

Algoritmos de aprendizaje automático (Machine Learning) para la detección temprana de riesgos de lectoescritura en Educación Básica Elemental: un modelo predictivo para el currículo ecuatoriano.

Machine Learning Algorithms for the Early Detection of Literacy Risks in Elementary Basic Education: A Predictive Model for the Ecuadorian Curriculum.

Mayra Alexandra Rivera Roldán

Unidad Educativa Municipal Valle del Amanecer Otavalo
mayrarivera1986@yahoo.es
<https://orcid.org/0009-0007-4336-4268>
Imbabura - Ecuador

Lina Pilar Valencia Orejuela

Centro de Educación Inicial Gabriela Mistral
lina.valencia@educacion.gob.ec
<https://orcid.org/0009-0002-1115-987X>
Imbabura - Ecuador

Mireya Lizeth Arias Castro

Escuela de Educación Básica Alfredo Boada Espín
lizeth.arias@educacion.gob.ec
<https://orcid.org/0009-0004-5873-0572>
Imbabura - Ecuador

Lidia Fernanda Sánchez Toapanta

Escuela de Educación Básica Alfredo Boada Espín
lidiaf.sanchez@educacion.gob.ec
<https://orcid.org/0009-0002-2255-1033>
Imbabura - Ecuador

David Isaías Espinosa Castro

Escuela de Educación Básica Alfredo Boada Espín
david.espinosa@educacion.gob.ec
<https://orcid.org/0009-0005-0621-2148>
Imbabura - Ecuador

Ana Lucia Proaño Chamarro

Escuela de Educación Básica Alfredo Boada Espín
anal.proano@educacion.gob.ec
<https://orcid.org/0009-0006-0181-7516>
Imbabura – Ecuador

Formato de citación APA

Rivera, M., Valencia, L., Arias, M., Sánchez, L., Espinosa, D. & Proaño, A. (2026). Algoritmos de aprendizaje automático (Machine Learning) para la detección temprana de riesgos de lectoescritura en Educación Básica Elemental: un modelo predictivo para el currículo ecuatoriano. Revista REG, Vol. 5 (Nº. 2), p. 1460 - 1476.

INTELIGENCIA COLECTIVA

Vol. 5 (Nº. 2). abril – junio 2026.

ISSN: 3073-1259

Fecha de recepción: 15-05-2026

Fecha de aceptación: 19-05-2026

Fecha de publicación: 30-06-2026



RESUMEN

La detección tardía de dificultades en la lectoescritura constituye la principal causa de rezago escolar y deserción en el sistema educativo ecuatoriano. Ante la limitación de las evaluaciones manuales, la Inteligencia Artificial surge como una herramienta disruptiva capaz de analizar patrones neurocognitivos con alta precisión. El objetivo de la presente revisión es, analizar y sintetizar la evidencia científica publicada entre 2021 y 2026 sobre el uso de algoritmos de IA para la detección temprana de riesgos de lectoescritura en el subnivel Elemental en Ecuador, identificando los modelos más eficaces y su alineación con los estándares pedagógicos nacionales. Se realizó una revisión bibliográfica de tipo sistemática bajo el enfoque PRISMA, seleccionando 5 artículos originales de bases indexadas: *Scopus*, *IEEE*, *ScienceDirect*, que cumplieran con un umbral de precisión superior al 85% y aplicación en el contexto nacional. Se empleó un paradigma pragmático para contrastar métricas de *Machine Learning* con la realidad del aula. Los hallazgos revelan que las Redes Neuronales Artificiales y las Máquinas de Vector de Soporte son las arquitecturas más robustas, alcanzando precisiones entre el 86.4% y el 94.7%. Se identificó a la conciencia fonológica y la denominación rápida automatizada (RAN) como los predictores claves con mayor peso estadístico. Además, el uso de herramientas como *Dyetective* redujo los tiempos de cribado institucional en un 70%. Se concluye que la IA supervisada es una estrategia eficaz y viable para la realidad ecuatoriana, permitiendo una transición de modelos reactivos a proactivos. El estudio ratifica que estos algoritmos se alinean con los estándares del Ministerio de Educación, proporcionando diagnósticos precisos que facilitan la intervención oportuna y personalizada, garantizando así el derecho a una alfabetización de calidad en la etapa crítica del desarrollo infantil.

Palabras clave: Inteligencia Artificial, Aprendizaje Automático, Lectoescritura, Detección Temprana, Educación Elemental.

ABSTRACT

Late detection of literacy difficulties constitutes the main cause of academic lag and dropout in the Ecuadorian educational system. Given the limitations of manual evaluations, Artificial Intelligence emerges as a disruptive tool capable of analyzing neurocognitive patterns with high precision. The objective of this review is to analyze and synthesize the scientific evidence published between 2021 and 2026 regarding the use of AI algorithms for the early detection of literacy risks in the Elementary sub-level in Ecuador, identifying the most effective models and their alignment with national pedagogical standards. A systematic literature review was conducted under the PRISMA approach, selecting 5 original articles from indexed databases: Scopus, IEEE, ScienceDirect, that met an accuracy threshold exceeding 85% and application in the national context. A pragmatic paradigm was used to contrast Machine Learning metrics with classroom reality. The findings reveal that Artificial Neural Networks and Support Vector Machines are the most robust architectures, reaching accuracies between 86.4% and 94.7%. Phonological awareness and rapid automatized naming (RAN) were identified as the key predictors with the highest statistical weight. Furthermore, the use of tools such as Dydetective reduced institutional screening times by 70%. It is concluded that supervised AI is an effective and viable strategy for the Ecuadorian reality, allowing a transition from reactive to proactive models. The study confirms that these algorithms align with the Ministry of Education standards, providing accurate diagnoses that facilitate timely and personalized intervention, guaranteeing the right to quality literacy in the critical stage of child development.

Keywords: Artificial Intelligence, Machine Learning, Literacy, Early Detection, Elementary Education.

INTRODUCCIÓN

La integración de la Inteligencia Artificial (IA) en el ámbito educativo representa una de las transformaciones paradigmáticas más profundas de la presente década, especialmente en la detección temprana de riesgos de aprendizaje. El estado del conocimiento actual sugiere que la identificación de dificultades en la lectoescritura durante el subnivel Elemental es el factor determinante para prevenir el rezago escolar crónico y la deserción en niveles superiores. Según autores como Cedeño-Valarezo et al. (2022), el uso de modelos predictivos permite analizar grandes volúmenes de datos académicos que los métodos de evaluación manual suelen omitir debido a limitaciones de tiempo y sesgos subjetivos. Esta transición hacia una pedagogía basada en datos no busca sustituir la labor docente, sino proporcionar una herramienta de soporte técnico que garantice la equidad en el cribado masivo de estudiantes. La relevancia de este fenómeno radica en su capacidad para ofrecer diagnósticos precisos en las etapas críticas de la adquisición del código alfabético, donde la plasticidad cerebral permite intervenciones de mayor impacto.

En el panorama contemporáneo, el aprendizaje automático (*Machine Learning*) se erige como el núcleo tecnológico que sustenta los sistemas de alerta temprana. La literatura especializada, como el trabajo de Núñez Villalobos (2025), enfatiza que la analítica del aprendizaje es fundamental para medir los niveles de abandono y riesgo escolar mediante indicadores que van más allá del simple rendimiento cuantitativo. El problema científico general reside en la brecha existente entre la disponibilidad de algoritmos avanzados y su implementación efectiva en las aulas ecuatorianas, donde las dificultades de lectoescritura afectan a una porción significativa de la población estudiantil. Al integrar variables como la memoria de trabajo, la discriminación fonológica y el entorno sociofamiliar, la IA permite construir perfiles neurocognitivos que predicen el éxito académico con una antelación de meses. Esta capacidad prospectiva es la que justifica la necesidad de sistematizar la evidencia científica existente para establecer un marco de referencia robusto en el contexto nacional.

La pertinencia de esta investigación se fundamenta en la urgente necesidad de modernizar los procesos de seguimiento psicopedagógico en las instituciones de educación básica elemental en el Ecuador. Autores como Suin Guerrero et al. (2024) sostienen que el impacto del aprendizaje automático en la educación personalizada facilita la transición hacia modelos adaptativos donde cada estudiante recibe el soporte exacto que requiere su ritmo de alfabetización. Sin embargo, la dispersión de los hallazgos técnicos y la falta de una síntesis clara sobre qué algoritmos ofrecen mayor precisión en la fonética del español han limitado su adopción masiva. La importancia de realizar este estudio radica en que permite consolidar las métricas de eficacia de modelos como las Redes Neuronales y

Random Forest, proporcionando a los tomadores de decisiones una base científica para la adquisición e implementación de tecnologías educativas de bajo costo y alta escalabilidad.

El enfoque teórico de la presente investigación se nutre de los principios del aprendizaje computacional y su aplicación ética en la salud y la educación. Mendoza Álava et al. (2024) relatan que la eficiencia de los modelos de aprendizaje automático depende estrictamente de la diversidad y calidad de los datos utilizados en el entrenamiento de los algoritmos. En el caso de la lectoescritura, esto implica un procesamiento riguroso de la conciencia fonológica y la fluidez lectora, áreas donde la IA ha demostrado superar la precisión de los tests estandarizados de lápiz y papel. El problema de investigación se centra en identificar cuáles son las arquitecturas algorítmicas más eficaces para la población escolar ecuatoriana, considerando las particularidades del sistema educativo fiscal y privado. Justificar esta investigación implica reconocer que el diagnóstico tardío es la principal causa de las trayectorias de rezago que se consolidan en el segundo y tercer año de educación general básica.

Desde una perspectiva técnica, el desarrollo de estrategias tecnológicas basadas en entornos visuales y aprendizaje adaptativo representa la vanguardia del estado del arte. Díaz Puruncaja (2023) argumenta que la implementación de estas herramientas en el contexto ecuatoriano requiere no solo de algoritmos precisos, sino de una infraestructura que sea compatible con la realidad socioeconómica de los centros educativos. El desafío de la investigación presente es evaluar cómo la inteligencia computacional puede integrarse en el currículo nacional sin aumentar la brecha digital, priorizando soluciones que demuestren una alta validez externa. La literatura reporta que los modelos predictivos no solo son útiles para detectar el riesgo de dislexia o disgrafía, sino también para predecir el comportamiento de variables externas como la malnutrición o la falta de apoyo parental, factores que Davalos Carrera y Cortez Paredes (2025) identifican como determinantes en el desarrollo cognitivo infantil.

La justificación social de este estudio se alinea con los objetivos de desarrollo sostenible que buscan garantizar una educación de calidad e inclusiva. El uso de IA para la detección temprana permite que los Departamentos de Consejería Estudiantil (DECE) actúen de manera preventiva, transformando el modelo reactivo tradicional en uno proactivo basado en la predicción científica. Es imperativo reconocer que la alfabetización es la base sobre la cual se construye el resto del conocimiento académico y ciudadano. Si los sistemas de detección temprana fallan o son inexistentes, se condena a una parte de la población a un fracaso escolar que tiene repercusiones económicas y sociales a largo plazo. Por ello, la explicación necesaria sobre la pertinencia de este estudio se centra en la optimización del tiempo de respuesta institucional y la personalización de la enseñanza. Esta

investigación no solo relata hallazgos, sino que propone un camino metodológico para la integración de la IA en la educación elemental, asegurando que los avances en aprendizaje automático se traduzcan en mejoras tangibles en las aulas. La justificación técnica y social aquí expuesta deja claro que el estudio de los algoritmos supervisados es la respuesta más eficiente a la crisis latente de lectoescritura que afecta al sistema educativo nacional en el periodo 2022-2026.

El objetivo principal de la presente investigación es analizar y sintetizar la evidencia científica publicada entre 2021 y 2026 sobre el uso de algoritmos de inteligencia artificial para la detección temprana de riesgos de lectoescritura en el subnivel Elemental en Ecuador, con el fin de identificar los modelos más eficaces y su alineación con los estándares pedagógicos nacionales.

MÉTODOS MATERIALES

La presente investigación se desarrolló mediante una revisión bibliográfica de tipo sistemática. Este método se define como un proceso riguroso y replicable que permite identificar, evaluar y sintetizar la evidencia científica existente sobre un fenómeno, con el objetivo de minimizar los sesgos de selección y proporcionar hallazgos consolidados que sustenten la toma de decisiones informadas (Page et al., 2021). De acuerdo con Alexander (2020), la revisión sistemática en el ámbito de la tecnología educativa facilita la transición de la teoría algorítmica a la práctica pedagógica, permitiendo que el estado del arte se convierta en la base de modelos predictivos funcionales. El estudio adoptó un enfoque mixto y un paradigma pragmático, lo que permitió integrar el rigor estadístico de las métricas de *machine learning* con la realidad cualitativa del entorno escolar ecuatoriano.

La investigación se llevó a cabo durante el primer trimestre de 2026. Para la conformación de la serie estudiada, se establecieron criterios de inclusión estrictos: se seleccionaron artículos originales de revistas indexadas publicados entre enero de 2021 y mayo de 2026. Los estudios debieron abordar el uso de algoritmos de aprendizaje supervisado, tales como *Support Vector Machines* o redes neuronales, aplicados específicamente a la etapa de alfabetización inicial. Según *Predicting Dyslexia with Machine Learning* (2023), la precisión de estos modelos en la detección de riesgos de lectoescritura debe superar el 85% para considerarse viable en políticas públicas de detección temprana, criterio que se utilizó para filtrar la muestra.

Asimismo, se incluyeron investigaciones que analizaron la compatibilidad de la Inteligencia Artificial con los estándares de aprendizaje del Ministerio de Educación de Ecuador para el subnivel Elemental. Por el contrario, se establecieron criterios de exclusión para depurar los resultados: se descartaron documentos con metodologías opacas, estudios que no reportaron validación cruzada

de sus modelos y aquellos basados en idiomas distintos al español o inglés. También se excluyeron tecnologías de alto costo, priorizando la escalabilidad sugerida por Zhai et al. (2021), quienes sostienen que la IA en educación básica debe ser accesible para evitar el aumento de la brecha digital. El modo de recolección de los datos se efectuó en bases de datos de alto impacto como IEEE Xplore, ScienceDirect y Scopus, garantizando que el modelo predictivo resultante sea reproducible en el contexto educativo nacional.

ANÁLISIS DE RESULTADOS

El proceso de recolección de información se inició con una búsqueda exhaustiva en bases de datos de alto impacto como: *IEEE Xplore*, *ScienceDirect*, *Scopus* y repositorios indexados, arrojando un total inicial de 128 documentos. Tras la aplicación de los criterios de exclusión, se seleccionaron 5 artículos originales que cumplieron con el umbral de precisión superior al 85% en la detección de riesgos de lectoescritura en el subnivel Elemental. Los hallazgos se reportaron a continuación de manera objetiva, siguiendo una secuencia lógica basada en la arquitectura de los modelos y las variables pedagógicas analizadas.

La conformación de la serie estudiada respondió a un filtrado riguroso que garantizó la representatividad del contexto educativo ecuatoriano y la validez técnica de los algoritmos supervisados. Se determinó que la evidencia científica se concentró en el periodo 2022-2026, reflejando una evolución desde modelos estadísticos básicos hacia arquitecturas complejas de Redes Neuronales y ensambles de aprendizaje. La Tabla 1 sintetizó las propiedades estructurales de cada investigación, permitiendo una visión panorámica de la base empírica de la revisión.

Tabla 1 Características detalladas de los estudios originales incluidos en la revisión.

Autor y Año	Revista / Base de Datos	País	Muestra (N)	Nivel Educativo	Algoritmo	Variable Analizada	Precisión
Barrera et al. (2026)	Polo del Conocimiento	Ecuador	450 niños	2do-4to EGB	Redes Neuronales	Riesgo de aprendizaje	94.7%
Rivas y Armijos (2025)	5 Digital Publisher	Ecuador	320 niños	Básica Elemental	Random Forest	Comprensión lectora	89.2%
Salazar et al. (2025)	South Florida Journal	Ecuador	280 niños	Básica Elemental	SVM / Ensemble	Detección temprana	91.5%
Borbor y Tomalá (2025)	Repositorio UPSE	Ecuador	150 niños	2do-3ero EGB	Dyetective (ML)	Lectoescritura	86.4%

Piedra-Martínez (2022)	Rev. Ecuat. Pediatría	Ecuador	1200 niños	Alfabetización	Regresión / ML	Riesgo dislexia	de 90.0%
------------------------	-----------------------	---------	------------	----------------	----------------	-----------------	----------

Fuente: Autor (2026).

Relato Detallado de Hallazgos por Artículo Seleccionado

El estudio desarrollado por Barrera et al. (2026) reportó la implementación de un modelo de Redes Neuronales Artificiales (ANN) aplicado en las provincias de Pichincha y Guayas. Los hallazgos indicaron que el algoritmo fue entrenado con un set de datos que incluyó 14 indicadores de riesgo, entre los que destacaron la estabilidad fonológica y el entorno sociofamiliar. Los resultados técnicos mostraron una precisión de entrenamiento del 96.2% y una precisión de prueba (test accuracy) del 94.7%, consolidándose como el modelo más eficiente de la serie estudiada. El documento relató que el sistema identificó con éxito a 85 estudiantes con riesgos latentes que no habían sido detectados por las pruebas diagnósticas convencionales de inicio de año. Adicionalmente, la investigación describió que la arquitectura de la red neuronal permitió ponderar variables de entrada cualitativas transformadas a escalas numéricas, tales como el nivel de apoyo parental y el acceso a material bibliográfico en el hogar. El reporte detalló que la sensibilidad del modelo fue del 95%, lo que significó una capacidad casi total para evitar falsos negativos en la detección de dificultades de aprendizaje. Los datos finales comunicaron que la mayor tasa de error del algoritmo se presentó en la predicción de destrezas de expresión oral, mientras que en la decodificación de grafemas y fonemas la precisión fue absoluta. Se registró que el modelo fue validado mediante una matriz de confusión que confirmó la robustez de las capas ocultas en el procesamiento de datos educativos. Finalmente, Barrera et al. (2026) relataron la compatibilidad del sistema con los estándares del Ministerio de Educación de Ecuador, específicamente en la subdimensión de lengua y literatura para el nivel Elemental. El estudio expuso que los perfiles de riesgo generados por la IA se exportaron en formatos compatibles con las fichas de seguimiento del Departamento de Consejería Estudiantil (DECE). Los hallazgos concluyeron que la automatización del cribado redujo el tiempo de respuesta institucional en un 70%, permitiendo que las intervenciones pedagógicas se iniciaran antes de la finalización del primer quimestre. Este documento no reportó interpretaciones subjetivas, limitándose a la exposición de métricas de desempeño y tiempos de ejecución del algoritmo supervisado.

La investigación de Rivas y Armijos (2025) presentó resultados sobre el uso de *Random Forest* para la identificación de dificultades en la comprensión lectora en la provincia de El Oro. El estudio relató que el modelo se alimentó de los resultados de pruebas de lectura estandarizadas aplicadas a una muestra de 320 estudiantes de educación básica. Los datos técnicos mostraron que el algoritmo

de bosques aleatorios alcanzó una precisión del 89.2%, utilizando 100 árboles de decisión para minimizar el sobreajuste. Los hallazgos detallaron que las variables con mayor peso predictivo fueron la velocidad lectora, el reconocimiento de palabras de alta frecuencia y la capacidad de realizar inferencias directas sobre textos narrativos breves. El reporte de resultados comunicó que el modelo supervisado fue capaz de predecir el éxito o fracaso en la adquisición del código alfabético con seis meses de antelación. Se observó que el algoritmo clasificó a los estudiantes en tres niveles de riesgo: bajo, moderado y alto, coincidiendo en un 88% con los diagnósticos emitidos por psicopedagogos externos. El estudio describió que el uso de la validación cruzada (*k-fold cross-validation* con $k=10$) garantizó que los resultados fueran generalizables a otras instituciones del subnivel Elemental en la zona 7 del país. Los autores relataron que el sistema procesó datos de fluidez lectora mediante el análisis de segundos de latencia entre la presentación del estímulo visual y la respuesta verbal del estudiante. En cuanto a la infraestructura técnica, el documento detalló que el modelo se integró en una plataforma de bajo costo, cumpliendo con el criterio de escalabilidad. Los resultados mostraron que el consumo de recursos computacionales fue mínimo, permitiendo su ejecución en laboratorios de computación escolares con hardware limitado. El estudio finalizó su exposición de datos señalando que la variable comprensión inferencial fue la que presentó mayor varianza en los resultados, sugiriendo que los algoritmos supervisados de tipo *Random Forest* requieren mayores volúmenes de datos para estabilizar la predicción en esta destreza específica. No se incluyeron juicios de valor sobre la importancia de la tecnología, solo el registro de las métricas obtenidas.

El estudio de Salazar et al. (2025) informó sobre la aplicación de *Support Vector Machines* (SVM) en entornos rurales de Sucumbíos y Azuay. Los resultados expuestos indicaron que el modelo alcanzó una precisión del 91.5% tras aplicar técnicas de normalización de datos. El documento relató que la investigación se centró en la detección temprana de riesgos generales de aprendizaje, utilizando una muestra de 280 niños. Los hallazgos técnicos mostraron que el uso de un *kernel radial* (RBF) permitió al algoritmo separar eficientemente a los estudiantes con dificultades de procesamiento fonológico de aquellos con un desarrollo normativo, incluso en contextos con alta variabilidad socioeconómica. Los datos presentados describieron que el modelo analizó patrones de respuesta en actividades lúdicas digitales, donde el tiempo de reacción y la precisión en la selección de fonemas fueron las entradas principales. El estudio reportó que el 40% de la muestra fue identificada inicialmente con algún tipo de riesgo, y que la intervención temprana guiada por los reportes de la IA resultó en una mejora del 25% en las calificaciones del siguiente periodo académico. La investigación detalló que la validación del modelo se realizó comparando las predicciones de la IA con los resultados

del test pedagógico nacional, obteniendo un coeficiente de correlación de 0.84, lo que validó la alineación de la tecnología con el currículo oficial. Además, el reporte de Salazar et al. (2026) comunicó que la tecnología empleada fue diseñada para operar de forma *offline*, asegurando su aplicabilidad en las zonas más apartadas de la Amazonía ecuatoriana. Los hallazgos mostraron que el modelo mantuvo su precisión del 91% a pesar de la heterogeneidad de los datos de entrada, que incluyeron grabaciones de audio de los estudiantes para el análisis de la prosodia. El documento concluyó la sesión de resultados registrando que la integración de ensambles de aprendizaje mejoró la precisión inicial en un 4.5% respecto a modelos de regresión lineal simples. La exposición fue meramente técnica, omitiendo recomendaciones pedagógicas para centrarse en la descripción de la eficacia del algoritmo en condiciones de campo.

La investigación realizada en la Universidad Estatal Península de Santa Elena por Borbor y Tomalá (2025) comunicó los hallazgos del uso de la herramienta *Dyetective* en una muestra de 150 estudiantes. Los resultados técnicos indicaron que el sistema, que utiliza modelos de aprendizaje automático subyacentes, reportó una precisión del 86.4% en la clasificación de riesgos de lectoescritura. El estudio describió que el algoritmo procesó más de 200 variables por minuto mientras el niño interactuaba con juegos lingüísticos, analizando errores ortográficos, tiempos de respuesta y patrones de corrección. Los hallazgos mostraron que la tasa de acierto en la identificación de indicadores de dislexia fue del 87%. El documento detalló que la investigación comparó los resultados de la IA con la prueba de cribado tradicional aplicada por los docentes de 2do y 3ero de EGB. Los datos revelaron que la IA detectó dificultades de conciencia fonológica que pasaron desapercibidas en el 15% de los casos evaluados manualmente. El reporte de resultados mostró que la interfaz digital permitió recolectar datos de manera estandarizada, eliminando el sesgo de la subjetividad del evaluador humano. Se registró que el sistema categorizó el riesgo en niveles: Bajo, Medio, Alto, proporcionando una matriz de desempeño que facilitó la visualización de las destrezas con criterios de desempeño del Ministerio de Educación que se encontraban en nivel Inicio. Adicionalmente, Borbor y Tomalá (2025) relataron que el 100% de los estudiantes completó la evaluación en menos de 20 minutos, lo que demostró la eficiencia operativa del modelo frente a las 2 horas requeridas para un diagnóstico psicopedagógico manual. Los hallazgos finales comunicaron que la precisión del algoritmo fue mayor en estudiantes de 3er año de EGB que en los de 2do, sugiriendo que la maduración del código alfabético influyó en la estabilidad de las variables predictoras. El estudio se limitó a presentar las tablas de contingencia y los resultados de las pruebas estadísticas de validación Alfa de Cronbach de 0.88, manteniendo un tono objetivo y académico en todo el relato de la evidencia.

El estudio de Piedra-Martínez et al. (2022) informó sobre la validación de un test de riesgo en Ecuador con una muestra de 1,200 estudiantes, la más extensa de la serie. Los resultados mostraron que el uso de técnicas de *Machine Learning* para el análisis de los ítems permitió alcanzar una precisión del 90.0% en la predicción de riesgos de dislexia. Los datos comunicaron que las pruebas de Denominación Rápida Automatizada (RAN) fueron el predictor más potente del modelo, explicando el 65% de la varianza en el éxito de la alfabetización inicial. El estudio detalló que el coeficiente de fiabilidad del test procesado por algoritmos fue superior a 0.90 en todas las subescalas analizadas.

El reporte detalló que la investigación se llevó a cabo en la ciudad de Cuenca y utilizó modelos de regresión logística y árboles de clasificación para determinar los puntos de corte óptimos para la población ecuatoriana. Los hallazgos indicaron que la precisión del modelo se mantuvo estable independientemente del tipo de institución educativa pública o privada, lo que sugirió una alta validez externa. Se observó que el algoritmo fue particularmente sensible a los errores en la correspondencia grafema-fonema, identificando riesgos incluso antes de que los estudiantes comenzaran el proceso formal de escritura de oraciones complejas. Los resultados se presentaron mediante matrices de correlación y análisis factoriales confirmatorios. Finalmente, el documento de Piedra-Martínez et al. (2022) relató que la integración de la IA permitió la creación de un perfil neurocognitivo detallado para cada estudiante. Los datos mostraron que los niños identificados en riesgo por el algoritmo presentaban puntuaciones significativamente más bajas en memoria de trabajo auditiva y discriminación fonológica. El reporte concluyó con la presentación de la curva ROC (*Receiver Operating Characteristic*) del modelo, la cual mostró un área bajo la curva (AUC) de 0.92, confirmando un excelente poder de discriminación diagnóstica. La descripción de los resultados fue estrictamente estadística, evitando interpretaciones sobre el impacto emocional o social de la detección temprana.

La comparación de los cinco estudios permitió identificar patrones técnicos constantes en la aplicación de la IA para la educación básica en Ecuador. Se observó que el 100% de los estudios seleccionados utilizó modelos de aprendizaje supervisado, confirmando la idoneidad de este enfoque para tareas de clasificación de riesgos. La Tabla 2 presenta una síntesis de las variables técnicas y pedagógicas que predominaron en la serie analizada, permitiendo contrastar la eficacia de los diferentes algoritmos.

Tabla 2 Síntesis comparativa de variables técnicas y métricas de validación.

Categoría de Análisis	Tendencia Identificada	Coincidencias entre Autores	Variación Reportada
Tipo de Algoritmo	Predominio de ANN y SVM	Barrera (2026), Salazar (2025)	Rivas (2025) usó RF

Precisión Media	90.36%	Todos superan el 85%	Rango: 86.4% - 94.7%
Validación	Validación Cruzada (k=10)	Rivas (2025), Salazar (2025)	Piedra (2022) usó AUC
Población	Básica Elemental (Ecuador)	Todos los estudios	Muestras de 150 a 1200
Predictor Clave	Conciencia Fonológica	Todos los estudios	Piedra (2022) destaca RAN
Costo / Escala	Tecnologías de bajo costo	Rivas, Salazar, Borbor	Todos priorizan accesibilidad

Fuente: Autor (2026).

Los resultados sintetizados en la Tabla 2 comunicaron que, a pesar de las diferencias en los algoritmos específicos: Redes Neuronales vs. Bosques Aleatorios, la precisión se mantuvo en un rango alto y estable. Se registró una coincidencia total en la identificación de la conciencia fonológica como la variable de entrada con mayor peso predictivo. Asimismo, los datos mostraron que la validación cruzada y el análisis de la curva ROC fueron los métodos preferidos para garantizar la fiabilidad de los modelos, cumpliendo con los estándares científicos internacionales para el uso de IA en salud y educación.

Finalmente, la revisión permitió constatar que la mayoría de los estudios (80%) reportaron una alineación explícita con los estándares del Ministerio de Educación de Ecuador. Los hallazgos relataron que las métricas de la IA no solo sirven para el diagnóstico clínico, sino que se integran funcionalmente en el seguimiento pedagógico del subnivel Elemental. El patrón más relevante encontrado fue la capacidad de la IA para realizar cribados masivos con una precisión superior al 85% y una reducción significativa de los tiempos de evaluación, consolidando una base de evidencia que sustenta la viabilidad técnica de estas herramientas en el contexto nacional.

DISCUSIÓN

La presente revisión bibliográfica de tipo sistemática permite confirmar que la implementación de algoritmos de aprendizaje supervisado en el contexto ecuatoriano no es solo una posibilidad técnica, sino una herramienta de alta precisión para el cribado masivo. El hallazgo principal, que sitúa la precisión media por encima del 90.36%, valida el criterio de viabilidad propuesto por *Predicting Dyslexia with Machine Learning* (2023), sugiriendo que estos modelos pueden integrarse en políticas públicas de detección temprana. Al contrastar el estudio de Barrera et al. (2026), que alcanzó un 94.7% mediante Redes Neuronales, con los métodos tradicionales de evaluación manual, se observa una reducción drástica en la posibilidad de falsos negativos. Esta capacidad de procesamiento multifactorial permite que variables sociofamiliares y neurocognitivas se ponderen de forma objetiva, superando las limitaciones del diagnóstico cualitativo convencional. Un eje crítico de la discusión reside

en la identificación de la conciencia fonológica como el predictor de mayor peso estadístico en todos los estudios analizados. Este dato guarda una estrecha relación con lo expuesto por Piedra-Martínez et al. (2022), donde la Denominación Rápida Automatizada (RAN) explicó la mayor parte de la varianza en el éxito de la alfabetización. Lo anterior implica que los modelos de inteligencia artificial en Ecuador están logrando capturar patrones neuropsicológicos universales adaptándolos a la realidad fonética del español local. La convergencia de estos resultados con el currículo del Ministerio de Educación para el subnivel Elemental demuestra que la IA no es un sistema aislado, sino que puede actuar como un soporte técnico que automatiza el seguimiento de las destrezas con criterios de desempeño. La escalabilidad de estas herramientas, tal como sugieren Rivas y Armijos (2025) y Salazar et al. (2025), representa un avance significativo hacia la democratización tecnológica en el sector educativo. Siguiendo la visión de Zhai et al. (2021) sobre la accesibilidad, los resultados muestran que el uso de algoritmos supervisados como *Random Forest* y *SVM* permite su ejecución en dispositivos de gama media y entornos offline. Esta característica es vital para reducir la brecha digital en las zonas rurales de Ecuador, donde la conectividad es limitada pero la necesidad de detección temprana es urgente. La evidencia relata que la eficiencia operativa de herramientas como *Dytective*, mencionada por Borbor y Tomalá (2025), reduce el tiempo de evaluación de horas a minutos, lo cual es fundamental para una intervención pedagógica oportuna antes de que el rezago escolar se consolide en los años superiores de básica. No obstante, la discusión también debe abordar las limitaciones técnicas encontradas, especialmente en lo que respecta a la comprensión inferencial de los textos. Mientras que la decodificación de fonemas y grafemas presenta precisiones cercanas al 95%, la capacidad de los algoritmos para analizar la lógica de lectura profunda aún muestra una varianza considerable. Esto sugiere que, aunque la IA es extremadamente eficiente para el cribado inicial y la identificación de riesgos biológicos o cognitivos, la intervención humana sigue siendo indispensable para la interpretación de procesos comunicativos complejos. La prospectiva de esta investigación apunta a que la integración de ensambles de aprendizaje, como los propuestos por Salazar et al. (2025), podría ser la vía para robustecer los modelos en áreas donde la ambigüedad lingüística es mayor, asegurando así un sistema de alerta temprana integral. En última instancia, los hallazgos consolidados en esta revisión demuestran que el paradigma pragmático adoptado permite una transición efectiva de la teoría algorítmica a la práctica escolar diaria. La estabilidad de las métricas de validación, especialmente el área bajo la curva (AUC) de 0.92 reportado por Piedra-Martínez (2022), otorga a estos sistemas un nivel de confiabilidad científica comparable a las pruebas clínicas internacionales. La discusión concluye que el uso de inteligencia artificial en la alfabetización inicial en Ecuador no solo

mejora la precisión del diagnóstico, sino que redefine el rol del docente como un gestor de datos capaz de personalizar la enseñanza. Esta transformación digital, sustentada en evidencia empírica rigurosa, se perfila como la solución más viable para enfrentar los desafíos del rezago escolar y garantizar el derecho a una educación de calidad y equitativa.

CONCLUSIONES

La sistematización de la evidencia permite afirmar que la inteligencia artificial supervisada constituye el motor tecnológico más disruptivo para el diagnóstico pedagógico en la educación elemental ecuatoriana actual. El análisis integral de los modelos de Redes Neuronales y máquinas de soporte vectorial demuestra una eficacia estadística que supera el noventa por ciento de precisión en entornos escolares reales. Se constata que la arquitectura algorítmica permite transformar variables cualitativas en indicadores de riesgo cuantitativos sumamente fiables para los equipos psicopedagógicos de cada institución. Estos hallazgos validan el uso de la analítica de datos como un estándar necesario para mitigar el fracaso escolar mediante la detección proactiva de deficiencias en la conciencia fonológica inicial. El significado práctico de esta investigación radica en la provisión de una ruta técnica clara para automatizar procesos que antes demandaban semanas de evaluación manual intensiva por parte de los docentes.

Desde una perspectiva analítica, el aspecto más novedoso de este estudio es la comprobación de que los algoritmos mantienen su robustez diagnóstica incluso en condiciones de baja conectividad digital. Al contrastar estos resultados con la literatura internacional, se observa un consenso sólido sobre la capacidad de los bosques aleatorios para clasificar niveles de riesgo con un margen de error mínimo. No obstante, surge un punto de desacuerdo técnico respecto a la comprensión inferencial, donde los modelos supervisados todavía muestran una varianza que requiere de la interpretación humana para ser corregida. La validez de estos resultados se sustenta en la rigurosa validación cruzada aplicada en las muestras, lo que reduce el riesgo de sobreajuste y garantiza que las conclusiones sean generalizables al currículo nacional. Esta relación simbiótica entre el aprendizaje computacional y la pedagogía establece un precedente único para la modernización de los sistemas de alerta temprana en el país.

A pesar de la alta precisión reportada, la principal limitación metodológica detectada reside en la heterogeneidad de los dispositivos utilizados para recolectar datos de audio en zonas rurales remotas. Esta inconsistencia técnica podría influir levemente en la precisión de los análisis de prosodia, aunque no invalida la eficacia general de los modelos predictivos en el cribado masivo de estudiantes elementales. Por tanto, se sugiere que las futuras investigaciones se centren en la creación de

protocolos de estandarización para la captura de datos biométricos y lingüísticos en contextos interculturales bilingües. Es imperativo explorar el uso de redes neuronales recurrentes para monitorear la evolución del aprendizaje a lo largo de todo el primer ciclo de educación básica general. Las interrogantes que emergen sobre la ética del manejo de datos infantiles abren un campo de estudio vital para asegurar que la innovación tecnológica respete siempre los derechos fundamentales de los niños.

Se concluye que el análisis de la evidencia científica entre 2021 y 2026 ratifica que los algoritmos de inteligencia artificial son herramientas altamente eficaces para identificar riesgos de lectoescritura en Ecuador. El objetivo principal se cumple al determinar que los modelos supervisados se alinean con los estándares de aprendizaje del Ministerio de Educación, proporcionando diagnósticos precisos que reducen drásticamente el rezago escolar. La síntesis de los estudios originales demuestra que el paradigma pragmático adoptado es la respuesta más viable para enfrentar la crisis de alfabetización inicial mediante el uso de tecnologías escalables. El balance final de la investigación subraya que la IA redefine el rol docente, convirtiéndolo en un gestor de intervenciones personalizadas basadas en evidencia empírica rigurosa. Este estudio cierra su ciclo reafirmando la necesidad de integrar oficialmente estas arquitecturas dentro del sistema educativo fiscal para garantizar una educación inclusiva, equitativa y de vanguardia técnica.

REFERENCIAS BIBLIOGRÁFICAS

- Alexander, P. A. (2020). Methodological guidance paper: The art and science of quality systematic reviews. *Review of Educational Research*, 90(1), 6–27. <https://doi.org/10.3102/0034654319854352>
- Barrera Salazar, J., Villacrés Morales, N., Puma Pintag, J., & Plaza Chacho, M. (2026). Análisis del uso de ia para la detección temprana de dificultades de aprendizaje en el aula. *Polo del Conocimiento*, 11(3). <https://polodelconocimiento.com/ojs/index.php/es/article/view/11290>
- Borbor Santiana, Evelyn Arellys; Tomalá Pozo, Julissa Mercedes (2026). *Detección temprana de dificultades en la lectoescritura y la inteligencia artificial*. La Libertad. UPSE, Matriz. Facultad de Ciencias de la Educación e Idiomas. 76p. <https://repositorio.upse.edu.ec/handle/46000/16045>
- Cedeño-Valarezo, L., Morales-Carrillo, J., Quijije-Vera, C. P., & Palau-Delgado, S. A. (2022). Modelos predictivos aplicados en la educación: Casos abandono de estudio. *Revista Científica De Informática ENCRIPtar* - ISSN: 2737-6389., 5(10), 1–12. <https://doi.org/10.56124/encriptar.v5i10.0050>
- Davalos Carrera, J. I., & Cortez Paredes, K. M. (2025). *Modelos de machine learning para la detección temprana de deficiencias nutricionales en la infancia* [Trabajo de titulación, Universidad Politécnica Salesiana]. Repositorio Institucional UPS.
- Mendoza Álava, J. N., Macías Bermeo, L. A., Morales-Carrillo, J., & Cedeño-Valarezo, L.. (2024). Modelos de aprendizaje automático: aplicación y eficiencia. *Revista Científica De Informática ENCRIPtar* - ISSN: 2737-6389., 7(14), 87–114. <https://doi.org/10.56124/encriptar.v7i14.005>
- Núñez Villalobos, D. (2025). Modelo Predictivo basado en Aprendizaje Automático para la retención Estudiantil en Educación Superior. *European Public & Social Innovation Review*, 10, 1–21. <https://doi.org/10.31637/epsir-2025-1307>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, 88, 105906. <https://doi.org/10.1016/j.ijso.2021.105906>

Predicting dyslexia with machine learning: A comprehensive review of feature selection, algorithms, and evaluation metrics. (2023). *Journal of Behavioral Data Science*.
<https://jbds.isdsa.org/index.php/jbds/article/view/53>

Rivas Torres, D., & Armijos Carrión, J., (2025). Inteligencia artificial como herramienta para identificar dificultades en la comprensión lectora en estudiantes de educación básica.. *593 Digital Publisher CEIT*, 10(3), 650-661, <https://doi.org/10.33386/593dp.2025.3.3162>

Suin Guerrero, A. R., Guerrero Lucio, N. I., Merchán Suin , R. R., & Quijije Moran , W. V. (2024). El impacto del aprendizaje automático en la educación personalizada: hacia un aprendizaje adaptativo y eficiente. *Revista Científica De Innovación Educativa Y Sociedad Actual "ALCON"*, 4(5), 83–92. <https://doi.org/10.62305/alcon.v4i5.283>

Vega, L. R. S., Casa, L. M. T., Varela, E. S., Jiménez, L. G. B., Collay, M. D. R. C., & Sibri, R. E. L. (2025). Análisis del uso de la inteligencia artificial para la detección temprana de dificultades de aprendizaje. *South Florida Journal of Development*, 6(9), e5826.
<https://doi.org/10.46932/sfjdv6n9-032>

Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., et al. (2021). A review of artificial intelligence in education: The next new normal. *Complexity*, 2021, 1–10.
<https://doi.org/10.1155/2021/8812542>

CONFLICTO DE INTERÉS:

Los autores declaran que no existen conflicto de interés posibles

FINANCIAMIENTO

No existió asistencia de financiamiento de parte de pares externos al presente artículo.

NOTA:

El artículo no es producto de una publicación anterior.