilidad y presencia de ruido.

En este trabajo se aplican técnicas de aprendizaje automático, concretamente Random Forest, Máquinas de Vectores Soporte con función de base radial (SVM-RBF) y XGBoost, para la detección del TDAH mediante señales EEG. El rendimiento de los modelos se evalúa en términos de exactitud, precisión y sensibilidad. El objetivo es contribuir al desarrollo de herramientas computacionales que complementen la evaluación clínica tradicional, favoreciendo diagnósticos más objetivos, reproducibles y tempranos.

El artículo está estructurado de la siguiente manera. La Sección 2 recoge algunos trabajos relacionados sobre la relación EEG y TDAH. La Sección 3 presenta la metodología empleada. La Sección 4 resume los resultados obtenidos. Finalmente, la Sección 5 presenta las conclusiones y trabajos futuros.

2. TDAH y señales EEG

Numerosos estudios han demostrado la utilidad del EEG para identificar patrones neurofisiológicos asociados al TDAH [Mao et al., 2025; López et al., 2025]. Se han observado incrementos en la actividad theta y alteraciones en las bandas alfa y beta [Adamou et al., 2020]. La evidencia muestra de forma consistente un aumento de la actividad theta frontal, relacionado con una menor activación cortical y dificultades para mantener la atención sostenida. Asimismo, se ha descrito una reducción relativa de la actividad beta, especialmente en regiones frontocentrales, asociada a déficits en el control ejecutivo, la inhibición conductual y la autorregulación cognitiva [Adamou et al., 2020; Loo et al., 2016]. Otro indicador ampliamente estudiado es la relación theta/beta (TBR), utilizada históricamente como biomarcador del TDAH debido al

aumento simultáneo de theta y la disminución de beta en determinados subgrupos clínicos [Snyder & Hall, 2006].

Además de las alteraciones espectrales clásicas, estudios recientes han identificado cambios en la conectividad funcional y la sincronización neuronal entre distintas regiones cerebrales. En particular, se han descrito anomalías en redes frontoparietales y circuitos implicados en la atención, el monitoreo de la respuesta y la memoria de trabajo, lo que sugiere una organización cerebral menos eficiente en individuos con TDAH [Kember et al., 2023; Mao et al., 2025]. Asimismo, se han observado diferencias en medidas de complejidad y entropía de la señal EEG, indicativas de una dinámica neuronal atípica durante tareas cognitivas y de control inhibitorio. En conjunto, estas evidencias respaldan el uso del EEG para detectar alteraciones neurofuncionales asociadas al TDAH y extraer características relevantes para modelos de clasificación basados en inteligencia artificial.

A pesar de su potencial, la evidencia también señala limitaciones importantes para su aplicación diagnóstica. El índice theta/beta (TBR), uno de los biomarcadores más estudiados, ha mostrado resultados inconsistentes y una considerable superposición entre pacientes con TDAH y sujetos control [van Dijk et al., 2020]. Además, metaanálisis recientes no han hallado asociaciones claras entre los subtipos neurofisiológicos identificados mediante EEG y las manifestaciones conductuales del trastorno [Boxum et al., 2025]. Estas limitaciones evidencian la necesidad de continuar investigando nuevas técnicas para la detección del TDAH basadas en señales EEG.

En los últimos años, numerosos estudios han aplicado técnicas de aprendizaje automático a señales EEG para apoyar el diagnóstico del TDAH. Mao et al. (2025) combinaron características espectrales, de entropía y conectividad funcional con algoritmos como Random Forest, XGBoost, CatBoost y LightGBM, incorporando además análisis de interpretabilidad mediante SHAP. Kim et al. (2025) emplearon XGBoost sobre registros EEG de 19 canales con validación Leave-One-Subject-Out, identificando características relevantes principalmente en bandas beta. Por su parte, la revisión sistemática de López et al. (2025) señala que las características espectrales, de conectividad y complejidad de la señal son los biomarcadores más utilizados, siendo las máquinas de vectores soporte uno de los clasificadores más habituales. Asimismo, Wei et al. (2025) observaron que los métodos basados en aprendizaje profundo pueden capturar patrones complejos del EEG y superar a los clasificadores tradicionales en rendimiento. Sin embargo, estos enfoques suelen requerir mayores recursos computacionales, más datos de entrenamiento y ofrecen una menor interpretabilidad, factores especialmente relevantes en aplicaciones clínicas.

Hernández-Capistrán et al. (2023) destacan la elevada heterogeneidad metodológica existente entre los estudios sobre detección del TDAH mediante aprendizaje automático, especialmente en el preprocesamiento, la selección de características, la validación y las métricas utilizadas. En este contexto, el presente trabajo realiza una comparación homogénea entre Random Forest, XGBoost y SVM-RBF bajo un protocolo común de extracción de características y validación agrupada por sujeto, permitiendo evaluar

objetivamente su comportamiento relativo en la detección del TDAH mediante EEG. Además, se incorpora un análisis de relevancia de características que facilita la interpretación neurofisiológica de los resultados.

3. Metodología

3.1. Dataset

El conjunto de datos utilizado corresponde a un archivo CSV de señales EEG para clasificación binaria TDAH/Control [Nasrabadi et al., 2026]. Participaron 61 niños con TDAH y 60 controles sanos, de entre 7 y 12 años. Los participantes con TDAH habían sido diagnosticados por un psiquiatra y habían recibido tratamiento con Ritalin durante un máximo de seis meses, mientras que los controles no presentaban antecedentes psiquiátricos, epilepsia ni conductas de riesgo. El EEG se registró siguiendo el sistema internacional 10-20 mediante 19 canales (Fz, Cz, Pz, C3, T3, C4, T4, Fp1, Fp2, F3, F4, F7, F8, P3, P4, T5, T6, O1 y O2) y una frecuencia de muestreo de 128 Hz. Los electrodos A1 y A2, situados en los lóbulos de las orejas, se utilizaron como referencia. Dado que la atención visual constituye uno de los déficits característicos del TDAH, el protocolo experimental se basó en una tarea visual en la que los niños debían contar personajes mostrados en imágenes de dibujos animados. El número de personajes se seleccionaba aleatoriamente entre 5 y 16 y las imágenes se presentaban de forma continua, apareciendo una nueva inmediatamente después de cada respuesta. Por ello, la duración del registro EEG dependía del tiempo de respuesta de cada participante.

En el dataset cada fila representa una muestra temporal y cada sujeto se identifica mediante la columna ID. La evaluación se ha planteado por sujeto para evitar que ventanas de un mismo paciente aparezcan simultáneamente en entrenamiento y test. La Tabla 1 muestra los valores más representativos del dataset.

Tabla 1: Datos descriptivos del dataset

Campo	Valor
Columnas	21
Sujetos	121
Variables EEG	19 canales: Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T7, T8, P7, P8, Fz, Cz, Pz.
Variables no EEG	Class: etiqueta de clase binaria, con valores ADHD y Control. ID: identificador de sujeto/paciente usado como grupo en la validación cruzada.
Frecuencia de muestreo	128 Hz

3.2. Preprocesamiento

Primero se realizó una limpieza del dataset, eliminando filas sin ID o Class y normalizando las etiquetas a 0=Control y 1=ADHD. Después se llevó a cabo una segmentación por sujeto empleando ventanas deslizantes, sin mezclar muestras entre sujetos. Se emplearon ventanas deslizantes de 1920 muestras, equivalentes a 15 segundos de señal EEG a una frecuencia de muestreo de 128 Hz, con un solapamiento del 50%, correspondiente a un paso de 960 muestras (7,5

segundos). Se generaron señales auxiliares de entrada con características temporales y espectrales por canal; 458 variables por epoch. Intencionadamente no se aplicaron técnicas de reducción de dimensionalidad (PCA, ICA, selección de variables, etc.), para evitar introducir factores adicionales en la comparativa y centrarse en la capacidad de los algoritmos sin dependencias de otras técnicas.

3.3. Técnicas empleadas

Para la clasificación de las señales se compararon tres técnicas de aprendizaje automático: Random Forest, SVM-RBF y XGboost.

Random Forest es un algoritmo de aprendizaje automático basado en el enfoque de conjuntos que construye múltiples árboles de decisión durante el entrenamiento y combina sus predicciones para obtener un modelo más robusto. Formalmente, dado un conjunto de entrenamiento $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$, el algoritmo genera B subconjuntos mediante muestreo con reemplazo (*bootstrap*), $\mathcal{D}_b \subset \mathcal{D}$. Para cada subconjunto se entrena un árbol de decisión $h_b(x)$. Además, en cada nodo del árbol, en lugar de considerar todas las variables p , se selecciona aleatoriamente un subconjunto de tamaño $m \ll p$, lo que introduce aleatoriedad adicional y reduce la correlación entre árboles. El predictor global del modelo puede expresarse como:

$$\hat{f}(x) = \frac{1}{B} \sum_{b=1}^B h_b(x) \quad (1)$$

en el caso de regresión, o bien mediante una regla de votación mayoritaria en clasificación.

En problemas de clasificación, la predicción final se obtiene como:

$$\hat{y} = \text{mod}\{h_1(x), h_2(x), \dots, h_B(x)\} \quad (2)$$

donde $\text{mode}(\cdot)$ representa la clase más frecuente. Esta estrategia reduce la varianza del modelo en comparación con un único árbol, lo que puede interpretarse como una descomposición del error general donde el ensamblado mejora la estabilidad del estimador:

$$\text{Var}(\hat{f}(x)) \approx \frac{1}{B} \rho \sigma^2 + \frac{1-\rho}{B} \sigma^2 \quad (3)$$

donde ρ es la correlación entre árboles y σ^2 la varianza de los árboles individuales.

XGBoost (*Extreme Gradient Boosting*) es un algoritmo de aprendizaje supervisado basado en técnicas de *boosting*, donde un conjunto de modelos débiles (típicamente árboles de decisión) se entrena secuencialmente para construir un modelo fuerte. A diferencia de métodos como Random Forest, donde los árboles se generan de forma independiente, en XGBoost cada nuevo modelo intenta corregir los errores cometidos por los anteriores. Dado un conjunto de datos $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n$, el modelo predice la salida como una suma de K funciones:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in \mathcal{F} \quad (4)$$

donde $\mathcal{F}$ representa el espacio de árboles de decisión. El entrenamiento se formula como un problema de optimización que minimiza una función de pérdida regularizada:

$$\mathcal{L} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (5)$$

donde $l(\cdot)$ es una función de pérdida (por ejemplo, log-loss en clasificación) y $\Omega(f_k)$ es un término de regularización que penaliza la complejidad de los árboles para evitar sobreajuste.

La optimización se realiza mediante un enfoque de expansión de Taylor de segundo orden sobre la función de pérdida, lo que permite aproximar el problema en cada iteración t :

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \quad (6)$$

donde g_i y h_i son el gradiente y el hessiano, respectivamente. Esta formulación permite un entrenamiento eficiente y preciso, optimizando la estructura del árbol en cada iteración. Además, XGBoost incorpora técnicas avanzadas como regularización L_1 y L_2 , manejo de valores perdidos y paralelización, lo que mejora tanto su rendimiento como su capacidad de generalización.

Las SVM son modelos de aprendizaje supervisado ampliamente utilizados en tareas de clasificación y regresión. Su objetivo principal es encontrar el hiperplano óptimo que maximiza el margen entre clases en un espacio de características. Dado un conjunto de entrenamiento $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n$, con $y_i \in \{-1, +1\}$, la SVM lineal busca resolver el siguiente problema de optimización:

$$\min_{w, b, \xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (7)$$

sujeto a:

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, \xi_i \geq 0 \quad (8)$$

donde w es el vector normal al hiperplano, b el sesgo, ξ_i son variables de holgura que permiten errores de clasificación, y C es un parámetro de regularización que controla el equilibrio entre maximizar el margen y minimizar los errores. Sin embargo, en muchos problemas reales — como el análisis de señales EEG — los datos no son linealmente separables, lo que requiere proyectarlos a un espacio de mayor dimensión.

Para abordar esta limitación, se emplea un kernel que permite operar implícitamente en un espacio de alta dimensión mediante una función núcleo $K(x_i, x_j)$. En el caso del kernel de base radial (RBF o gaussiano), uno de los más utilizados, se define como:

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2} \quad (9)$$

donde $\gamma > 0$ es un parámetro que controla la influencia de cada instancia de entrenamiento. La función de decisión de la SVM con kernel RBF puede escribirse como:

$$f(x) = \text{sign} \left(\sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \right) \quad (10)$$

donde α_i son los coeficientes obtenidos al resolver el problema dual. Este enfoque permite modelar fronteras de decisión altamente no lineales, adaptándose a la complejidad de los datos.

3.4. Selección de hiperparámetros

La Tabla 2 recoge los hiperparámetros empleados en cada una de las técnicas. Los hiperparámetros utilizados fueron escogidos a partir de experimentos preliminares realizados sobre el conjunto de entrenamiento.

Tabla 2: Configuración de las técnicas de aprendizaje automático

Técnica	Parámetros de configuración
Random Forest	n_estimators=100; max_depth=10; criterion=entropy; max_features=sqrt; bootstrap=True; class_weight=balanced; min_samples_leaf=1; n_jobs=-1; random_state=42.
SVM RBF	StandardScaler; kernel=rbf; C=10.0; gamma=scale; probability=True; class_weight=balanced; random_state=42.
XGBoost	n_estimators=200; max_depth=4; learning_rate=0.05; subsample=0.8; colsample_bytree=0.8; objective=binary:logistic; eval_metric=logloss; tree_method=hist; random_state=42.

4. Resultados

La Tabla 3 muestra el rendimiento de los clasificadores, comparando las métricas estándar del rendimiento de clasificación: Exactitud, Precisión, Sensibilidad y F1.

Tabla 3: Comparación del rendimiento de los clasificadores

Método	Exactitud	Precisión	Sensibilidad	F1
Random Forest	0.7986 ± 0.0582	0.7927 ± 0.0572	0.7886 ± 0.0582	0.7872 ± 0.0578
XGBoost	0.7737 ± 0.0579	0.7781 ± 0.0548	0.7737 ± 0.0579	0.7723 ± 0.0567
SVM RBF	0.7537 ± 0.0423	0.7573 ± 0.0425	0.7537 ± 0.0423	0.7520 ± 0.0448

Los resultados obtenidos evidencian diferencias consistentes en el rendimiento de los modelos evaluados. En particular, el algoritmo Random Forest alcanza el mejor desempeño global, registrando una exactitud de 0.7986 ± 0.0582 y un F1-score de 0.7872 ± 0.0578 , superando de forma sistemática a los demás clasificadores en todas las métricas consideradas. Por su parte, XGBoost presenta un rendimiento ligeramente inferior, con una exactitud de 0.7737 ± 0.0579 y un F1-score de 0.7723 ± 0.0567 , lo que indica que, si bien constituye una alternativa competitiva, no logra igualar la capacidad predictiva del Random Forest en este contexto. En contraste, el modelo SVM con kernel RBF obtiene los valores más bajos, con una exactitud de 0.7537 ± 0.0423 y un F1-score

de 0.7520 ± 0.0448 , evidenciando una menor capacidad de discriminación.

Un aspecto destacable es la consistencia entre métricas, ya que los valores de precisión, sensibilidad y exactitud son muy similares en los tres modelos. Este comportamiento sugiere la ausencia de sesgos relevantes hacia falsos positivos o negativos, así como un adecuado equilibrio entre clases. Además, SVM presenta las menores desviaciones estándar, indicando una mayor estabilidad entre particiones de validación cruzada. Sin embargo, esta robustez no compensa su menor rendimiento medio. En conjunto, los resultados muestran que Random Forest ofrece el mejor equilibrio entre capacidad predictiva y consistencia, situándose como la opción más adecuada para la tarea planteada. Aunque XGBoost presenta un rendimiento competitivo, no supera a Random Forest, mientras que SVM, pese a su menor variabilidad, muestra una precisión global más limitada.

Adicionalmente a las métricas cuantitativas, también se han estudiado las curvas ROC (Figura 1). Esta figura recoge las curvas ROC correspondientes a las tres técnicas. XGboost se representa en naranja, RBF-SVM en verde y Random Fores en azul.

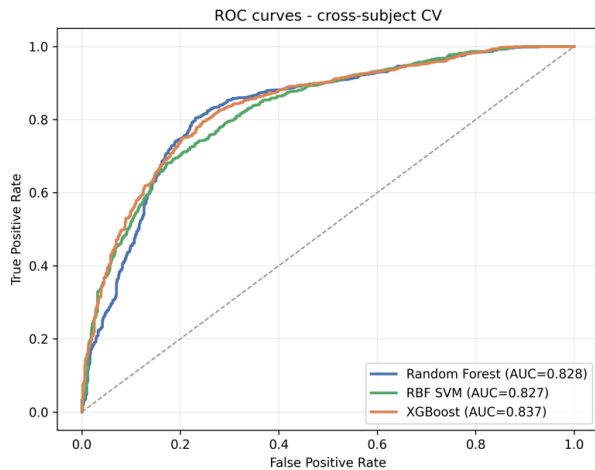


Figura 1: Comparativa de curvas ROC

En el modelo Random Forest, se observa una curva ROC claramente desplazada hacia la esquina superior izquierda, lo que indica una elevada tasa de verdaderos positivos (TPR) manteniendo bajas tasas de falsos positivos (FPR). Este comportamiento es coherente con los altos valores de exactitud y F1-score previamente reportados, confirmando la sólida capacidad discriminativa del modelo.

En el modelo XGBoost la curva ROC obtenida presenta un perfil similar al de Random Forest, aunque ligeramente más alejado de la región óptima. Este resultado indica que, si bien XGBoost mantiene un buen equilibrio entre sensibilidad y especificidad, su capacidad de discriminación es marginalmente inferior, en línea con las diferencias observadas en las métricas de rendimiento.

El modelo SVM con kernel RBF, exhibe una curva ROC más próxima a la diagonal. Este patrón sugiere una menor capacidad para distinguir entre clases, lo que se traduce en un rendimiento inferior en términos de exactitud y F1-score. No obstante, la forma más suave de la curva podría reflejar una mayor estabilidad del modelo frente a variaciones en los datos.

En conjunto, las curvas ROC corroboran los resultados numéricos de la Tabla 3, confirmando que Random Forest es el modelo con mayor capacidad discriminativa, seguido de XGBoost, mientras que SVM presenta un rendimiento más limitado. La consistencia entre ambos tipos de evaluación (métricas y curvas ROC) refuerza la validez de los resultados y sugiere que las diferencias observadas son robustas a la elección del umbral de clasificación.

Las figuras 2 y 3 muestran el análisis por orden de importancia de las características de entrada.

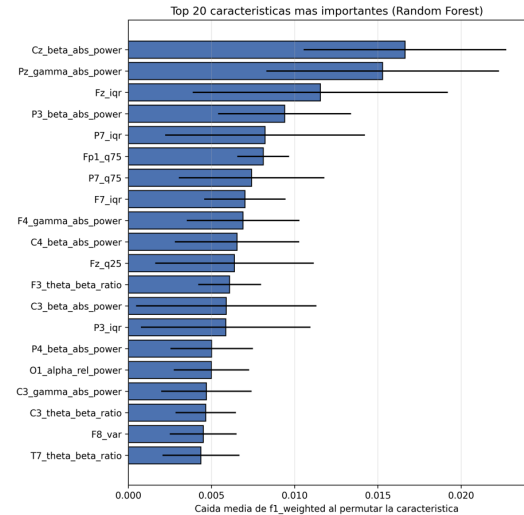


Figura 2: Características ordenadas por orden de importancia para Random Forest

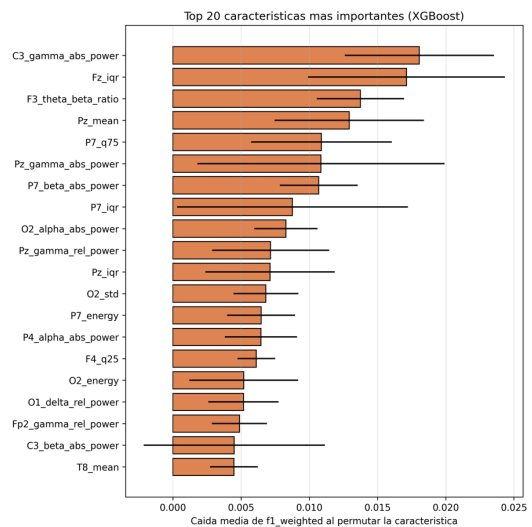


Figura 3: Características ordenadas por orden de importancia para XGBoost

En el modelo Random Forest, la característica más relevante es *Cz_beta_abs_power*, seguida de *Pz_gamma_abs_power*. Estos resultados indican que el modelo se apoya principalmente en la potencia absoluta de las bandas beta y gamma. La elevada importancia de *Cz_beta_abs_power* es especialmente significativa, ya que la actividad beta se relaciona con procesos atencionales y de control cognitivo, frecuentemente alterados en el TDAH. Del mismo modo, la relevancia de *Pz_gamma_abs_power* sugiere que la actividad de alta frecuencia en regiones parietales aporta información útil para discriminar entre sujetos con y sin TDAH.

En el caso de XGBoost, la característica con mayor impacto es *C3_gamma_abs_power*, seguida por *Fz_iqr*, *F3_theta_beta_ratio* y *Pz_mean*. A diferencia de Random Forest, donde predominan las bandas beta, XGBoost concede una importancia muy elevada a variables relacionadas con la banda gamma y con la razón theta/beta, uno de los biomarcadores más estudiados en la literatura sobre TDAH. La presencia de *F3_theta_beta_ratio* entre las variables más importantes indica que el modelo es capaz de extraer información relevante a partir de relaciones entre bandas frecuenciales, más allá de las potencias absolutas individuales.

Entre las coincidencias más relevantes destacan *Fz_iqr*, *P7_iqr*, *P7_q75*, *F3_theta_beta_ratio*, *C3_beta_abs_power*, *C3_gamma_abs_power* y *Pz_gamma_abs_power*, lo que sugiere que estas variables contienen información consistente para la detección del TDAH independientemente del algoritmo utilizado. La recurrencia de características asociadas a los canales frontales y parietales respalda la relevancia funcional de las redes frontoparietales en este trastorno. La presencia reiterada de medidas gamma, beta, theta/beta e índices de variabilidad en ambos modelos refuerza la hipótesis de que la combinación de información espectral y estadística constituye una estrategia eficaz para la identificación automática del TDAH a partir de señales EEG.

5. Conclusiones

Los resultados obtenidos confirman el potencial de las señales EEG combinadas con técnicas de aprendizaje automático para la detección del TDAH. Random Forest presentó el mejor desempeño global, alcanzando una exactitud de 0.7886 ± 0.0582 y un F1-score de 0.7872 ± 0.0578 , superando consistentemente a XGBoost y SVM-RBF. XGBoost mostró resultados competitivos (exactitud de 0.7737 ± 0.0579), mientras que SVM-RBF obtuvo el menor rendimiento (exactitud de 0.7537 ± 0.0423), reflejando una menor capacidad discriminativa. Sin embargo, los resultados obtenidos deben interpretarse como una comparación exploratoria entre modelos, no se emplearon métodos sistemáticos de optimización de hiperparámetros. Los resultados muestran potencial para apoyar el diagnóstico, sin embargo, las métricas obtenidas son insuficientes para aplicaciones clínicas autónomas.

Como líneas futuras cabe destacar la evaluación comparada de otro tipo de técnicas de aprendizaje automático, como las basadas en aprendizaje profundo, la evaluación de técnicas de reducción de la dimensionalidad, la optimización sistemática de hiperparámetros, y el empleo de métodos de agregación de aprendizaje como los ensembles.

Referencias

Adamou, M., Fullen, T., & Jones, S. L. (2020). EEG for diagnosis of adult ADHD: A systematic review with narrative analysis. *Frontiers in Psychiatry*, 11, 871.

Ahire, N., Awale, R. N., & Wagh, A. (2025). Electroencephalogram (EEG) based prediction of attention deficit hyperactivity disorder (ADHD) using machine learning. *Applied Neuropsychology: Adult*, 32(4), 966-977.

American Psychiatric Association. (2022). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5-TR*. American Psychiatric Publishing

Boxum, M., Voetterl, H., van Dijk, H., Gordon, E., DeBeus, R., Arnold, L. E., & Arns, M. (2025). Challenging the diagnostic value of theta/beta ratio: Insights from an EEG subtyping meta-analytical approach in ADHD. *Applied Psychophysiology and Biofeedback*, 50(4), 655-666.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

Garcia-Ausin, R., & Sierra-Garcia, J. E. (2024). Holistic intervention proposal for students with ADHD. In *EDULEARN24 Proceedings* (pp. 9118-9124). IATED.

Garcia-Ausin, R., Pizarro, J. P., de la Fuente, R., & Sierra-Garcia, J. E. (2026). Stigma associated with ADHD in educational settings: an examination of effective practices and recommendations for mitigation. *INTED2026 Proceedings*, 0517.

Hernández-Capistran, J., Sánchez-Morales, L. N., Alor-Hernández, G., Bustos-López, M., & Sánchez-Cervantes, J. L. (2023). Machine and deep learning algorithms for ADHD detection: A review. *Innovations in machine and deep learning: Case studies and applications*, 163-191.

Kember, J., Stepien, L., Panda, E., & Tekok-Kilic, A. (2023). Resting-state EEG dynamics help explain differences in response control in ADHD. *PLOS ONE*, 18(10), e0277382.

Koutsoklenis, A., & Honkasilta, J. (2023). ADHD in the DSM-5-TR: What has changed and what has not. *Frontiers in psychiatry*, 13, 1064141.

Lenartowicz, A., & Loo, S. K. (2014). Use of EEG to diagnose ADHD. *Current psychiatry reports*, 16(11), 498.

Loo, S. K., Lenartowicz, A., & Makeig, S. (2016). Research review: Use of EEG biomarkers in child psychiatry research—current state and future directions. *Journal of Child Psychology and Psychiatry*, 57(1), 4-17

López, C. Q., Vera, V. D. G., & Quintero, M. J. R. (2025). Diagnosis of ADHD in children with EEG and machine learning: Systematic review and meta-analysis. *Clinical and Health*, 36(2), 109-121.

Mao, Y., Qi, X., He, L., Wang, S., Wang, Z., & Wang, F. (2025). Advanced machine learning techniques reveal multidimensional EEG abnormalities in children with ADHD: a framework for automatic diagnosis. *Frontiers in Psychiatry*, 16, 1475936.

Nasrabadi, A. M., Allahverdy, A., Samavati, M., Mohammadi, M.R. (2026). EEG dataset for ADHD. Kaggle. <https://www.kaggle.com/datasets/danizo/eeg-dataset-for-adhd>

Slater, J., Joobar, R., Koborsy, B. L., Mitchell, S., Sahlas, E., & Palmer, C. (2022). Can electroencephalography (EEG) identify ADHD subtypes? A systematic review. *Neuroscience & Biobehavioral Reviews*, 139, 104752.

Snyder, S. M., & Hall, J. R. (2006). A meta-analysis of quantitative EEG power associated with attention-deficit hyperactivity disorder. *Journal of Clinical Neurophysiology*, 23(5), 441-456.

van Dijk, H., deBeus, R., Kerson, C., et al. (2020). Different spectral analysis methods for the theta/beta ratio do not distinguish ADHD from controls. *Applied Psychophysiology and Biofeedback*, 45(3), 165-173.

Wei, L. L. Y., Ibrahim, A. A. A., & Alfred, R. (2025, August). Deep Learning Approach to EEG-based Attention Deficit Hyperactivity Disorder (ADHD) Detection: An Empirical Comparison of Ensemble Classifiers. In *2025 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAET)* (pp. 49-54). IEEE.