



Aplicación de análisis CHAID en la evaluación de gravedad tumoral mamaria: utilidad diagnóstica por edad

Application of CHAID analysis in the assessment of breast tumor severity:
diagnostic utility by age

Alberto Guevara Tirado^{1*} <https://orcid.org/0000-0001-7536-7884>

¹Universidad Científica del Sur. Lima, Perú.

*Correspondencia. Correo electrónico: albertoguevara1986@gmail.com

RESUMEN

Introducción: El cáncer de mama es causa frecuente de mortalidad. La mamografía y modelos como CHAID pueden mejorar el diagnóstico temprano al estructurar variables clínicas y radiológicas interactivas.

Objetivo: Identificar y jerarquizar las características mamográficas asociadas a la gravedad tumoral según el grupo etario mediante un análisis CHAID.

Métodos: Estudio observacional transversal de una base de datos secundaria clínico-radiológica de 961 registros, de los cuales 956 mujeres fueron incluidas en el análisis final. Variables analizadas: gravedad tumoral, edad, forma del nódulo, márgenes y densidad. Se utilizó un modelo de árbol de decisión mediante detección automática de interacciones por *ji* cuadrado (CHAID), incorporando validación cruzada y estimando sensibilidad, especificidad y exactitud.

Resultados: El modelo CHAID fue más específico en < 50 años (89,5 %) y más sensible en ≥ 50 años (77,8 %). En cuanto a patrones estructurales, en el grupo < 50 años la variable más discriminante fue la forma del nódulo, destacando la forma irregular como fuertemente asociada a malignidad (76,3 %), seguida de márgenes no circunscritos y densidad iso. En ≥ 50 años, el árbol





se ramificó inicialmente por la morfología de los márgenes, especialmente los espiculados, que se asociaron con alta probabilidad de malignidad (83,3 %), seguidos por forma irregular y baja densidad.

Conclusiones: El modelo CHAID evidenció una utilidad diagnóstica diferenciada por grupo etario, con mayor precisión en la identificación de lesiones benignas en mujeres menores de 50 años y mejor capacidad para detectar malignidad en mujeres de 50 años o más.

Palabras clave: árboles de decisión; grupos de edad; mamografía; neoplasias de la mama; toma de decisiones asistida por computador.

ABSTRACT

Introduction: Breast cancer is a frequent cause of mortality. Mammography and models such as CHAID may improve early diagnosis by structuring interactive clinical and radiological variables.

Objective: To identify and rank mammographic features associated with tumor severity according to age group using CHAID analysis.

Methods: A cross-sectional observational study was conducted using a secondary clinic-radiological database comprising 961 records, of which 956 women were included in the final analysis. The variables analyzed included tumor severity, age, nodule shape, margins, and density. A decision tree model based on chi-squared automatic interaction detection (CHAID) was applied, incorporating cross-validation and estimating sensitivity, specificity, and accuracy.

Results: The CHAID model showed higher specificity in women aged <50 years (89.5%) and higher sensitivity in those aged ≥ 50 years (77.8%). Regarding structural patterns, in the <50-year group, the most discriminative variable was nodule shape, with irregular shape strongly associated with malignancy (76.3%), followed by non-circumscribed margins and iso-density. In women aged ≥ 50 years, the tree initially split according to margin morphology, particularly spiculated margins, which were associated with a high probability of malignancy (83.3%), followed by irregular shape and low density.





Conclusions: The CHAID model demonstrated age-specific diagnostic utility, with greater accuracy in identifying benign lesions in women younger than 50 years and improved ability to detect malignancy in women aged 50 years or older.

Keywords: age groups; breast neoplasms; decision making, computer-assisted; decision trees; mammography.

Recibido: 24/07/2025

Aprobado: 14/01/2026

INTRODUCCION

El cáncer de mama es una de las principales causas de morbilidad y mortalidad en mujeres a nivel mundial, con más de 2 millones de casos nuevos detectados el año 2022.⁽¹⁾ La detección temprana permite mejorar el pronóstico, reducir la agresividad del tratamiento y mitigar el impacto social y económico de la enfermedad.⁽²⁾ En este contexto, la mamografía sigue siendo el principal método de tamizaje y evaluación inicial, dada su accesibilidad y eficacia para detectar alteraciones morfológicas sutiles en el tejido mamario.⁽³⁾

Entre las características radiológicas asociadas con mayor riesgo de malignidad destacan la forma irregular del nódulo, los márgenes espiculados y la edad avanzada.⁽⁴⁾ No obstante, estas variables no actúan de manera independiente, sino que se combinan e interactúan de forma compleja, lo que podría dificultar el juicio diagnóstico, sobre todo en entornos con recursos limitados o menor experiencia interpretativa.

Frente a este reto, los modelos automatizados como los árboles de decisión permiten estructurar estas interacciones de manera lógica y visual. En particular, el algoritmo de árboles de decisión basados en detección automática de interacciones mediante *ji* cuadrado CHAID (*Chi-squared Automatic Interaction Detection*, en inglés) ofrece una herramienta útil para jerarquizar variables categóricas, identificar patrones clave y adaptar la evaluación diagnóstica según subgrupos etarios.





El presente estudio tuvo como objetivo evaluar la utilidad del análisis CHAID para identificar y jerarquizar las características mamográficas asociadas a la gravedad tumoral (benigna o maligna), con especial énfasis en la utilidad diagnóstica diferenciada según el grupo etario (< 50 años vs. ≥ 50 años). Los resultados permitirán determinar si este modelo de aprendizaje supervisado puede contribuir a mejorar la clasificación inicial de lesiones sospechosas, proporcionando criterios visuales de apoyo para el triaje clínico y radiológico.

MÉTODOS

Diseño y población

Estudio observacional transversal, basado en una base de datos secundaria y de acceso abierto que contiene información clínico-radiológica de 961 mujeres con lesiones mamográficas, recolectadas entre los años 2003 y 2006 en un instituto universitario de radiología en Alemania. La base de datos incluye pacientes con hallazgos sospechosos en mamografía digital, quienes fueron sometidas a biopsia para la confirmación diagnóstica histopatológica.

La población analizada estuvo conformada por mujeres con diagnóstico confirmado de masa mamaria benigna o maligna, con edades comprendidas entre 18 y 96 años. A partir de esta población, se realizó una estratificación por edad en dos grupos etarios (≤ 49 años y ≥ 50 años), con el objetivo de identificar diferencias en los patrones asociados a la gravedad tumoral mediante análisis CHAID en función de la edad.

No se realizó muestreo, ya que se incluyó la totalidad de los registros disponibles con información completa en las variables de interés. Tras la exclusión de casos con datos incompletos, la población final analizada estuvo constituida por 956 pacientes. Este diseño permitió evaluar asociaciones entre características mamográficas (forma, márgenes y densidad) y la edad, en relación con el desenlace diagnóstico (benigno o maligno) dentro de cada subgrupo etario.

Variables y mediciones

Se utilizaron cinco variables principales registradas en la base de datos de masas mamográficas, de las cuales una fue la variable dependiente y cuatro actuaron como variables independientes. La





variable dependiente fue la gravedad tumoral, clasificada de forma binaria como benigna (0) o maligna (1), en función del diagnóstico histopatológico definitivo obtenido tras biopsia. Esta variable representó el desenlace que el modelo CHAID buscó determinar.

Entre las variables independientes se incluyó la edad de las pacientes, registrada en años como variable numérica discreta. La edad también fue empleada para el análisis estratificado en dos grupos (≤ 49 años y ≥ 50 años), con el fin de explorar diferencias en el rendimiento diagnóstico del modelo según el grupo etario. Esta segmentación se basó en evidencia clínica que describe variaciones relevantes en las características mamarias, el patrón de presentación tumoral y el desempeño diagnóstico de las técnicas de imagen a lo largo de la edad adulta, particularmente en torno a la transición climaterica, entendida como un proceso progresivo y no como un evento puntual.⁽⁵⁾ Dichas variaciones pueden influir en la expresión radiológica de las lesiones y, en consecuencia, en la capacidad del modelo para identificar y jerarquizar características asociadas a la gravedad tumoral.

Las otras tres variables correspondieron a atributos BI-RADS de las masas mamográficas. La forma del nódulo (*shape*) fue una variable categórica codificada en cuatro categorías: redonda (1), ovalada (2), lobulada (3) e irregular (4). Por su parte, la morfología de los márgenes (*margin*) también fue categórica, con cinco posibles clasificaciones: circunscrita (1), microlobulada (2), oscurecida (3), mal definida (4) y espiculada (5). Finalmente, la densidad de la masa (*density*) se registró como una variable ordinal con cuatro niveles: alta (1), iso (2), baja (3) y con contenido graso (4).

Análisis estadístico

Para identificar patrones asociados a la gravedad tumoral en mujeres con sospecha de cáncer de mama, se utilizó el algoritmo CHAID. Este método permite construir árboles de decisión dividiendo los datos en grupos similares; se basa en pruebas estadísticas de *ji* cuadrado aplicadas a variables categóricas u ordinales.⁽⁶⁾ Su principal ventaja es que muestra de forma clara y visual cómo distintas combinaciones de características influyen en el diagnóstico, facilitando la interpretación clínica incluso sin conocimientos estadísticos avanzados.

El modelo fue optimizado mediante validación cruzada de 10 particiones, una técnica que divide la base de datos en 10 subconjuntos para entrenar y probar el modelo en distintas combinaciones,





lo que permite estimar su capacidad de generalización a nuevos casos. Se calcularon dos tipos de error: el error de validación cruzada, que refleja el desempeño en datos no utilizados para construir el modelo, y el error de resustitución, que indica qué tan bien clasifica los datos con los que fue entrenado.

Para evaluar el rendimiento diagnóstico, se calcularon indicadores clásicos como sensibilidad (capacidad para detectar casos malignos), especificidad (capacidad para identificar casos benignos) y exactitud global (proporción total de clasificaciones correctas), todos obtenidos a partir de las matrices de clasificación generadas por los árboles CHAID. Todo el procesamiento estadístico se realizó con el software IBM-SPSS versión 27.

Aspectos bioéticos

Este estudio utilizó una base de datos secundaria de acceso abierto, la cual se encuentra anonimizada y está licenciada bajo *Creative Commons* (CC BY 4.0), lo que permite su utilización con fines académicos siempre que se cite adecuadamente a los autores. Al tratarse de datos secundarios y desidentificados, no fue necesaria la aprobación por un comité de ética ni el contacto directo con personas participantes, en concordancia con la Declaración de Helsinki y las directrices éticas aplicables a investigaciones con datos previamente recolectados. La base de datos y sus especificaciones pueden consultarse en el siguiente enlace:

<https://archive.ics.uci.edu/dataset/161/mammographic+mass>

RESULTADOS

La distribución de los nódulos mamarios analizados mostró que la forma más frecuente fue la irregular, seguida por las redondas y ovaladas, mientras que las lobuladas fueron menos comunes. En cuanto a los márgenes, predominó el tipo circunscrito, seguido por los mal definidos y oscurecidos, y en menor proporción se observaron márgenes microlobulados y espiculados. La mayoría de los nódulos presentaron una densidad baja, siendo escasos los de densidad alta o con contenido graso. El 53,25 % de los casos fueron clasificados como malignos, mientras que el 46,8





% correspondieron a lesiones benignas. La edad de las pacientes osciló entre 18 y 96 años, con una media de 55,49 años y una desviación estándar de 14,48 años (tabla 1).

Tabla 1 - Distribución de características de los nódulos mamarios (n= 956)

Variable	Categoría	n	%
Forma	Redonda	224	23,43
	Ovalada	211	22,07
	Lobular	95	9,94
	Irregular	426	44,56
Márgenes	Circunscritos	357	37,34
	Microlobulado	24	2,51
	Oscurecido	116	12,13
	Mal definidos	198	20,71
	Espiculado	261	27,31
Densidad	Alta	16	1,67
	Iso	59	6,17
	Baja	869	90,87
	Contenido graso	12	1,29
Gravedad	Benigno	447	46,75
	Maligno	509	53,25

Edad (años): media= 55,49; mediana= 57; desviación estándar= 14,48; rango= 78 (18–96).

En mujeres de 49 años o menos, el árbol CHAID mostró que la variable más influyente fue la forma del tumor. Las lesiones de forma irregular presentaron una alta proporción de malignidad (63,4 %), mientras que aquellas con forma redonda, ovalada o lobular fueron mayoritariamente benignas (89,5 %). Dentro de este grupo benigno, la edad permitió subdivisiones: las mujeres de 32 años o menos mostraron 100 % de benignidad y, aunque la benignidad se mantuvo elevada en los otros rangos etarios, disminuyó ligeramente en el grupo de 39,1 a 44 años. En mujeres de 50 años o más, la variable más determinante fue el margen tumoral: los márgenes espiculados o microlobulados mostraron una alta proporción de malignidad (85,4 %), seguidos por los márgenes mal definidos u oscurecidos (70,8 %). Los márgenes circunscritos se asociaron mayoritariamente a tumores benignos (81 %). Dentro del grupo de márgenes mal definidos, la forma y posteriormente la edad





diferenciaron el riesgo: las formas irregulares presentaron mayor malignidad, especialmente en mujeres mayores de 59 años (87 %). El modelo evidenció interacciones relevantes entre edad, forma y márgenes para la determinación de la gravedad tumoral (Fig. 1).

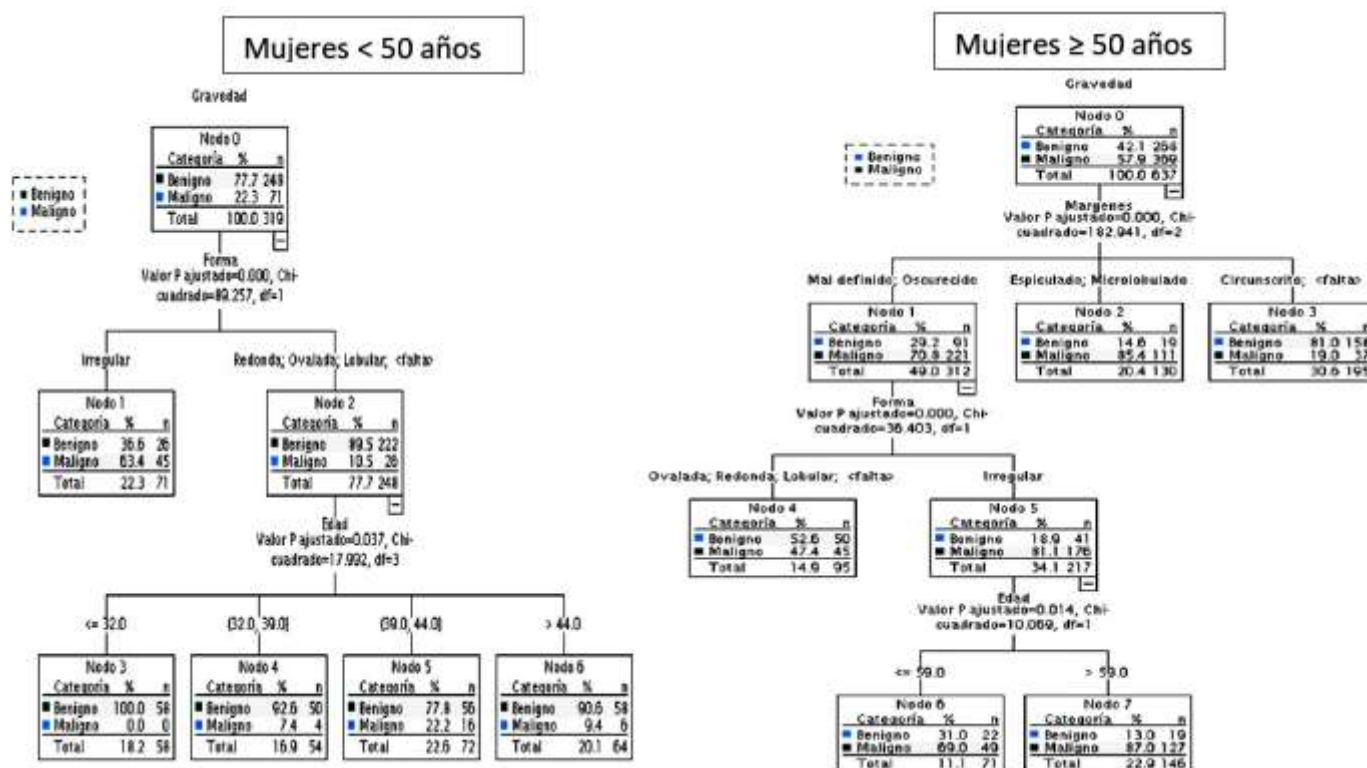


Fig. 1 - Árbol CHAID para determinar la gravedad tumoral mamaria según características morfológicas y edad.

El modelo CHAID mostró diferencias estructurales y de rendimiento según el grupo etario. En mujeres de 49 años o menos, el árbol incluyó como variables determinantes la forma del nódulo y la edad; generó un total de 7 nodos, de los cuales 5 fueron terminales, con una profundidad de 2 niveles. En mujeres de 50 años o más, se incorporaron tres variables: márgenes, forma y edad, con un modelo ligeramente más complejo, compuesto por 8 nodos totales, 5 terminales y una profundidad de 3 niveles. Ambos modelos utilizaron validación cruzada. El error de validación cruzada fue menor en mujeres jóvenes ($0,207 \pm 0,023$) en comparación con las mayores ($0,237 \pm$





0,017), lo que sugiere un mejor ajuste del modelo en este grupo. De forma similar, el error de resustitución también fue más bajo en el grupo joven ($0,163 \pm 0,021$ frente a $0,223 \pm 0,016$), lo que refuerza su mayor precisión en ese subgrupo (tabla 2).

Tabla 2 - Características del modelo CHAID por grupo etario

Característica del modelo	Mujeres ≤ 49 años	Mujeres ≥ 50 años
Variables independientes incluidas	Forma, Edad	Márgenes, Forma, Edad
Número total de nodos	7	8
Nodos terminales	5	5
Profundidad del árbol	2	3
Validación cruzada	Sí	Sí
Error de validación cruzada	$0,207 \pm 0,023$	$0,237 \pm 0,017$
Error de resustitución	$0,163 \pm 0,021$	$0,223 \pm 0,016$

El análisis del rendimiento diagnóstico del modelo CHAID reveló diferencias según el grupo etario. En mujeres de 49 años o menos, el modelo clasificó correctamente al 89,50 % de los casos benignos y al 63,40 % de los malignos; alcanzó una sensibilidad de 63,40 %, una especificidad de 89,50 % y una exactitud global del 83,70 %, con un error de validación cruzada de $20,7 \% \pm 2,30 \%$. En contraste, en mujeres de 50 años o más, el modelo presentó una mejor capacidad para identificar correctamente los casos malignos (sensibilidad: 77,80 %) pero menor especificidad (77,60 %), con una exactitud global del 77,70 % y un error de validación cruzada de $23,70 \% \pm 1,70 \%$. Estos resultados sugieren que el modelo CHAID fue más preciso en detectar lesiones benignas en mujeres jóvenes, mientras que mostró mayor eficacia para identificar malignidad en mujeres mayores (tabla 3).





Tabla 3 - Rendimiento diagnóstico del modelo CHAID en la determinación de gravedad tumoral mamaria por edad

Indicador	≤ 49 años	≥ 50 años
Benignos determinados correctamente	222	208
Benignos clasificados como malignos	26	60
Malignos determinados correctamente	45	287
Malignos clasificados como benignos	26	82
Clasificación correcta (%)	89,50 % (Benigno)	77,60 % (Benigno)
	63,40 % (Maligno)	77,80 % (Maligno)
Sensibilidad (%)	63,40 %	77,80 %
Especificidad (%)	89,50 %	77,60 %
Exactitud global (%)	83,70 %	77,70 %
Error de validación cruzada	20,70 % ± 2,3 %	23,70 % ± 1,70 %

DISCUSIÓN

El presente estudio empleó un modelo de árbol de decisión CHAID para identificar patrones jerárquicos asociados a la gravedad tumoral (benigno vs. maligno) en mujeres evaluadas por lesiones mamarias, con estratificación por grupos etarios. Este tipo de modelo permite detectar interacciones significativas entre variables categóricas, generando reglas de decisión fácilmente interpretables. Si bien factores de riesgo como la forma de la lesión, los márgenes y la edad son ampliamente conocidos en la literatura,⁽⁷⁾ el modelo CHAID permitió estructurarlos en una secuencia lógica de decisión, ofreciendo así una herramienta potencialmente útil para la práctica clínica, especialmente en contextos donde se requiere una interpretación visual rápida de datos complejos.

En mujeres menores de 50 años, el análisis identificó a la forma irregular como principal determinante de malignidad, con una tasa de clasificación correcta del 63,4 % para los casos malignos. Por el contrario, las formas redonda, ovalada o lobulada se asociaron con bajo riesgo, y dentro de este subgrupo, la edad permitió una discriminación adicional del riesgo. Este patrón sugiere que, en pacientes jóvenes, la morfología predomina como criterio diagnóstico, pero la edad puede afinar la sospecha clínica en lesiones de apariencia benigna. Esto podría explicarse por el





comportamiento biológico más agresivo de los tumores en mujeres jóvenes, que tienden a ser menos diferenciados, con mayor proliferación celular y características histológicas asociadas a peor pronóstico.^(8,9)

En contraste, en mujeres de 50 años o más, la variable más discriminativa fue la característica del margen de la lesión, destacando los márgenes espiculados y microlobulados como fuertemente asociados a malignidad (85,4 %), seguidos por márgenes mal definidos u oscurecidos (70,8 %). En este grupo, la forma y la edad agregaron valor pronóstico solo dentro de ciertos subgrupos, lo que sugiere que la interpretación radiológica en mujeres mayores debe priorizar el análisis de los márgenes tumorales como indicador clave. Esto puede atribuirse a que, con el envejecimiento, el tejido mamario sufre una involución fibroglandular, lo que mejora la visibilidad de las estructuras internas y permite una detección más precisa de patrones infiltrativos, como los márgenes espiculados, típicos de carcinomas invasivos de crecimiento radial.^(10,11)

La capacidad del modelo para reflejar estos patrones jerárquicos diferenciados por edad representa uno de sus principales aportes. A diferencia de los modelos tradicionales de regresión, los árboles de decisión como CHAID permiten visualizar combinaciones secuenciales de variables, facilitando la toma de decisiones clínicas especialmente en escenarios con limitaciones de tiempo o de experiencia especializada. Estos hallazgos también reflejan diferencias en el comportamiento tumoral entre grupos etarios: en mujeres jóvenes, los tumores malignos tienden a presentar una arquitectura más desorganizada e infiltrativa,⁽¹²⁾ lo que se traduce radiológicamente en formas irregulares; mientras que, en mujeres mayores, los cambios progresivos del tejido mamario por involución grasa y fibrosis aumentan la visibilidad de los márgenes espiculados o mal definidos, signos clásicos de invasión estromal.⁽¹³⁾ Asimismo, su alta precisión general (83,7 % en < 50 años y 77,7 % en ≥ 50 años) respalda su uso como herramienta complementaria en el triaje radiológico, al capturar estas diferencias fisiopatológicas de manera estructurada y accesible.

Estudios previos han demostrado la utilidad de modelos similares en la predicción de características tumorales o pronóstico. Por ejemplo, *Nam YH*⁽¹⁴⁾ aplicó un modelo de árbol para predecir el subtipo triple negativo de cáncer de mama, hallando que variables como la edad y la terapia neoadyuvante son determinantes importantes, alcanzando una precisión del 87,6 %. Asimismo, *Ponnuraja C* y





otros⁽¹⁵⁾ evaluaron la supervivencia en cáncer de mama usando árboles de decisión en la base SEER; encontraron tasas de precisión por encima del 95 % en ciertos grupos etarios de alto riesgo, validando su utilidad como herramienta de apoyo clínico.

Aunque los factores identificados en el presente estudio no son nuevos, su organización en reglas de decisión claras y contextualizadas por edad confiere al modelo un valor aplicado significativo. Las implicaciones del modelo CHAID son relevantes en contextos de diagnóstico temprano y apoyo a la evaluación diagnóstica, en la medida en que puede apoyar la interpretación mamográfica inicial al resaltar combinaciones de características asociadas a mayor probabilidad de malignidad según el grupo etario. En este sentido, los patrones observados sugieren que la forma irregular en mujeres de menor edad y la presencia de márgenes espiculados o mal definidos en mujeres de mayor edad podrían tener un mayor peso relativo en la evaluación diagnóstica, incluso cuando otros signos resultan menos evidentes. Asimismo, su representación visual podría facilitar la comprensión del riesgo mamográfico y constituir una herramienta complementaria en entornos con recursos limitados o menor experiencia interpretativa, sin sustituir el juicio clínico ni los protocolos diagnósticos establecidos.

Sin embargo, también se abren interrogantes para futuras investigaciones. ¿Qué mecanismos biológicos explican la mayor relevancia de la forma del nódulo en mujeres jóvenes y de los márgenes en mayores de 50 años? ¿Podrían otros factores como el tejido mamario, la densidad hormonal o el subtipo histológico influir en la expresión radiológica del tumor según la edad? Asimismo, ¿cómo se comportarían estos modelos en poblaciones con diferentes características genéticas, étnicas o con acceso desigual a la mamografía digital? Estas preguntas plantean la necesidad de estudios longitudinales, multicéntricos y con integración de datos moleculares para comprender a fondo la interacción entre edad, morfología tumoral y agresividad oncológica.

Entre las limitaciones del estudio se reconoce la ausencia de validación externa del modelo, así como la falta de inclusión de variables clínicas adicionales, como antecedentes familiares, hallazgos clínicos o resultados histopatológicos detallados. Asimismo, los resultados deben interpretarse en el marco de los protocolos diagnósticos y consideraciones éticas vigentes, sin implicar modificaciones en los criterios establecidos para la indicación de procedimientos





invasivos. Futuros estudios podrían explorar el desempeño del modelo CHAID en combinación con métodos de aprendizaje automático más complejos, así como evaluar su comportamiento en poblaciones diversas, con el objetivo de fortalecer su validez externa y su utilidad como herramienta complementaria de apoyo a la interpretación diagnóstica.

En conclusión, el modelo de árbol de decisión CHAID demostró ser una herramienta útil para estructurar reglas diagnósticas interpretables en la clasificación de lesiones mamarias benignas y malignas, identificando combinaciones jerárquicas de variables radiológicas relevantes según grupo etario. En mujeres menores de 50 años, la forma de la lesión fue el principal determinante de malignidad, mientras que en mujeres de 50 años o más, los márgenes espiculados y microlobulados destacaron como determinantes dominantes.

Aunque los factores implicados son conocidos en la literatura médica, el modelo aporta valor clínico al traducirlos en rutas de decisión claras, específicas y adaptables al contexto clínico. Esta aproximación puede facilitar la toma de decisiones en el tamizaje y evaluación inicial del cáncer de mama, especialmente en entornos con recursos limitados o menor experiencia radiológica. Se recomienda validar el modelo en otras poblaciones y explorar su integración en herramientas digitales de apoyo diagnóstico.

REFERENCIAS BIBLIOGRAFICAS

1. Arnold M, Morgan E, Rungay H, Mafra A, Singh D, Laversanne M, et al. Current and future burden of breast cancer: Global statistics for 2020 and 2040 [Internet]. *Breast*. 2022;66:15–23. DOI: <https://doi.org/10.1016/j.breast.2022.08.010>
2. Barrios CH. Global challenges in breast cancer detection and treatment [Internet]. *Breast*. 2022;62(Suppl 1):S3–6. DOI: <https://doi.org/10.1016/j.breast.2022.02.003>
3. Brown AL, Vijapura C, Patel M, De La Cruz A, Wahab R. Breast cancer in dense breasts: Detection challenges and supplemental screening opportunities [Internet]. *Radiographics*. 2023;43(10):e230024. DOI: <https://doi.org/10.1148/rg.230024>





4. Schopp JG, Polat DS, Arjmandi F, Hayes JC, Ahn RW, Sullivan K, et al. Imaging challenges in diagnosing triple-negative breast cancer [Internet]. *Radiographics*. 2023;43(10):e230027. DOI: <https://doi.org/10.1148/rg.230027>
5. Patel N, Shah S, Chacha S. A prospective comparative study of breast cancer in pre- and postmenopausal women [Internet]. *Natl J Physiol Pharm Pharmacol*. 2025;15(2):1. DOI: <https://doi.org/10.5455/njppp.2025.v15.i2.15>
6. Van Diepen M, Franses PH. Evaluating chi-squared automatic interaction detection [Internet]. *Inf Syst*. 2006;31(8):814–31. DOI: <https://doi.org/10.1016/j.is.2005.03.002>
7. Xu H, Xu B. Breast cancer: Epidemiology, risk factors and screening [Internet]. *Chin J Cancer Res*. 2023 [acceso: 24/07/2025];35(6):565–83. Disponible en: <https://doi.org/10.21147/j.issn.1000-9604.2023.06.02>
8. Henzler M, Willborn KC, Janni W, Huober J, Lukac S, Otremba B, et al. Oncologic outcomes of young breast cancer patients according to tumor biology [Internet]. *Cancers (Basel)*. 2025;17(8):1333. DOI: <https://doi.org/10.3390/cancers17081333>
9. Fabiano V, Mandó P, Rizzo M, Ponce C, Coló F, Loza M, et al. Breast cancer in young women presents with more aggressive pathologic characteristics: Retrospective analysis from an Argentine national database [Internet]. *JCO Glob Oncol*. 2020;6:639–46. DOI: <https://doi.org/10.1200/JGO.19.00228>
10. Li W, Zhao X, Han Q, Ren C, Gao S, Liu Y, et al. Relationship between breast tissue involution and breast cancer [Internet]. *Front Oncol*. 2025;15:1420350. DOI: <https://doi.org/10.3389/fonc.2025.1420350>
11. Lee SH, Moon WK. Glandular tissue component on breast ultrasound in dense breasts: A new imaging biomarker for breast cancer risk [Internet]. *Korean J Radiol*. 2022;23(6):574–80. DOI: <https://doi.org/10.3348/kjr.2022.0099>
12. Zhu JW, Charkhchi P, Adekunle S, Akbari MR. What is known about breast cancer in young women? [Internet]. *Cancers (Basel)*. 2023;15(6):1917. DOI: <https://doi.org/10.3390/cancers15061917>





13. Shroff NK, Posleman Monetto FE. Stromal fibrosis of the breast and the associated radiological findings [Internet]. Cureus. 2021;13(6):e15995. DOI: <https://doi.org/10.7759/cureus.15995>
14. Nam YH. A study on the factors and prediction model of triple-negative breast cancer for public health promotion [Internet]. Diagnostics (Basel). 2023;13(22):3486. DOI: <https://doi.org/10.3390/diagnostics13223486>
15. Ponnuraja C, Lakshmanan BC, Srinivasan V, Prasanth BK. Decision tree classification and model evaluation for breast cancer survivability: A data mining approach [Internet]. Biomed Pharmacol J. 2017;10(1):281–9. DOI: <https://doi.org/10.13005/bpj/1107>

Conflictos de interés

Se declara que no existen conflictos de interés.

Información financiera

El autor declara que no hubo subvenciones involucradas en este trabajo

Declaración de disponibilidad de datos

UCI Machine Learning Repository. Mammographic Mass Dataset. Disponible en: <https://archive.ics.uci.edu/dataset/161/mammographic+mass>. Los datos están disponibles según los términos de la licencia Creative Commons BY 4.0.