



REVISIÓN

Predicción de la sobrevida global y libre de recaída en pacientes con cáncer de mama triple negativo a través de agrupaciones basadas en el aprendizaje automático sobre datos clínicos



Juan Pablo Alzate-Granados* y Luis Fernando Niño

Departamento de Patología, Facultad de Medicina, Universidad Nacional de Colombia - Sede Bogotá, Bogotá, Colombia

Departamento de Ingeniería de Sistemas e Industrial, Facultad de Ingeniería, Grupo de Investigación LISI, Universidad Nacional de Colombia - Sede Bogotá, Bogotá, Colombia

Recibido el 7 de enero de 2025; aceptado el 9 de marzo de 2025

Disponible en Internet el 4 de abril de 2025

PALABRAS CLAVE

Cáncer de mama triple negativo;
Aprendizaje automático;
Sobrevida global;
Recaída;
Biomarcadores

Resumen

Introducción: el cáncer de mama triple negativo (TNBC) representa del 15 al 20% de los casos de cáncer de mama, caracterizándose por la ausencia de receptores hormonales y HER2, lo que limita las opciones terapéuticas a la quimioterapia citotóxica. Este subtipo presenta alta agresividad, recaídas frecuentes y peor pronóstico. La integración de datos clínicos y moleculares mediante algoritmos de aprendizaje automático ofrece una oportunidad para mejorar la predicción de desenlaces clínicos y personalizar el manejo de esta enfermedad. **Objetivo:** desarrollar modelos predictivos para estimar la sobrevida global (SG) y libre de recaída (SLR) en pacientes con TNBC.

Métodos: se realizó un estudio de cohorte retrospectiva con 4.808 pacientes diagnosticadas entre 2012 y 2024. Las pacientes fueron agrupadas usando el algoritmo *k-prototypes*, que maneja datos mixtos. Se analizaron variables demográficas, clínicas, biomarcadores y desenlaces clínicos. Los análisis multivariados y modelos de riesgos proporcionales de Cox evaluaron la asociación entre los clúster y los desenlaces de SG y SLR.

Resultados: se identificaron 4 clústeres de pacientes. El grupo de mayor riesgo (clúster 3) presentó la mayor mortalidad (42,3%; HR = 1,94; IC 95%: 1,63-2,30) y recaída (54,25%; HR = 1,68; IC 95%: 1,45-1,95), mientras que el clúster 0 mostró los mejores desenlaces (mortalidad del 22,51%). Factores como mutaciones en *PIK3CA* (HR = 1,535; $p=0,001$) y *TP53* (HR = 1,180; $p=0,023$) se asociaron con desenlaces adversos.

Discusión: los hallazgos refuerzan la utilidad de la estratificación basada en clúster para predecir desenlaces y guiar intervenciones personalizadas en el TNBC.

© 2025 SESPM. Publicado por Elsevier España, S.L.U. Se reservan todos los derechos, incluidos los de minería de texto y datos, entrenamiento de IA y tecnologías similares.

* Autor para correspondencia.

Correo electrónico: jpalzategr@unal.edu.co (J.P. Alzate-Granados).

KEYWORDS

Triple-negative breast cancer;
Machine learning;
Overall survival;
Recurrence;
Biomarkers

Predicting overall and recurrence-free survival in patients with triple-negative breast cancer using machine learning-based clustering of clinical data

Abstract

Introduction: Triple-negative breast cancer (TNBC) accounts for 15%–20% of all breast cancers and is characterised by the absence of hormone and HER2 receptors, limiting therapeutic options to cytotoxic chemotherapy. This subtype is highly aggressive, with frequent relapses and a poor prognosis. The integration of clinical and molecular data using machine learning algorithms offers an opportunity to improve the prediction of clinical outcomes and personalise the management of this disease.

Objective: To develop predictive models to estimate overall survival (OS) and relapse-free survival (RFS) in patients with TNBC.

Methods: A retrospective cohort study was conducted with 4,808 patients diagnosed between 2012 and 2024. Patients were grouped using the k-prototypes algorithm for mixed data. Demographic and clinical variables, biomarkers, and clinical outcomes were analysed. Multivariate analyses and Cox proportional hazards models were used to assess the association between clusters and OS and RFS outcomes.

Results: Four clusters of patients were identified. The highest risk group (cluster 3) had the highest mortality (42.3%; HR: 1.94; 95% CI: 1.63–2.30) and relapse (54.25%; HR: 1.68; 95% CI: 1.45–1.95), whereas cluster 0 had the best outcomes (mortality of 22.51%). Factors such as mutations in PIK3CA (HR: 1.535; $p=0.001$) and TP53 (HR: 1.180; $p=0.023$) were associated with adverse outcomes.

Discussion: The results support the utility of cluster-based stratification to predict outcomes and guide personalised interventions in TNBC.

© 2025 SESPM. Published by Elsevier España, S.L.U. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Introducción

El cáncer de mama triple negativo (TNBC, por sus siglas en inglés) es una forma de cáncer de mama caracterizada por la ausencia de expresión de los receptores hormonales de estrógeno (ER) y progesterona (PR), así como del receptor HER2 del factor de crecimiento epidérmico humano. Este perfil molecular lo distingue de otros subtipos y se asocia con una biología tumoral particularmente agresiva. Representa aproximadamente entre el 15 y el 20% de todos los casos de cáncer de mama, lo que lo convierte en una entidad de relevancia epidemiológica y clínica. El TNBC afecta de manera desproporcionada a mujeres más jóvenes y a ciertos grupos raciales y étnicos, lo que sugiere la influencia de factores genéticos y ambientales específicos^{1,2}.

A diferencia de otros subtipos de cáncer de mama, las opciones terapéuticas para el TNBC son limitadas, debido a la ausencia de las dianas moleculares que posibilitan el uso de terapias hormonales o dirigidas contra HER2. En la práctica clínica, la quimioterapia citotóxica sigue siendo la principal estrategia terapéutica, a pesar de sus limitaciones en términos de eficacia y tolerancia a largo plazo. Este subtipo se caracteriza, además, por una alta tasa de recaídas y una peor sobrevida global en comparación con otros tipos de cáncer de mama. Estas características resaltan la necesidad de desarrollar enfoques innovadores y personalizados para su manejo^{3,4}.

La creciente disponibilidad de datos clínicos, moleculares y genómicos presenta una oportunidad para mejorar la

comprensión de la heterogeneidad del TNBC y sus implicaciones pronósticas y terapéuticas. Sin embargo, los enfoques actuales aún no logran explotar completamente esta riqueza de información para generar modelos predictivos que sean aplicables en la práctica clínica, subrayando la urgencia de avanzar en este campo^{5,6}.

A pesar de los avances significativos en la caracterización molecular y genética del TNBC, persisten desafíos importantes que limitan la predicción precisa de los desenlaces clínicos en esta población de pacientes. Los modelos actuales de predicción de sobrevida y recurrencia se basan predominantemente en enfoques estadísticos convencionales, como modelos de regresión, que están diseñados para manejar un número limitado de variables y asumen relaciones lineales entre estas. Estas limitaciones metodológicas dificultan la captura de la complejidad inherente de las interacciones entre múltiples factores clínicos, moleculares y epidemiológicos^{7,8}.

Además, muchos de los sistemas de predicción existentes se centran exclusivamente en subgrupos de variables específicas, como biomarcadores genómicos o características histológicas, sin integrar datos clínicos holísticos que incluyan factores demográficos, comorbilidades y características del tratamiento. Esta falta de integración impide identificar subgrupos clínicamente relevantes que puedan beneficiarse de intervenciones personalizadas.

Otro vacío importante radica en la falta de herramientas capaces de manejar la heterogeneidad molecular y clínica del TNBC. A pesar de los esfuerzos por clasificar el TNBC en

subtipos moleculares basados en perfiles transcriptómicos, estos enfoques no han sido ampliamente validados ni aplicados en contextos clínicos debido a su complejidad y a la ausencia de estándares consistentes. Esto limita su utilidad para guiar decisiones terapéuticas en tiempo real⁹.

Finalmente, la falta de métodos que utilicen técnicas avanzadas de aprendizaje automático, capaces de identificar patrones ocultos y relaciones no lineales en datos de alta dimensión, representa una barrera crítica para el desarrollo de modelos predictivos más precisos y generalizables. La implementación de tales modelos podría transformar el manejo del TNBC, permitiendo una estratificación más precisa de los pacientes y el diseño de estrategias de tratamiento individualizadas¹⁰.

El objetivo principal es desarrollar modelos predictivos que combinen las agrupaciones generadas con datos históricos para estimar la probabilidad de supervivencia global (SG) y libre de recaída (SLR) en pacientes con TNBC. Esto se logrará mediante la integración de datos clínicos, demográficos y terapéuticos, utilizando algoritmos de aprendizaje automático.

Para abordar esta problemática se emplearán algoritmos no supervisados para generar agrupaciones basadas en características clínicas y moleculares, utilizando datos de una cohorte retrospectiva de pacientes con TNBC.

Métodos

Diseño del estudio

Se realizó un estudio de cohorte retrospectiva para evaluar las asociaciones entre características clínicas, demográficas y biomarcadores seleccionados con la SG y la SLR en pacientes con TNBC. Este diseño permitió aprovechar datos históricos para identificar factores predictivos y desenlaces clínicos en esta población (fig. 1).

Criterios de selección de la población

Se incluyeron pacientes mayores de 18 años con diagnóstico confirmado de TNBC, realizado entre 2012 y 2024, a través de estudios histopatológicos que determinaron la ausencia de receptores de estrógeno, progesterona y HER2. Las pacientes debían haber recibido tratamiento primario con intención curativa y contar con al menos 3 años de seguimiento clínico documentado. Se excluyeron los casos con estadios metastásicos al diagnóstico inicial y aquellos con datos clínicos incompletos o inconsistentes.

Variables

Las variables dependientes analizadas fueron la SG, definida como el tiempo desde el diagnóstico hasta el fallecimiento por cualquier causa; y la SLR, que incluyó el intervalo entre el inicio del tratamiento y la recaída clínicamente documentada. Las variables independientes abarcaron características sociodemográficas, como edad y lugar de residencia; factores clínicos, como índice de masa corporal, comorbilidades registradas en el índice de Charlson, y tipo de tratamiento recibido (cirugía, quimioterapia, radioterapia). Además, se incluyeron biomarcadores relevantes como estado mutacional de BRCA1/2, niveles de Ki67, expresión de PD-L1 y HER2-low.

Fuentes de información y medición

La información provino de bases de datos de centros oncológicos especializados y registros electrónicos de una aseguradora nacional. Los datos fueron extraídos mediante sistemas estructurados y sometidos a validación mediante la metodología SEMMA, que incluyó limpieza, selección y evaluación. Las mediciones de las variables se realizaron siguiendo protocolos estandarizados.

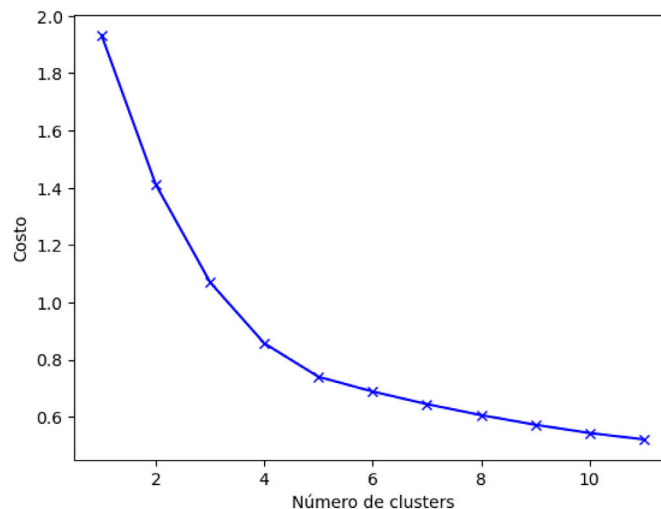


Figura 1 Método del código para determinar el número óptimo de clusters.

Cálculo del tamaño de la muestra

Se utilizó la totalidad de la cohorte disponible, compuesta por aproximadamente 4.000 pacientes. Este enfoque maximizó la potencia estadística y minimizó los errores de tipo II, aprovechando la naturaleza retrospectiva del diseño que permite el análisis de datos preexistentes.

Análisis estadístico

El análisis descriptivo incluyó medidas de tendencia central y dispersión para variables continuas y frecuencias absolutas y proporciones para variables categóricas. Las comparaciones entre pacientes con y sin desenlaces se realizaron utilizando pruebas t de Student o U de Mann-Whitney para variables continuas, y chi-cuadrado o prueba exacta de Fisher para variables categóricas. Para las variables dependientes, se ajustaron modelos de regresión de riesgos proporcionales de Cox, con estimación de Hazard ratios (HR) ajustados.

Uso del algoritmo *k-prototypes*

Para identificar subgrupos homogéneos dentro de la población con TNBC, se empleó el algoritmo de *k-prototypes*. Este método combina la distancia euclidiana para variables continuas y la distancia de coincidencia simple para variables categóricas, permitiendo manejar de forma integrada datos clínicos mixtos. El proceso incluyó la selección de un número inicial de grupos (*k*), la inicialización aleatoria de centroides y la iteración hasta lograr una convergencia estable. El número óptimo de grupos se determinó utilizando el método del codo (*Elbow Method*). Las variables seleccionadas para el análisis fueron previamente estandarizadas y sometidas a técnicas de reducción de dimensionalidad para evitar redundancias y optimizar el rendimiento del modelo.

Consideraciones éticas

El estudio fue aprobado por el Comité de Ética de la Universidad Nacional de Colombia y se garantizó el anonimato de los datos según la normativa nacional e internacional. La información utilizada fue previamente anonimizada y los investigadores no tuvieron acceso a identificadores personales.

Resultados

El estudio incluyó un total de 4.808 pacientes con TNBC. La mediana de edad fue de 51 años (P25%-P75%: 42-59 años), con un IMC mediano de 26,2 (23,03-29,57). El puntaje de Charlson presentó una mediana de 4³⁻⁶. En cuanto al estado funcional, el 81,8% de los pacientes obtuvo la máxima puntuación en el índice de Karnofsky (100), mientras que el 38,91% presentó un ECOG de 0, reflejando buen desempeño funcional.

La mayoría de los pacientes recibió tumorectomía (26,98%) o mastectomía simple (22,5%), y el 54,58% recibió quimioterapia adyuvante. Los agentes más utilizados fueron

taxanos (49,6%) y antraciclinas (39,18%). La radioterapia convencional fue empleada en el 59,55% de los casos, mientras que el 73,54% recibió esquemas de dosis densas. Los desenlaces mostraron una mortalidad del 28,81% y una recaída en el 40,95%, con un tiempo de seguimiento mediano de 8,46 años (6,72-10,25) (tabla 1).

La clasificación TNM mostró que la mayoría presentaron tumores T3 (39,6%) o T4 (31,68%). En la evaluación nodal, el 34,05% fueron clasificados como N2 y el 31,28% como N3. El estadio tumoral más frecuente fue IIIC (31,87%), seguido por IIIA (22,59%) y IIIB (19,22%) (tablas 1 y 2).

El análisis de biomarcadores indicó que el 87,48% de los pacientes tenían niveles de Ki67 mayores o iguales al 15%. Las mutaciones en BRCA 1/2 se detectaron en el 50,08% de los casos, mientras que mutaciones en TP53 estuvieron presentes en el 15,08%. El 5,82% de los pacientes presentó características HER2-low, y la expresión mediana de PD-L1 fue del 2,522% (P25%-P75%: 1,2885-4,7955).

Se seleccionaron 4 clústeres como el número óptimo para la agrupación, ya que balancea adecuadamente la reducción del costo con la simplicidad del modelo (fig. 1).

La distribución de los pacientes en 4 clústeres evidenció características distintivas que permiten describirlos según la gravedad de la condición clínica y la respuesta al tratamiento. El clúster 0, denominado «Grupo de bajo riesgo y alta funcionalidad», agrupa principalmente a pacientes jóvenes (edad promedio: 40,10 años), con baja carga de comorbilidad (Charlson: 4,42) y excelente estado funcional (Karnofsky: 100, ECOG: 0). Estos pacientes, aunque con menor prevalencia de comorbilidades graves, reciben tratamientos agresivos como esquemas de dosis densas. El clúster 1, identificado como «Grupo de moderado riesgo con alta respuesta a inmunoterapia», incluye pacientes con edad intermedia (48,67 años), moderada comorbilidad (Charlson: 4,68) y alta expresión de PD-L1, lo que favorece su respuesta a tratamientos inmunoterapéuticos, como atezolizumab y pembrolizumab.

El clúster 2, denominado «Grupo de alto riesgo y alta carga de comorbilidad», está compuesto por pacientes de mayor edad (60,98 años) y con elevada carga de comorbilidades (Charlson: 5,35), destacando la prevalencia de EPOC, enfermedad vascular periférica y antecedentes familiares. Este grupo requiere un manejo intensivo, reflejado en un mayor número de ciclos prequirúrgicos. Finalmente, el clúster 3, descrito como «Grupo de muy alto riesgo con alta discapacidad funcional», presenta la mayor carga de comorbilidades (Charlson: 6,11) y un marcado deterioro funcional (ECOG 3-4). Además, se caracteriza por estadios avanzados (predominancia de estadio IIIC), alta tasa de mutación BRCA 1/2 (71,03%) y complejidad en el manejo debido a la agresividad del cáncer y la menor expresión de PD-L1 en comparación con el clúster 1.

La tabla 3 y la figura 2 muestran la asociación entre los clústeres y los desenlaces de muerte y recaída en los pacientes con TNBC. En términos de mortalidad, el clúster 0 (grupo de bajo riesgo y alta funcionalidad) sirvió como referencia, con un 22,51% de muertes. Comparativamente, los pacientes en el clúster 1 tuvieron un riesgo ligeramente mayor de muerte (HR= 1,08; IC 95%: 0,89-1,32), aunque esta diferencia no fue significativa. En el clúster 2, el riesgo de muerte aumentó significativamente (HR= 1,52; IC 95%: 1,34-1,72), con una tasa de mortalidad del 33,2%. El clúster 3

Tabla 1 Características de los pacientes

| Variable | Mediana | P25%-P75% |
|----------------------------------|----------|-------------------|
| <i>Edad</i> | 51 | 42-59 |
| <i>Peso (Kg)</i> | 68,62 | 60,975-76,28 |
| <i>Talla (cm)</i> | 161,69 | 156,97-166,06 |
| <i>IMC</i> | 26,20454 | 23,03-29,57 |
| <i>Charlson</i> | 4 | 3-6 |
| <i>Karnofsky</i> | | |
| | <i>n</i> | <i>Porcentaje</i> |
| 40 | 1 | 0,02 |
| 50 | 13 | 0,27 |
| 60 | 99 | 2,06 |
| 70 | 57 | 1,19 |
| 80 | 329 | 6,84 |
| 90 | 376 | 7,82 |
| 100 | 3.933 | 81,8 |
| <i>Escala ECOG</i> | | |
| | <i>n</i> | <i>Porcentaje</i> |
| 0 | 1.871 | 38,91 |
| 1 | 360 | 7,49 |
| 2 | 1.649 | 34,3 |
| 3 | 301 | 6,26 |
| 4 | 627 | 13,04 |
| <i>Clasificación TNM</i> | | |
| <i>T</i> | | |
| | <i>n</i> | <i>Porcentaje</i> |
| T0 | 8 | 0,17 |
| T1 | 499 | 10,38 |
| T2 | 821 | 17,08 |
| T3 | 1.904 | 39,6 |
| T4 | 1.523 | 31,68 |
| Tx | 16 | 0,33 |
| TiS | 37 | 0,77 |
| <i>N</i> | | |
| | <i>n</i> | <i>Porcentaje</i> |
| N0 | 1.141 | 23,73 |
| N1 | 452 | 9,4 |
| N2 | 1.637 | 34,05 |
| N3 | 1.504 | 31,28 |
| Nx | 74 | 1,54 |
| | <i>n</i> | <i>Porcentaje</i> |
| <i>Diseminación Leptomenínea</i> | 63 | 1,31 |
| <i>Estadio tumoral</i> | | |
| | <i>n</i> | <i>Porcentaje</i> |
| 0 | 37 | 0,78 |
| IA | 258 | 5,47 |
| IIA | 500 | 10,6 |
| IIB | 447 | 9,47 |
| IIIA | 1.066 | 22,59 |
| IIIB | 907 | 19,22 |
| IIIC | 1.504 | 31,87 |
| <i>Niveles de Ki67</i> | | |
| | <i>n</i> | <i>Porcentaje</i> |
| Mayor o igual a 15% | 4.206 | 87,48 |
| Menor a 15% | 602 | 12,52 |
| <i>Mutación BRCA 1/2</i> | | |
| | <i>n</i> | <i>Porcentaje</i> |
| No | 2.400 | 49,92 |
| Sí | 2.408 | 50,08 |

Tabla 1 (continuación)

| Variable | Mediana | P25%-P75% |
|--|----------------|-------------------|
| <i>Otras mutaciones</i> | | |
| | <i>n</i> | <i>Porcentaje</i> |
| No | 3.804 | 79,12 |
| PIK3CA | 182 | 3,79 |
| PTEN | 97 | 2,02 |
| TP53 | 725 | 15,08 |
| HER2-low | | |
| | <i>n</i> | <i>Porcentaje</i> |
| No | 4.528 | 94,18 |
| Sí | 280 | 5,82 |
| | <i>Mediana</i> | <i>P25%-P75%</i> |
| Expresión PD-L1 (%) | 2,522 | 1,2885-4,7955 |
| ECOG: Eastern Cooperative Oncology Group; TNM: tumor (T), node (N) y metástasis (M). | | |

presentó el mayor riesgo de muerte (HR= 1,94; IC 95%: 1,63-2,3), con una tasa de mortalidad del 42,3%.

En cuanto a recaídas, el clúster 0 mostró una tasa de recaída del 34,43%, utilizada como referencia. El clúster 1 presentó un riesgo ligeramente superior (HR= 1,11; IC 95%: 0,95-1,3), sin diferencias estadísticamente significativas. El clúster 2 mostró un incremento significativo en el riesgo de recaída (HR= 1,37; IC 95%: 1,24-1,52), con una tasa de recaída del 45,33%. Finalmente, el clúster 3 registró el mayor riesgo de recaída (HR= 1,68; IC 95%: 1,45-1,95), con un 54,25% de los pacientes experimentando este desenlace.

Discusión

En este estudio, utilizamos algoritmos de aprendizaje automático no supervisados para identificar subgrupos homogéneos de pacientes con TNBC y evaluar su asociación con la SG y la SLR. Los resultados mostraron que las agrupaciones generadas permitieron estratificar a los pacientes según el riesgo clínico y desenlaces, evidenciando diferencias significativas en la mortalidad y recaída entre los 4 clústeres identificados. Los grupos de mayor riesgo (clústeres 2 y 3) presentaron una alta carga de comorbilidades, deterioro funcional y peor pronóstico clínico, mientras que los grupos de menor riesgo (clústeres 0 y 1) reflejaron mejores desenlaces y respuesta a ciertos tratamientos, como inmunoterapia.

Los hallazgos de este estudio refuerzan lo descrito en investigaciones previas acerca de la heterogeneidad biológica y clínica del TNBC y su influencia en los desenlaces. Se ha establecido que características clínicas como la edad avanzada y una alta carga de comorbilidades incrementan el riesgo de mortalidad y recaída¹¹. Por otro lado, los pacientes con la característica HER2-low también mostraron una reducción significativa en el riesgo de mortalidad y recaída, destacando su potencial como un subgrupo clínicamente manejable, según lo evidenciado en estudios recientes sobre la eficacia de terapias dirigidas^{12,13}.

En cuanto a mutaciones específicas, los resultados confirmaron que la presencia de mutaciones en *PIK3CA* y *TP53* incrementa sustancialmente el riesgo de desenlaces adversos. Estos resultados concuerdan con investigaciones

Tabla 2 Tratamientos y resultados de los pacientes

| Tipo de cirugía | n | Porcentaje |
|--|----------|-------------------|
| Cuadrantectomía | 890 | 18,51 |
| Mastectomía oncoplástica | 586 | 12,19 |
| Mastectomía radical modificada | 824 | 17,14 |
| Mastectomía simple | 1.082 | 22,5 |
| Otras | 129 | 2,68 |
| Tumorectomía | 1.297 | 26,98 |
| Tipo de quimioterapia | n | Porcentaje |
| Adyuvante | 2.624 | 54,58 |
| Neoadyuvante | 2.184 | 45,42 |
| Nombre de la quimioterapia | n | Porcentaje |
| Antraciclinas | 1.884 | 39,18 |
| Otros | 539 | 11,21 |
| Taxanos | 2.385 | 49,6 |
| | Mediana | P25%-P75% |
| Número de ciclos prequirúrgicos | 6 | 3-6 |
| Radioterapia | n | Porcentaje |
| Convencional | 2.863 | 59,55 |
| Hiperfraccionada | 480 | 9,98 |
| Hipofraccionada | 1.250 | 26 |
| No | 215 | 4,47 |
| Inmunoterapia | n | Porcentaje |
| Atezolizumab | 453 | 9,42 |
| No | 3.713 | 77,23 |
| Otros | 98 | 2,04 |
| Pembrolizumab | 544 | 11,31 |
| Esquemas de dosis densas | n | Porcentaje |
| No | 1.272 | 26,46 |
| Sí | 3.536 | 73,54 |
| Uso de platinos | n | Porcentaje |
| Carboplatino | 2.318 | 48,21 |
| Cisplatino | 1.493 | 31,05 |
| No | 916 | 19,05 |
| Otros | 17 | 0,35 |
| Oxaliplatino | 64 | 1,33 |
| Desenlaces | n | Porcentaje |
| Muerte | 1.385 | 28,81 |
| Recaída | 1.969 | 40,95 |
| | Mediana | P25%-P75% |
| Tiempo de seguimiento (años) | 8,46 | 6,72-10,25 |
| Tiempo a la recaída (años) | 8,145 | 6,32-9,98 |

que identifican estas alteraciones genéticas como indicadores de agresividad tumoral y resistencia a tratamientos estándar. No obstante, la interpretación de estos resultados debe considerar las diferencias en los métodos de detección de mutaciones entre estudios, lo que podría limitar la comparabilidad de los resultados^{14,15}.

Por otro lado, los hallazgos relacionados con las intervenciones terapéuticas proporcionan evidencia

adicional para el debate en curso sobre la personalización del tratamiento en el TNBC. Las mastectomías radicales, particularmente la radical modificada, redujeron significativamente el riesgo de mortalidad (HR= 0,741; IC 95%: 0,633–0,868; $p<0,001$) y recaída (HR= 0,771; IC 95%: 0,659–0,902; $p=0,001$). Estas observaciones refuerzan estudios que sugieren que este tipo de abordaje quirúrgico podría ser más beneficioso en pacientes seleccionados^{16,17}. Asimismo, los pacientes del clúster 1, con alta expresión de PD-L1, respondieron favorablemente a tratamientos inmunoterapéuticos como atezolizumab y pembrolizumab, en línea con estudios que destacan la eficacia de estas terapias en subgrupos moleculares específicos. Sin embargo, los resultados para los grupos de alto riesgo, como el clúster 3, resaltan las limitaciones de los tratamientos agresivos en pacientes con deterioro funcional severo (ECOG ≥ 3), lo que pone de manifiesto la importancia de considerar no solo las características tumorales sino también las condiciones generales del paciente al planificar estrategias terapéuticas¹⁸.

En contraste con algunos reportes, nuestros resultados no evidenciaron un beneficio significativo del tratamiento con esquemas de quimioterapia neoadyuvante en la reducción de la recaída, sugiriendo que factores como la selección de pacientes y características del tumor podrían influir en la eficacia de esta intervención. Esto subraya la necesidad de estudios adicionales que evalúen la interacción entre los tratamientos y las características individuales de los pacientes con TNBC².

Entre las principales limitaciones del estudio se encuentran su diseño retrospectivo y la dependencia de bases de datos históricas, lo que podría introducir sesgos en la calidad y completitud de los datos. Además, la exclusión de pacientes con datos incompletos podría limitar la generalización de los hallazgos a poblaciones más diversas. Sin embargo, el uso de una gran cohorte de más de 4.000 pacientes fortalece la validez estadística de los resultados y permite una evaluación consistente de la heterogeneidad del TNBC. Asimismo, la integración de datos clínicos y moleculares mediante técnicas avanzadas de aprendizaje automático representa una fortaleza metodológica que supera las limitaciones de los enfoques convencionales.

Los hallazgos de este estudio subrayan la importancia de utilizar enfoques basados en aprendizaje automático para identificar subgrupos de pacientes con TNBC y guiar estrategias terapéuticas personalizadas. La estratificación basada en clúster no solo mejora la predicción de desenlaces, sino que también puede optimizar la asignación de recursos clínicos y el diseño de ensayos clínicos dirigidos. En el ámbito de la investigación, este enfoque ofrece una plataforma para validar y explorar biomarcadores emergentes y su interacción con terapias innovadoras, como inmunoterapias y tratamientos dirigidos.

Conclusión

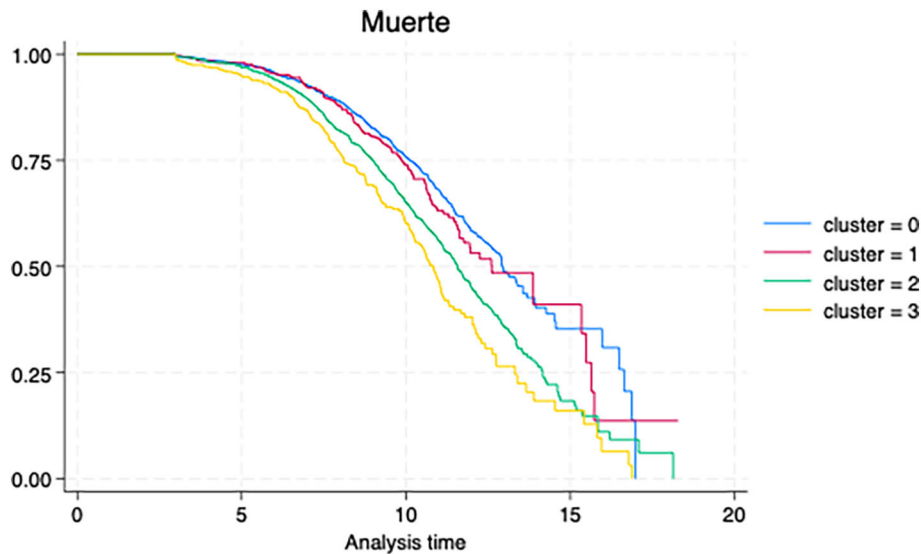
Los resultados de este estudio demuestran que los algoritmos no supervisados, específicamente el método de *k-prototypes*, permiten una estratificación efectiva de pacientes con TNBC en subgrupos clínicamente diferenciados con implicaciones críticas para la predicción

Tabla 3 Asociación de muerte y recaída por los clústeres

| Variable | | Clúster 0 (n=1.937) | Clúster 1 (n=508) | Clúster 2 (n=1.928) | Clúster 3 (n=435) |
|----------|-------------|------------------------|----------------------|------------------------|----------------------|
| Muerte | n (%)* | 436 (22,51) | 125 (24,61) | 640 (33,2) | 184 (42,3) |
| | HR (IC 95%) | Ref. | 1,08 (0,89 - 1,32) | 1,52 (1,34-1,72) | 1,94 (1,63-2,3) |
| Recaída | n (%) | 667 (34,43) | 192 (37,8) | 874 (45,33) | 236 (54,25) |
| | HR (IC 95%) | Ref. | 1,11 (0,95-1,3) | 1,37 (1,24-1,52) | 1,68 (1,45-1,95) |

Ref.: Categoría de referencia

* Valor de p menor a 0,05.

**Figura 2** Sobrevida global entre los diferentes clústeres.

de la SG y SLR. La identificación de 4 clústeres, con variaciones significativas en las tasas de mortalidad y recaída, subraya la heterogeneidad de esta enfermedad y la necesidad de enfoques personalizados. Particularmente, el clúster 3, caracterizado por alto riesgo y deterioro funcional, presentó la mayor mortalidad y recaídas, mientras que el clúster 0 evidenció los mejores desenlaces clínicos.

Consentimiento informado

Este estudio fue realizado utilizando datos previamente anonimizados, garantizando la protección de la identidad y privacidad de los participantes.

Financiación

Este trabajo no recibió financiación específico de ninguna entidad pública, privada o sin ánimo de lucro.

Contribución de los autores

Juan Pablo Alzate Granados: Concepción del estudio, análisis de datos y redacción del manuscrito. Luis Fernando Niño:

Supervisión del análisis estadístico, revisión crítica del manuscrito y apoyo logístico en la recolección de datos.

Conflicto de intereses

Los autores declaran no tener conflictos de intereses relacionados con esta investigación.

Bibliografía

1. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest.* 2011;121(7):2750–67.
2. Jézéquel P, Loussouarn D, Guérin-Charbonnel C, Campion L, Vanier A, Gouraud W, et al. Gene-expression molecular subtyping of triple-negative breast cancer tumours: importance of immune response. *Breast Cancer Res BCR.* 2015(17):43.
3. Quist J, Mirza H, Cheang MCU, Telli ML, O'Shaughnessy JA, Lord CJ, et al. A four-gene decision tree signature classification of triple-negative breast cancer: implications for targeted therapeutics. *Mol Cancer Ther.* 2019;18(1):204–12.
4. Garrido-Castro AC, Lin NU, Polyak K. Insights into molecular classifications of triple-negative breast cancer: improving patient selection for treatment. *Cancer Discov.* 2019;9(2):176–98.

5. Gong Y, Ji P, Yang YS, Xie S, Yu TJ, Xiao Y, et al. Metabolic-pathway-based subtyping of triple-negative breast cancer reveals potential therapeutic targets. *Cell Metab.* 2021;33(1): 51–64.e9.
6. Gregório AC, Lacerda M, Figueiredo P, Simões S, Dias S, Moreira JN. Therapeutic implications of the molecular and immune landscape of triple-negative breast cancer. *Pathol Oncol Res POR.* 2018;24(4):701–16.
7. Carlino F, Diana A, Piccolo A, Ventriglia A, Bruno V, De Santo I, et al. Immune-based therapy in triple-negative breast cancer: from molecular biology to clinical practice. *Cancers.* 2022;14 (9):2102.
8. Hudis CA, Gianni L. Triple-negative breast cancer: an unmet medical need. *Oncologist.* 2011;16(S1):1–11.
9. Denkert C, Liedtke C, Tutt A, Von Minckwitz G. Molecular alterations in triple-negative breast cancer—the road to new treatment strategies. *Lancet.* 2017;389(10087):2430–42.
10. Jerusalem G, Collignon J, Schroeder H, Lousberg L. Triple-negative breast cancer: treatment challenges and solutions. *Breast Cancer Targets Ther.* 2016;93.
11. Yu G, Li X, He TF, Gruosso T, Zuo D, Souleimanova M, et al. Predicting relapse in patients with triple negative breast cancer (TNBC) using a deep-learning approach. *Front Physiol.* 2020(11): 511071.
12. Quist J, Taylor L, Staaf J, Grigoriadis A. Random forest modelling of high-dimensional mixed-type data for breast cancer classification. *Cancers.* 2021;13(5):991.
13. Wang DY, Done SJ, McCreedy DR, Boerner S, Kulkarni S, Leong WL. A new gene expression signature, the ClinicoMolecular triad Classification, may improve prediction and prognostication of breast cancer at the time of diagnosis. *Breast Cancer Res.* 2011;13(5):R92.
14. Bai Y, Cole K, Martinez-Morilla S, Ahmed FS, Zugazagoitia J, Staaf J, et al. An open-source, automated tumor-infiltrating lymphocyte algorithm for prognosis in triple-negative breast cancer. *Clin Cancer Res.* 2021;27(20):5557–65.
15. Groheux D, Ferrer L, Vargas J, Martineau A, Teixeira L, Menu P, et al. Multimodal machine learning model prediction of complete pathological response to neoadjuvant chemotherapy in triple-negative breast cancer. *J Clin Oncol.* 2022;40 (16_suppl):601.
16. Verghese G, Li M, Liu F, Lohan A, Cherian N, Gazinska P, et al. Abstract P5-01-01: multiscale deep learning framework to capture systemic immune features in lymph nodes predictive of triple negative breast cancer outcome. *Cancer Res.* 2023;83 (5_Supplement) P5-01-01.
17. Chen Z, Wang M, De Wilde RL, Feng R, Su M, Torres-de La Roche LA, et al. A machine learning model to predict the triple negative breast cancer immune subtype. *Front Immunol.* 2021 (12):749459.
18. Xu Y, Ju L, Tong J, Zhou C, Yang J. Supervised machine learning predictive analytics for triple-negative breast cancer death outcomes. *Onco Targets Ther.* 2019(12):9059–67.