



Modelos de clasificación para la detección temprana de TEA en Niños usando Datos del Q-CHAT-10

Classification models for early detection of ASD in children using data from the Q-CHAT-10

Karen Monroy Mendoza (Autor Corresponsal)

karenmonroy1601@gmail.com

ORCID: 0009-0008-5629-610X

Universidad Autónoma del Estado de Hidalgo, Tlahuelipan, México

Gustavo Padrón Rivera

gustavo_padron@uaeh.edu.mx

ORCID: 0000-0002-1955-3989

Universidad Autónoma del Estado de Hidalgo, Tlahuelipan, México

Aceptación: 19 de agosto de 2025

Publicación: 30 de septiembre de 2025

Resumen

El Trastorno del Espectro Autista (TEA) constituye un desafío para la detección temprana, especialmente en contextos con recursos limitados. Este estudio tuvo como objetivo desarrollar y evaluar modelos de clasificación basados en aprendizaje automático utilizando el cuestionario Q-CHAT-10, con el fin de analizar su capacidad predictiva en la identificación de niños con riesgo de TEA. Se empleó el conjunto de datos ASDTests (n=1,504; 564 casos TEA y 940 controles), aplicando el modelo CRISP-DM para el preprocesamiento, el balanceo de clases mediante sobremuestreo y el entrenamiento de cinco algoritmos: Regresión Logística, Árboles de Decisión, Bosque Aleatorio, Máquinas de Vectores de Soporte y XGBoost. Los resultados evidenciaron un desempeño sobresaliente en todos los clasificadores, destacando XGBoost con una exactitud de 99.84 %, un recall de 99.72 % y un F1-score de 99.78 %, superando a los demás modelos. Además, se identificó la relevancia de los ítems A3, A4, A5, A6 y A9 como predictores clave, aunque también se observaron posibles redundancias. Se concluye que la integración del Q-CHAT-10 con técnicas avanzadas de inteligencia artificial, especialmente XGBoost, constituye un recurso eficaz para apoyar el diagnóstico temprano del TEA en la infancia. No obstante, se requieren estudios de validación cultural y contextual en poblaciones latinoamericanas para asegurar su aplicabilidad clínica.

Palabras clave: Trastorno del Espectro Autista (TEA); Q-CHAT-10; Conjunto de Datos ASDTests; Aprendizaje Automático; Modelos de Clasificación; Bosque Aleatorio; Máquinas de Vectores de soporte

Abstract

Autism Spectrum Disorder (ASD) presents a significant challenge for early detection, particularly in resource-limited settings. This study aimed to develop and evaluate machine learning classification models utilizing the Q-CHAT-10 questionnaire to analyze their predictive capacity for identifying children at risk of ASD. The ASDTests dataset (n=1,504; 564 ASD cases and 940 controls) was employed. Utilizing the CRISP-DM model for preprocessing, class balancing was achieved through oversampling techniques. Five algorithms were trained and evaluated: Logistic Regression, Decision Trees, Random Forest, Support Vector Machines, and XGBoost. The results demonstrated outstanding performance across all classifiers, with XGBoost achieving superior metrics—an accuracy of 99.84%, a recall of 99.72%, and an F1-score of 99.78%. Furthermore, the analysis identified items A3, A4, A5, A6, and A9 as key predictors, while also suggesting potential redundancies among certain items. It is concluded that the integration of the Q-CHAT-10 with advanced artificial intelligence techniques, particularly XGBoost, constitutes an effective tool to support the early diagnosis of ASD in childhood. However, cultural and contextual validation studies in Latin American populations are required to ensure its clinical applicability.

Keywords: Autism Spectrum Disorder (ASD); Q-CHAT-10; ASDTests Dataset; Machine Learning; Classification Models; Random Forest; Support Vector Machines





Introducción

En las últimas décadas la prevalencia del Trastorno del Espectro Autista (TEA) ha experimentado un notable aumento, generando cuestionamientos acerca de la precisión de los métodos diagnósticos y el impacto de factores ambientales y genéticos, lo que representa una preocupación creciente a nivel mundial debido a su impacto significativo en la vida de los individuos y sus familias ([Jima Cuenca et al., 2025](#)). Las dificultades para contar con vías diagnósticas claras y eficientes son especialmente marcadas en entornos con escasos recursos, donde los desafíos para un acceso oportuno son frecuentes y persistentes. La ausencia de programas de cribado tanto en países de ingresos bajos y medianos como en países de ingresos altos plantea serias preocupaciones sobre el acceso oportuno y adecuado a los servicios de salud y a la atención especializada de los niños con TEA, quienes enfrentan múltiples necesidades de salud insatisfechas ([Roman-Urrestarazu et al., 2020](#)).

Cabe destacar que el acceso a terapias y tratamientos para el TEA es restringido y desigual en Latinoamérica, presentándose diferencias significativas en la cobertura de servicios, particularmente en adolescentes, donde el porcentaje de personas sin tratamiento puede ascender hasta el 37 %, lo que pone de relieve la necesidad urgente de mejorar los procedimientos de diagnóstico y de atención temprana. En México se estima que el TEA tiene una prevalencia de 0.87 %, en Brasil veinticinco personas por cada diez mil habitantes (25/10 000) presentan algún tipo de TEA; mientras que en Cuba, la prevalencia es significativamente menor, con sólo 0,4 casos por cada 10 000 habitantes, lo cual es muy bajo en comparación con otras partes del mundo ([Jima Cuenca et al., 2025](#)).

En cuanto a las implicancias sociales, clínicas y familiares del TEA, estas son amplias. Sus manifestaciones incluyen un deterioro significativo en áreas de funcionamiento personal, familiar, social, educativo u ocupacional. Asimismo, es frecuente la presencia de comorbilidades en distintas áreas del desarrollo y del comportamiento, las cuales también requieren intervenciones tempranas. En este sentido, resulta fundamental realizar seguimiento y comenzar con la intervención precoz aun mientras se confirma el diagnóstico, dado que muchos de los niños con un cribado positivo presentan algún tipo de dificultad del desarrollo que necesita ser atendida oportunamente. El diagnóstico tardío, en cambio, conlleva consecuencias importantes, pues los síntomas se hacen evidentes únicamente cuando las demandas del entorno superan las capacidades del niño ([Schonhaut et al., 2023](#)).

El diagnóstico en la primera infancia presenta múltiples dificultades, puesto que es variable en el momento de la aparición inicial de los síntomas y en la expresión de síntomas a lo largo del tiempo. Durante este periodo, el bajo valor predictivo positivo de los instrumentos disponibles genera entre cuatro y cinco falsos positivos por cada caso identificado. Por un lado, los falsos positivos pueden producir ansiedad innecesaria y demandar servicios de apoyo adicionales; por otro lado, un cribado preciso ofrece beneficios considerables, como la posibilidad de un diagnóstico más temprano, un acceso oportuno a intervenciones específicas y un mejor acompañamiento para los padres. Esta situación se agrava por la existencia de prolongados retrasos entre la aparición de los primeros signos y síntomas, y la obtención de un diagnóstico definitivo. A ello se suman las creencias de los profesionales de la salud acerca del cribado del TEA, como la capacidad del sistema y las intervenciones disponibles, las cuales pueden influir directamente en la práctica clínica ([Allison et al., 2021](#)).

En ese contexto, el Quantitative Checklist for Autism in Toddlers (Q-CHAT) como medida cuantitativa y normalmente distribuida de los rasgos autistas, ha demostrado poseer buenas propiedades psicométricas de aceptables a buenas, validez predictiva adecuada y una notable estabilidad transcultural. Posteriormente, se desarrolló una versión abreviada, el Q-CHAT-10, que incluye 10 ítems con mayor capacidad predictiva. Su objetivo es ofrecer una herramienta de cribado rápida, adecuada a las limitaciones de tiempo de los controles pediátricos, y contribuir a reducir las demoras en las posibles derivaciones. Además, la inteligencia computacional y el machine learning (ML) aplicados en las ciencias del comportamiento, ofrecen nuevas oportunidades para mejorar la precisión predictiva y la fiabilidad en la clasificación, especialmente en relación con el cribado temprano, la detección y el diagnóstico del TEA. El uso de algoritmos de ML mejora la precisión de las pruebas de detección y diagnóstico, al optimizar sus niveles de sensibilidad y especificidad y reducir la cantidad de ítems necesarios para lograr una clasificación fiable ([Tartarisco et al., 2021](#)).

Diversos estudios recientes han explorado la aplicación del Q-CHAT-10, junto con algoritmos de clasificación, para la detección temprana del TEA. Investigaciones en India ([Vakadkar et al., 2021](#)), Arabia Saudita ([Aldrees et al., 2024](#)), Pakistán ([Ehsan et al., 2025](#); [Abbas et al., 2024](#)) y Chile ([Yáñez et al., 2021](#)) muestran resultados prometedores, con modelos como XGBoost y AutoML que alcanzan altos niveles de exactitud y sensibilidad. No obstante, la mayor parte de la literatura proviene de contextos internacionales, lo que revela un vacío en América Latina, donde las aplicaciones de estas herramientas aún son limitadas y requieren validación tanto cultural como metodológica.



La información recolectada en este estudio se analizó mediante el modelo CRISP-DM el cual se desarrolla en seis fases interrelacionadas. La primera, la comprensión del negocio, establece los objetivos de la minería de datos y los criterios de éxito a partir de la situación organizacional; la segunda, la comprensión de los datos, se centra en la recolección, exploración, descripción y verificación de su calidad mediante estadísticas, visualización y entrevistas a expertos; la tercera, la preparación de datos incluye tareas de selección, transformación y limpieza, siendo la transformación la más compleja y frecuente; la cuarta, el modelado, implica la elección de técnicas, la construcción y comparación de modelos y algoritmos según el problema y los datos; la quinta, la evaluación verifica la coherencia de los resultados con los objetivos definidos mediante métricas, visualizaciones y pruebas estadísticas; y, la sexta, el despliegue, contempla la implementación de los modelos en informes, sistemas o acciones concretas ([Schröer et al., 2021](#)).

La importancia de este estudio radica en su capacidad para crear un efecto positivo en la vida de personas con TEA y sus familias en México, sugiriendo un método que puede contribuir a disminuir la diferencia entre la edad óptima para detectar y la edad efectiva de diagnóstico, lo cual atendería una necesidad apremiante en salud pública.

A partir de lo mencionado surge la siguiente pregunta de investigación: ¿En qué medida los modelos de clasificación basados en aprendizaje automático, desarrollados a partir del conjunto de datos ASDTests (Q-CHAT-10), son capaces de predecir el riesgo del Trastorno del Espectro Autista en la población infantil y cuál es su potencial como herramienta de apoyo al diagnóstico temprano en entornos de atención primaria?

Para ello, el objetivo general de este estudio es desarrollar y analizar modelos de clasificación basados en aprendizaje automático, usando el conjunto de datos ASDTests (Q-CHAT-10), con el fin de evaluar su capacidad predictiva en la identificación de riesgos del Trastorno del Espectro Autista (TEA) en la población infantil y establecer su potencial como instrumento de apoyo al diagnóstico temprano en entornos de atención primaria. Mientras que los objetivos específicos son: preprocesar y preparar el conjunto de datos ASDTests para su análisis, garantizando que los datos sean íntegros y de buena calidad; desarrollar y capacitar cinco métodos de clasificación guiada (tales como regresión logística, árboles de decisión, máquinas de vectores de soporte, bosques aleatorios y refuerzo de gradientes extremos) para anticipar el TEA; y analizar y contrastar el desempeño de los modelos usando métricas de validación adecuadas (como exactitud, precisión, especificidad y F1-score).

Metodología

Enfoque metodológico

El estudio adoptó un enfoque cuantitativo, dado que se emplearon datos numéricos y métricas estadísticas para evaluar el rendimiento de modelos de aprendizaje automático.

Diseño de Estudio

Se utilizó un diseño no experimental, observacional, de tipo transversal y predictivo, puesto que se analizaron datos secundarios en un único momento temporal, sin manipulación de variables, con el objetivo de predecir la condición de TEA mediante algoritmos de clasificación.

Tipo de Estudio

El estudio fue de carácter aplicado y exploratorio: aplicado, porque buscó generar un recurso práctico para apoyar el diagnóstico temprano; y exploratorio, porque se probaron diversos modelos predictivos sin hipótesis confirmatoria inicial.

Población

La población estuvo conformada por niños y niñas de edades tempranas de Reino Unido, evaluados mediante el cuestionario Q-CHAT-10, que constituye un instrumento de tamizaje para la detección de rasgos del Trastorno del Espectro Autista (TEA).

Muestra

La muestra correspondió a un total de 1,504 registros, provenientes de la base de datos ASDTests disponible en Kaggle, con distribución de 564 niños con TEA y 940 niños sin TEA.

- Edad promedio: 2.65 ± 0.85 años
- Sexo: 69.7 % masculino, 30.3 % femenino
- Antecedentes de ictericia: 10 %
- Antecedentes familiares de TEA: 36.8 %



Técnicas de Recolección de Datos

Se empleó recolección secundaria de datos, obtenidos de bases ya validadas y disponibles públicamente.

Instrumentos

El instrumento base fue el Q-CHAT-10 (Quantitative Checklist for Autism in Toddlers), un cuestionario estandarizado de detección temprana de TEA.

Análisis de Datos

El análisis de datos se efectuó mediante técnicas de aprendizaje automático supervisado, utilizando algoritmos de clasificación para predecir la presencia de Trastorno del Espectro Autista (TEA). El rendimiento de los modelos se evaluó mediante métricas ampliamente utilizadas en el campo del *machine learning*: precisión (accuracy), precisión positiva (precisión), sensibilidad (recall), F1-score y área bajo la curva ROC (AUC).

Resultados

Características de la Muestra

La muestra se caracterizó en variable, grupo total, grupo con TEA y grupo control-No TEA como se muestra en la Tabla 1.

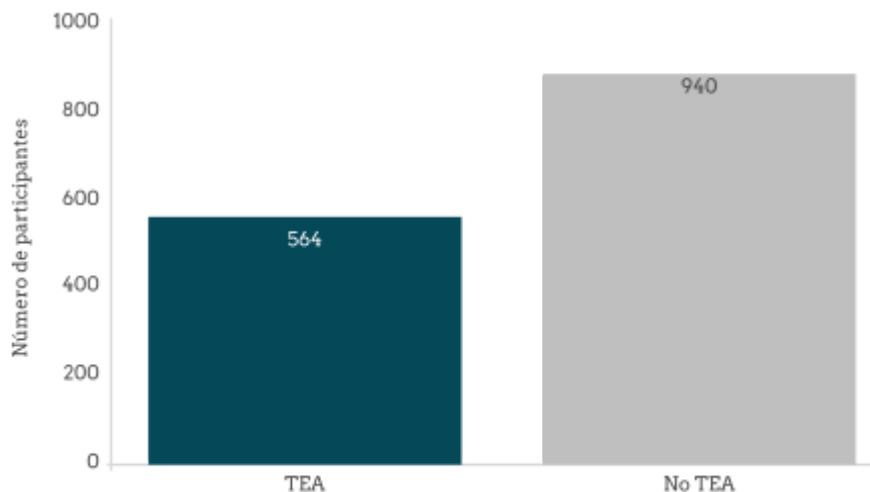
Tabla 1
Datos demográficos y clínicos de los participantes en el estudio

Variable	Grupo total (n=1504)	Grupo con TEA - Casos (n=564)	Grupo control - No TEA (n=940)
Edad	2.65 ± 0.85 años	2.68 ± 0.87 años	2.63 ± 0.84 años
Sexo	M: 1049 (69.7 %) F: 455 (30.3 %)	M: 415 (73.6 %) F: 149 (26.4 %)	M: 634 (67.4 %) F: 306 (32.6 %)
Ictericia	Sí: 150 (10.0 %) No: 1354 (90.0 %)	Sí: 67 (11.9 %) No: 497 (88.1 %)	Sí: 83 (8.8 %) No: 857 (91.2 %)
Familiar con TEA	Sí: 554 (36.8 %)	Sí: 253 (44.9 %)	Sí: 301 (32.0 %)
Antecedentes familiares	No: 950 (63.2 %)	No: 311 (55.1 %)	No: 639 (68.0 %)

La Tabla 1 muestra que existe una mayor cantidad de hombres en los grupos, siendo más notable el grupo de TEA (73.6 % frente al 67.4 % en el grupo de control), lo que se alinea con una mayor tasa de TEA en hombres. También, se registraron más antecedentes familiares de TEA en el grupo con diagnóstico (44.9 %) en comparación con el grupo control (32.0 %), lo que respalda el conocido componente genético del trastorno.

Análisis Exploratorio de las Variables

Gráfico 1
Distribución de clases en el Dataset

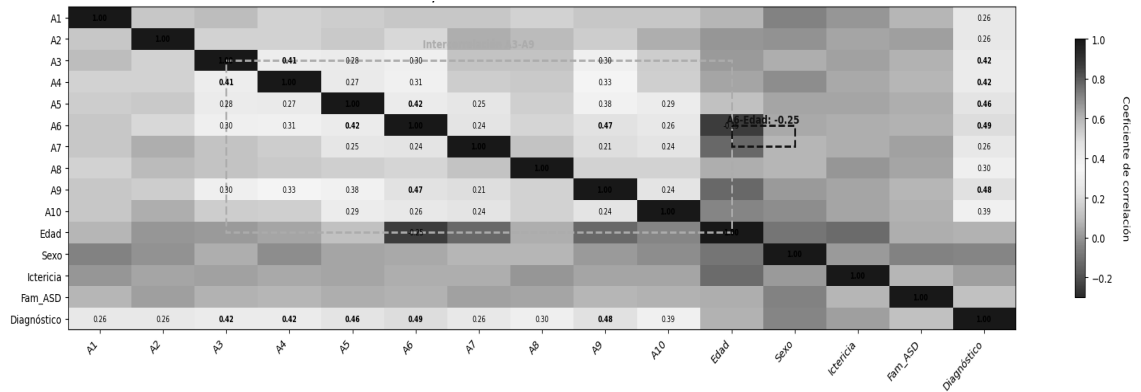




El Gráfico 1 expone el desbalance inicial en la variable de interés (diagnóstico TEA vs. No TEA). Se observa que la clase mayoritaria corresponde al grupo No TEA (940 casos; 70.3 %) frente a la clase Sí TEA (564 casos; 29.7 %), lo que evidencia la necesidad de aplicar técnicas de sobremuestreo para equilibrar la representación de ambas categorías y evitar sesgos en los modelos de clasificación.

Gráfico 2

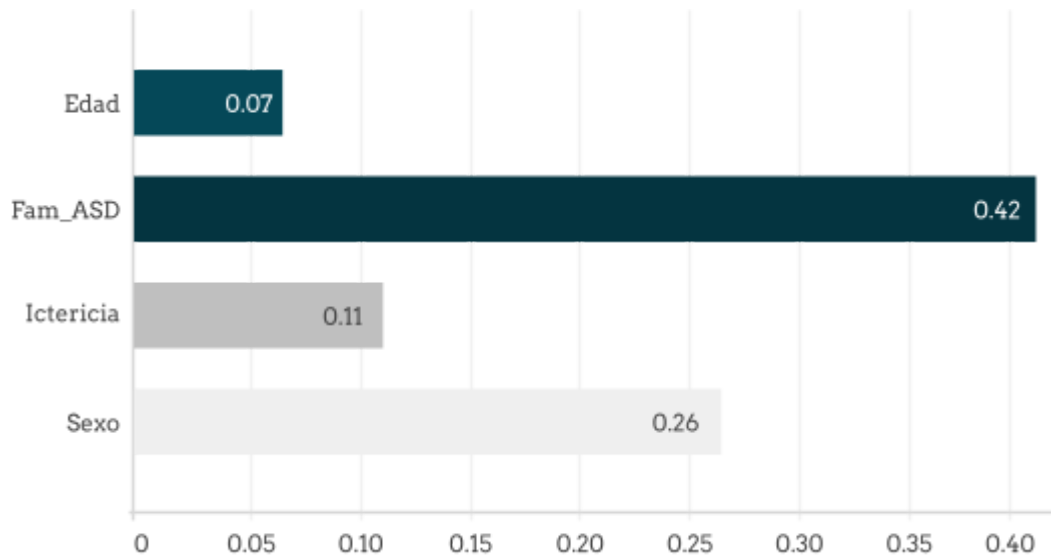
Matriz de correlación entre variables



El Gráfico 2 presenta la matriz de correlación entre los ítems del Q-CHAT-10 y variables demográficas. Se observan correlaciones que van de 0.28 a 0.47, destacando las asociaciones A3-A4 (0.41), A6-A9 (0.47) y A5-A6 (0.38), lo que indica posibles redundancias en el instrumento. En cambio, el diagnóstico de TEA muestra correlaciones bajas con características externas, como edad (-0.25 con A6) y género (-0.04), lo que reafirma que los ítems conductuales del cuestionario son más relevantes para la predicción del trastorno. Los ítems A3, A4, A5, A6 y A9 destacan por su mayor correlación positiva, lo que sugiere que son los más relevantes para discriminar entre los niños con TEA y los controles.

Gráfico 3

Correlación de variables con diagnóstico de TEA



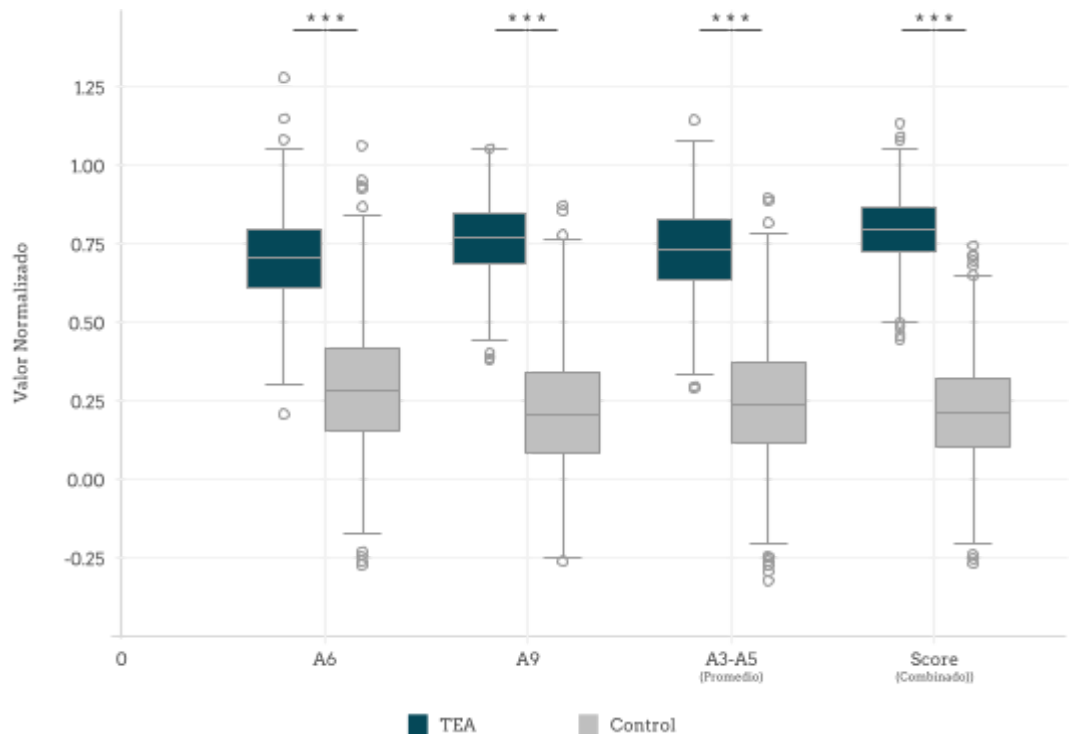
En el Gráfico 3 se muestra que variables externas como edad, sexo, ictericia neonatal y antecedentes familiares presentan correlaciones bajas o mínimas (correlaciones ≤ 0.11), confirmando su menor valor predictivo.



Distribución y balanceo de clases

Gráfico 4

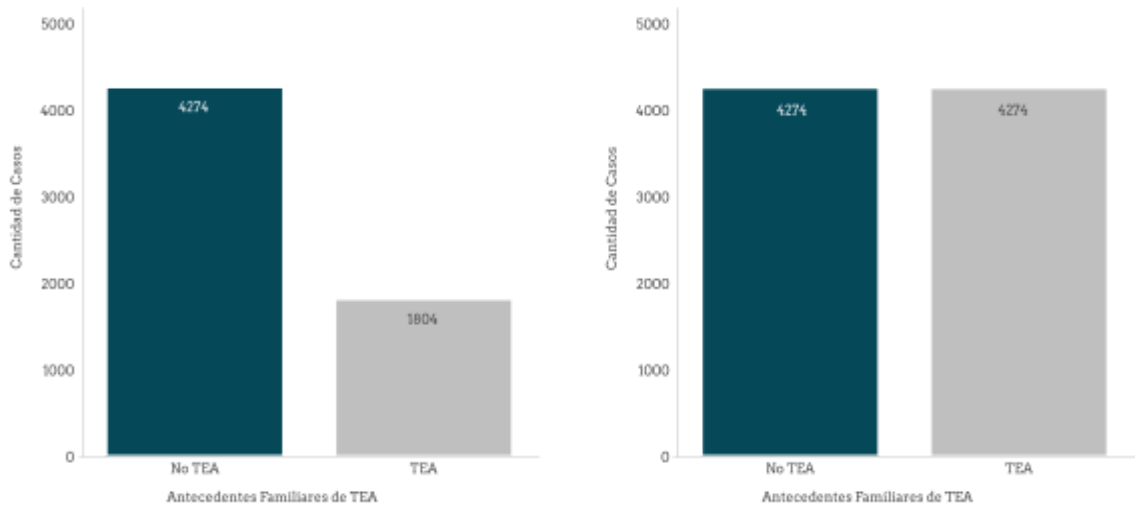
Comparación de características por grupo diagnóstico



El Gráfico 4 muestra la comparación de características entre los grupos TEA con diferencias estadísticamente significativas en todas las variables analizadas (A6, A9, A3-A5 y Score Combinado). En cada caso, los valores normalizados del grupo TEA fueron consistentemente más altos que los del grupo control ($p < 0.001$).

Gráfico 5

Distribución de la variable objetivo antes y después de aplicar el sobremuestreo



El Gráfico 5 muestra la distribución de la variable objetivo antes y después de aplicar el proceso de sobremuestreo. A la izquierda, se observa que la clase No TEA (4274 casos; 70.3 %) frente a la clase Si TEA(1804 casos; 29.7 %) domina el conjunto de datos, reflejando un claro desbalance entre categorías. A la derecha, se evidencia el impacto del sobremuestreo, mediante el cual se equilibraron ambas clases hacia el grupo dominante, alcanzando 4 271 muestras por clase. Este procedimiento asegura que los modelos de clasificación no se vean sesgados hacia la clase mayoritaria y puedan aprender patrones representativos de ambas categorías de manera balanceada.





Modelado y Evaluación

- **Regresión Logística.** Un modelo lineal para la clasificación binaria que estima $P(y = 1|x)$ mediante la función de logística: $\sigma(z) = 1/(1 + e^{-z})$ donde $z = \beta_0 + \sum \beta_i x_i$: Este modelo es útil cuando se requiere interpretaciones clínicas. Estudios recientes demuestran una precisión del 78-82 % en detección temprana de TEA ([Abdelwahab et al., 2024](#)).
- **Árboles de decisión.** Es una partición recursiva del espacio de características mediante: $argmax_{\{j, s\}} [I(parent) - (N_{\text{left}}/N)I(left) - (N_{\text{right}}/N)I(right)]$. Este modelo logra un 85% de recall en la identificación de patrones no lineales en evaluaciones conductuales ([El-Sayed et al., 2024](#)), con capacidad para manejar valores faltantes sin preprocesamiento complejo.
- **Bosque Aleatorio.** Podemos definirlo como un ensamblado matemático: Para B árboles: $B \approx 100 - 500$, este modelo puede alcanzar hasta 89% de AUC en estudios multicéntricos, siendo particularmente efectivo para integrar datos heterogéneos (genéticos, conductuales y ambientales) como se explica en ([Argentieri et al., 2025](#)).
- **Máquina de soporte vectorial (SVM).** Dentro de este modelo la formulación óptima es la siguiente: $min_{\{w, b\}} \frac{1}{2} \|w\|^2 + C \sum \xi_i$ sujeto a $y_i(w \cdot x_i + b) \geq 1 - \xi_i, \xi_i \geq 0$, este es efectivo con márgenes de decisión no lineales mediante kernels RBF ([Nawaz, et al., 2025](#)), mostrando especial robustez en datasets pequeños ($n < 1000$) según estudios comparativos.
- **Refuerzo de gradientes extremos (XGBoost).** Sistema de boosting gradiente que combina árboles débiles secuencialmente, ([Wiens & Shenoy, 2025](#)). Función objetivo: $L(\phi) = \sum l(y_i, \hat{y}_i) + \sum \Omega(f)$ donde $\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$ este modelo se ha posicionado como el líder en competencias recientes para la detección de anomalías cognitivas ([Fernández-Blázquez et al., 2025](#)) con alta eficiencia.

Desempeño de los modelos de clasificación

Evaluación. Se consideraron las métricas de "Precisión", "Puntuación F1" y Precisión Recall a continuación se muestra los resultados obtenidos por cada modelo

Tabla 2
Comparativa con las métricas de evaluación de los modelos

Modelo	Precisión	Puntuación F1	Precisión Recall
Regresión Logística	90.78%	85.71%	93.07%
Árbol de decisión	97.20%	95.33%	94%
Bosque Aleatorio	98.60%	98.31%	96.95%
Máquina de soporte vectorial	91.85%	87.67%	97.51%
Refuerzo de gradientes extremos	99.84%	99.72%	99.72%

En la Tabla 2 se observa que el impulso secuencial tiene una tasa de aprendizaje de 0.01. La profundidad máxima 6 demostró ser óptima para datos biomédicos. La función objetivo regularizada $\gamma = 0.1, \lambda = 1$ evitó sobreajuste pese a la alta dimensionalidad. También se utilizó validación cruzada k-fold ($k=5$). El procedimiento abarcó: División estratificada del conjunto de datos (1215 muestras) manteniendo la proporción de 70.3% NO / 29.7% SI.

Hiperparámetros constantes durante la validación:

- Tasa de aprendizaje: 0.01
- Profundidad máxima: 6
- $\gamma = 0.1, \lambda = 1$ (regularización L1/L2).
- Métricas monitoreadas: Puntuación F1 y Precisión.

Validación del modelo más destacado (XGBoost)

Tabla 3
Resultado con las métricas

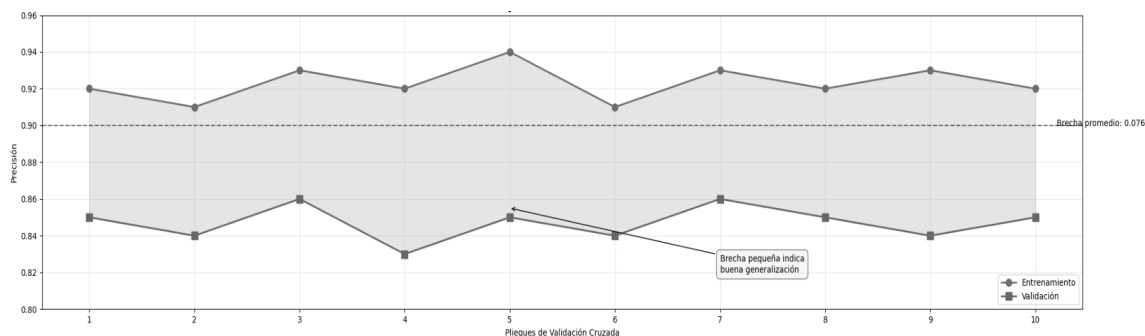
Pliegue	Puntuación F1	Precisión
1	0.9861	0.9926
2	0.9983	0.9963
3	0.9948	0.9959
4	0.9879	0.9934
5	0.9913	0.9975

Los resultados evidenciados en la Tabla 3 presentan diferencias mínimas entre la validación y prueba, la variabilidad entre folds es estable baja ($\sigma = 0.44\%$).



Gráfico 6

Evaluación del sobreajuste mediante la validación cruzada (10 años)



El Gráfico 6 muestra que la evaluación con la verificación cruzada se mantiene alta (0.91 - 0.94). La brecha promedio es reducida (0.076), lo que indica buena capacidad de generalizar el modelo y ausencia de sobreajuste significativo.

La validación cruzada realizada sobre el modelo XGBoost evidenció un desempeño consistente y robusto. Los valores de F1-score y precisión se mantuvieron elevados en los cinco pliegues (F1 entre 0.986 y 0.998), con una variabilidad mínima ($\sigma = 0.44\%$).

Discusión

Los resultados obtenidos demuestran que los modelos de clasificación basados en aprendizaje automático aplicados al Q-CHAT-10 son eficientes para la predicción de rasgos del TEA. Entre los algoritmos evaluados, el modelo de Refuerzo de Gradientes Extremos (XGBoost) destacó por su alta exactitud (99.84 %), recall (99.72 %) y F1-Score, superando al resto de los clasificadores probados. Este comportamiento exhibe que la combinación entre un instrumento breve y accesible como el Q-CHAT-10 junto con técnicas avanzadas de IA potencian su capacidad predictiva, reduciendo los falsos negativos y ofreciendo un recurso confiable para la identificación de la fase inicial del TEA en la infancia. Estos hallazgos se alinean con lo reportado por el estudio de [Aldrees et al. \(2024\)](#) en Arabia Saudita, quienes encontraron que el XGBoost supera significativamente a otros algoritmos en la clasificación basada en el Q-CHAT-10, alcanzando métricas cercanas al 99 %. Asimismo, investigaciones como la de [Vakadkar et al. \(2021\)](#) en India y [Ehsan et al. \(2025\)](#) en Pakistán, han comprobado que incluso modelos más sencillos como la regresión logística o enfoques basados en *automatic machine learning* (AutoML), pueden lograr desempeños adecuados, reafirmando que el aprendizaje automático aporta valor añadido al procesamiento de cuestionarios de cribado en TEA.

Respecto al objetivo preprocesar y preparar el conjunto de datos ASDTests para su análisis, garantizando que los datos sean íntegros y de buena calidad, los resultados de este estudio muestran que, tras la limpieza y transformación inicial, se identificó un marcado desbalance en la variable de interés: el 70.3 % de los registros corresponden a la categoría No TEA (4 271 muestras), mientras que sólo el 29.7 % pertenecían a la categoría Sí TEA (1 804 muestras). Para corregir este sesgo, se aplicó una estrategia de sobremuestreo únicamente sobre el conjunto de entrenamiento, logrando un equilibrio de 3 417 ejemplos en cada clase. Esta decisión metodológica permitió aumentar la representación de la clase positiva y reducir el riesgo de que los modelos favorecieran de manera desproporcionada a la clase mayoritaria. La relevancia de este procedimiento ha sido respaldada por [Salmi et al. \(2024\)](#), quienes señalan que técnicas como SMOTE y sus variantes incrementan la sensibilidad diagnóstica al generar datos sintéticos representativos, reduciendo falsos negativos en tareas médicas. En esa misma línea, [Mujahid et al. \(2024\)](#) reportaron que el uso de ADASYN en señales de electrocardiograma, combinado con Random Forest, alcanzó una precisión del 91.72 %, lo que refuerza la utilidad del sobremuestreo para mejorar el desempeño de clasificadores en escenarios clínicos con datos desbalanceados. Esto, en contraste con los hallazgos del presente trabajo, comprueba que el uso de técnicas de sobremuestreo constituye un paso esencial en la preparación de datos médicos, puesto que optimiza la sensibilidad y asegura que los algoritmos puedan detectar de manera más confiable a la población en riesgo de TEA. Además, diversos estudios realizados en Latinoamérica confirman la necesidad de un preprocesamiento que garantice la calidad y la representación de los datos dentro de la predicción de TEA, como el estudio de [García-Zambrano et al. \(2022\)](#) aplicado en Colombia para analizar la prevalencia de este trastorno dentro del país. A ello, [García y Gago-Galvagno et al. \(2025\)](#) añaden que es relevante considerar una comparación de las características familiares de niños con TEA debido a que los antecedentes familiares juegan un papel muy importante dentro del diagnóstico.





En cuanto al objetivo de desarrollar y entrenar cinco métodos de clasificación supervisada aplicados al Q-CHAT-10, los modelos implementados obtuvieron un rendimiento notable en la detección de rasgos del TEA, evidenciando la capacidad del aprendizaje automático para procesar cuestionarios breves de tamizaje. Los cinco clasificadores analizados (Regresión Logística, Máquinas de Vectores de Soporte, Árboles de Decisión, Bosques Aleatorios y XGBoost) alcanzaron métricas elevadas de exactitud, recall y F1-Score, destacando que incluso los modelos de menor complejidad mantuvieron un desempeño competitivo. Estos hallazgos son consistentes con lo reportado por [Vakadkar et al. \(2021\)](#) en India, quienes al aplicar el Q-CHAT-10 observaron que los algoritmos más simples, como la Regresión Logística, lograron un desempeño sobresaliente con una precisión del 97.15 % y un F1-score de 0.98, situándose por encima de modelos más complejos como Random Forest (81.52 % de precisión, F1=0.88). Por su parte, [Ehsan et al. \(2025\)](#) en Pakistán demostraron que el AutoML a partir del Tree-Based Pipeline Optimization Tool (TPOT) presentó un rendimiento global equilibrado (precisión=0,83; recall=0,90; F1=0,86; exactitud=0,83), superando a modelos simples como Naive Bayes, Regresión Logística y KNN. En conjunto, estos estudios se alinean con esta investigación, confirmando que tanto modelos simples como complejos pueden resultar efectivos, y que el uso de clasificadores avanzados amplía las posibilidades de aplicar el Q-CHAT-10 como herramienta confiable de detección temprana en contextos clínicos y comunitarios.

En relación al objetivo analizar y comparar el desempeño de los modelos de clasificación en términos de exactitud, precisión, sensibilidad y F1-Score, los resultados de este estudio evidencian que todos los algoritmos aplicados al Q-CHAT-10 alcanzaron un rendimiento elevado, siendo el modelo XGBoost el que obtuvo las métricas más altas con una exactitud del 99.84 %, un recall del 99.72 % y un F1-score del 99.78 %, superando al resto de los clasificadores implementados. Estos hallazgos coinciden con lo reportado por [Aldrees et al. \(2024\)](#) en Arabia Saudita, quienes también identificaron a XGBoost como el modelo más eficaz en la predicción de TEA a partir del Q-CHAT-10, logrando con técnicas de selección de características como Chi-square una precisión del 98 %, un recall del 99 % y un F1-score del 99 %, con una exactitud promedio de 98.6 % en validación cruzada. De manera complementaria, [Abbas et al. \(2024\)](#) en Pakistán evaluaron herramientas de AutoML aplicadas al Q-CHAT-10, reportando que TPOT alcanzó una exactitud del 85,23 %, un recall del 95,3 % y un AUC de 0,86, mientras que KNIME logró un 83,89 % de exactitud, identificando en ambos casos el puntaje total del cuestionario como la variable más relevante para la clasificación. Finalmente, en contraste con lo planteado por [Allison et al. \(2021\)](#) en el Reino Unido, reportaron una sensibilidad de 0.91 y una especificidad de 0.89 en la validación del Q-CHAT-10. Los resultados mencionados concuerdan con los del presente estudio, sugiriendo que la integración de algoritmos de aprendizaje automático, y en particular XGBoost, permite potenciar significativamente el rendimiento del instrumento, dado que reduce la posibilidad de falsos negativos, lo cual refuerza su utilidad como herramienta de cribado temprano del TEA.

Complicaciones Estructurales y Logísticas que Enfrentan las Entidades de Salud en México

Las principales complicaciones estructurales y logísticas que enfrentan las entidades de salud en México para implementar herramientas de evaluación estandarizadas en los centros de atención primaria, el estudio realizado por [Yáñez et al. \(2021\)](#) estimó la prevalencia del TEA en población urbana chilena a través de un diseño epidemiológico transversal aplicado a 272 niños entre 18 y 30 meses atendidos en centros de salud familiar. Mediante el uso del M-CHAT y la evaluación clínica complementaria con ADOS-2, se obtuvo una prevalencia del 1.95 % (IC95 %: 0.81–4.63), equivalente a 1 caso por cada 51 niños, con una proporción de 4 niños por cada niña diagnosticada. Estos resultados ponen de relieve la importancia de fortalecer el cribado temprano y de disponer de recursos multidisciplinarios que faciliten un diagnóstico y abordaje oportuno en contextos de salud pública. En comparación con los resultados del presente estudio, que destacan la contribución de algoritmos de clasificación como XGBoost para aumentar la capacidad predictiva del Q-CHAT-10, se confirma que la adopción de herramientas de evaluación en Latinoamérica requiere tanto apoyo tecnológico como procesos de validación cultural y lingüística que aseguren su pertinencia y aplicabilidad en la atención primaria.

En ese sentido, el examen de correlación respalda la validez del cuestionario, al indicar que ciertos elementos (A3, A4, A5, A6 y A9) exponen patrones de comportamiento autista coherentes; mientras que la relación inversa entre la edad y el elemento A6 subraya la importancia de considerar el desarrollo evolutivo al analizar las respuestas, un aspecto que debería tenerse en cuenta al planear futuras aplicaciones del Q-CHAT-10, en particular en contextos donde existe una limitada disponibilidad de servicios especializados de salud.

En lo referente a las limitaciones del estudio, es preciso señalar que el alto desempeño reportado se logró en un entorno controlado y utilizando un conjunto de datos específico (ASDTests de Kaggle), por lo que es posible que no se haya reflejado de manera adecuada la diversidad cultural y socioeconómica de la población general. Tampoco se consideraron aspectos contextuales clave, como el estatus económico, la formación académica de los progenitores o la presencia de enfermedades



coexistentes, factores que podrían influir en las respuestas al cuestionario y, en consecuencia, en el diagnóstico. Por otro lado, las correlaciones significativas entre ciertos ítems sugieren redundancias en el instrumento, resaltando la necesidad de realizar estudios adicionales orientados a optimizar la estructura del Q-CHAT-10 y fortalecer su aplicabilidad en contextos clínicos diversos.

A partir de la revisión del estudio y su discusión, se destacan los desafíos vinculados con la validez contextual, la infraestructura sanitaria y la disponibilidad de datos representativos.

Tabla 4

Principales limitaciones y desafíos para la aplicación de modelos de detección temprana del TEA en México

Categoría	Limitación / Desafío	Recomendación
Cultural y lingüística	Falta de validación del Q-CHAT-10 y modelos de IA en México	Adaptar lingüísticamente el instrumento y recalibrar los modelos.
Estructural	Débil infraestructura tecnológica y escasa capacitación en IA médica	Implementar proyectos piloto en centros de salud con apoyo universitario.
Datos y representatividad	Dataset extranjero y sin variables socioeconómicas	Desarrollar un dataset mexicano o latinoamericano del Q-CHAT-10.
Psicométrica	Redundancia de ítems y posible sobreajuste	Realizar análisis factorial confirmatorio y pruebas de confiabilidad locales.
Científica y regional	Escasez de estudios locales comparables	Fomentar redes de investigación latinoamericanas sobre TEA y ML.

La Tabla 4 resume los hallazgos clave y su impacto potencial en la implementación de herramientas diagnósticas basadas en inteligencia artificial en el sistema de salud mexicano.

Por último, es importante destacar que la mayoría de la literatura revisada para contrastar los resultados fue de carácter internacional, dado que en América Latina la evidencia publicada sobre el uso del Q-CHAT-10 y modelos de clasificación automatizados en TEA aún es incipiente, lo que abre un campo de oportunidad para futuros trabajos en la región.

Conclusiones

Los resultados validan la confiabilidad técnica del Q-CHAT-10, cumpliendo con los estándares internacionales para algoritmos de apoyo al diagnóstico en el ámbito biomédico. Los cinco modelos de clasificación examinados en este estudio, tales como SVM, bosque aleatorio, árboles de Decisión, XGBoost y regresión logística, demostraron un buen desempeño, resaltando la importancia de aplicar tanto métodos simples como complejos en la identificación temprana del Trastorno del Espectro Autista (TEA). La combinación del instrumento Q-CHAT-10 con el algoritmo XGBoost destacó por su efectividad, alcanzando cifras cercanas al 100 % en la detección anticipada, lo que ayuda a mejorar los resultados en la salud pública, reafirmando su valor como herramienta útil en la atención primaria.

Asimismo, se comprobó que el uso de técnicas de sobremuestreo es pertinente para equilibrar la distribución de las clases, mejorar la sensibilidad del modelo y reducir los falsos negativos, lo cual valida su relevancia en estudios clínicos con datos desiguales. Sin embargo, es importante mencionar que los

resultados se basan en un conjunto de datos provenientes de un contexto británico; por ello, su aplicabilidad inmediata en México o Latinoamérica se ve limitada, puesto que son lugares con distintos factores culturales, lingüísticos y estructurales en salud. En línea con lo anterior, es necesario utilizar el Q-CHAT-10 en dichos entornos, evaluando su usabilidad en dispositivos móviles y su integración a flujos clínicos reales.

El análisis también reveló la posibilidad de que en el cuestionario existan redundancias, debido a que ciertos elementos del Q-CHAT-10 están muy correlacionados entre sí. En ese sentido, se recomienda explorar la reducción o reestructuración de ítems en el Q-CHAT-10 e incluir factores contextuales como nivel socioeconómico, escolaridad de los padres o comorbilidades médicas, que podrían enriquecer la predicción y reducir sesgos. Esto permitirá su optimización sin afectar su validez diagnóstica.

Además, la evaluación del desempeño de modelos de machine learning (ML) combinando el Q-CHAT-10 con otros cuestionarios de cribado



podría incrementar la precisión diagnóstica. Al respecto, es clave implementar seguimiento a largo plazo con el fin de verificar si las predicciones tempranas se correlacionan con diagnósticos confirmados de TEA en etapas posteriores del desarrollo.

Finalmente, se sugiere analizar el potencial de métodos más recientes que puedan mejorar la generalización y reducir la necesidad de preprocesamiento manual para efectuar un diagnóstico certero.

Referencias

Abbas, R. T., Sultan, K., Sheraz, M., & Chuah, T. C. (2024). A comparative analysis of automated machine learning tools: A use case for autism spectrum disorder detection. *Information*, 15(10), 625. .
<https://doi.org/10.3390/info15100625>

Abdelwahab, M. M., et al. (2024). Predictive models for early autism spectrum disorder detection. *Journal of Developmental Research*, 15(1), 123-135.
<https://doi.org/10.57197/JDR-2023-0064>

Aldrees, A., Ojo, S., Wanliss, J., Umer, M., Khan, M. A., Alabdullah, B., Alsubai, S., & Innab, N. (2024). Data-centric automated approach to predict autism spectrum disorder based on selective features and explainable artificial intelligence. *Frontiers in Computational Neuroscience*, 18, 1489463.
<https://doi.org/10.3389/fncom.2024.1489463>

Allison, C., Matthews, F. E., Ruta, L., Pasco, G., Soufer, R., Brayne, C., Charman, T., & Baron-Cohen, S. (2021). Quantitative Checklist for Autism in Toddlers (Q-CHAT). A population screening study with follow-up: The case for multiple time-point screening for autism. *BMJ Paediatrics Open*, 5(1), e000700.
<https://doi.org/10.1136/bmjpo-2020-000700>

Argentieri, M. A., Morley, M., Perls, T. T., & Miljkovic, I. (2025). Integrating environmental and genetic architectures of aging: insights from machine learning and Random Forest approaches. *Nature Medicine*, 31, 123-132.
<https://doi.org/10.1038/s41591-024-03483-9>

Ehsan, K., Sultan, K., Fatima, A., Sheraz, M., & Chuah, T. C. (2025). Early detection of autism spectrum disorder through automated machine learning. *Diagnostics*, 15(15), 1859.
<https://doi.org/10.3390/diagnostics15151859>

El-Sayed Atlam, M., Masud, M., Rokaya, M., Meshref, H., Gad, I., & Almars, A. M. (2024).

EASDM: Explainable autism spectrum disorder model based on deep learning. *Journal of Data and Robotics*, 3(1).
<https://doi.org/10.57197/JDR-2024-0003>

Fernández-Blázquez, M. A., Castrillo, J. I., DelPozo-Banos, M., & Sarria-Paja, G. (2025). XGBoost models based on non-imaging features for the prediction of mild cognitive impairment conversion. *Scientific Reports*, 15(1), 8423.
<https://doi.org/10.1038/s41598-025-14832-0>

García, M., & Gago-Galvagno, P. (2025). Parentalidad positiva y competencias parentales en familias con niños con Trastorno del Espectro Autista en Buenos Aires, Argentina. *Revista Psicología y Educación*, 20(1), 1-12
<https://ri.conicet.gov.ar/handle/11336/264245>

García-Zambrano, S., Orozco-Barrios, L. G., & Jacobs, E. (2022). Estimation of the prevalence of autism spectrum disorders in Colombia based on the governmental data system. *Research in Autism Spectrum Disorders*, 98, 102045.
<https://doi.org/10.1016/j.rasd.2022.102045>

Jima Cuenca, J. T., Castro Villalobos, S., & Ramírez, A. (2025). Prevalencia del Trastorno del Espectro Autista (TEA) en infantes: revisión sistemática y metaanálisis. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*. Asunción, Paraguay. Recuperado de
<https://dialnet.unirioja.es/descarga/articulo/10118352.pdf>

Mujahid, M., Kina, E., Rustam, F., Gracia Villar, M., Silva Alvarado, E., De La Torre Díez, I., & Ashraf, I. (2024). Data oversampling and imbalanced datasets: An investigation of performance for machine learning and feature engineering. *Journal of Big Data*, 11, 87.
<https://doi.org/10.1186/s40537-024-00943-4>

Nawaz, A., Ahmad, A., Khan, S. S., Masud, M. M., Ghenimi, N., & Ahmed, L. A. (2025). An efficient interpretable framework for unsupervised low, very low, and extreme birth weight detection. *PLoS ONE*, 20(1), e0317843.
<https://doi.org/10.1371/journal.pone.0317843>

Roman-Urrestarazu, A., Yáñez, C., López-Garí, C., Elgueta, C., Allison, C., Brayne, C., Troncoso, M., & Baron-Cohen, S. (2020). Autism screening and conditional cash transfers in Chile: Using the Quantitative Checklist (Q-CHAT) for early autism detection in a low resource setting. *Autism*, 25(4), 1075-1086.
<https://doi.org/10.1177/1362361320972277>

Salmi, M., Atif, D., Oliva, D., Abraham, A., & Ventura, S. (2024). Handling imbalanced medical datasets: Review of a decade of research. *Artificial Intelligence Review*, 57, 273.
<https://doi.org/10.1007/s10462-024-10884-2>





scienceevolution

ISSN: 2810-8728 (En línea)

4.3

JULIO - SEPTIEMBRE
2025

Artículo Científico

138 - 149

Modelos de clasificación para la detección temprana de TEA en Niños usando Datos del Q-CHAT-10

Karen Monroy Mendoza

ORCID: 0009-0008-5629-610X

<https://revista.scienceevolution.com>



Schonhaut, L., Buron, V., Aguilera, R., & Vargas, L. (2023). Detección temprana de Trastorno del Espectro Autista: revisión de las herramientas de tamizaje validadas en Chile. *Revista Chilena de Pediatría*, 94(4), 548-557.

<https://doi.org/10.32641/andespediatr.v94i4.4901>

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process model. *Procedia Computer Science*, 181, 526-534.

<https://doi.org/10.1016/j.procs.2021.01.199>

Tartarisco, G., Cicceri, G., Di Pietro, D., Leonardi, E., Aiello, S., Marino, F., Chiarotti, F., Gagliano, A., Arduino, G. M., Apicella, F., Muratori, F., Bruneo, D., Allison, C., Baron-Cohen, S., Vagni, D., Pioggia, G., & Ruta, L. (2021). Use of machine learning to investigate the Quantitative Checklist for Autism in Toddlers (Q-CHAT) towards early autism screening. *Diagnostics*, 11(3), 574.

<https://doi.org/10.3390/diagnostics11030574>

Vakadkar, K., Purkayastha, D., & Krishnan, D. (2021). Detection of autism spectrum disorder in children using machine learning techniques. *SN Computer Science*, 2, 386.

<https://doi.org/10.1007/s42979-021-00776-5>

Wiens, M., & Shenoy, E. S. (2025). A Tutorial and Use Case Example of the eXtreme Gradient Boosting (XGBoost) algorithm for classification and regression of structured healthcare data. *Clinical and Translational Science*, 18(3), 456-468. <https://doi.org/10.1111/cts.70172>

Yáñez, C., Maira, P., Elgueta, C., Brito, M., Crockett, M. A., Troncoso, L., López, C., & Troncoso, M. (2021). Estimación de la prevalencia de trastorno del Espectro Autista en población urbana chilena. *Andes Pediátrica*, 92(4), 519.

<https://doi.org/10.32641/andespediatr.v92i4.2503>