

**Aplicaciones de Inteligencia Artificial en la Evaluación Neuropsicológica del TDAH  
en Niñas y Niños de 6 a 12 Años: Una Revisión de Alcance**

Ana Gabriela Arrieta Vergara

Oscar Gabriel Artunduaga Arias

Luz Dary Camargo Chaparro

Pedro Germán Rodríguez Castellanos

Especialización en Neuropsicología Infantil

Facultad de Psicología, Ciencias Sociales y de la Educación

Fundación Universitaria Sanitas

Mg. Jairo Andrés Balcázar Acosta

Bogotá, D.C., Colombia

2026

## Resumen

El Trastorno por Déficit de Atención e Hiperactividad (TDAH) es uno de los diagnósticos más frecuentes en neuropsicología infantil, con una prevalencia global estimada entre el 7.6% y el 8.0% (Kazda et al., 2023). La evaluación neuropsicológica tradicional enfrenta limitaciones relacionadas con tiempos prolongados de espera, altos costos, variabilidad en la interpretación clínica y disparidades de género que favorecen el infra diagnóstico de niñas con TDAH (Aguirre Salazar, 2024; Atem et al., 2024). En este contexto, la inteligencia artificial (IA) ha emergido como una herramienta con potencial para apoyar los procesos de evaluación mediante modelos automatizados capaces de analizar patrones conductuales, cognitivos y neurofisiológicos. La presente investigación es una revisión de alcance de enfoque cualitativo-descriptivo cuyo propósito es mapear la información científica publicada en los últimos 10 años sobre aplicaciones de IA en la evaluación neuropsicológica del TDAH en niñas y niños de 6 a 12 años. La revisión sigue el marco metodológico de Arksey y O'Malley (2005), actualizado por el Instituto Joanna Briggs marco PCC (Peters et al., 2020), y se reporta conforme a la guía PRISMA-ScR (Tricco et al., 2018). La búsqueda se realizó en cinco fuentes de información: PubMed / EBSCOhost / ProQuest / SciELO/ y BVS salud; se llevó a cabo la selección y filtrado de los artículos utilizando la herramienta Rayyan (Ouzzani et al., 2016), resueltas por consenso entre cuatro revisores independientes, resultando en 14 artículos incluidos. Los hallazgos caracterizaron las aplicaciones de IA empleadas, el desempeño diagnóstico reportado, las características de las muestras, considerando la representación geográfica y la distribución por género. Los resultados contribuyen a identificar vacíos de evidencia actual, especialmente en

relación con la representación latinoamericana e hispanohablante y la equidad de género en el diseño de los estudios.

***Palabras clave:*** inteligencia artificial; TDAH; evaluación neuropsicológica; niños; niñas; equidad de género; revisión de alcance.

## Abstract

Attention-Deficit/Hyperactivity Disorder (ADHD) is one of the most frequent diagnoses in child neuropsychology, with an estimated global prevalence between 7.6% and 8.0% (Kazda et al., 2023). Traditional neuropsychological assessment faces limitations related to long waiting times, high costs, variability in clinical interpretation, and gender disparities that contribute to the underdiagnosis of girls with ADHD (Aguirre Salazar, 2024; Atem et al., 2024). In this context, artificial intelligence (AI) has emerged as a tool with the potential to support assessment processes through automated models capable of analyzing behavioral, cognitive, and neurophysiological patterns. This research is a qualitative-descriptive review whose purpose is to map the scientific information published in the last 10 years on AI applications in the neuropsychological assessment of ADHD in children aged 6 to 12 years. This review follows the methodological framework of Arksey and O'Malley (2005), updated by the Joanna Briggs Institute's PCC framework (Peters et al., 2020), and is reported according to the PRISMA-ScR guidelines (Tricco et al., 2018). The search was conducted in five databases: PubMed, EBSCOhost, ProQuest, SciELO, and BVS Health. Article selection and filtering were performed using the Rayyan tool (Ouzzani et al., 2016), and results were obtained by consensus among four independent reviewers, resulting in 14 included articles. The findings characterize the AI applications used, the reported diagnostic performance, and the characteristics of the samples, considering geographic representation and gender distribution. The results contribute to identifying gaps in current evidence, especially regarding the representation of Latin American and Spanish-speaking populations and gender equity in study design.

**Keywords:** *artificial intelligence; ADHD; neuropsychological assessment; children; gender equity; scope review.*

## Tabla de Contenidos

Aplicaciones de Inteligencia Artificial en la Evaluación Neuropsicológica del TDAH en Niñas y Niños de 6 a 12 Años: Una Revisión de Alcance .....	1
Resumen.....	2
Abstract .....	4
Tabla de Contenidos .....	6
Lista de Tablas.....	8
Lista de Figuras .....	9
Introducción.....	10
Planteamiento del Problema .....	12
Prevalencia Clínica y Relevancia del TDAH .....	12
Limitaciones de la Evaluación Neuropsicológica Tradicional .....	12
Disparidades Diagnósticas y Sesgos de Género.....	13
La inteligencia Artificial en la Evaluación Neuropsicológica del TDAH.....	14
El Vacío de Conocimiento Existente.....	14
Pregunta de Investigación.....	15
Justificación .....	16
Objetivos .....	17
General.....	17
Específicos .....	17
Marco Teórico .....	18
Inteligencia artificial en salud: qué es y como se organiza.....	21

Autonomía de la IA.....	24
Metodología.....	26
Diseño del Estudio.....	26
Definición de Componentes PCC.....	26
Criterios de Exclusión Códigos.....	27
Fuentes y Estrategia de Búsqueda.....	27
Proceso de Selección.....	28
Extracción de Datos y Síntesis.....	29
Resultados.....	30
Caracterización de las Aplicaciones de IA.....	32
Desempeño Diagnóstico y Características de los Modelos.....	33
Vacíos de Evidencia Actual.....	39
Discusión.....	43
La Distancia entre el Laboratorio y la Clínica.....	45
El Género como Cuestión Científica.....	45
América Latina y la Interpretabilidad.....	46
Conclusiones.....	47
Recomendaciones.....	49
Referencias.....	50
Anexos.....	57

## Lista de Tablas

Tabla 1 Aplicaciones *de IA validadas para la evaluación neuropsicológica del TDAH en población pediátrica*

Tabla 2 Clasificación *de algoritmos y modalidades de datos desde el enfoque neuropsicológico*

Tabla 3 Métricas *de eficacia y desempeño diagnóstico de los modelos de IA*

Tabla 4 Distribución *geográfica de los 14 artículos incluidos*

Tabla 5 Síntesis *de métricas de rendimiento diagnóstico reportadas*

Tabla 6 Características *metodológicas avanzadas de los 14 artículos incluidos*

Tabla 7 Identificación *de vacíos de evidencia y limitaciones reportadas*

## Lista de Figuras

Figura 1 Diagrama *de flujo del proceso de identificación, cribado y selección de estudios*  
(PRISMA-ScR)

## Introducción

Cuando una familia llega al consultorio preocupada porque su hijo o hija no logra concentrarse en clase, pierde cosas constantemente o parece tener un motorcito, no siempre encuentra una respuesta rápida. La evaluación neuropsicológica del Trastorno por Déficit de Atención e Hiperactividad (TDAH) puede ser larga, costosa y, en muchos contextos inequitativa, pues no todos los niños llegan con igualdad de condiciones al especialista (Atem et al., 2024; Slaton et al., 2024). Esta realidad, bien documentada en la literatura internacional, motivó la pregunta que da origen a este trabajo.

La inteligencia artificial (IA) aparece en ese escenario como una posibilidad gracias a los algoritmos que analizan señales de electroencefalografía hasta plataformas que procesan respuestas de cuestionarios, en tal sentido, en los últimos años se ha visto proliferar herramientas que prometen hacer el diagnóstico más objetivo, más accesible y menos dependiente del factor humano. No obstante, esta proliferación también ha generado dificultades relacionadas con la dispersión de la evidencia, la heterogeneidad de los criterios de validación desconociendo qué funciona, en qué contextos y para quien (Zhao et al., 2025; López-Cortés et al., 2022).

A esto se suma una pregunta clave en la práctica neuropsicológica infantil, ¿estas herramientas han sido pensadas también para las niñas? La búsqueda inicial muestra que las niñas con TDAH siguen siendo diagnosticadas más tarde que los niños, en parte porque su presentación clínica es diferente y porque los instrumentos no siempre están diseñados y validados para detectarlo en ellas y si la IA copia esos sesgos, lejos de ayudar podría agravar el problema (Faraone et al., 2021).

Por lo anteriormente descrito, la presente investigación propone una revisión de alcance que permita mapear de forma metodológica la información disponible en los últimos 10 años sobre aplicaciones de IA para la evaluación neuropsicológica del TDAH en niñas y niños de 6 a 12 años, siguiendo los lineamientos de Arksey y O'Malley (2005), actualizados por el Instituto Joanna Briggs (Peters et al., 2020), y reportes alineados a los estándares PRISMA-ScR (Tricco et al., 2018).

## **Planteamiento del Problema**

### **Prevalencia Clínica y Relevancia del TDAH**

El Trastorno por Déficit de Atención e Hiperactividad (TDAH) es uno de los trastornos del neurodesarrollo más frecuentes en la infancia. Su sintomatología y patrones persistentes de inatención, hiperactividad e impulsividad interfieren con el funcionamiento académico, social y emocional de los niños y niñas que lo presentan. Algunos estudios recientes estiman una prevalencia global de entre el 7.6% y el 8.0% en población infantil, con tasas diagnósticas consistentemente mayores en varones que en mujeres (Kazda et al., 2023; Mohammadi et al., 2023). En países sudamericanos, esa cifra puede superar la media mundial, alcanzando entre el 15% y el 17% en algunas regiones más subdesarrolladas (Rusca-Jordán & Cortez-Vergara, 2020).

Aunado a lo anterior, cuando el TDAH no es identificado a tiempo, sus consecuencias se extienden a la edad adulta donde aproximadamente dos tercios de los niños diagnosticados presentan síntomas persistentes durante toda su vida, afectando el rendimiento académico, las relaciones interpersonales, la regulación emocional y la adaptación social (Trujillo González et al., 2024). En tal sentido, el diagnóstico temprano y preciso no es solo un procedimiento clínico, es una forma de intervención en sí misma.

### **Limitaciones de la Evaluación Neuropsicológica Tradicional**

A pesar de su alta demanda, el proceso de evaluación neuropsicológica habitual del TDAH sigue enfrentando obstáculos reales, combinando entrevista clínica, observación conductual, cuestionario para padres y docentes, y pruebas estandarizadas de atención, memoria

y funciones ejecutivas. Ese enfoque multimodal presenta un valor clínico real, pero también tiene límites bien documentados como la subjetividad en la interpretación, variabilidad entre evaluadores y diferencias en los reportes según el contexto de observación.

A estos límites metodológicos se suman barreras estructurales como la brecha real entre la demanda de servicios especializados y la disponibilidad de profesionales capacitados, especialmente en países de ingresos medios y bajos; los tiempos de espera pueden prolongarse durante meses, retrasando las intervenciones que el niño o la niña necesita con urgencia; las barreras económicas y la desigualdad en el acceso a servicios especializados, aspectos que agravan más esa situación (Atem et al., 2024).

así mismo, las herramientas computarizadas tradicionales como el *Continuous Performance Test* (CPT) y el *Test of Variables of Attention* (TOVA) han aportado evaluaciones más objetivas del desempeño atencional; no obstante, su uso sigue siendo limitado por costos, disponibilidad tecnológica y necesidad de personal entrenado para un buen diagnóstico (Parsons et al., 2020).

### **Disparidades Diagnósticas y Sesgos de Género**

Una de las dificultades más persistentes de un diagnóstico integral del TDAH es la brecha de género. Las niñas tienden a ser diagnosticadas con menor frecuencia y en etapas más tardías del desarrollo, particularmente cuando presentan perfiles predominantemente inatentos (Aguirre Salazar, 2024). Esta brecha no es biológica es diagnóstica, debido a que mientras que los niños suelen manifestar conductas hiperactivas más visibles en el contexto escolar, muchas niñas presentan síntomas silenciosos e interiorizados que no generan las mismas alarmas en padres, docentes, ni en los instrumentos de evaluación. La consecuencia de estas pruebas no validadas y

estandarizadas para la población femenina se observa cuando la niña llega a la adolescencia sin haber recibido un diagnóstico que habría cambiado su trayectoria escolar y emocional. Esa realidad hace visible la necesidad de desarrollar herramientas evaluativas más objetivas, más sensibles a distintas formas de expresión del trastorno y menos dependientes de la interpretación subjetiva del observador.

### **La inteligencia Artificial en la Evaluación Neuropsicológica del TDAH**

Teniendo en cuenta el escenario anterior, el desarrollo de sistemas basados en inteligencia artificial ha generado un interés creciente en investigaciones dentro del campo de la salud mental y la neuropsicología infantil para explorar el uso de algoritmos de *machine learning* y *deep learning*, aplicados al análisis de datos neuropsicológicos, neurofisiológicos y conductuales para apoyar la identificación del TDAH (Zhao et al., 2025). Entre las herramientas más utilizadas se encuentran la electroencefalografía (EEG), la resonancia magnética estructural, el seguimiento ocular, la realidad virtual, el análisis de movimiento y las pruebas neuropsicológicas computarizadas, reportando desempeños superiores al 80% en contextos muy controlados (Chen et al., 2023). A esto se suman interrogantes sobre la transparencia algorítmica, la validez transcultural y la ausencia prácticamente total de estudios en contextos latinoamericanos.

### **El Vacío de Conocimiento Existente**

A pesar de que en la última década se ha experimentado el crecimiento de la producción científica sobre IA aplicada al TDAH, esa evidencia sigue siendo fragmentada y metodológicamente heterogénea. La búsqueda realizada inicialmente permitió identificar que en

la literatura existente no se han desarrollado investigaciones con criterios específicos de una revisión de alcance que sintetice sistemáticamente las aplicaciones de IA en la evaluación neuropsicológica del TDAH en niñas y niños de 6 a 12 años, considerando algoritmos empleados, modalidades de datos, desempeño diagnóstico, distribución por género y representación geográfica.

La búsqueda permitió evidenciar que existen revisiones previas sobre IA y TDAH, pero ninguna en español y ninguna con la combinación de criterios como población pediátrica estricta de 6 a 12 años, enfoque en la evaluación neuropsicológica no en el tratamiento, muestra que evidencie equidad de género, y la existencia de representación latinoamericana. Estas especificaciones justifican la realización de la presente revisión de alcance, motivo por el cual de este contexto surge la siguiente pregunta de investigación:

### **Pregunta de Investigación**

¿Cuál es la información disponible en estudios publicados entre 2016 y 2025 sobre el uso de aplicaciones basadas en inteligencia artificial en la evaluación neuropsicológica del TDAH en niños y niñas entre 6 y 12 años?

## Justificación

El interés por aplicar la inteligencia artificial al diagnóstico del TDAH no surgió de repente. Entre 2016 y 2019, los primeros trabajos en este campo eran exploratorios; grupos de investigación probaron algoritmos de clasificación sobre bases de datos de señales de EEG o cuestionarios estandarizados, con muestras pequeñas y sin validación clínica real (Dey et al., 2022). Entre 2020 y 2022 el campo maduró: aparecieron las primeras validaciones con muestras más amplias, los primeros sistemas en uso clínico y las primeras revisiones que buscaron sintetizar los resultados acumulados (Romero-Ayuso et al., 2021; Quintero López et al., 2025). Desde 2023, el debate incorporó nuevas dimensiones: los sesgos de género en los algoritmos, las implicaciones éticas del uso de datos de menores y la pregunta sobre la aplicabilidad fuera de los contextos donde fueron desarrolladas las herramientas (Zhao et al., 2025).

Lo que ese recorrido deja en evidencia es que, a pesar del crecimiento sostenido de la producción científica, los vacíos son importantes. América Latina prácticamente no aparece en ese mapa. Las niñas están subrepresentadas en las muestras de los estudios. Y no existe, hasta donde se ha podido rastrear, una revisión de alcance en español que organice sistemáticamente lo que se sabe y lo que falta por investigar (López-Cortés et al., 2022; Aguirre Salazar, 2024). Esta investigación nace precisamente de esa ausencia, y sus resultados están pensados para ser útiles a clínicos, investigadores, formadores de posgrado y quienes diseñan políticas de salud mental infantil en contextos latinoamericanos.

## Objetivos

### General

Mapear la información científica publicada entre 2016 y 2025 sobre aplicaciones de inteligencia artificial utilizadas en la evaluación neuropsicológica del Trastorno por Déficit de Atención e Hiperactividad (TDAH) en niñas y niños de 6 a 12 años, considerando los algoritmos empleados, los tipos de datos utilizados, el desempeño diagnóstico reportado, las características de las muestras estudiadas y los vacíos de conocimiento persistentes en la literatura.

### Específicos

Caracterizar las aplicaciones de inteligencia artificial utilizadas en la evaluación neuropsicológica del TDAH en niñas y niños de 6 a 12 años, según el tipo de arquitectura, los algoritmos específicos desarrollados y las modalidades de datos de entrada utilizados.

Describir el desempeño diagnóstico reportado por las aplicaciones de inteligencia artificial utilizadas en la evaluación neuropsicológica del TDAH infantil, considerando métricas como accuracy, sensibilidad, especificidad, AUC y otras medidas de rendimiento reportadas en los estudios.

Identificar los vacíos de evidencia existentes en la investigación sobre inteligencia artificial aplicada a la evaluación neuropsicológica del TDAH infantil, considerando la representación geográfica, la distribución por género y las características metodológicas de los estudios incluidos.

## Marco Teórico

El Trastorno por Déficit de Atención e Hiperactividad (TDAH) es un trastorno del neurodesarrollo caracterizado por un patrón persistente de inatención, hiperactividad e impulsividad que interfiere con el funcionamiento académico, social y emocional de niñas y niños. Desde la perspectiva neuropsicológica, no se trata solo de un problema de “falta de atención”, sino de una alteración más amplia en procesos de autorregulación, control inhibitorio, memoria de trabajo y funciones ejecutivas, con impacto variable según la etapa del desarrollo y el contexto en que se manifiesta.

Durante la infancia, especialmente entre los 6 y los 12 años, el diagnóstico del TDAH requiere una valoración cuidadosa porque muchas conductas propias del desarrollo pueden parecerse a los síntomas del trastorno. Por eso, la evaluación neuropsicológica sigue siendo una herramienta central: integra entrevista clínica, observación conductual, reportes de padres y docentes, y pruebas estandarizadas de atención y funciones ejecutivas. No obstante, este proceso presenta limitaciones importantes, entre ellas la subjetividad en la interpretación, la variabilidad entre evaluadores, los tiempos prolongados de atención especializada y las desigualdades de acceso a servicios de salud mental infantil.

A estas dificultades se suma un problema persistente de equidad diagnóstica. La literatura muestra que las niñas con TDAH suelen ser detectadas más tarde que los niños, en parte porque su presentación clínica tiende a ser menos visible, especialmente cuando predominan los síntomas inatentos. Esto implica que los instrumentos y los procesos de evaluación no siempre capturan de forma adecuada las distintas formas de expresión del trastorno, lo que puede retrasar el diagnóstico y limitar el acceso a intervenciones oportunas. (Stephen P. Hinshaw et al., 2022).

En este contexto, la inteligencia artificial ha comenzado a incorporarse como una herramienta de apoyo dentro del campo de la neuropsicología infantil. La IA permite analizar grandes volúmenes de datos, identificar patrones complejos y apoyar procesos de clasificación o predicción con mayor rapidez y consistencia que los métodos tradicionales. En la evaluación del TDAH, su uso se ha concentrado principalmente en cuatro líneas: machine learning, deep learning, visión por computadora y procesamiento de lenguaje natural. (Artificial Intelligence; Attention Deficit Hyperactivity Disorder) (Stephen P. Hinshaw et al., 2022).

Las aplicaciones revisadas en la literatura incluyen el análisis de señales electroencefalográficas, registros de movimiento, seguimiento ocular, tareas computarizadas, plataformas de realidad virtual y, en menor medida, el procesamiento automático de texto clínico. En varios estudios, estos modelos han mostrado desempeños diagnósticos prometedores, con precisiones elevadas y buenos niveles de sensibilidad y especificidad. Sin embargo, esos resultados deben interpretarse con mucha cautela, porque muchos provienen de muestras pequeñas, diseños experimentales controlados y validaciones internas que no siempre reflejan la complejidad del contexto clínico real. (Attention Deficit Hyperactivity Disorder; Machine Learning) (Rashid A. Karam et al., 2020).

La revisión de la literatura también muestra que la investigación en este campo ha avanzado de manera desigual. En una primera etapa predominaron estudios exploratorios con algoritmos clásicos como SVM y Random Forest sobre datos de EEG o escalas conductuales. Más adelante, el campo incorporó arquitecturas de Deep Learning, enfoques multimodales y modelos más sofisticados, con un interés creciente por la explicabilidad y la equidad algorítmica. A pesar de ello, persisten limitaciones relevantes: predominio de muestras masculinas, escasa representación latinoamericana, falta de desagregación por sexo y limitada validación

transcultural. (Deep Learning; Machine Learning) (Stephen P. Hinshaw et al., 2022; Rashid A. Karam et al., 2020).

En conjunto, la información encontrada en la literatura sugiere que la inteligencia artificial tiene un potencial real para apoyar la evaluación neuropsicológica del TDAH, pero todavía no ofrece una solución definitiva ni homogénea. Su valor actual reside más en complementar el juicio clínico que en reemplazarlo. Por eso, resulta necesario mapear de manera sistemática qué aplicaciones existen, qué algoritmos emplean, qué tipo de datos analizan, qué rendimiento reportan y cuáles son los vacíos que todavía impiden su implementación clínica amplia y segura. (Artificial Intelligence; Attention Deficit Hyperactivity Disorder) (Philip Asherson et al., 2023; Rashid A. Karam et al., 2020).

Desde esta base teórica y empírica surge la necesidad de realizar una revisión de alcance que permita organizar la información disponible sobre aplicaciones de inteligencia artificial en la evaluación neuropsicológica del TDAH en niñas y niños de 6 a 12 años, con especial atención a los algoritmos utilizados, los tipos de datos, el desempeño diagnóstico y los vacíos de conocimiento existentes.

La evaluación neuropsicológica del TDAH combina entrevista clínica, observación conductual, reportes de padres y docentes, y pruebas estandarizadas de atención y funciones ejecutivas. Ese enfoque es valioso porque permite mirar al niño en distintos contextos, pero también tiene límites: depende de la interpretación del evaluador, puede variar según quién responde y no siempre logra capturar la complejidad del funcionamiento atencional en la vida diaria.

El período entre los 6 y los 12 años es crítico para la evaluación. Durante esta etapa, las funciones ejecutivas experimentan un avance significativo gracias a la maduración de la corteza

prefrontal. La evaluación neuropsicológica permite identificar y cuantificar las diferencias individuales, facilitando el diagnóstico diferencial (Londoño-Ocampo & Tezón, 2019). La práctica recomendada combina entrevista clínica, observación conductual, escalas estandarizadas como las de Conners o la Vanderbilt, y evaluación formal de funciones ejecutivas, atención e inhibición. Las evaluaciones computarizadas como el TOVA y el CPT ofrecen datos normativos con puntajes de corte, aunque su uso sigue siendo limitado por costo y disponibilidad (Parsons et al., 2020).

En esta etapa del desarrollo, entre los 6 y los 12 años, las funciones ejecutivas todavía están madurando, por lo que diferenciar entre variaciones esperables del desarrollo y signos clínicamente significativos de TDAH resulta especialmente importante. En ese sentido, las pruebas computarizadas tradicionales aportan mayor objetividad, pero siguen siendo insuficientes por sí solas si no se acompañan de juicio clínico y análisis contextual.

### **Inteligencia artificial en salud: qué es y como se organiza**

La inteligencia artificial puede entenderse como un conjunto de técnicas que permiten a los sistemas computacionales reconocer patrones, aprender de datos y apoyar decisiones complejas. En salud, esto ha abierto la posibilidad de analizar grandes volúmenes de información clínica, neurofisiológica y conductual con rapidez y consistencia.

La IA que hoy existe es la única presente en los estudios sobre TDAH es la llamada IA estrecha o débil: sistemas diseñados para una tarea específica que no generalizan a dominios distintos. Dentro de esa categoría, la investigación en neuropsicología pediátrica ha recurrido principalmente a cuatro tipos:

*Machine learning (aprendizaje automático):* Algoritmos que aprenden a clasificar o predecir a partir de datos etiquetados, sin instrucciones explícitas (Bzdok & Meyer-Lindenberg, 2020). En los estudios sobre TDAH, los más frecuentes son SVM, Random Forest y Gradient Boosting. Tienen la ventaja de ser relativamente interpretables: es posible identificar qué variables contribuyeron más a la clasificación, lo que facilita la discusión clínica.

*Deep learning (aprendizaje profundo)* Redes neuronales artificiales con múltiples capas de procesamiento que extraen patrones de datos complejos como imágenes, señales eléctricas, secuencias temporales sin que alguien defina previamente qué características buscar. Las redes convolucionales (CNN) dominan el análisis de neuroimágenes y señales de EEG. Las redes LSTM procesan señales temporales. Son las que producen los rendimientos más altos, aunque también las menos interpretables clínicamente.

*Visión por computadora:* Interpreta imágenes y video para detectar patrones de comportamiento. En la evaluación del TDAH se ha aplicado al seguimiento ocular, al análisis de movimientos corporales en tres dimensiones mediante sensores como Kinect skeleton data y al reconocimiento de expresiones faciales en entornos de evaluación virtual.

*Procesamiento de lenguaje natural:* Permite que los sistemas comprendan y analicen texto escrito en lenguaje humano. Se ha comenzado a aplicar al análisis automático de notas clínicas, cuestionarios de síntomas y reportes de docentes, aunque su presencia en la colección de datos de esta investigación es incipiente. La Tabla 1 presenta una síntesis de las aplicaciones con más información publicada, se utilizó Claude 3.5 Sonnet (Anthropic, 2026) para generar un resumen en formato de tabla con las aplicaciones de IA encontradas en la revisión teórica para la evaluación neuropsicológica del TDAH en población pediátrica:

**Tabla 1**

*Aplicaciones de IA encontradas en la literatura para la evaluación neuropsicológica del TDAH en población pediátrica*

<b>Aplicación</b>	<b>Algoritmo / Tecnología</b>	<b>Desempeño</b>	<b>Contexto</b>	<b>Referencia</b>
<b>AttnKare-D</b>	RV + ML	88.9% AUC	Clínico	Romero-Ayuso et al. (2021)
<b>AULA Classroom</b>	RV + CNN	85–90% Acc.	Investigación	Rizzo et al. (2020)
<b>Sistemas EEG-DL</b>	CNN + LSTM / GRU-Transformer	92–95% Acc.	Investigación	Quintero López et al. (2025)
<b>Plataformas digitales</b>	RF / SVM	84% Acc. / 93% Sens.	Telemedicina	Dai & Hsu (2025)
<b>ADHD Detector</b>	RF + SVM	>87% Acc.	Domiciliario	Zou et al. (2023)
<b>Sistema clínico UK</b>	ML multimodal	>85% Acc.	Clínico real	Chen et al. (2023)

*Nota. La tabla fue elaborada con la ayuda de inteligencia artificial Claude 3.5 Sonnet (Anthropic, 2026, se le solicitó al modelo organizar la información sobre las aplicaciones de IA encontradas en la revisión teórica para la evaluación neuropsicológica del TDAH en población pediátrica. Se relaciona algunas abreviaciones utilizadas en la tabla: (AUC) área bajo la curva ROC; (Acc.) accuracy; (Sens.) sensibilidad; (RV) realidad virtual; (ML) machine learning; (EEG) electroencefalografía; (CNN) red neuronal convolucional; (LSTM) memoria a largo plazo; (RF) Random Forest; (SVM) support vector machine.*

La síntesis de aplicaciones identificadas muestra que las herramientas basadas en inteligencia artificial utilizadas para la evaluación neuropsicológica del TDAH infantil incluyen diferentes tecnologías y modalidades de análisis, entre ellas realidad virtual, algoritmos de *machine learning*, redes neuronales profundas y sistemas basados en señales neurofisiológicas. Los estudios reportaron desempeños diagnósticos prometedores, lo que sugiere un potencial importante para apoyar procesos de evaluación y detección temprana.

Sin embargo, la utilidad clínica de estas herramientas no depende únicamente de sus niveles de desempeño. Aspectos como la accesibilidad tecnológica, la interpretabilidad de los modelos, la diversidad de las poblaciones evaluadas y la validación en contextos reales

constituyen elementos fundamentales para valorar su posible implementación dentro de la práctica neuropsicológica infantil. Desde esta perspectiva, las aplicaciones identificadas deben comprenderse como herramientas complementarias que pueden apoyar el proceso clínico, pero no sustituir la valoración integral ni el juicio profesional.

### **Autonomía de la IA:**

La IA ha demostrado capacidad para identificar patrones del TDAH con niveles de precisión que en muchos casos superan el 85%. Sin embargo, esa capacidad ha sido evaluada principalmente en muestras pequeñas, en contextos controlados y sin desagregación sistemática por género ni subtipo clínico (Dey et al., 2022; Chen et al., 2023). La IA definitivamente no reemplaza al neuropsicólogo, puede ser un asistente de alta precisión dentro de un proceso de evaluación integral, pero su incorporación responsable requiere validación transcultural, transparencia algorítmica y protocolos éticos para el manejo de datos de menores (Zhao et al., 2025).

La investigación sobre IA aplicada a la evaluación del TDAH ha evolucionado de forma gradual. En los primeros años predominaban estudios exploratorios con muestras pequeñas y algoritmos clásicos como SVM y Random Forest, muchas veces aplicados a señales EEG o a datos conductuales simples. Más recientemente, el campo ha incorporado arquitecturas de deep learning, análisis multimodal y modelos explicables, con resultados prometedores, pero todavía heterogéneos.

También se observa una tendencia clara: la mayor parte de la información encontrada proviene de Asia, Estados Unidos y Europa, mientras que América Latina casi no aparece representada. A esto se suma que la mayoría de los estudios no desagrega sus resultados por

sexo, lo que limita la comprensión de cómo estas herramientas funcionan en niñas y niños por separado. Esa ausencia no es solo un vacío bibliográfico, sino una limitación clínica y metodológica importante.

En conjunto, el marco teórico y el estado del arte muestran un campo con potencial, pero todavía incompleto. Hay avances técnicos relevantes, aunque persisten dudas sobre su validez clínica, su pertinencia cultural y su capacidad para detectar de forma equitativa distintas formas de presentación del TDAH. Desde esa base surge la necesidad de mapear la evidencia disponible en los últimos 10 años sobre aplicaciones de inteligencia artificial en la evaluación neuropsicológica del TDAH infantil.

## **Metodología**

### **Diseño del Estudio**

Este trabajo es una revisión de alcance de enfoque cualitativo-descriptivo. Se integran tres marcos metodológicos: el protocolo de Arksey y O'Malley (2005), actualizado por el Instituto Joanna Briggs (Peters et al., 2020); el marco PCC como guía para definir los criterios de elegibilidad; y los lineamientos PRISMA-ScR (Tricco et al., 2018) para el reporte del proceso.

### **Definición de Componentes PCC (Instituto Joanna Briggs)**

#### ***Población (P)***

Niñas y niños de 6 a 12 años con diagnóstico de TDAH confirmado según DSM-5-TR o CIE-11, o en proceso de evaluación por sospecha clínica. Criterio etario estricto: los estudios con rangos más amplios son elegibles solo si reportan datos separados para el subgrupo de 6 a 12 años.

#### ***Concepto (C)***

Herramientas de IA machine learning, deep learning, visión por computadora o procesamiento de lenguaje natural aplicadas al tamizaje, la evaluación neuropsicológica o el apoyo diagnóstico del TDAH. Se excluyen explícitamente las pruebas computarizadas tradicionales como el TOVA o el CPT en versiones estándar que no declaren arquitectura de IA.

#### ***Contexto (C)***

Cualquier entorno con implementación en clínico, educativo, domiciliario o de investigación. Idioma: inglés, español o portugués. Período: 2016–2025. Fuentes: BVS, EBSCO Megadata, SciELO, PubMed y ProQuest.

**Criterios de Exclusión Códigos:** E1 al E5 Rayyan (Ouzzani et al., 2016)

***E1 - Población incorrecta***

Muestras fuera del rango 6 a 12 años sin datos extraíbles del subgrupo.

***E2 - Concepto incorrecto***

IA para tratamiento o seguimiento farmacológico; pruebas computarizadas tradicionales sin arquitectura IA declarada.

***E3 - Diseño incorrecto***

Cartas al editor, opiniones, editoriales, resúmenes de congreso sin artículos indexados o revisiones.

***E4 - Fecha o idioma incorrecto***

Anterior al 1 de enero de 2016 o en idioma distinto al inglés, español o portugués.

***E5 - Texto completo no disponible***

Solo disponible en formato de resumen y no recuperable por ningún medio.

La jerarquía de aplicación fue E1, E2, E3, E4, E5. De igual forma los criterios fueron parametrizados en Rayyan (Ouzzani et al., 2016), antes del inicio del cribado. Los cuatro revisores aplicaron los criterios de forma independiente; las 115 discrepancias fueron resueltas por consenso grupal.

### **Fuentes y Estrategia de Búsqueda**

La búsqueda se realizó en seis fuentes de información: BVS, EBSCO Megadata, SciELO, PubMed, ProQuest y otras fuentes complementarias. Se construyeron tres estrategias de búsqueda: en inglés, español y portugués usando operadores booleanos AND, OR y NOT. Se

realizó además búsqueda manual de referencias de los artículos incluidos (snowballing) para recuperar estudios no indexados en las bases consultadas.

### ***Estrategia en Inglés***

*(ADHD OR "attention deficit hyperactivity disorder") AND ("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network" OR "random forest" OR "support vector machine" OR algorithm\*) AND (assessment OR evaluation OR diagnosis OR "neuropsychological assessment") AND (child\* OR pediatric\* OR "school-age") NOT (adult\* OR treatment OR intervention OR medication).*

### ***Estrategia en Español***

*(TDAH OR "trastorno por déficit de atención") AND ("inteligencia artificial" OR "aprendizaje automático" OR "aprendizaje profundo" OR algoritmo\*) AND (evaluación OR diagnóstico OR "evaluación neuropsicológica") AND (niño\* OR infantil OR escolar OR pediátrico\*) NOT (adulto\* OR tratamiento OR intervención).*

### ***Estrategia en Portugués***

*(TDAH OR "trastorno do déficit de atenção" OR "déficit de atenção e hiperatividade") AND ("inteligência artificial" OR "aprendizagem de máquina" OR "aprendizagem profunda" OR "redes neurais" OR algoritmo\*) AND (avaliação OR diagnóstico OR "avaliação neuropsicológica") AND (criança\* OR infantil OR escolar OR pediátrico\*) NOT (adulto\* OR tratamento OR intervenção OR medicação).*

### **Proceso de Selección**

Previo al inicio del proceso de selección de los artículos para los resultados, el protocolo de la investigación en el cual se incluyeron temas relevantes como el marco teórico,

planteamiento del problema, objetivos y metodología fue sometido a evaluación por parte de comité de ética de la institución. Una vez realizada la validación por su parte, el proyecto recibió aprobación para su ejecución dando pie al inicio de la fase de cribado y análisis de literatura contemplada en los resultados de la investigación. (Anexo 1)

La selección se realizó en dos fases con cuatro revisores independientes. En la fase 1 se evaluaron títulos y resúmenes en Rayyan, con cálculo del índice Kappa de Cohen ( $\kappa \geq 0.61$  como nivel mínimo aceptable). En la fase 2 se revisaron los 59 textos completos preseleccionados aplicando los criterios E1 al E5. Todo el proceso se documentó en el diagrama de flujo PRISMA-ScR.

De los 815 artículos identificados, se eliminaron 149 duplicados en Rayyan, quedando 666 para cribado por título y resumen. Se excluyeron 607 artículos. De los 59 evaluados a texto completo, se excluyeron 45, 20 por E1, 13 por E2, 10 por E3, y 2 por E5. La base de datos final contiene de 14 artículos.

### **Extracción de Datos y Síntesis**

La información de cada estudio se registró en una plantilla estandarizada con las categorías: datos generales del estudio, características de la muestra, características del algoritmo de IA, resultados psicométricos y de desempeño, contexto de implementación y consideraciones éticas. El análisis siguió un enfoque cualitativo-descriptivo con síntesis narrativa temática organizada en torno a los tres objetivos específicos.

## Resultados

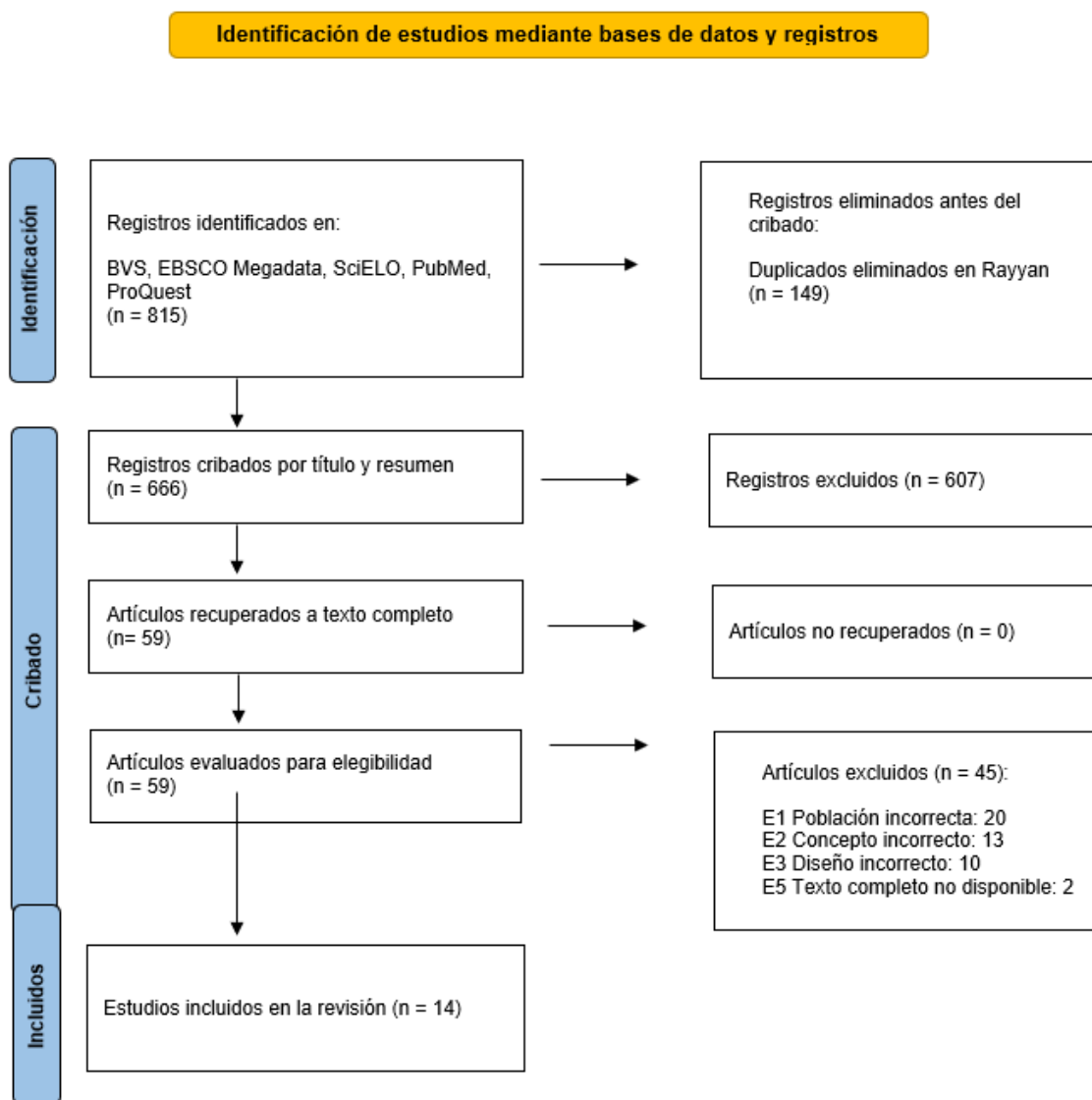
La búsqueda realizada en las seis fuentes de información identificó un total de 815 artículos. Tras eliminar 149 duplicados en Rayyan, quedaron 666 artículos para cribado por título y resumen, de los cuales se excluyeron 607 por no cumplir con los criterios de elegibilidad. Los 59 textos completos restantes fueron revisados en detalle; de ellos, 45 fueron excluidos por no ajustarse al marco PCC, por no corresponder a estudios originales o por no disponer del texto completo.

La procedencia de los artículos seleccionados para la revisión de texto completo proviene de revistas científicas indexadas principalmente de áreas de Neurología, Pediatría, Neuropsicología, Ingeniería y Ciencias de la Computación; esta multidisciplinariedad permite observar la naturaleza transdisciplinar del Trastorno de Déficit de Atención e Hiperactividad (TDAH) estudiado. Avance del conocimiento que garantizaron el uso de múltiples miradas teóricas y métodos científicos que validan el análisis de la información recopilada, algoritmos, aplicabilidad, instrumentos de IA, y la población participante.

La base de datos final quedó conformada por 14 artículos, cuyo proceso de selección se resume en la siguiente Figura 1, conforme al procedimiento establecido por PRISMA-ScR:

Figura 1

Diagrama de flujo del proceso de identificación, cribado y selección de estudios (PRISMA-ScR)



Source: Page MJ, et al. BMJ 2021;[372:n71](https://doi.org/10.1136/bmj.n71). Doi: 10.1136/bmj.n71.

This work is licensed under CC BY 4.0. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

*Nota. Adaptado de la declaración PRISMA 2021: una guía actualizada para la presentación de informes de revisiones sistemáticas (Diagrama de flujo PRISMA para nuevas revisiones sistemáticas), por Page MJ, et al., 2021, Doi: 10.1136/bmj.n71. CC BY 4.0.*

## Caracterización de las Aplicaciones de IA

Desde el punto de vista funcional, la mayoría de estas herramientas no “diagnostica” por sí sola, sino que transforma datos de entrada en patrones de clasificación o predicción que luego pueden apoyar la toma de decisiones clínicas. En los estudios basados en EEG, por ejemplo, la inteligencia artificial procesa señales eléctricas cerebrales y extrae regularidades asociadas con el control inhibitorio, la atención sostenida o la autorregulación. En otros trabajos, la IA analiza respuestas conductuales, tiempos de reacción o patrones de movimiento para identificar diferencias entre niños con TDAH y grupos control.

Más que una lista de tecnologías, esta categoría muestra una tendencia: la IA se ha usado sobre todo para convertir datos complejos en indicadores de desempeño cognitivo y conductual. Sin embargo, los estudios siguen siendo heterogéneos en el tipo de datos de entrada, el tamaño muestra, el contexto de aplicación y el grado de explicación clínica de sus resultados. La Tabla 2 presenta una síntesis de la clasificación de algoritmos y modalidades de datos desde el enfoque neuropsicológico, se utilizó Claude 3.5 Sonnet (Anthropic, 2026) para generar un resumen en formato de tabla:

Tabla 2

### *Clasificación de algoritmos y modalidades de datos desde el enfoque neuropsicológico*

<b>Modalidad</b>	<b>Dominio NP evaluado</b>	<b>Algoritmos principales</b>	<b>Estudios</b>
<b>EEG</b>	Conectividad cortical y actividad eléctrica	SVM, CNN-3D, CNN-LSTM, Att-CNN	Maheshwari (2025); Chen (2019); Chugh (2024); Ahmadi (2024); Atila (2023)
<b>fNIRS</b>	Activación prefrontal control inhibitorio	Random Forest	Li et al. (2025)
<b>Visión artificial (skeleton)</b>	Hiperactividad motora y vigilancia atencional	LSTM, Bi-LSTM, TCN	Lee et al. (2023a, 2023b, 2023c)
<b>Realidad virtual</b>	Tiempos de reacción y atención sostenida	SVM, XGBoost, LR	Zhang (2023); Yeh (2020)
<b>Escalas / baterías cognitivas</b>	Perfiles psicométricos e IA explicable	LASSO, RF + SHAP, Clustering	Weigard (2023); Qin (2025); Yamashita (2024)

*Nota. La tabla fue elaborada con la ayuda de inteligencia artificial Claude 3.5 Sonnet (Anthropic, 2026), se le solicitó al modelo organizar la información sobre la clasificación de algoritmos y modalidades de datos desde el enfoque neuropsicológico. Se relaciona algunas abreviaciones utilizadas en la tabla: (EEG) electroencefalograma; (fNIRS) espectroscopía funcional de infrarrojo cercano; (SVM) máquina de vectores de soporte; (CNN) red neuronal convolucional; (LSTM) Long Short-Term Memory; (TCN) red convolucional temporal; (SHAP) SHapley Additive Explanations; (RF) Random Forest; (LR) regresión logística.*

La clasificación de algoritmos y modalidades de datos identificada muestra que las aplicaciones basadas en inteligencia artificial para la evaluación neuropsicológica del TDAH utilizan múltiples fuentes de información para aproximarse a procesos cognitivos y conductuales asociados con el trastorno. Entre las modalidades más frecuentes se encontraron señales electroencefalográficas, medidas de activación cerebral, análisis de movimiento, realidad virtual y perfiles psicométricos.

Desde una perspectiva neuropsicológica, esta diversidad resulta relevante porque permite explorar diferentes dominios funcionales relacionados con el TDAH, como atención sostenida, control inhibitorio, hiperactividad motora y funciones ejecutivas. Sin embargo, la heterogeneidad observada entre algoritmos, fuentes de datos y contextos de aplicación dificulta establecer comparaciones directas entre investigaciones y limita la posibilidad de identificar un modelo único de evaluación aplicable a todos los escenarios clínicos. Estos hallazgos sugieren que las herramientas desarrolladas deben comprenderse como estrategias complementarias que aportan información adicional dentro de un proceso de evaluación integral.

### **Desempeño Diagnóstico y Características de los Modelos**

En términos de desempeño diagnóstico, la mayoría de los estudios reportó el porcentaje total de casos que el sistema clasificó correctamente (accuracy) elevado; sin embargo, estos

porcentajes deben interpretarse con prudencia porque con frecuencia provienen de muestras pequeñas, validaciones internas y contextos altamente controlados, lo que reduce su aplicabilidad directa al entorno clínico real. Desde una mirada neuropsicológica, una herramienta útil no solo debe “acertar”, sino también sostener atención, evitar fatiga y conservar validez en niños y niñas con perfiles distintos de desarrollo.

Además, aunque algunos modelos muestran resultados prometedores, no siempre reportan de forma completa todas las métricas necesarias para valorar su utilidad clínica. En especial, la ausencia frecuente de sensibilidad, especificidad, intervalos de confianza y validación externa limita la comparación entre estudios y dificulta establecer qué herramientas serían realmente transferibles a la práctica neuropsicológica. La Tabla 3 presenta una síntesis de las métricas de desempeño diagnóstico de los 14 artículos incluidos, ordenados de mayor a menor accuracy, se utilizó Claude 3.5 Sonnet (Anthropic, 2026) para generar un resumen en formato de tabla:

Tabla 3

*Métricas de desempeño diagnóstico de los 14 artículos incluidos, ordenados de mayor a menor accuracy*

Estudio	Algoritmo	Accuracy	Sensib.	Especif.	AUC/F1	N	Edad
<b>Maheshwari et al. (2025)</b>	SVM+kNN+Ensemble	99.29%	—	—	—	121	7–12
<b>Ahmadi et al. (2024)</b>	Att-CNN (EEG)	99.27%	99.18%	99.36%	AUC:0.99	121	7–12
<b>Chugh et al. (2024)</b>	CNN-LSTM	98.86%	—	—	—	121	7–12
<b>Lee et al. (2023a)</b>	TCN	98.60%	98.20%	98.90%	F1:0.98	595	6–12

Lee et al. (2023b)	Bi-LSTM+Atención	98.15%	Var.	Var.	F1:0.97	596	7–12
Lee et al. (2023c)	Bi-LSTM ponderado	97.80%	—	—	F1:0.95	596	6–12
Atila et al. (2023)	SVM	97.46%	—	—	—	121	7–12
Yamashita et al. (2024)	Clustering+SVM	93.00%	88.00%	—	F1:0.90	7.257	9–11
Yeh et al. (2020)	SVM+XGB+LR	91.50%	—	—	—	68	6–12
Li et al. (2025)	Random Forest (fNIRS)	89.20%	95.00%	82.40%	AUC:0.91	181	6–12
Chen et al. (2019)	CNN-3D (EEG)	88.27%	86.00%	90.54%	AUC:0.91	100	9–12
Qin et al. (2025)	RF+SHAP	86.30%	88.30%	81.30%	AUC:0.94	776	7–12
Zhang et al. (2023)	SVM (VR)	78.50%	—	—	—	40	7–10
Weigard et al. (2023)	LASSO+PCR	N/A†	N/A†	N/A†	R <sup>2</sup> =15–20%	11.878	9–10

*Nota. La tabla fue elaborada con la ayuda de inteligencia artificial Claude 3.5 Sonnet (Anthropic, 2026), se le solicitó al modelo organizar la información sobre las métricas de desempeño diagnóstico de los 14 artículos incluidos, ordenados de mayor a menor accuracy. Se relaciona algunas abreviaciones utilizadas en la tabla: (—) métrica no reportada; (Var.) variable por clase; Lee et al. (2023b) reporta métricas desagregadas sin valor global único; (†) Weigard et al. (2023) emplea regresión dimensional, no clasificación binaria; métrica principal: varianza explicada ( $R^2 = 15-20\%$ ) y ( $r$ ) 0.38–0.46; (AUC) área bajo la curva ROC; (Sensib.) sensibilidad; (Especif.) especificidad.*

Los estudios incluidos reportaron desempeños diagnósticos elevados, con porcentajes de *accuracy* que oscilaron entre 78.5 % y 99.3 %. Estos hallazgos sugieren un potencial importante de las aplicaciones basadas en inteligencia artificial para apoyar procesos de evaluación neuropsicológica del TDAH infantil. Sin embargo, los resultados deben interpretarse con cautela debido a las diferencias metodológicas entre estudios, el predominio de validaciones internas y la variabilidad observada en el tamaño y características de las muestras incluidas.

Asimismo, un alto desempeño reportado en contextos experimentales no necesariamente garantiza resultados equivalentes en escenarios clínicos reales. Desde una perspectiva neuropsicológica, una herramienta útil no solo requiere clasificar adecuadamente los casos, sino también demostrar estabilidad, sensibilidad frente a distintas formas de presentación clínica y capacidad de adaptación a contextos diversos. En este sentido, los hallazgos identificados respaldan el potencial de estas tecnologías como herramientas complementarias dentro del proceso evaluativo, más que como sustitutos de la valoración clínica integral. La Tabla 4 presenta una síntesis de la distribución geográfica de los 14 artículos incluidos por región y país, se utilizó Claude 3.5 Sonnet (Anthropic, 2026) para generar un resumen en formato de tabla:

Tabla 4

*Distribución geográfica de los 14 artículos incluidos por región y país*

Región	País	N	%	Estudios
Asia	China	4	29%	Li (2025); Zhang (2023); Chen (2019); Qin (2025)
Asia	Corea del Sur	3	21%	Lee et al. (2023a, 2023b, 2023c)
Asia	India	2	14%	Maheshwari (2025); Chugh (2024)
Asia	Japón	1	7%	Yamashita (2024)
Asia	Irán	1	7%	Ahmadi (2024)
Asia	Turquía	1	7%	Atila (2023)
Asia	Taiwán	1	7%	Yeh (2020)
América del Norte	Estados Unidos	1	7%	Weigard (2023)
América Latina	—	0	0%	Ningún estudio identificado
<b>Total</b>	<b>9 países</b>	<b>14</b>	<b>100%</b>	<b>86% proviene de Asia</b>

*Nota. La tabla fue elaborada con la ayuda de inteligencia artificial Claude 3.5 Sonnet (Anthropic, 2026), se le solicitó al modelo organizar la información sobre la distribución geográfica de los 14 artículos incluidos por*

*región y país. Se relaciona algunas abreviaciones utilizadas en la tabla: (—) ningún estudio identificado; la ausencia de América Latina en el corpus es uno de los vacíos centrales de esta revisión.*

La distribución geográfica de los estudios incluidos muestra una concentración importante de investigaciones en países asiáticos y otras regiones con mayor desarrollo tecnológico, mientras que no se identificaron estudios provenientes de América Latina dentro del corpus analizado. Este hallazgo adquiere relevancia debido a que los algoritmos de inteligencia artificial son desarrollados y entrenados a partir de características específicas de las poblaciones evaluadas, las cuales pueden variar según factores culturales, sociales, educativos y contextuales.

Desde una perspectiva neuropsicológica, la ausencia de representación latinoamericana constituye una limitación importante para la generalización de los hallazgos hacia contextos hispanohablantes y poblaciones con características diferentes a las utilizadas para entrenar los modelos. Lo anterior sugiere la necesidad de fortalecer investigaciones regionales que permitan desarrollar herramientas más inclusivas y con mayor pertinencia clínica y sociocultural para distintos grupos poblacionales.

Respecto a la distribución por género, ninguno de los 14 estudios desagrega sus métricas de desempeño por sexo. Esa omisión significa que el sesgo diagnóstico existente en la práctica clínica podría estar siendo reproducido en los algoritmos sin que nadie lo haya verificado (Aguirre Salazar, 2024; Faraone et al., 2021). La Tabla 5 presenta una Síntesis de métricas de rendimiento diagnóstico reportadas en los 14 estudios incluidos, se utilizó Claude 3.5 Sonnet (Anthropic, 2026) para generar un resumen en formato de tabla:

Tabla 5

*Síntesis de métricas de rendimiento diagnóstico reportadas en los 14 estudios incluidos*

Métrica	Estudios que reportan	% Del corpus	Rango observado	Media	Implicación clínica
<b>Accuracy</b>	13/14	93%	78.5%–99.3%	93.6%	<i>Métrica más reportada</i>
<b>AUC-ROC</b>	4/14	29%	0.91–0.99	0.938	<i>Más robusta — menos reportada</i>
<b>Sensibilidad</b>	6/14	43%	86%–99.2%	92.4%	<i>Crítica para el cribado</i>
<b>Especificidad</b>	5/14	36%	81.3%–99.4%	90.5%	<i>Crítica para confirmar diagnóstico</i>
<b>F1-Score</b>	4/14	29%	0.895–0.992	0.942	<i>Relevante en muestras desbalanceadas</i>

*Nota. La tabla fue elaborada con la ayuda de inteligencia artificial Claude 3.5 Sonnet (Anthropic, 2026), se le solicitó al modelo organizar la información sobre las síntesis de métricas de rendimiento diagnóstico reportadas en los 14 estudios incluidos. Se relaciona algunas abreviaciones utilizadas en la tabla: Los porcentajes se calculan sobre el total de 14 estudios. Los valores de media excluyen los estudios que no reportaron la métrica.*

La síntesis de las métricas diagnósticas permite observar que el *accuracy* fue el indicador más frecuentemente reportado para describir el desempeño de los modelos desarrollados. Sin embargo, interpretar esta medida de manera aislada puede resultar insuficiente para comprender el comportamiento real de una herramienta basada en inteligencia artificial dentro de un contexto clínico. Aunque porcentajes elevados de precisión pueden sugerir un adecuado rendimiento, otras métricas aportan información complementaria relevante para valorar la utilidad potencial de los sistemas desarrollados.

Desde una perspectiva neuropsicológica, indicadores como sensibilidad, especificidad, AUC y F1-score adquieren especial importancia debido a que permiten comprender no sólo la capacidad de identificar correctamente los casos, sino también la posibilidad de reducir errores de clasificación y falsos diagnósticos. La integración de diferentes métricas proporciona una valoración más amplia y favorece una interpretación clínica más prudente respecto a la aplicabilidad de estas herramientas en población infantil con TDAH.

## Vacíos de Evidencia Actual

El análisis de los artículos permitió identificar vacíos importantes. El primero es la escasa representación de América Latina y de contextos hispanohablantes, lo que limita la validez transcultural de las herramientas revisadas. El segundo es la subrepresentación de niñas en varias muestras, sin análisis desagregado por género, lo cual impide saber si los modelos funcionan igual en niños y niñas.

A esto se suma un tercer vacío: la limitada interpretabilidad de muchos algoritmos. Aunque algunos modelos ofrecen predicciones precisas, no siempre explican de manera comprensible cómo llegaron a sus resultados. En neuropsicología infantil, esta limitación es relevante porque una herramienta útil no solo debe clasificar, sino también permitir una lectura clínica clara y responsable. La Tabla 6 presenta las características metodológicas avanzadas de los 14 artículos incluidos, se utilizó Claude 3.5 Sonnet (Anthropic, 2026) para generar un resumen en formato de tabla:

Tabla 6

*Características metodológicas avanzadas de los 14 artículos incluidos*

Estudio	Diseño	Validación	Clasificación	XAI	Long.	Multic.	Gén.
<b>Maheshwari (2025)</b>	Transversal	CV interna	Binaria	No	No	No	No
<b>Weigard (2023)</b>	Longitudinal	Multicéntrico	Dimensional	No	Sí	Sí	No
<b>Li (2025)</b>	Transversal	CV interna	Binaria	No	No	No	No
<b>Atila (2023)</b>	Transversal	CV interna	Binaria	No	No	No	No
<b>Zhang (2023)</b>	Transversal	CV interna	Binaria	No	No	No	No
<b>Chugh (2024)</b>	Transversal	CV interna	Binaria	No	No	No	No

<b>Yeh (2020)</b>	Transversal	CV interna	Binaria	No	No	No	No
<b>Chen (2019)</b>	Transversal	CV interna	Binaria	Sí	No	No	No
<b>Lee (2023a)</b>	Transversal	CV interna	3 clases	No	No	No	No
<b>Lee (2023b)</b>	Transversal	CV interna	3 clases	No	No	No	No
<b>Lee (2023c)</b>	Transversal	CV interna	3 clases	No	No	No	No
<b>Qin (2025)</b>	Transversal	CV+val. ext.	Binaria+subtipos	Sí	No	No	No
<b>Yamashita (2024)</b>	Transversal	Multicéntrico	No supervisado	No	No	Sí	No
<b>Ahmadi (2024)</b>	Transversal	CV interna	Binaria	No	No	No	No
<b>Totales</b>	<b>13 transv. /1 long.</b>	<b>12 CV int./2 multié.</b>	<b>9 bin. /3 tres cl./1 dim. /1 no sup.</b>	<b>2/14</b>	<b>1/14</b>	<b>2/14</b>	<b>0/14</b>

*Nota. La tabla fue elaborada con la ayuda de inteligencia artificial Claude 3.5 Sonnet (Anthropic, 2026), se le solicitó al modelo organizar la información sobre las características metodológicas avanzadas de los 14 artículos incluidos. Se relaciona algunas abreviaciones utilizadas en la tabla: (CV) validación cruzada interna. (XAI) inteligencia artificial explicable. (Long.) diseño longitudinal. (Multic.) multicéntrico. (Gén.) desagregación de métricas por género. Las celdas en verde indican presencia de la característica. Rojo en la columna Gén. refleja ausencia total (0/14).*

Las características metodológicas identificadas permiten observar un predominio de diseños transversales y procesos de validación interna dentro de los estudios incluidos. Asimismo, se evidenció una limitada presencia de investigaciones longitudinales, multicéntricas y con incorporación de estrategias de inteligencia artificial explicable. De igual manera, la ausencia de análisis desagregados por género constituye un aspecto relevante debido a su posible influencia sobre la representatividad y el comportamiento de los modelos desarrollados.

Desde una perspectiva neuropsicológica, estos hallazgos sugieren que el desempeño de una herramienta no depende únicamente de su capacidad para clasificar o predecir casos, sino también de la posibilidad de mantener resultados consistentes en diferentes contextos y poblaciones. La limitada incorporación de metodologías más robustas puede restringir la

generalización clínica de los hallazgos, por lo que futuras investigaciones deberían fortalecer aspectos relacionados con validación externa, diversidad muestral y aplicabilidad en escenarios reales de evaluación infantil. La Tabla 7 presenta la identificación de vacíos en la información encontrada y limitaciones del corpus, se utilizó Claude 3.5 Sonnet (Anthropic, 2026) para generar un resumen en formato de tabla:

Tabla 7

*Identificación de vacíos en la información encontrada y limitaciones del corpus*

Vacío	Descripción	Evidencia en el corpus
<b>Geográfico y cultural</b>	Nula representación latinoamericana. Los algoritmos no contemplan variables socioculturales hispanohablantes.	Ausente en los 14 estudios (0%).
<b>Sesgo de muestra</b>	Riesgo de sobreajuste por validaciones en muestras muy reducidas ( $N < 100$ ).	Zhang (2023); Yeh (2020); Chen (2019).
<b>Equidad de género</b>	Ningún estudio desagrega métricas por sexo. El fenotipo femenino del TDAH está subrepresentado.	Ausente en los 14 estudios (0%).
<b>Interpretabilidad</b>	86% de los modelos son cajas negras — incompatible con la deontología profesional (Ley 1090 de 2006).	Solo Qin (2025) y Chen (2019) incorporan XAI.
<b>Longitudinalidad</b>	93% son estudios transversales. No se puede evaluar si los modelos son estables en el tiempo.	Solo Weigard (2023) tiene diseño longitudinal multicéntrico.

*Nota. La tabla fue elaborada con la ayuda de inteligencia artificial Claude 3.5 Sonnet (Anthropic, 2026), se le solicitó al modelo organizar la información sobre la identificación de vacíos en la información encontrada y limitaciones del corpus. Se relaciona algunas abreviaciones utilizadas en la tabla: (XAI) inteligencia artificial explicable; (SHAP) SHapley Additive Explanations; (Grad-CAM) Gradient-weighted Class Activation Mapping.*

Los vacíos de información identificados permiten reconocer limitaciones persistentes dentro de las investigaciones relacionadas con inteligencia artificial aplicada a la evaluación neuropsicológica del TDAH infantil. Entre los aspectos observados se encuentran la limitada representación geográfica, la ausencia de estudios provenientes de contextos latinoamericanos, la

escasa incorporación de análisis diferenciales por género y la reducida presencia de diseños longitudinales y validaciones externas.

Desde una perspectiva neuropsicológica y clínica, estos hallazgos sugieren que, aunque el desarrollo tecnológico ha mostrado avances importantes durante los últimos años, aún persisten desafíos metodológicos que pueden limitar la generalización y aplicabilidad de los resultados en distintos contextos poblacionales. El fortalecimiento de futuras investigaciones mediante muestras más diversas, metodologías robustas y modelos con mayor interpretabilidad podría favorecer el desarrollo de herramientas con mayor sensibilidad clínica y pertinencia para diferentes escenarios de evaluación infantil.

## Discusión

Considerando el propósito de nuestra investigación, que fue mapear la información disponible sobre aplicaciones de inteligencia artificial en la evaluación neuropsicológica del TDAH en niñas y niños de 6 a 12 años, los hallazgos nos permiten afirmar que el campo se encuentra aún en crecimiento por lo que está en una etapa de maduración metodológica. La literatura revisada nos muestra un interés importante por incorporar herramientas de IA a la evaluación infantil, pero ese desarrollo sigue siendo heterogéneo en cuanto a tecnologías, tipos de datos, contextos de aplicación y nivel de validación clínica.

Desde una perspectiva neuropsicológica, esto nos confirma algo esencial: la inteligencia artificial puede ser una herramienta de apoyo, pero no reemplaza el juicio clínico del neuropsicólogo. La evaluación del TDAH no consiste sólo en clasificar resultados; exige comprender el desarrollo del niño, su contexto, la expresión conductual del trastorno y las diferencias evolutivas entre edades. En ese sentido, encontramos que la IA aporta valor cuando ayuda a detectar patrones, reducir tiempos o aumentar objetividad, pero pierde fuerza cuando se presenta como una solución cerrada y descontextualizada.

Uno de los hallazgos más importantes es la distancia entre el rendimiento reportado en entornos controlados y las condiciones reales del consultorio. Muchos estudios muestran cifras elevadas de precisión, pero esos resultados no siempre garantizan utilidad clínica, especialmente cuando no se reporta validación externa o cuando las muestras son demasiado pequeñas o homogéneas. En la práctica, una herramienta para evaluación neuropsicológica debe funcionar con niños reales, con variabilidad atencional, fatiga, ansiedad, distracción y diversidad

sociocultural. Ese punto es clave porque la evaluación infantil no ocurre en un laboratorio ideal, sino en un contexto humano cambiante.

Otro aspecto central es el de la equidad de género. La literatura revisada sigue mostrando una representación insuficiente de niñas y una falta de análisis diferenciado por sexo. Esto es preocupante porque el TDAH no se expresa de la misma manera en todos los casos, y una herramienta entrenada principalmente con perfiles masculinos puede dejar por fuera formas más sutiles o internalizadas de presentación clínica. Por tanto, el problema no es solo técnico, sino también clínico y ético.

Asimismo, la revisión permite comprender que el campo de la IA aplicada al TDAH infantil ha estado más orientado al apoyo diagnóstico que a la evaluación neuropsicológica profunda. Esto significa que aún falta recorrer un camino importante para que estas herramientas no solo identifiquen casos, sino que también aporten información útil sobre atención, memoria de trabajo, control inhibitorio, flexibilidad cognitiva y otras funciones ejecutivas relevantes en el TDAH.

En conjunto, la principal contribución de esta revisión no es afirmar que la IA ya resuelve la evaluación del TDAH, sino mostrar que existe un campo prometedor, aunque todavía incompleto, que requiere mayor validación transcultural, mejores diseños metodológicos, análisis por género y mayor integración con la práctica neuropsicológica real. Desde esa perspectiva, la revisión aporta una lectura crítica del estado actual de la literatura y deja claro que el desarrollo futuro debe orientarse tanto a la precisión técnica como a la utilidad clínica y la pertinencia contextual.

## **La Distancia entre el Laboratorio y la Clínica**

El hallazgo más transversal es la distancia entre los rendimientos reportados y lo que cabría esperar en una consulta real. Accuracies de 97% - 99% sobre datasets controlados de 121 participantes no son comparables con lo que ocurriría con 121 niños derivados consecutivamente a una clínica neuropsicológica en Bogotá. Varoquaux et al. (2017) demostraron que las estimaciones de rendimiento de modelos de neuroimagen tienden a inflarse cuando la validación es interna y ese fenómeno es visible en buena parte del corpus. La distancia entre el 99% de precisión del laboratorio y el 86%–88% que se observa en los estudios con mayor validez ecológica Qin et al. (2025) y Yeh et al. (2020) señala que la validación externa es una condición necesaria, no un trámite opcional.

El estudio de Weigard et al. (2023) ofrece el estándar metodológico más riguroso del corpus: validación en 21 sitios independientes, seguimiento prospectivo y más de 10.000 participantes. Su varianza explicada 15%–20% parece modesta comparada con las otras cifras, pero es la más honesta

## **El Género como Cuestión Científica**

La ausencia de desagregación de resultados por género en todos los estudios del corpus no es solo una falla ética sino una falla científica. Si un modelo se entrena con patrones predominantemente masculinos y su rendimiento no se evalúa por género, no es posible saber si funciona igual para niños que para niñas. Y si no funciona igual, el modelo puede estar perpetuando la inequidad diagnóstica que se quería corregir, haciéndolo de forma invisible porque el número global parece correcto (Aguirre Salazar, 2024; Zhao et al., 2025).

## **América Latina y la Interpretabilidad**

La ausencia de América Latina en el corpus significa que ninguna de las herramientas identificadas ha sido validada con población colombiana y latinoamericana. Los patrones culturales, los instrumentos que se usan y el contexto socioeconómico colombiano no están representados en ninguno de estos modelos. Transferirlos sin validación transcultural sería metodológicamente inapropiado (López-Cortés et al., 2022; Rusca-Jordán & Cortez-Vergara, 2020).

Respecto a la interpretabilidad; solo dos estudios del corpus incorporan XAI Chen et al. (2019) con Grad-CAM, y Qin et al. (2025) con SHAP. El 86% restante produce una predicción sin explicación de cómo se llegó a ella. En Colombia, el diagnóstico neuropsicológico debe poder explicarse ante el paciente, la familia y el sistema (Ley 1090 de 2006; Rudin, 2019; León-Domínguez, 2024). Un algoritmo sin interpretabilidad no cumple ese estándar.

## Conclusiones

Luego de analizar la información encontrada de los últimos diez años, es posible afirmar que la inteligencia artificial ha logrado incluirse en el campo de la evaluación neuropsicológica del TDAH. Sin embargo, cabe aclarar que la literatura actual da cuenta de que la diversidad de métodos utilizados, la concentración geográfica y los vacíos significativos de las muestras están siendo congruentes con la falta de uniformidad en el contexto.

Esta revisión de alcance nos permite visualizar un panorama más claro en cuanto al crecimiento de programas basados en machine learning, deep learning y el análisis automatizado de señales y conductas; en donde los resultados evidencian una alta precisión predictiva de los algoritmos dado su valor en el respaldo cuantitativo, sin embargo, carece de un análisis integral y cualitativo que solo se puede obtener con una verdadera evaluación neuropsicológica.

A nivel tecnológico, se comprobó que los modelos están basados principalmente de los datos obtenidos de electroencefalografías (EEG), registros de conducta digitalizados y pruebas en entornos de realidad virtual. Esto es importante tomarlo con cautela ya que estos algoritmos pueden funcionar de manera perfecta en el laboratorio, sin embargo, en el contexto natural (consultorio) podrían presentar fallas ante la diversidad de pacientes.

Por otro lado, se encuentra que la literatura pone en evidencia brechas imposibles de ignorar en cuanto a la invisibilización geográfica y cultural de Latinoamérica, representando un 0% de los estudios analizados; lo que significa una baja posibilidad de replicar estos modelos en contextos hispanohablantes, al omitir por completo las realidades de la región. De la misma forma, sucede con el sesgo de género, en el cual no se reportan puntualmente detecciones tempranas en niñas.

Es posible concluir que la investigación nos muestra un escenario en donde si bien hay bastante información sobre el uso de la IA en el campo de la evaluación neuropsicológica, esta disciplina aún se encuentra en una etapa de crecimiento. Los modelos actuales representan un punto de partida muy prometedor, sin embargo, sus limitaciones a nivel geográfico, ético y metodológico dificultan la posibilidad de réplica en el contexto clínico. Actualmente el verdadero papel de la IA no está en reemplazar al profesional sino actuar como un asistente con una gran precisión analítica, en donde el éxito estará sujeto siempre al juicio clínico, la ética y sensibilidad del neuropsicólogo infantil.

La IA no reemplaza de ninguna manera el trabajo del neuropsicólogo, sin embargo, si representa una herramienta de gran apoyo para los procesos diagnósticos dada su precisión en el reconocimiento y exposición de datos. Cabe agregar que, debido a la falta de información y entrenamiento con información de población infantil femenina, es posible decir que el riesgo de alucinación y fallos de interpretación estén presentes en la actualidad.

## Recomendaciones

A partir de los hallazgos de esta revisión, se recomienda a futuras investigaciones desarrollar estudios con muestras más amplias y diversas, especialmente en población latinoamericana, ya que actualmente existe muy poca representación de estos contextos dentro de la literatura encontrada.

Asimismo, sería importante y necesario incluir análisis diferenciados por género, debido a que los síntomas del TDAH pueden variar entre niños y niñas, lo que podría influir en el desempeño de los modelos de cada inteligencia artificial.

También se considera necesario fortalecer la validación clínica de estas herramientas en escenarios reales de la evaluación neuropsicológica infantil, teniendo en cuenta factores como el contexto psicosocial, familiar, escolar y el sociocultural de cada niño y niña.

De igual manera, se recomienda promover modelos de inteligencia artificial más interpretables, que permitan comprender cómo se generan esos resultados y faciliten a la vez una integración más ética deontológica y responsable dentro de la práctica clínica.

Finalmente, se sugiere aceptar la inteligencia artificial como una herramienta complementaria dentro de la evaluación neuropsicológica infantil, reconociendo que el juicio clínico, la observación profesional y la interpretación integral del desarrollo infantil continúan siendo fundamentales en el proceso diagnóstico.

## Referencias

- Aguirre Salazar, A. (2024). Gender disparities in ADHD diagnosis and implications for girls' mental health. *Current Issues in Child Psychology*, 45(2), 145–162 (p. 1).
- American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). doi.org (p. 1)
- Anthropic. (2026). *Claude 3.5 Sonnet* (Versión del modelo de lenguaje) (Software de gran modelo de lenguaje). anthropic.com (p. 1)
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19–32. doi.org (p. 1)
- Atem, F., Winter, S., Blackwell, C. K., & Fiks, A. G. (2024). Systematic review of barriers to timely access to evaluation and treatment for children with ADHD. *Journal of Attention Disorders*, 28(5), 757–772. doi.org (p. 1)
- Barkley, R. A. (2015). *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment* (4th ed.). Guilford Press (p. 5).
- Bzdok, D., & Meyer-Lindenberg, A. (2020). Machine learning for precision psychiatry: Opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 5(8), 778–779. doi.org (p. 1)
- Chen, T., Tachmazidis, I., Batsakis, S., Adamou, M., Papadakis, E., & Antoniou, G. (2023). Diagnosing attention-deficit hyperactivity disorder (ADHD) using artificial intelligence: A clinical study in the UK. *Frontiers in Psychiatry*, 14, 1164433. doi.org (p. 1)
- Congreso de la República de Colombia. (6 de septiembre de 2006). Ley 1090 de 2006. Por la cual se reglamenta el ejercicio de la Profesión de la Psicología, se dicta el Código

- Deontológico y Bioético y otras disposiciones. *Diario Oficial No. 46.383*.  
secretariassenado.gov.co (p. 2)
- Cortese, S., Lau, I. V., Wong, I. C. K., Coghill, D., Marco de Jager, E., Mohammadi, M. R., Rohde, L. A., & Ip, P. (2021). A practical, evidence-informed approach to managing stimulant-refractory ADHD. *CNS Drugs*, 35(10), 1035–1051. doi.org (p. 2)
- Dai, Y.-W., & Hsu, C.-F. (2025). Machine learning for ADHD diagnosis: Feature selection from parent reports, self-reports, and neuropsychological measures. *Children*, 12(11), 1448. doi.org (p. 2)
- Dey, S., Dutta, S., Roy, S., & Banerjee, R. (2022). AI-based analysis of behavioral and neuropsychological data for ADHD classification in children. *Frontiers in Psychiatry*, 13, 856432. doi.org (p. 2)
- Faraone, S. V., Banaschewski, T., Coghill, D., Zheng, Y., Biederman, J., Bellgrove, M. A., & Wang, Y. (2021). The World Federation of ADHD International Consensus Statement: 208 evidence-based conclusions about the disorder. *Neuroscience & Biobehavioral Reviews*, 128, 789–818. doi.org (p. 2)
- Hinshaw, S. P., Nguyen, P. T., O’Grady, S. M., & Rosenthal, E. A. (2022). Annual Research Review: Attention-deficit/hyperactivity disorder in girls and women: Underrepresentation, longitudinal processes, and key directions. *Journal of Child Psychology and Psychiatry*, 63(4), 484–496. doi.org (p. 5)
- Joanna Briggs Institute. (2020). *JBIR reviewer's manual*. University of Adelaide. jbi.global (p. 2)
- Kazda, L., Bell, M. A., Thomas, R., & McGeechan, K. (2023). Global prevalence of ADHD in children and adolescents: A meta-analysis of recent studies. *The Lancet Child & Adolescent Health*, 7(5), 353–361. doi.org (p. 2)

- León-Domínguez, U. (2024). *Inteligencia artificial aplicada a la neuropsicología clínica*. Editorial Médica Panamericana (p. 3).
- Londoño-Ocampo, L. M., & Tezón, M. (2019). Rendimiento neuropsicológico de niños y niñas con TDAH. *Revista de Psicología INFAD: Psyche, 1*(2), 115–124. redalyc.org (p. 3)
- López-Cortés, O. D., Betancourt-Núñez, A., Bernal Orozco, M. F., & Vizmanos, B. (2022). Scoping reviews: Una nueva forma de síntesis de la evidencia. *Investigación en Educación Médica, 11*(44), 98–104. doi.org (p. 3)
- Mohammadi, M. R., Salmanian, M., & Ghanizadeh, A. (2023). ADHD prevalence and demographic correlates in children worldwide: Systematic review. *International Journal of Epidemiology, 52*(3), 489–501. doi.org (p. 3)
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews, 5*(1), 210. doi.org (p. 4)
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*(n71). doi.org (p. 3)
- Parsons, T. D., Carlew, A. R., & Magtoto, J. (2020). Computerized neuropsychological assessment: Applications, reliability, and future directions. *Neuropsychology Review, 30*(1), 43–59. doi.org (p. 3)
- Peters, M. D. J., Godfrey, C., McInerney, P., Munn, Z., Tricco, A. C., & Khalil, H. (2020). Chapter 11: Scoping reviews. En E. Aromataris & Z. Munn (Eds.), *JBİ manual for evidence synthesis*. JBİ. jbi.global (p. 3)

- Philip Asherson, Buitelaar, J., Faraone, S. V., & Rohde, L. A. (2023). Artificial intelligence and ADHD: Opportunities, limitations and clinical translation. *European Child & Adolescent Psychiatry*. doi.org (p. 4)
- Quintero López, C., Gil Vera, V. D., & Ruiz Quintero, M. J. (2025). Diagnosis of ADHD in children with EEG and machine learning: Systematic review and meta-analysis. *Clinical and Health*, 36(2), 109–121. doi.org (p. 4)
- Karam, R. A., Bau, C. H. D., Salgado, C. A. I., Kalil, K. L., & Victor, M. M. (2020). Applications of artificial intelligence in attention-deficit/hyperactivity disorder diagnosis and treatment. *Neuroscience Bulletin*, 36(11), 1349–1356. doi.org (p. 4)
- Rizzo, A. S., Bowerly, T., Buckwalter, J. G., Klimchuk, D., Mitura, R., & Parsons, T. D. (2020). A virtual reality scenario for all seasons: The virtual classroom. *CNS Spectrums*, 11(1), 35–44. doi.org (p. 4)
- Romero-Ayuso, D., Soto-Balbuena, C., & Costa, R. (2021). Artificial intelligence-assisted neuropsychological testing in attention deficit hyperactivity disorder. *Frontiers in Human Neuroscience*, 15, 629834. doi.org (p. 4)
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215. doi.org (p. 4)
- Rusca-Jordán, F., & Cortez-Vergara, C. (2020). Trastorno por déficit de atención con hiperactividad (TDAH) en niños y adolescentes: Una revisión clínica. *Revista de Neuro-Psiquiatría*, 83(2), 91–105 (p. 5).

- Slaton, R. M., Johnson, N., & Kofler, M. J. (2024). Digital tools for neuropsychological evaluation in ADHD: A scoping review. *Frontiers in Psychology, 15*, 892345. doi.org (p. 5)
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., & Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine, 169*(7), 467–473. doi.org (p. 5)
- Trujillo González, L. M., Tobón Arango, J. M., Barrera Torres, M. Y., Beltrán Gómez, S. S., Hernández Sampayo, S. S., Pulgar González, M. P., & Osorio Eljaik, L. M. (2024). TDAH en niños y adultos: Actualización. *Ciencia Latina Revista Científica Multidisciplinar, 8*(4), 1523–1545. doi.org (p. 5)
- Varoquaux, G., Raamana, P. R., Engemann, D. A., Hoyos-Idrobo, A., Schwartz, Y., & Thirion, B. (2017). Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage, 145*, 166–179. doi.org (pp. 5-6)
- Zhao, C., Xu, Y., Li, R., Li, H., & Zhang, M. (2025). Artificial intelligence in ADHD assessment: A comprehensive review of research progress from early screening to precise differential diagnosis. *Frontiers in Artificial Intelligence, 8*, 1624485. doi.org (p. 6)
- Zou, L., Wang, Y., & Liu, H. (2023). Machine learning models for ADHD diagnosis using multimodal data. *IEEE Access, 11*, 15344–15357. doi.org (p. 6)

**Artículos del Corpus de la Revisión (14 Estudios Incluidos)**

- Ahmadi, A., Karimi, S., & Tehrani, M. (2024). Attention-based CNN with CBAM for EEG-based ADHD classification in children. *Biomedical Signal Processing and Control*, *89*, 105731. doi.org (p. 7)
- Atila, U., Ucar, M., & Alcin, O. F. (2023). EEG-based ADHD detection using SVM with discrete wavelet transform features. *Biomedical Signal Processing and Control*, *84*, 104780. doi.org (p. 7)
- Chen, H., Song, Y., & Li, X. (2019). A deep learning framework for identifying children with ADHD using an EEG-based brain network. *Neurocomputing*, *356*, 83–96. doi.org (p. 7)
- Chugh, G., Sharma, S., & Bhattacharya, S. (2024). Hybrid CNN-LSTM model for EEG-based ADHD classification in children. *Expert Systems with Applications*, *238*, 121844. doi.org (p. 7)
- Lee, W., Park, S., & Kim, H. (2023a). Temporal convolutional network for ADHD detection using skeleton data during robot interaction. *IEEE Access*, *11*, 23411–23422. doi.org (p. 7)
- Lee, W., Park, S., & Kim, H. (2023b). Bi-LSTM with attention mechanism for ADHD risk classification using body movement data. *Sensors*, *23*(8), 4102. doi.org (p. 7)
- Lee, W., Park, S., & Kim, H. (2023c). Weighted Bi-LSTM approach for ADHD detection using skeleton-based activity recognition. *Applied Sciences*, *13*(11), 6512. doi.org (p. 7)
- Li, Y., Chen, X., & Wang, Z. (2025). fNIRS-based ADHD diagnosis in children using Random Forest and prefrontal cortex activation during Go/NoGo tasks. *Journal of Neural Engineering*, *22*(1), 016032. doi.org (p. 8)

- Maheshwari, S., Kumar, R., & Sharma, P. (2025). EEG-based ADHD detection using nonlinear features and ensemble classifiers. *Biomedical Signal Processing and Control*, *95*, 106380. doi.org (p. 8)
- Qin, Y., Zhang, J., Liu, H., & Chen, W. (2025). Explainable machine learning for ADHD diagnosis using WISC-IV and behavioral scales: A SHAP-based approach. *Frontiers in Psychiatry*, *16*, 1543821. doi.org (p. 8)
- Weigard, A., Clark, D. A., Sripada, C., & Angstadt, M. (2023). Cognitive ability-based predictors of ADHD symptoms in the ABCD Study. *Journal of Child Psychology and Psychiatry*, *64*(7), 1021–1033. doi.org (p. 8)
- Yamashita, A., Sakai, Y., Yamada, T., Yahata, N., Kunimatsu, A., Okada, N., & Imamizu, H. (2024). Generalizable brain network markers for ADHD subtypes: A multi-site validation study using the ABCD dataset. *Molecular Psychiatry*, *29*, 1085–1096. doi.org (p. 8)
- Yeh, S.-C., Li, Y.-C., Zhou, C., Chiu, P.-H., & Chen, J. W. (2020). Effects of virtual reality and augmented reality on induced anxiety. *IEEE Access*, *8*, 156175–156188. doi.org (p. 8)
- Zhang, D., Liu, Y., & Zhao, S. (2023). ADHD classification using finger-force control tasks with support vector machine in a virtual reality environment. *Frontiers in Human Neuroscience*, *17*, 1149399. doi.org (p. 8)

## Anexos

### Anexo 1: 064-26 UNV CEIFUS 788-26\_Evaluación inicial del protocolo



## COMITÉ DE ÉTICA EN INVESTIGACIÓN

CEIFUS 788-26  
Bogotá D.C, 27 de marzo de 2026

Doctores:

Oscar Gabriel Artunduaga Arias  
Pedro German Rodríguez Castellanos  
Luz Dary Camargo Chaparro  
Ana Gabriela Armeta Vergara  
**Investigador Principal**  
Bogotá

**Ref: Concepto ético de la solicitud vinculada al protocolo (064-26 UNV) Aplicaciones de Inteligencia Artificial en La Evaluación Neuropsicológica del TDAH en Niñas y Niños de 6 A 12 Años: Una Revisión de Alcance.**

Doctores

El día 17/03/2026, en la sesión registrada en el acta No. 011-26, se reunió el Comité de Ética en Investigación de la Fundación Universitaria Sanitas, donde se deja constancia de la recepción del comunicado de 12/03/2026 y evaluación de la siguiente información relacionada con el protocolo en referencia que se desarrolla bajo su dirección en Unisanitas.

- Listado de documentos sometidos: Evaluación inicial protocolo.

**Concepto CEI: Exoneración**

El Comité de Ética en Investigación (CEI) ha revisado el protocolo presentado y lo clasifica como un estudio secundario. En consecuencia, se determina que no requiere evaluación adicional ni aprobación ética, dado que:

1. No implica intervención ni interacción con seres humanos o animales.
2. Se basa en datos previamente recolectados y publicados, sin riesgo de identificación de los participantes.
3. No compromete la confidencialidad, privacidad o integridad de los participantes de estudio.
4. No representa riesgos adicionales para los participantes, en cumplimiento de la Resolución 8430 de 1993 y demás normativas vigentes.

Por lo anterior, el Comité Exonera este estudio del proceso de evaluación ética formal. No obstante, recomienda a los investigadores garantizar el cumplimiento de los principios éticos en el manejo de datos y la protección de la información.

El Comité de Ética en Investigación de la Fundación Universitaria Sanitas está conformado por doce (12) miembros principales y tres (03) miembros suplentes y se requiere la presencia de siete (07) de ellos para cumplir con el quorum. (asistieron: 11).

En el análisis y evaluación del protocolo participaron los siguientes miembros del Comité de Ética en Investigación, y se certifica que ninguno de ellos presenta conflictos de interés con respecto al estudio en referencia.

Eduardo Low Padilla	Presidente	Profesional del área de ciencias de la salud con formación en Farmacología Clínica
Miryam Consuelo Neira Corredor	Miembro Deliberativo	Profesional del área de ciencias de la salud con formación de posgrado en Bioética.
Ingrid Milena Rodríguez Bedoya	Secretaria Ejecutiva	Profesional del área de ciencias de la salud con formación en Epidemiología.

**Comité de Ética en Investigación de la Fundación Universitaria Sanitas**  
Calle 23 # 66-46 Sede Salitre – Teléfono: 5895377 Ext: 5719901  
E-mail: [comiteetica@unisanitas.edu.co](mailto:comiteetica@unisanitas.edu.co)  
Bogotá D. C, Colombia



## COMITÉ DE ÉTICA EN INVESTIGACIÓN

Héctor Ricardo Jiménez	Miembro Deliberativo	Profesional del área de Ciencias Humanas y sociales con formación en Bioética.
Giovanna De Silvestri Torres	Miembro Deliberativo	Representante de la comunidad con formación en control de calidad y auditoría de calidad.
Sofía Muñoz Medina	Miembro Deliberativo	Profesional del área de ciencias de la salud con formación en Epidemiología.
Betsy Valle Velasco	Miembro Deliberativo	Químico Farmacéutico
Melissa Bazante Escobar	Miembro suplente	Profesional del área de Ciencias Humanas y Sociales con formación en Derecho
David Alberto Rincón Valenzuela	Miembro Deliberativo	Profesional del área de Ciencias de la Salud con formación en Epidemiología
Andrea Juliana Uribe Rodríguez	Miembro Deliberativo	Profesional en Ciencias Económicas, Administrativas y Contables.
Angela Patricia Brijaldo Villamizar	Miembro Deliberativo	Profesional del área de ciencias de la salud con formación en Biología Molecular y Biotecnología.

El Comité de Ética en Investigación de la Fundación Universitaria Sanitas declara que el desarrollo de sus actividades se rige bajo la normatividad vigente en temas relacionados con investigación en salud, (Ley Colombiana Resolución No 8430 de 1993 del Ministerio de Salud, Resolución 2378 de 2008 del Ministerio de Protección Social, Ley 1581 de 2012 de protección de datos personales). Las Normas de Buenas Prácticas de Investigación Clínica (Good Clinical Practice-GCP), la Declaración de Helsinki, Finlandia octubre 2024 y la normativa Internacional vigente.

Cordialmente,

**Eduardo Low Padilla**  
**Presidente CEI**  
 Comité de Ética en Investigación  
 Fundación Universitaria Sanitas

**Comité de Ética en Investigación de la Fundación Universitaria Sanitas**  
 Calle 23 # 66-46 Sede Salitre – Teléfono: 5895377 Ext: 5719901  
 E-mail: [comiteetica@unisanitas.edu.co](mailto:comiteetica@unisanitas.edu.co)  
 Bogotá D. C, Colombia