Revista de Gastroenterología de México 91 (2026) 68–75

REVISTA DE GASTROENTEROLOGÍA DE MÉXICO
www.elsevier.es/rgmx

ARTÍCULO ORIGINAL

Desarrollo de un modelo de aprendizaje automático para predecir recurrencia en el cáncer de colon derecho: estudio piloto retrospectivo unicéntrico

R. Zayas-Bórquez, J. Canto-Losa, E. Posadas-Trujillo, N. Salgado-Nesme y O. Santes*

Departamento de Cirugía Colorrectal, Instituto Nacional de Ciencias Médicas y Nutrición «Salvador Zubirán», Ciudad de México, México

Recibido el 29 de abril de 2025; aceptado el 22 de julio de 2025
Disponible en Internet el 30 de septiembre de 2025

PALABRAS CLAVE
Cáncer de colon derecho;
Recurrencia a distancia;
Random forest;
Ganglios linfáticos;
Modelo predictivo clínico

Resumen
Introducción y objetivo: El cáncer de colon derecho (CCD) se distingue por sus características clínicas y patrones de recurrencia únicos. Este estudio tuvo como objetivo desarrollar un modelo predictivo de recurrencia a distancia en los pacientes con CCD operados con intención curativa, empleando el algoritmo de aprendizaje automático *random forest* a partir de variables clínicas e histopatológicas.
Materiales y métodos: Se realizó un análisis retrospectivo de 64 pacientes tratados entre 2016-2024. Se incluyeron variables como edad, sexo, invasión linfo-vascular y número de ganglios linfáticos evaluados (transformado para interpretación inversa). Se empleó sobremuestreo para balancear los datos y se construyó un modelo *random forest* para predecir recurrencia a distancia (definida como aquella documentada a partir de los 6 meses posteriores a la cirugía). Se evaluó su desempeño mediante precisión, sensibilidad, puntuación F1 y área bajo la curva ROC (AUC).
Resultados: El modelo obtuvo un AUC de 0.76 en el conjunto de prueba y AUC promedio de 0.81 en validación cruzada, con sensibilidad del 75% y especificidad del 100%. Las variables más relevantes fueron: menor cosecha ganglionar, edad avanzada, sexo masculino e invasión linfática/vascular. Un modelo simplificado con estas 4 variables mantuvo una precisión del 95%. Se construyó una escala clínica de puntajes acumulativos que clasificó a los pacientes en grupos de bajo y alto riesgo, con tasas de recurrencia a distancia del 8.3 y 56.3%, respectivamente.
Conclusión: El modelo predictivo mostró una capacidad robusta para estratificar el riesgo de recurrencia a distancia, respaldando el uso de algoritmos de aprendizaje automático como herramienta complementaria en el manejo individualizado del CCD.
© 2025 Asociación Mexicana de Gastroenterología. Publicado por Masson Doyma México S.A. Este es un artículo Open Access bajo la CC BY-NC-ND licencia (http://creativecommons.org/licencias/by-nc-nd/4.0/).

* Autor para correspondencia. Vasco de Quiroga 15, Belisario Domínguez Secc 16, Tlalpan, 14080 Ciudad de México, México. Teléfono: 5554870900 Ext. 2133.
Correo electrónico: oscar.santesj@incmnsz.mx (O. Santes).

https://doi.org/10.1016/j.rgmx.2025.07.011
0375-0906/© 2025 Asociación Mexicana de Gastroenterología. Publicado por Masson Doyma México S.A. Este es un artículo Open Access bajo la CC BY-NC-ND licencia (http://creativecommons.org/licencias/by-nc-nd/4.0/).

KEYWORDS
Right-sided colon cancer;
Distant recurrence;
Random forest;
Lymph nodes;
Clinical predictive model

Development of a machine learning model for predicting right-sided colon cancer recurrence: A retrospective single center pilot study

Abstract
Introduction and aim: Right-sided colon cancer (RSCC) is characterized by distinct clinical features and recurrence patterns. Our study aimed to develop a predictive model for distant recurrence in patients with RSCC who underwent curative surgery, employing the random forest machine learning algorithm, based on clinical and histopathologic variables.
Materials and methods: A retrospective analysis of 64 patients treated within the time frame of 2016-2024 was conducted. The variables included age, sex, lymphovascular invasion, and number of lymph nodes evaluated (transformed for inverse interpretation). Oversampling was employed to balance the dataset and a random forest model for predicting distant recurrence (defined as that occurring at least six months after surgery) was constructed. Its performance was evaluated through accuracy, sensitivity, F1 score, and area under the ROC curve (AUC).
Results: The model achieved an AUC of 0.76 in the test set and a mean AUC of 0.81 in cross-validation, with 75% sensitivity and 100% specificity. The most relevant variables were low lymph node harvest, older age, male sex, and lymphovascular invasion. A simplified model with those four variables maintained 95% accuracy. A clinical risk scale based on cumulative scores was developed that classified patients into low-risk and high-risk groups, with distant recurrence rates of 8.3% and 56.3%, respectively.
Conclusion: The predictive model showed a robust capacity for stratifying the distant recurrence risk, supporting the use of machine learning algorithms as a complementary tool in the individualized management of RSCC.
© 2025 Asociación Mexicana de Gastroenterología. Published by Masson Doyma México S.A. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introducción

El cáncer colorrectal (CCR) constituye una de las principales causas de morbimortalidad a nivel mundial, posicionándose como el tercer tipo de cáncer más frecuente y la segunda causa de muerte por cáncer. Dentro de esta entidad, el cáncer de colon derecho (CCD) se distingue por presentar características clínicas, moleculares y pronósticas particulares respecto del colon izquierdo, incluyendo un comportamiento más agresivo y una mayor propensión a recurrencia sistémica[1,2]. Las tasas de recurrencia —local o a distancia— en CCD oscilan entre el 10 y el 30% según diferentes series internacionales, condicionadas por el estadio patológico, la calidad de la resección quirúrgica y la presencia de factores histológicos adversos[3].

La identificación precoz de los pacientes con alto riesgo de recurrencia a distancia es fundamental para guiar estrategias de seguimiento posoperatorio, individualizar la indicación de tratamientos adyuvantes y mejorar los desenlaces oncológicos a largo plazo. En este sentido, el desarrollo de herramientas predictivas confiables representa una prioridad clínica.

Los avances recientes en técnicas de aprendizaje automático (AA) han permitido construir modelos predictivos más precisos que los métodos estadísticos convencionales, al ser capaces de integrar simultáneamente múltiples variables clínicas, patológicas e incluso moleculares[4,5]. Algoritmos como *random forest* (RF) han demostrado alta eficacia para clasificar y predecir eventos clínicos en oncología, destacando por su resistencia al sobreajuste, su robustez ante datos desbalanceados y su capacidad para capturar relaciones no lineales entre predictores[6,7]. Además, tienen el valor añadido de identificar variables accesibles desde la práctica clínica como sustitutos indirectos de la agresividad tumoral o de la carga biológica de la enfermedad, incluso en ausencia de estadificación TNM completa.

En este contexto, resulta clínicamente pertinente explorar si las variables de obtención rutinaria como edad, sexo, número de ganglios evaluados e invasión linfovascular pueden ser integradas en modelos predictivos para anticipar la recurrencia a distancia en los pacientes con CCD[8,9]. La incorporación de estos modelos, sustentados en AA, podría aportar una herramienta adicional de estratificación individual del riesgo, orientando una medicina personalizada más precisa y eficiente. Por lo tanto, el objetivo de este estudio fue desarrollar y validar internamente un modelo de predicción de recurrencia a distancia en CCD, utilizando el algoritmo RF a partir de variables clínicas e histopatológicas recolectadas en un centro especializado.

Materiales y métodos

Se llevó a cabo un estudio observacional retrospectivo en un solo centro, con el objetivo de identificar factores asociados a la recurrencia del CCD y construir un modelo predictivo basado en AA. El periodo de inclusión abarcó de 2016 a 2024, e incorporó un total de 64 pacientes con diagnóstico confirmado de CCR derecho. El presente manuscrito fue elaborado conforme a los lineamientos de la guía Strengthening the Reporting of Observational Studies in Epidemiology

(STROBE), con el propósito de asegurar transparencia metodológica y validez interna.

Todos los pacientes fueron sometidos a cirugía con intención curativa en el mismo centro, siguiendo criterios homogéneos de tratamiento quirúrgico y vigilancia oncológica. Se excluyeron aquellos con cirugía paliativa, enfermedad metastásica al momento del diagnóstico, enfermedad residual o información clínico-patológica incompleta. Los tumores estuvieron localizados exclusivamente en colon derecho (ciego, colon ascendente o ángulo hepático). Esto permitió conformar una cohorte clínicamente homogénea.

Se incluyeron inicialmente las siguientes variables clínico-patológicas: edad (años), sexo (masculino/femenino), invasión linfática, invasión vascular, invasión perineural, grado histológico, estado de los márgenes quirúrgicos, número total de ganglios linfáticos evaluados, así como la estadificación TNM reconstruida retrospectivamente a partir de los componentes T, N y M. Estas variables fueron seleccionadas por su relevancia clínica y epidemiológica previamente reportada en relación con recurrencia tumoral.

Se consideraron únicamente a los pacientes con seguimiento clínico mínimo de 24 meses, o con recurrencia a distancia documentada antes de ese umbral. La recurrencia se definió operativamente como la detección de enfermedad a distancia igual o superior a los 6 meses posteriores a la cirugía, confirmada mediante estudios de imagen. Este umbral temporal fue adoptado para evitar la inclusión de enfermedad residual, siguiendo criterios empleados en literatura reciente. Las recurrencias identificadas antes de los 6 meses fueron clasificadas como enfermedad residual y, por lo tanto, excluidas del análisis.

Desarrollo del modelo predictivo

Para la construcción del modelo de predicción se utilizó el algoritmo de clasificación RF, seleccionado por su capacidad para manejar relaciones no lineales entre variables, su bajo riesgo de sobreajuste en comparación con otros métodos y su robustez ante desequilibrios en la proporción de clases.

Previo al entrenamiento, los datos fueron estandarizados para asegurar un comportamiento homogéneo entre predictores. Dado que los casos con recurrencia fueron minoritarios, se aplicó sobremuestreo *(upsampling)* de dicha clase para balancear el conjunto de entrenamiento. Posteriormente, se realizó una partición estratificada aleatoria del conjunto balanceado, asignando 70% de los casos al grupo de entrenamiento y 30% al de prueba. Se empleó una semilla aleatoria fija para asegurar reproducibilidad.

Con el objetivo de evitar sobreajuste *(overfitting)* asociado al uso de múltiples variables en una cohorte limitada, se utilizó validación cruzada estratificada (5-*fold*) y se construyó un modelo clínico simplificado con solo 4 predictores: edad, sexo, invasión linfática/vascular y número total de ganglios linfáticos evaluados. Esta última variable fue tratada como factor de riesgo inverso, de modo que una menor cosecha ganglionar se interpretó como mayor riesgo de recurrencia.

Se reconstruyó retrospectivamente la estadificación clínica TNM con base en los componentes T, N y M. Esta clasificación fue incluida en los análisis iniciales del modelo; sin embargo, mostró una capacidad predictiva muy limitada (AUC = 0.438, sensibilidad = 0%), por lo que fue excluida del modelo final. La validación cruzada confirmó su bajo desempeño en comparación con el modelo clínico propuesto.

Con el fin de minimizar el riesgo de sobreajuste inherente al uso de múltiples predictores en muestras pequeñas, se implementaron varias medidas específicas: preselección de variables clínicamente relevantes, construcción de un modelo simplificado con solo 4 predictores, uso de validación cruzada estratificada (5-*fold*) y análisis de curva de aprendizaje. Estas estrategias permitieron mantener la estabilidad y la reproducibilidad del modelo, reduciendo la posibilidad de ajuste excesivo a los datos de entrenamiento.

Evaluación del modelo y construcción de la escala clínica

El modelo fue evaluado mediante métricas estándar de clasificación: precisión global, sensibilidad, especificidad, puntaje F1 y área bajo la curva ROC (AUC). Se empleó matriz de confusión para identificar errores sistemáticos y confirmar la estabilidad del modelo. La curva de aprendizaje mostró convergencia desde los 30 casos, con baja diferencia entre error de entrenamiento y validación, lo que sugiere una baja probabilidad de sobreajuste.

Con base en la importancia relativa de las variables, se diseñó una escala de riesgo clínico por puntajes acumulativos. Se establecieron puntos de corte para estratificar a los pacientes en 2 grupos de riesgo:

- Bajo riesgo: 0-10 puntos
- Alto riesgo: 11-20 puntos

Las tasas de recurrencia observadas fueron del 8.3% en el grupo de bajo riesgo y del 56.3% en el grupo de alto riesgo, evidenciando una clara diferenciación pronóstica. Esta clasificación permitió traducir los resultados del modelo a una herramienta clínica simple y reproducible.

Análisis estadístico

Las variables continuas fueron expresadas como media $\pm$ desviación estándar o mediana (RIC), según distribución evaluada con la prueba de Shapiro-Wilk. Las variables categóricas se presentaron como frecuencias absolutas y porcentajes. Para comparaciones entre grupos (recurrencia vs. no recurrencia), se utilizaron pruebas t de Student o U de Mann-Whitney (según normalidad), y Chi-cuadrado o exacta de Fisher para variables categóricas. Se consideró estadísticamente significativo un valor de $p < 0.05$.

La importancia de cada variable en el modelo fue estimada mediante el índice de ganancia de impureza *(Gini importance)*. Se comparó el rendimiento del modelo clínico con modelos que incorporaban la clasificación TNM, sin observar mejoras significativas.

Todos los análisis se realizaron en Python v.3.11.5, utilizando las bibliotecas Pandas, NumPy, Scikit-Learn, Matplotlib y Seaborn.

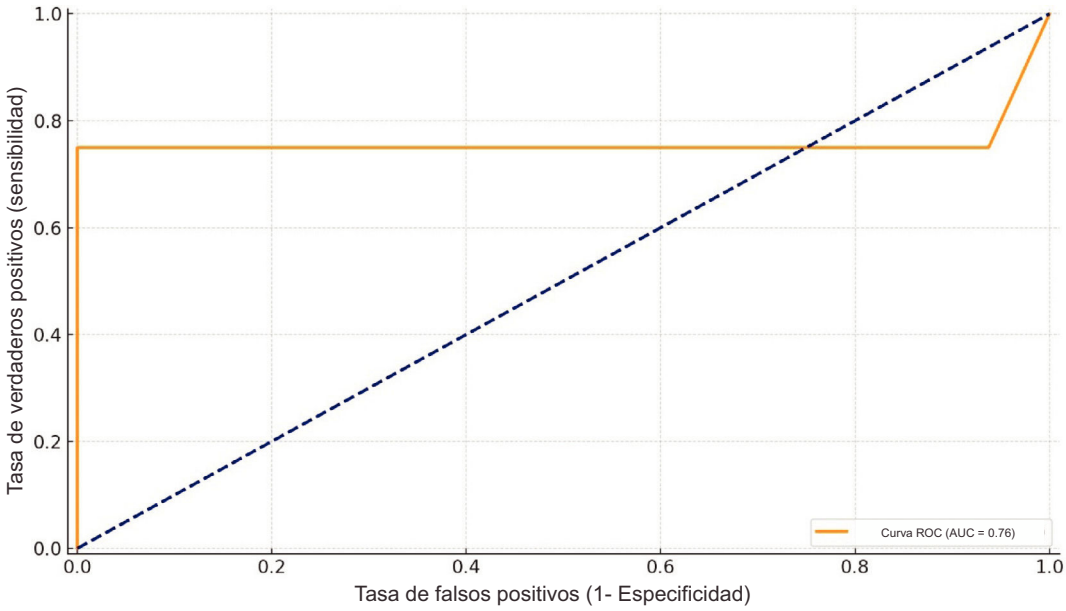

Figura 1 Curva ROC del modelo *random forest* simplificado. Representación de la sensibilidad frente a 1 - especificidad. El área bajo la curva (AUC) fue de 0.76 en el conjunto de la prueba, lo que refleja una buena capacidad discriminativa para identificar los pacientes con recurrencia a distancia.

Resultados

Se incluyeron 64 pacientes con diagnóstico confirmado de CCD operados con intención curativa. Todos contaban con seguimiento oncológico mínimo de 24 meses o presentaron recurrencia documentada dentro de ese periodo. El desenlace primario fue la recurrencia a distancia, definida como aquella confirmada por estudios de imagen a partir de los 6 meses posteriores a la cirugía.

El modelo predictivo desarrollado mediante el algoritmo RF alcanzó una precisión global del 95% en el conjunto de prueba. La curva ROC mostró un AUC de 0.76, lo que indica una capacidad discriminativa aceptable para identificar los pacientes con alto riesgo de recurrencia a distancia (fig. 1). En términos clínicos, esto implica una probabilidad del 76% de que el modelo clasifique correctamente a un paciente con recurrencia frente a uno sin recurrencia.

El análisis de importancia de variables reveló que el predictor más influyente fue la menor cosecha ganglionar (valor inverso del número de ganglios evaluados), con una importancia relativa de 0.44. Le siguieron la edad (0.35), la invasión linfática/vascular (0.15) y el sexo masculino (0.06) (figs. 2 y 3). Este hallazgo refuerza la interpretación clínica de que una menor cantidad de ganglios linfáticos evaluados se asocia con mayor riesgo de recurrencia tumoral.

Para optimizar la aplicabilidad del modelo, se construyó una versión simplificada basada exclusivamente en estas cuatro variables. Este modelo redujo la complejidad sin sacrificar precisión, alcanzando una precisión del 95%, sensibilidad del 75%, especificidad del 100%, y un F1 *score* de 0.86. La validación cruzada estratificada (5-*fold*) mostró un AUC promedio de 0.81, con baja desviación estándar, lo que indica consistencia del desempeño en múltiples subconjuntos de datos (fig. 4), siendo finalmente AUC de 0.76 en test set y AUC promedio de 0.81 en validación cruzada.

En contraste, un modelo construido exclusivamente con la estadificación TNM como única variable obtuvo un AUC de 0.438, con sensibilidad del 0% y F1 *score* de 0.0, reflejando un valor predictivo nulo. Incluso al integrar la clasificación TNM con otras variables clínicas, el modelo solo alcanzó un AUC de 0.77 y una sensibilidad del 25%. Estos resultados contrastan con el desempeño del modelo clínico simplificado, que superó de forma significativa a la estadificación anatómica convencional en todas las métricas evaluadas.

Escala clínica de predicción de recurrencia

A partir del modelo simplificado, se diseñó una escala clínica basada en un sistema de puntaje acumulativo proporcional a la importancia relativa de cada variable en el modelo RF. Este sistema permitió clasificar a los pacientes en 2 niveles de riesgo clínico diferenciados (tablas 1 y 2):

- Bajo riesgo (0-10 puntos): tasa de recurrencia observada del 8.3% (4/48)
- Alto riesgo (11-20 puntos): tasa de recurrencia observada del 56.3% (9/16)

Esta estratificación binaria demostró una capacidad discriminativa robusta y clínicamente relevante. La progresión de tasas de recurrencia observadas valida el uso potencial de esta escala como herramienta práctica para personalizar el seguimiento posquirúrgico, intensificando la vigilancia en pacientes de alto riesgo o considerando estrategias terapéuticas adicionales.

Discusión

El objetivo de este estudio fue desarrollar un modelo clínico predictivo, utilizando AA, para anticipar la recurrencia

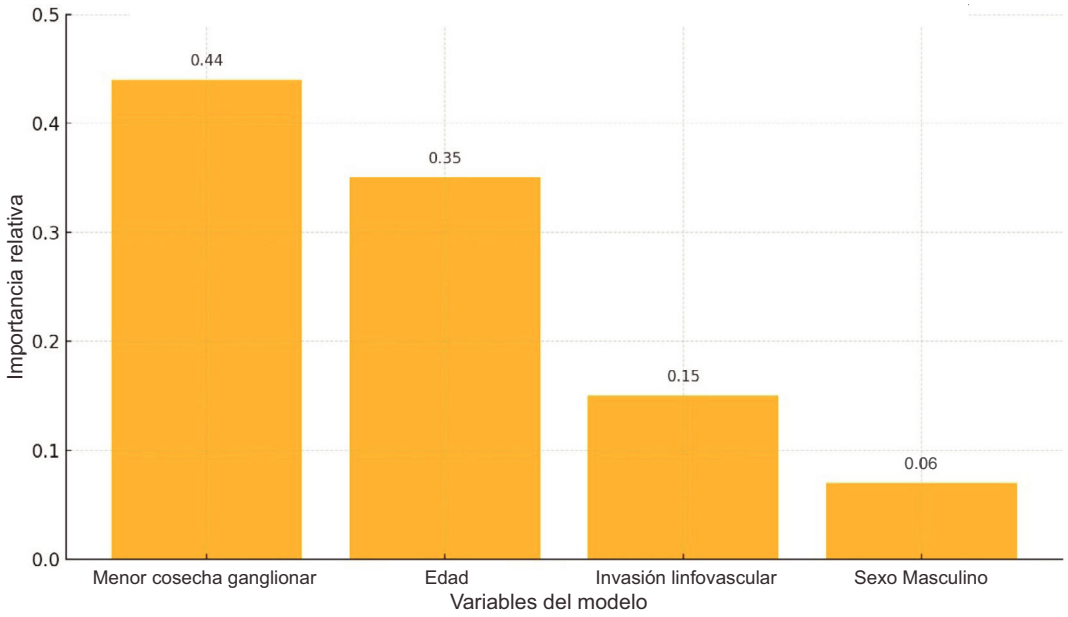

Figura 2 Importancia de las variables en el modelo *random forest*. La barra horizontal representa la importancia relativa de cada variable en la predicción de recurrencia a distancia. La menor cosecha ganglionar (inverso del número de ganglios linfáticos recabados) fue el predictor más influyente, seguida por la edad, la invasión linfática/vascular y el sexo masculino.

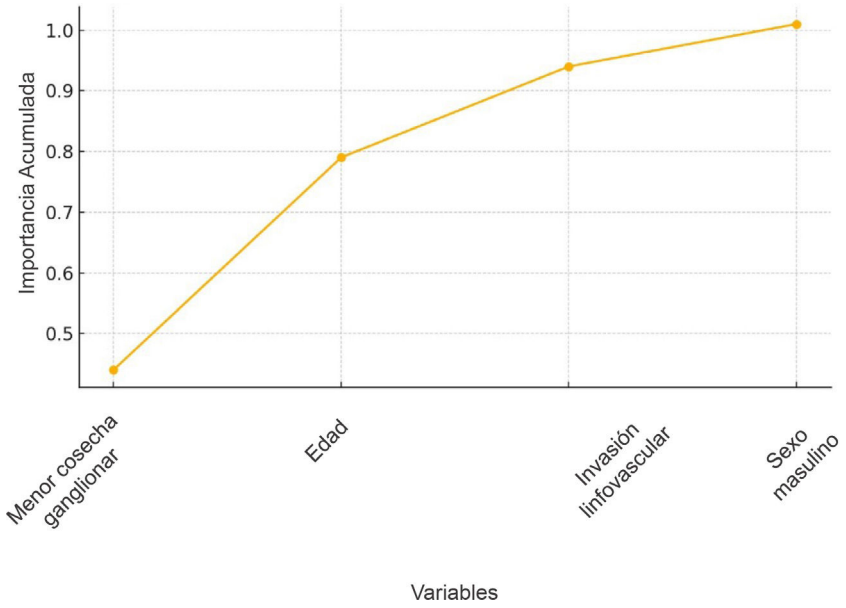

Figura 3 Importancia acumulada de las variables en el modelo *random forest*. Distribución proporcional de la contribución de cada variable al desempeño global del modelo. La menor cosecha ganglionar aportó el 44% de la capacidad predictiva total, mientras que edad, invasión linfática/vascular y sexo masculino representaron el resto de la varianza explicada.

a distancia en CCD. Los resultados obtenidos muestran que los hallazgos de este estudio demuestran que el modelo basado en RF ofrece un desempeño robusto para predecir la recurrencia a distancia en los pacientes con CCD. El modelo simplificado, compuesto por variables clínico-patológicas rutinarias, alcanzó una precisión general del 95%, una sensibilidad del 75%, especificidad del 100%, y un AUC de 0.76. Estos resultados reflejan una capacidad discriminativa adecuada y clínicamente útil, especialmente al considerar el tamaño muestral y la ausencia de estadificación TNM completa.

Las variables más relevantes identificadas por el modelo fueron: menor cosecha ganglionar, edad avanzada, presencia de invasión linfovascular y sexo masculino. Contrario a lo que sugieren los modelos tradicionales centrados en la positividad ganglionar, nuestro análisis identificó al número absoluto de ganglios linfáticos evaluados como el predictor más influyente. Esta variable podría reflejar no solo la

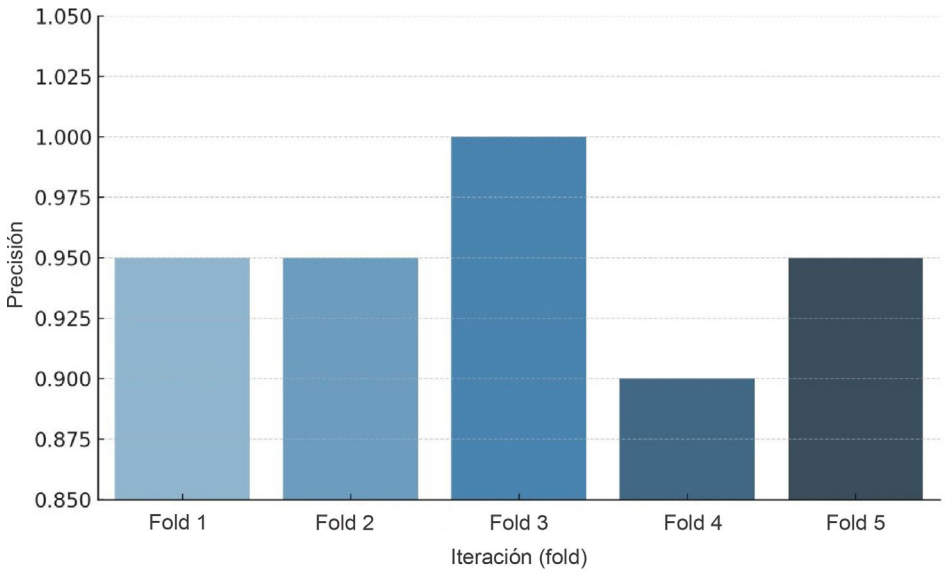

Figura 4 Resultados de validación cruzada del modelo *random forest* simplificado. Precisión obtenida en cada uno de los 5 *folds* de validación cruzada estratificada. El modelo mostró una precisión media del 95%, con AUC promedio de 0.81 y estabilidad entre iteraciones, lo que respalda su reproducibilidad y robustez a pesar del tamaño muestral limitado.

Tabla 1 Sistema de puntuación basado en características clínicas y patológicas

Variable	Criterio	Puntaje
Ganglios linfáticos evaluados	≥ 15 ganglios evaluados	1
	10-14 ganglios evaluados	4
	5-9 ganglios evaluados	7
	< 5 ganglios evaluados	10
Edad	≥ 70 años	4
	60-69 años	3
	50-59 años	2
	< 50 años	1
Invasión linfovascular	Presente	4
	Ausente	0
Sexo	Masculino	2
	Femenino	0

La puntuación total se obtiene sumando los puntajes de cada variable.

Tabla 2 Clasificación del riesgo de recurrencia

Puntaje total	Categoría de riesgo	Pacientes (n)	Recurrencias	Tasa de recurrencia
0-10	Riesgo bajo	48	4	8.3%
11-20	Riesgo alto	16	9	56.3%

Clasificación del riesgo de recurrencia según la puntuación total. Los pacientes se clasifican en 2 niveles de riesgo: bajo (0-10 puntos) y alto (11-20 puntos), según la puntuación acumulada derivada de factores clínicos y patológicos.

calidad técnica de la cirugía o la agresividad tumoral, sino también aspectos inmunológicos o anatómicos que condicionan la cosecha ganglionar y, por extensión, el pronóstico del paciente.

Estudios previos de AA han reportado desempeños superiores al TNM. Por ejemplo, Liu et al. (2023) lograron un AUC de 0.962 utilizando XGBoost para predecir recurrencia en CCR[1]. En nuestro estudio, el algoritmo RF logró resultados consistentes con menor complejidad computacional y mayor interpretabilidad clínica. Si bien el AUC fue inferior, esto puede atribuirse al tamaño de muestra y la selección estricta de variables fácilmente disponibles en la práctica clínica.

Desde el punto de vista de aplicabilidad, se construyó una escala clínica derivada del modelo simplificado. Esta permitió estratificar a los pacientes en 2 grupos de riesgo

clínicamente diferenciables: bajo (0-10 puntos, recurrencia del 8.3%) y alto (11-20 puntos, recurrencia del 56.3%). La magnitud de la diferencia observada entre los grupos respalda la validez del modelo y su utilidad potencial como herramienta de apoyo para decisiones oncológicas individualizadas, incluyendo la intensificación del seguimiento o el planteamiento de estrategias terapéuticas adyuvantes.

En contraste, el modelo basado exclusivamente en la clasificación TNM mostró un desempeño limitado en nuestro análisis (AUC = 0.438; sensibilidad = 0%; F1 *score* = 0.0). Incluso al combinarlo con variables clínicas, su capacidad predictiva continuó siendo inferior (AUC = 0.77; sensibilidad = 25%). Estos hallazgos se alinean con la evidencia actual, la cual respalda que, aunque la clasificación TNM es el estándar internacional para la estadificación y pronóstico en cáncer de colon, presenta limitaciones importantes en la predicción individual de desenlaces, particularmente recurrencia y supervivencia. Diversos estudios han demostrado que el desempeño predictivo del TNM es modesto cuando se utiliza de forma aislada; p. ej., el AUC para modelos basados exclusivamente en TNM suele oscilar entre 0.60 y 0.64, lo que indica una capacidad discriminativa limitada para predecir recurrencia o mortalidad a nivel individual[10–12]. Además, su sensibilidad para predecir recurrencia es subóptima, y los valores reportados en la literatura rara vez superan el 60%[13].

La incorporación de variables clínicas adicionales (como edad, sexo, grado tumoral, número de ganglios examinados, invasión vascular/perineural y marcadores moleculares como MSI, BRAF y KRAS) ha demostrado mejorar significativamente la capacidad predictiva de los modelos. Al integrar estas variables, el AUC puede incrementarse hasta 0.71-0.74, con una mejora consistente en la concordancia de los modelos[10,11,13]. No obstante, incluso estos modelos multivariables tradicionales muestran limitaciones en la predicción individual.

En los últimos años, modelos basados en inteligencia artificial y AA han demostrado superar significativamente tanto al sistema TNM como a los modelos clínicos convencionales. Estos enfoques permiten integrar grandes volúmenes de datos clínicos, patológicos, moleculares e incluso imágenes histopatológicas, alcanzando AUC superiores (p. ej., entre 0.77 y 0.86) y mejorando la sensibilidad y precisión en la predicción de desenlaces oncológicos[14,15]. Además, la importancia relativa de la estadificación clínica tradicional tiende a disminuir cuando se emplean variables adicionales y técnicas avanzadas de modelado[14,15].

Metodológicamente, se aplicaron estrategias robustas para mitigar el riesgo de sobreajuste, incluyendo: preselección de variables clínicamente relevantes, balanceo de clases por sobremuestreo, validación cruzada 5-*fold*, y diseño de un modelo simplificado con interpretabilidad clínica. La estabilidad del modelo fue confirmada con un AUC promedio de 0.81 en la validación cruzada.

Finalmente, se definió la recurrencia como nueva enfermedad detectada ≥ 6 meses después de la cirugía, en concordancia con estudios recientes que recomiendan este umbral para diferenciar recurrencia verdadera de enfermedad residual[16]. Esta definición fortalece la validez de los resultados y permite comparabilidad con la literatura internacional.

Conclusión

El modelo predictivo desarrollado mediante RF mostró un desempeño sólido para anticipar la recurrencia a distancia en los pacientes con CCD, con un AUC de 0.76, sensibilidad del 75% y especificidad del 100%. Las variables más relevantes fueron: menor cosecha ganglionar, edad, invasión linfovascular y sexo masculino.

A partir de estos hallazgos, se construyó una escala clínica simplificada que estratifica el riesgo de recurrencia en 2 niveles:

- Bajo riesgo (0-10 puntos): recurrencia observada del 8.3%
- Alto riesgo (11-20 puntos): recurrencia observada del 56.3%

Estos resultados respaldan el uso potencial de modelos de AA como herramientas clínicas accesibles para la estratificación individual del riesgo oncológico, superando la capacidad discriminativa de la estadificación TNM tradicional. El modelo fue validado internamente mediante validación cruzada estratificada, aunque se reconoce que su utilidad debe corroborarse externamente.

Limitaciones

Este estudio debe interpretarse como un análisis piloto con valor exploratorio e hipótesis-generador. Si bien el modelo mostró un desempeño robusto en la validación cruzada, sus hallazgos requieren ser replicados y validados externamente en cohortes más amplias antes de considerarse aplicables en la práctica clínica.

Presenta limitaciones inherentes a su diseño retrospectivo y al tamaño muestral reducido (n = 64), lo cual restringe la potencia estadística, la capacidad para detectar interacciones de orden superior y la generalización externa del modelo. Si bien los algoritmos de AA como RF permiten el entrenamiento con muestras pequeñas, el riesgo de sobreajuste *(overfitting)* permanece latente, a pesar de las medidas adoptadas (balanceo de clases, validación cruzada, simplificación del modelo).

Asimismo, al tratarse de un estudio unicéntrico, los pacientes incluidos podrían no representar completamente la variabilidad observada en otras instituciones o regiones. Si bien las variables empleadas son de obtención clínica estándar, se reconoce que las características demográficas o las prácticas institucionales locales podrían influir en los resultados. Por lo tanto, la validación externa en cohortes más diversas es imprescindible antes de considerar su aplicación generalizada.

Asimismo, el uso de variables derivadas de expedientes clínicos limita el control sobre la homogeneidad en la recolección de datos, y la estadificación TNM fue reconstruida retrospectivamente, lo cual puede introducir sesgos de clasificación.

Por estas razones, los resultados presentados deben considerarse exploratorios, y su aplicabilidad clínica requiere ser validada externamente en estudios prospectivos con diseño multicéntrico, muestra más amplia y mayor diversidad clínica. Solo así será posible confirmar el valor predictivo

de la escala y su impacto potencial en la mejora del manejo individualizado del CCD.

Financiación

Este estudio no recibió financiación específica de entidades públicas, comerciales o sin fines de lucro.

Consideraciones éticas

Dado el carácter retrospectivo y observacional del estudio, basado en expedientes clínicos anonimizados, no se requirió consentimiento informado individual. Se garantizó en todo momento la confidencialidad y el cumplimiento de los principios éticos de la Declaración de Helsinki, así como la normativa nacional vigente en investigación en salud. El protocolo cumplió con los criterios establecidos por COFEPRIS y el Reglamento de la Ley General de Salud en materia de investigación científica.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Referencias

1. Liu Y, Du W, Guo Y, et al. Identification of high-risk factors for recurrence of colon cancer following complete mesocolic excision: An 8-year retrospective study. PLoS One. 2023;18:e0289621, http://dx.doi.org/10.1371/journal.pone.0289621.
2. Xu Y, Ju L, Tong J, et al. Machine Learning Algorithms for Predicting the Recurrence of Stage IV Colorectal Cancer After Tumor Resection. Sci Rep. 2020;10:2519, http://dx.doi.org/10.1038/s41598-020-59115-y.
3. Lygre KB, Forthun RB, Høysæter T, et al. Assessment of postoperative circulating tumour DNA to predict early recurrence in patients with stage I-III right-sided colon cancer: Prospective observational study. BJS Open. 2024;8, http://dx.doi.org/10.1093/bjsopen/zrad146, zrad146.
4. Osterman E, Ekström J, Sjöblom T, et al. Accurate population-based model for individual prediction of colon cancer recurrence. Acta Oncologica. 2021;60:1241–9, http://dx.doi.org/10.1080/0284186X. 2021.1953138.
5. Malakorn S, Ouchi A, Hu CY, et al. Tumor sidedness, recurrence, and survival after curative resection of localized colon cancer. Clin Colorectal Cancer. 2021;20:353–E60, http://dx.doi.org/10.1016/j.clcc.2020.08.007.
6. Chen PC, Yeh YM, Lin BW, et al. A Prediction Model for Tumor Recurrence in Stage II-III Colorectal Cancer Patients: From a Machine Learning Model to Genomic Profiling. Biomedicines. 2022;10:340, http://dx.doi.org/10.3390/biomedicines10020340.
7. Mohammad K, Alexanderani M, Omar M, et al. Abstract 2584: Decoding colon cancer recurrence: Unveiling accurate predictions with attention-guided deep neural networks on histopathological whole slide images. Cancer Research. 2024;84 Suppl 6:2584, http://dx.doi.org/10.1158/1538-7445.AM2024-2584.
8. Achilonu OJ, Fabian J, Bebington B, et al. Predicting Colorectal Cancer Recurrence and Patient Survival Using Supervised Machine Learning Approach: A South African Population-Based Study. Front Public Health. 2021;29:694306, http://dx.doi.org/10.3389/fpubh.2021.778749.
9. Park JH, Kim MJ, Park SC, et al. Difference in Time to Locoregional Recurrence Between Patients with Right-Sided and Left-Sided Colon Cancers. Dis Colon Rectum. 2015;58:831–7, http://dx.doi.org/10.1097/DCR. 0000000000000426.
10. Dienstmann R, Mason MJ, Sinicrope FA, et al. Prediction of overall survival in stage II and III colon cancer beyond TNM system: A retrospective, pooled biomarker study. Ann Oncol. 2017;28, http://dx.doi.org/10.1093/annonc/mdx052.
11. Weiser MR, Gönen M, Chou JF, et al. Predicting survival after curative colectomy for cancer: Individualizing colon cancer staging. J Clin Oncol. 2011;29:4796–802, http://dx.doi.org/10.1200/JCO.2011.36.5080.
12. Sassun R, Sileo A, Ng JC, et al. Validated Integration of Tumor Deposits in N Staging for Prognostication in Colon Cancer. JAMA Surg. 2025;160:408–14, http://dx.doi.org/10.1001/jamasurg.2024.6729.
13. Weiser MR, Hsu M, Bauer PS, et al. Clinical Calculator Based on Molecular and Clinicopathologic Characteristics Predicts Recurrence Following Resection of Stage I-III Colon Cancer. J Clin Oncol. 2021;39:911–9, http://dx.doi.org/10.1200/JCO.20.02553.
14. Buk Cardoso L, Cunha Parro V, Verzinhasse Peres S, et al. Machine learning for predicting survival of colorectal cancer patients. Sci Rep. 2023;13:8874, http://dx.doi.org/10.1038/s41598-023-43248-x. Erratum in: Sci Rep. 2023;13:16049.
15. He B, Wang L, Zhou W, et al. A fusion model to predict the survival of colorectal cancer based on histopathological image and gene mutation. Sci Rep. 2025;15:9677, http://dx.doi.org/10.1038/s41598-025-91420-2.
16. Nors J, Iversen LH, Erichsen R, et al. Incidence of Recurrence and Time to Recurrence in Stage I to III Colorectal Cancer: A Nationwide Danish Cohort Study. JAMA Oncol. 2024;10:54–62, http://dx.doi.org/10.1001/jamaoncol.2023.5098.