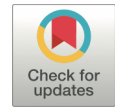


ARTÍCULO ORIGINAL



Inteligencia Artificial en Cirugía: Creación y validación de una calculadora colombiana de riesgo de mortalidad perioperatoria

Artificial Intelligence in Surgery: Development and validation of a Colombian perioperative mortality risk calculator

Nicolás Lozano-Suárez, MD, MSc¹ , Juan Felipe Salamanca-Lozano, Mat² ,
Paulo Andrés Cabrera-Rivera, MD, MSc³ , Leonardo Briceño-Ayala, MD, PhD⁴ ,
Carlos J. Pérez-Rivera, MD, MSc⁵ ; ColSOS Collaborative

- 1 Programa de especialización en Cirugía General, Universidad CES, Medellín, Colombia; Grupo de Investigación en Cirugía General, Fundación Cardioinfantil-Instituto de Cardiología, Bogotá, D.C., Colombia.
- 2 Matemático, Universidad Nacional de Colombia, AlMed, Bogotá, D.C., Colombia.
- 3 Departamento de Cirugía general, Cardioinfantil-Instituto de Cardiología, Bogotá, D.C., Colombia.
- 4 Escuela de Medicina y Ciencias de la Salud, Universidad del Rosario; Grupo de Investigación en Salud Pública, Escuela de Medicina y Ciencias de la Salud, Universidad del Rosario, Bogotá, D.C., Colombia.
- 5 Departamento de Cirugía general, Universidad El Bosque; Grupo de Investigación en Cirugía General, Fundación Cardioinfantil-Instituto de Cardiología, Bogotá, D.C., Colombia.

Trabajo ganador del Tercer puesto en el Concurso Nacional de Investigación en Cirugía “José Félix Patiño Restrepo”, categoría Médicos Residentes, Asociación Colombiana de Cirugía, 50° Congreso Semana Quirúrgica Nacional, agosto de 2024.

Resumen

Introducción. Se estima que ocurren 4,2 millones de muertes anuales en los primeros 30 días postoperatorios. La Comisión de Lancet en Cirugía Global resalta la importancia de medir y reducir esta mortalidad. Este estudio desarrolló una calculadora de mortalidad perioperatoria específica para la población colombiana, pretendiendo identificar e intervenir tempranamente los pacientes con alto riesgo.

Métodos. Se utilizaron datos del estudio multicéntrico ColSOS, en el que se incluyeron 3807 pacientes de 54 centros en Colombia. Se recopilaron variables clínicas, sociodemográficas y perioperatorias; se manejaron los datos faltantes con imputación múltiple. La selección de variables se realizó mediante análisis bivariado, regresión Lasso y *Recursive Feature Elimination* (RFE). Se compararon modelos predictivos utilizando regresión logística y XGBoost, evaluando su rendimiento con validación cruzada.

Resultados. El modelo XGBoost fue seleccionado por mostrar una mejor sensibilidad y menor número de falsos negativos que la regresión logística. Se destacó la importancia en la predicción de la clasificación ASA, enfermedad pulmonar obstructiva crónica, inestabilidad hemodinámica y urgencia del procedimiento. El modelo predijo mortalidad con un área bajo la curva (AUC) de 0,87.

Fecha de recibido: 27/07/2024 - Fecha de aceptación: 08/08/2024 - Publicación en línea: 08/11/2024

Correspondencia: Nicolás Lozano-Suárez, Calle 10A # 22 – 04, Medellín, Colombia. Teléfono: +57 3138926759.

Dirección electrónica: lozanos.nicolas@uces.edu.co

Citar como: Lozano-Suárez N, Salamanca-Lozano JF, Cabrera-Rivera PA, Briceño-Ayala L, Pérez-Rivera CJ; ColSOS Collaborative. Inteligencia Artificial en Cirugía: Creación y validación de una calculadora colombiana de riesgo de mortalidad perioperatoria. Rev Colomb Cir. 2025;40:67-78. <https://doi.org/10.30944/20117582.2741>

Este es un artículo de acceso abierto bajo una Licencia Creative Commons - BY-NC-ND <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>

Conclusión. El presente estudio ha desarrollado la primera calculadora de mortalidad perioperatoria diseñada para la población colombiana, incluyendo múltiples especialidades quirúrgicas. El modelo de *machine learning* seleccionado presenta una sensibilidad y especificidad que la hacen equiparable a las mejores herramientas internacionales. La implementación de esta herramienta permite identificar y manejar tempranamente a los pacientes en riesgo, con lo que se podría mejorar la atención quirúrgica en Colombia.

Palabras claves: inteligencia artificial; cirugía general; medición de riesgo; mortalidad hospitalaria; predicción.

Abstract

Introduction. It is estimated that 4.2 million deaths occur annually within the first 30 postoperative days. The Lancet Commission on Global Surgery highlights the importance of measuring and reducing this mortality. This study developed a specific perioperative mortality calculator for the Colombian population, aiming to identify and intervene early in high-risk patients.

Methods. Data from the multicenter ColSOS study were used, in which 3807 patients from 54 centers in Colombia were included. Clinical, sociodemographic, and perioperative variables were collected; missing data were handled with multiple imputations. Variable selection was performed through bivariate analysis, Lasso regression, and Recursive Feature Elimination (RFE). Predictive models were compared using logistic regression and XGBoost, evaluating their performance with cross-validation.

Results. The XGBoost model was selected because of its superior sensitivity and fewer false negatives than logistic regression. The importance in predicting ASA classification, COPD, hemodynamic instability, and urgency of the procedure was highlighted. The model predicted mortality with an area under the curve (AUC) of 0.87.

Conclusion. This study has developed the first perioperative mortality calculator designed for the Colombian population, including multiple surgical specialties. The selected machine learning model demonstrates sensitivity and specificity comparable to the best international tools. The implementation of this tool allows for early identification and management of patients at risk, which could improve surgical care in Colombia.

Keywords: artificial intelligence; general surgery; risk assessment; hospital mortality; forecasting.

Introducción

Se estima que en el mundo ocurren 4,2 millones de muertes anuales en los primeros 30 días posoperatorios, de las cuales casi la mitad se consideran prevenibles¹. Así mismo, las intervenciones quirúrgicas contribuyen aproximadamente al 30 % de la carga de enfermedad mundial¹. Por esta razón, se creó la Comisión de Lancet en Cirugía Global buscando mejorar la atención quirúrgica global y postularon seis indicadores para evaluar las fortalezas y debilidades de los sistemas de salud, reconociendo la necesidad de su intervención para mejorar la atención quirúrgica^{1,2}. La mortalidad perioperatoria es el cuarto indicador de esta comisión y ha sido estimado en distintas poblaciones alrededor del mundo³⁻⁶, incluyendo aproximaciones en nuestro país⁷⁻⁹.

Se considera que el primer gran paso es la medición y la estimación de la mortalidad perioperatoria, pero también se requiere la generación de intervenciones para mejorar este indicador y la atención quirúrgica. Algunos países, como Etiopía y Nigeria, han iniciado procesos de planes nacionales de cirugía^{10,11}, sin embargo, es necesario buscar estrategias que impacten el ámbito clínico. Entre las intervenciones utilizadas para reducir la mortalidad se encuentra la identificación de pacientes con un riesgo elevado de muerte. Aunque múltiples estudios han descrito modelos predictivos, estos suelen ser específicos para determinados procedimientos o categorías quirúrgicas¹² y muchos otros son solo formulados, pero no están disponibles para su uso o aplicación.

El Colegio Americano de Cirujanos (ACS) creó un modelo predictivo para distintos desenlaces posoperatorios, entre ellos mortalidad, primero por medio de una regresión logística¹² y posteriormente utilizando una metodología avanzada de *machine learning*¹³. Esta herramienta fue entrenada con 1.414.006 pacientes y presentó un área bajo la curva (AUC) de 0,94¹²⁻¹⁴.

Aunque esta herramienta está disponible en forma de calculadora, no fue entrenada ni validada en población colombiana. Por lo tanto, el objetivo de este estudio fue desarrollar una calculadora de mortalidad perioperatoria basada en un modelo predictivo específico para la población colombiana, para identificar de forma preoperatoria a los pacientes con alto riesgo de mortalidad y así poder intervenirlos o monitorizarlos tempranamente.

Métodos

Fuente de información y población

Los datos fueron obtenidos del estudio de cohorte prospectiva multicéntrico ColSOS, en el cual se recolectaron datos de 3807 pacientes de 54 centros alrededor de Colombia. Este estudio buscó describir los desenlaces perioperatorios de mortalidad y complicaciones intrahospitalarias hasta 30 días posteriores al procedimiento, la muerte o el egreso del paciente. Se incluyeron pacientes mayores de 18 años que fueron sometidos a un procedimiento quirúrgico en un quirófano, excluyendo procedimientos endoscópicos o por radiología intervencionista. La recolección de datos de distintos procedimientos y especialidades quirúrgicas se llevó a cabo entre mayo de 2022 y enero de 2023^{9,15}. Este estudio siguió la metodología TRIPOD para el reporte de este modelo predictivo¹⁶.

Variables

Se incluyeron variables clínicas y sociodemográficas como edad, sexo, comorbilidades, antecedentes de tabaquismo, índice de masa corporal (IMC), clasificación ASA (*American Society of Anesthesiologists*), tipo de afiliación al sistema de salud, estrato socioeconómico, estado de infección

y vacunación por COVID-19. También se consideraron variables perioperatorias como el procedimiento a realizar, descrito por los códigos únicos de procedimientos en salud (CUPS), la categoría quirúrgica (cirugía general, ortopedia, gineco-obstetricia, cirugía plástica, entre otros), el abordaje quirúrgico (abierto, video-asistido, robótico, endovascular, entre otros), la complejidad del procedimiento, y si el procedimiento fue electivo o urgente. Además, se registró la condición del paciente previo a la cirugía, dado por el estado hemodinámico y la presencia de paro cardíaco preoperatorio. Las definiciones de las variables incluidas en el estudio se pueden encontrar en el protocolo¹⁵.

La variable desenlace fue mortalidad perioperatoria, obtenida del cuarto indicador de la comisión de Lancet en cirugía global¹⁷. Esta se define como toda mortalidad secundaria a cualquier causa antes del alta hospitalaria, o dentro de un máximo de 30 días de estancia hospitalaria, en un paciente que se sometió a un procedimiento quirúrgico en un quirófano¹⁸.

Manejo e imputación de los datos

Las variables de IMC, esquema de vacunación y diagnóstico de COVID-19 fueron imputadas, dado que presentaban datos faltantes en un 18 %, 1,8 % y 1,8 %, respectivamente. Para ello se utilizaron dos técnicas de imputación múltiple: MICE (*Multivariate Imputation by Chained Equations*) para el caso del IMC y una imputación iterativa con clasificador de bosques aleatorios para las variables categóricas. Ambas técnicas se evaluaron usando validación cruzada (*cross-validation*), en la cual se imputaron datos faltantes sembrados de forma aleatoria. Se obtuvo una exactitud del 82 % en las variables categóricas y una mediana de error absoluto menor a 3 kg/m² en el índice de masa corporal, como se observa en la Figura 1.

Selección de variables

Se utilizaron tres técnicas para la selección de variables: análisis bivariado, regresión Lasso y *Recursive Feature Elimination* (RFE).



Figura 1. Distribución de error en la imputación del índice de masa corporal (IMC) en 5 muestras de 25 pacientes por cross-validation.

Fuente: Figura creada por los autores.

En el análisis bivariado, se evaluó la distribución de los datos cuantitativos con las pruebas de Kolmogorov-Smirnov y Shapiro-Wilk. Luego, se aplicaron las pruebas T de Student o U de Mann-Whitney, según la distribución. Los datos categóricos se analizaron con las pruebas de Chi-cuadrado y Kruskal-Wallis, considerando un valor de $p < 0,05$ como significativo. La regresión Lasso se realizó con una lambda obtenida mediante validación cruzada, optimizada por precisión balanceada y AUC (área bajo la curva). Este modelo penaliza la magnitud de los coeficientes. Es decir, a las variables que tienen poco o ningún efecto sobre mortalidad les da un valor de coeficiente 0, excluyéndolas del modelo. El RFE es un clasificador logístico para seleccionar el mejor modelo con n variables, desde una hasta todas las variables disponibles. Los modelos se compararon usando el Criterio de Información de Akaike (AIC), eligiendo el de menor AIC.

Finalmente, se compararon las variables seleccionadas por las tres metodologías, conservando aquellas preservadas por al menos dos de los tres métodos (Tabla 1).

Desarrollo de los modelos predictivos

Se utilizaron dos estrategias en el desarrollo de los modelos predictivos para mortalidad, seleccionadas en base a su interpretabilidad y el uso en calculadoras y modelos predictivos quirúrgicos

previos. Primero, se realizó una estimación mediante regresión logística binaria, incluyendo las variables seleccionadas y obteniendo el AUC y las curvas ROC (*Receiver Operating Characteristic*). Posteriormente, se utilizó el modelo de *machine learning* XGBoost, un modelo de clasificación binaria que crea múltiples árboles de decisión para predecir la probabilidad de que una observación pertenezca a una clase positiva. El modelo se ajustó para optimizar la sensibilidad, es decir, la proporción de verdaderos positivos entre los positivos, penalizando los falsos negativos y, por lo tanto, evitando al máximo la predicción de supervivencia en casos de mortalidad.

Por otro lado, se realizaron distintos modelos según el tratamiento realizado a la variable de tipo de cirugía (CUPS); esta clasificación tiene la capacidad de ser altamente específica pero también regresar a categorías más generales. Por lo cual, se evaluó la variable como procedimiento específico, procedimiento general y categoría de procedimiento. De igual manera, al tener algunos procedimientos con menos casos, se siguió la metodología propuesta por Bilimoria KY, et al.¹², de incluir los CUPS con un número mínimo de procedimientos, con punto de corte en 30 y 20 casos. Todos los análisis fueron realizados en el lenguaje de programación Python (Wilmington, Estados Unidos: Python Software Foundation) y R Studio 4.3.3 (Vienna, Austria: R Foundation for Statistical Computing).

Tabla 1. Variables seleccionadas para los modelos predictivos

Variables seleccionadas	Variables abandonadas
Edad	Diagnóstico previo COVID-19
Sexo	Dislipidemia
Índice de masa corporal (IMC)	Diabetes Mellitus
Hipertensión arterial (HTA)	Estrato socioeconómico
Enfermedad renal crónica (ERC)	Falla cardíaca crónica
Arritmia cardíaca	Paro cardíaco preoperatorio
Diagnóstico de COVID-19	Historia tabaquismo
Esquema de vacunación para COVID-19	Trastorno tiroideo
Enfermedad pulmonar obstructiva crónica (EPOC)	
Consumo de tabaco actual	
Clasificación de ASA	
Inestabilidad hemodinámica preoperatoria	
Complejidad del procedimiento	
Momento de intervención (Urgente o electivo)	
Afiliación al sistema de salud	
Tipo de abordaje quirúrgico	
Categoría del procedimiento	
Tipo de procedimiento por CUPS*	

* CUPS: códigos únicos de procedimientos en salud.

Fuente: Tabla creada por los autores.

Los modelos fueron comparados mediante validación cruzada, utilizando las medidas de sensibilidad, especificidad, precisión global y balanceada, AUC de la curva ROC con su gráfica, el índice de Brier y la distribución gráfica de las predicciones en una matriz de confusión. Las definiciones e interpretaciones de estas medidas se presentan en la Tabla 2. El mejor modelo fue seleccionado según estos estadísticos.

Resultados

Comparación de los modelos

La regresión logística mostró una sensibilidad del 76,5 %, especificidad del 83 %, precisión balanceada del 79,7 % y un AUC de 0,88. El modelo de *machine learning* XGBoost presentó una sensibilidad del 80 %, especificidad del 82 %, precisión balanceada del 81 % y un AUC de 0,87. Al evaluar la matriz de confusión, se observó que el modelo XGBoost tenía un menor número de falsos negativos y una mejor sensibilidad, lo cual se consideró prioritario; con un rendimiento similar en otros parámetros, se decidió seleccionar el modelo XGBoost.

Al evaluar el mejor tratamiento del tipo de procedimiento por CUPS, se encontró que al usar dos dígitos del código (categoría general del procedimiento), el modelo presentaba el mejor rendimiento, por lo que fue el modelo final elegido (Tabla 3).

Características del modelo

El modelo XGBoost escogido utilizó las 16 variables seleccionadas, dando gran importancia a las variables de clasificación de ASA, antecedente de EPOC, inestabilidad hemodinámica preoperatoria y si el procedimiento era urgente o electivo (Figura 2).

El modelo arroja una predicción de mortalidad entre 0 a 100 %. Al evaluar su comportamiento en un conjunto de prueba (20 % de los pacientes) se estimó un número significativo de falsos positivos, pero el 75 % de los mismos presentaban una predicción menor del 75 %, mientras que los pacientes verdaderos positivos tenían predominantemente valores superiores al 80 % (Figura 3). Por otro lado, los verdaderos negativos fueron identificados con una alta certeza, como era de esperarse al optimizarlo por sensibilidad.

Tabla 2. Definiciones de las métricas de evaluación de los modelos.

Métrica	Definición e interpretación
Sensibilidad	Mide la proporción de verdaderos positivos detectados entre todos los casos positivos reales ^{19,20} . Entre más cercano este a 100 % mejor predice los verdaderos positivos.
Especificidad	Mide la proporción de verdaderos negativos detectados entre todos los casos negativos reales ^{19,20} . Entre más cercano este a 100 % mejor predice los verdaderos negativos.
Precisión global	Mide la proporción de predicciones correctas (tanto verdaderos positivos como verdaderos negativos) entre el total de casos evaluados. Entre mayor sea el valor es mejor la herramienta ²⁰ .
Precisión balanceada	Es el promedio de la precisión en cada clase. Una precisión balanceada alta indica que el modelo tiene un buen desempeño en ambas clases, incluso en conjuntos de datos desbalanceados ²¹ .
Área bajo la curva (AUC)	Mide la capacidad de un modelo para distinguir entre clases positivas y negativas. Se obtiene al evaluar diferentes umbrales de predicción, trazando la sensibilidad frente a (1 – especificidad). Un modelo con un AUC cercano a 1 es excelente, indicando una alta capacidad de discriminación, mientras que un AUC de 0,5 sugiere un modelo sin capacidad predictiva, equivalente a realizar predicciones al azar ^{19,20} .
Índice de Brier	Promedio del cuadrado de la diferencia entre las predicciones y las observaciones. Un índice de Brier bajo indica que las predicciones coinciden bien con las observaciones, lo que sugiere que el modelo es bueno ²⁰ .

Fuente: Tabla creada por los autores.

Tabla 3. Métricas de los modelos predictivos para mortalidad perioperatoria.

métrica \ modelo	XGBoost						logit
	Procedimiento específico	Categoría procedimiento ★	Procedimiento general	Número de procedimientos			Procedimiento específico
				≥ 30	≥ 20		
Sensibilidad	84,7	80,4	91,7	81,8	83,1	90,3	76,5
Exactitud	70,1	81,6	67,2	77,8	74,6	62,4	82,9
ROC AUC	86,6	87,1	86,4	87,1	85,6	87,1	88,4
Precisión balanceada	77,2	81,0	79,2	79,8	78,8	76,1	79,7
Especificidad	69,8	81,7	66,7	77,7	74,4	61,9	83,0
Precisión	5,4	7,9	5,4	6,7	5,9	4,7	8,0
Brier	22,9	16,2	22,4	17,6	19,1	22,9	1,7

La estrella indica el modelo seleccionado y el patrón de colores indica que el valor más alto es más claro y el más bajo es más oscuro.

* ROC AUC: Receiver Operating Characteristic Área Bajo la Curva.

Fuente: Tabla creada por los autores.

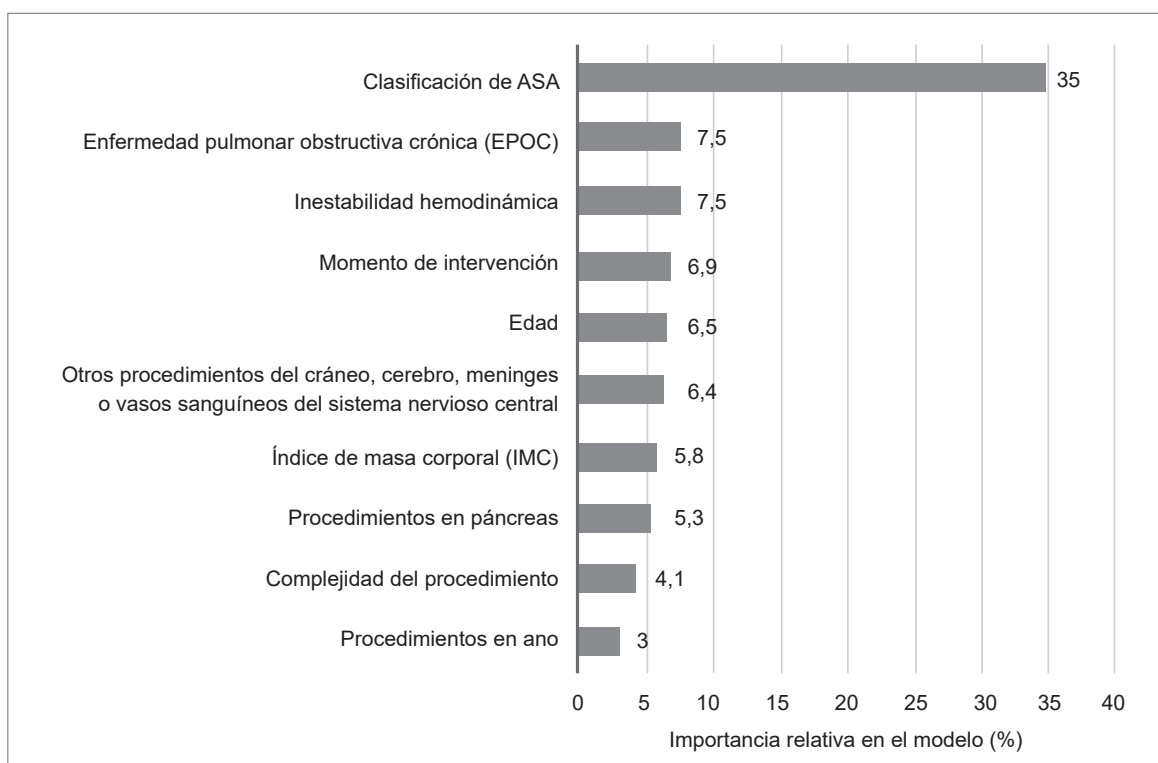


Figura 2. Importancia relativa de las variables con mayor contribución en el modelo predictivo.

Fuente: Figura creada por los autores.

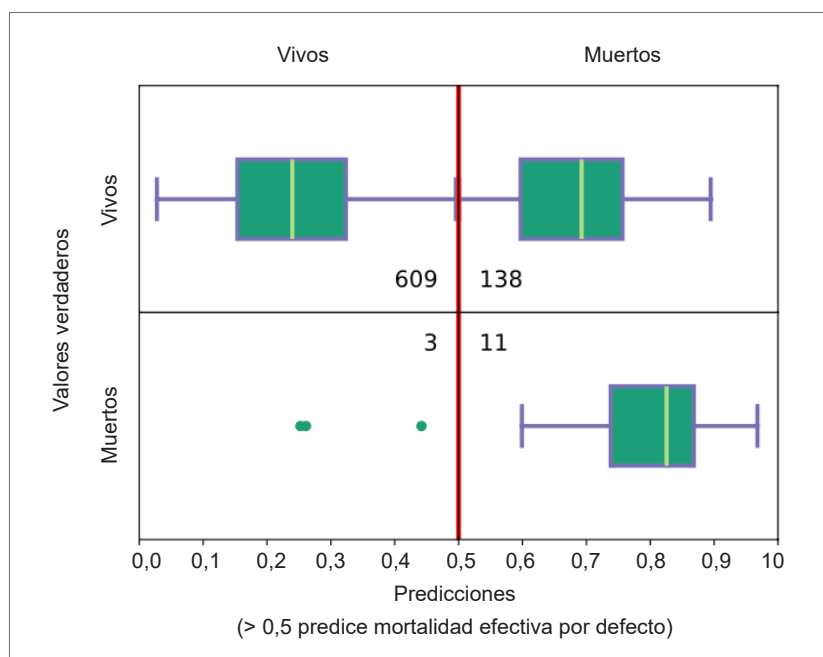


Figura 3. Matriz de confusión modelo XGBoost para mortalidad perioperatoria.

Fuente: Figura creada por los autores.

Al graficar la curva ROC para el modelo, se calculó un área bajo la curva de 87 % y un umbral óptimo de corte del 60 % para determinar si un paciente va a presentar una mortalidad perioperatorio o no (Figura 4).

Discusión

Unos 300 millones de procedimientos quirúrgicos se realizan globalmente, con un estimado de mortalidad de 1-4 % en el perioperatorio²². De hecho, sin la atención adecuada al cuidado quirúrgico, el periodo postoperatorio continuará siendo una causa prevenible de muerte y morbilidad²³. Es por esto por lo que la estimación de riesgos preoperatorios es de suma importancia para el cirujano, los pacientes y sus familiares. Hasta hace poco no había una herramienta disponible que funcionará para múltiples especialidades quirúrgicas, entrenada y validada en población colombiana, por lo que este estudio propuso la creación de una

calculadora de mortalidad perioperatoria basada y validada con datos de pacientes colombianos, obtenidos del estudio multicéntrico ColSOS¹⁵.

Los modelos predictivos involucran el uso de métodos y herramientas matemáticas y/o computacionales para crear modelos que permitan pronosticar desenlaces futuros²⁴. Aunque los modelos predictivos en la actualidad han ganado popularidad y se han extendido en la práctica clínica, el concepto de la predicción para apoyar las decisiones médicas ha sido usado desde tiempos Babilónicos^{25,26}. Esto mismo ocurre en cirugía, donde la literatura ha presentado varios modelos predictivos de mortalidad perioperatoria para procedimientos específicos.

Ramkumar T, et al.²⁷, compararon herramientas predictivas para la mortalidad postoperatoria en pacientes sometidos a resección colorrectal. Utilizaron datos de 347 pacientes (321 electivos y 26 urgentes) y evaluaron POSSUM (Clasificación

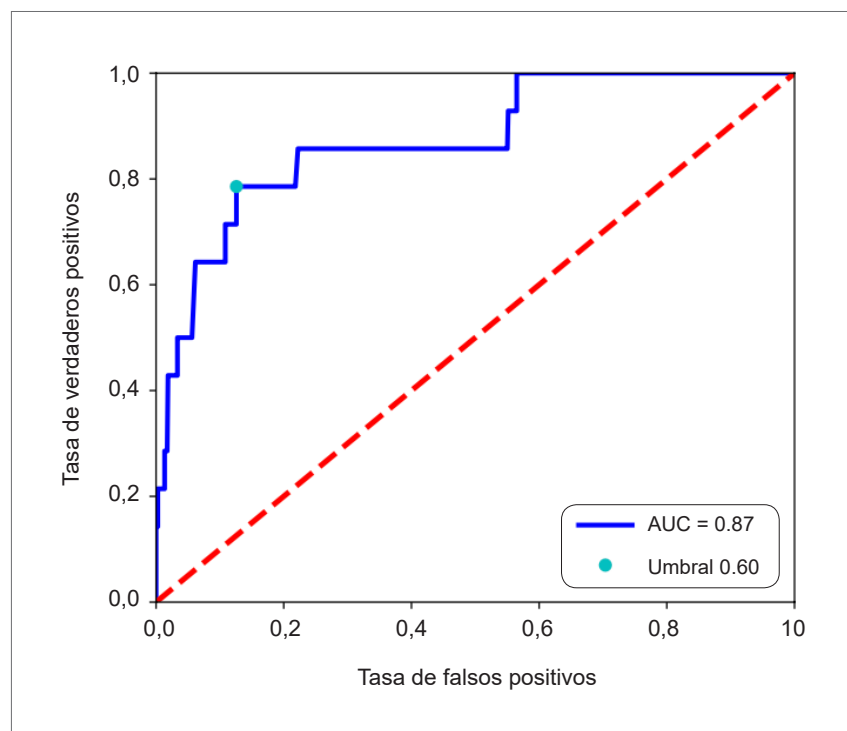


Figura 4. Curva ROC para modelo XGBoost.

Fuente: Figura creada por los autores.

de Severidad Fisiológica y Psicológica para la Enumeración de Mortalidad y Morbilidad), obteniendo un área bajo la curva (AUC) de 0,75, y P-POSSUM, que utiliza una regresión lineal con diferentes componentes para la misma clasificación, logrando una AUC de 0,749.

Otro ejemplo es el de Mathlouthi A, et al.²⁸, quienes desarrollaron un modelo predictivo para la mortalidad perioperatoria posterior a la reparación endovascular de aneurisma. El modelo se entrenó con datos de 36.133 pacientes, de los cuales el 1,1 % presentó el desenlace, obteniendo un estadístico C de 0,749. Así como estos modelos, encontramos predicciones específicas de mortalidad postoperatoria en pacientes cirróticos²⁹, en cirugía de trasplante³⁰, cirugía vascular³¹, cirugía cardiovascular³² y cirugía general³³, entre otros. Pero su utilidad es limitada dado que la mayoría son realizados en un único centro o con pacientes que tienen características muy específicas, que los hacen poco generalizables.

Mientras tanto, la calculadora de desenlaces posoperatorios diseñada por Bilimoria KY, et al.¹², sí incluye distintas especialidades quirúrgicas. Este modelo predictivo fue basado en los datos del ACS NSQIP, que es el programa del Colegio Americano de Cirujanos para la mejora de la calidad, incluyendo más de 1.400.000 pacientes. Es de resaltar que en este modelo se incluyeron los CPT (*Current Procedural Terminology*), que son equiparables a los CUPS utilizados en Colombia. El modelo obtenido presentó una AUC de 0,94 para mortalidad y 0,81 para morbilidad, lo cual indica un alto grado de discriminación.

Es importante hacer énfasis en que las herramientas predictivas son muy útiles, pero pueden ser utilizadas de manera incorrecta si son usadas en pacientes, poblaciones o contextos para las cuales no fueron creadas^{27,34}. Está es la razón por la que son tan relevantes los datos presentados en este estudio, dado que es una calculadora entrenada y diseñada para la población y el contexto colombianos.

Esta herramienta diseñada para mortalidad perioperatoria en Colombia presentó un AUC de 0,87, la cual indica una buena discriminación del desenlace. Al compararlo con los estudios iniciales

de la calculadora de desenlaces posoperatorios del ACS, se evidencia una menor discriminación, con un reporte de AUC de 0,94¹², pero validaciones posteriores evidencian un menor AUC, de 0,88¹³ y de 0,77¹⁴. Por lo que se considera que la discriminación del modelo presentado es equiparable a otras herramientas, lo que lo hace viable para ser estudiado y considerado en la práctica clínica.

El modelo predictivo que finalmente fue utilizado fue el XGBoost, que es un modelo derivado del *machine learning*. Esta comparación también fue llevada a cabo por Liu Y, et al.¹³, en donde se estimó el AUC de la predicción por medio de regresión logística y por medio del XGBoost, encontrando una discriminación ligeramente superior del XGBoost sobre la regresión (AUC 0,829 vs 0,825), por lo cual consideraron actualizar la calculadora de desenlaces posoperatorios del ACS al método de *machine learning*. Se considera, en base a nuestros resultados y la literatura internacional, que el uso de XGBoost es superior a la regresión logística, pero aun así, se aclara que la regresión presenta una adecuada discriminación y podría ser utilizada para la predicción de desenlaces.

Limitaciones y fortalezas

Es importante entender las limitaciones y fortalezas del estudio para el uso de la calculadora y el método de predicción. Los datos fueron entrenados y validados usando la base del estudio CoLSOS, un estudio que recogió datos a nivel nacional, pero con una preponderancia de la región Andina y poca afluencia de las zonas periféricas del país. Adicionalmente, incluyó únicamente pacientes mayores de 18 años, por lo que no está diseñada para uso pediátrico.

En cuanto a las fortalezas, se destaca una alta sensibilidad de la calculadora sin comprometer las otras métricas, lo que indica una priorización del modelo para categorizar correctamente los casos de mortalidad. Esto puede llevar en algunas ocasiones a un aumento de falsos positivos, que es donde disminuye la discriminación del modelo. Sin embargo, esto se ve atenuado por los riesgos elevados que otorga la calculadora en los verdaderos positivos comparado con los falsos positivos.

Dentro de las fortalezas más relevantes, se encuentra que la calculadora fue diseñada y entrenada con datos colombianos, por lo que genera predicciones más realistas en esta región, que otras calculadoras existentes; y está disponible para su uso (<https://app.ai-med.co>). Finalmente, el uso de inteligencia artificial para su clasificación destaca el análisis estadístico actualizado y complejo del modelo predictivo.

Conclusión

El presente estudio ha desarrollado y validado una calculadora de mortalidad perioperatoria específicamente diseñada para la población colombiana, utilizando datos del estudio multicéntrico ColSOS. Se compararon modelos predictivos de regresión logística y *machine learning* (XGBoost), encontrando que el modelo XGBoost ofreció una mejor sensibilidad y un menor número de falsos negativos, lo cual es crítico para la identificación temprana de pacientes en riesgo. El modelo XGBoost, con un AUC de 0,87, mostró un rendimiento similar a otras herramientas internacionales, pero con la ventaja de estar ajustado a las características y contextos específicos de la población colombiana.

La implementación de esta herramienta en la práctica clínica mejora significativamente la atención quirúrgica en Colombia, permitiendo la identificación y el manejo temprano de pacientes en riesgo, contribuyendo así a la reducción de la mortalidad perioperatoria. Esta calculadora no solo cumple con el cuarto indicador de la Comisión de Lancet en Cirugía Global, sino que también posiciona a Colombia como líder regional en el uso de herramientas avanzadas para la atención quirúrgica.

Agradecimientos

El presente estudio no podría haberse llevado a cabo sin el apoyo técnico y tecnológico de AIMed y ST&T Ltda., quienes fueron fundamentales para el desarrollo de la calculadora y el posicionamiento de la herramienta en una página web (<https://ai-med.co/>), haciéndola accesible para todos los cirujanos, médicos y demás personal de salud.

Consideraciones éticas

Consentimiento informado: El presente estudio se ciñó a las normas éticas internacionales y nacionales. La aprobación inicial de este estudio y sus respectivos subanálisis fueron dados por el Comité de Ética Institucional de la Fundación Cardioinfantil-Instituto de Cardiología (421 2021) y la Universidad del Rosario (DVO005 2011-CV1580). Es un estudio sin riesgo, según la Resolución 8430 de 1993 de Colombia, por lo que sólo se obtuvo consentimiento informado escrito u oral según lo definió el comité de ética de cada institución participante en el estudio ColSOS.

Conflictos de intereses: Los autores declararon que no tienen conflictos de interés.

Uso de tecnologías asistidas por Inteligencia Artificial: El estudio utiliza inteligencia artificial como método de análisis más no para la escritura de este manuscrito.

Fuentes de financiación: El estudio ColSOS, en el que se basa este trabajo, fue apoyado por la Universidad del Rosario y la Fundación Cardioinfantil-Instituto de Cardiología, con el número de subvención CTO 057-2021 y el ID del proyecto IV-FGV017, recibiendo US\$375,000 de cada uno, para un total de US\$750,000. El contacto para la financiación de este estudio es Melba Mercedes Mojica-Amaya, Profesional en Gestión de Proyectos de Investigación en la Universidad del Rosario, Bogotá, D.C., Colombia. Adicionalmente, se recibió apoyo técnico y tecnológico de AIMed y ST&T Ltda. para el desarrollo de la calculadora y su publicación en una página web, facilitando su uso y socialización entre los cirujanos y el personal de salud.

Contribución de los autores:

- Concepción y diseño del estudio: Nicolás Lozano-Suárez, Juan Felipe Salamanca-Lozano, Paulo Andrés Cabrera-Rivera, Leonardo Briceño-Ayala, Carlos J. Pérez-Rivera
- Adquisición de datos: Nicolás Lozano-Suárez, Carlos J. Pérez-Rivera; Colsos Collaborative.
- Análisis e interpretación de datos: Nicolás Lozano-Suárez, Juan Felipe Salamanca-Lozano, Paulo Andrés Cabrera-Rivera, Leonardo Briceño-Ayala
- Redacción del manuscrito: Nicolás Lozano-Suárez, Juan Felipe Salamanca-Lozano, Carlos J. Pérez-Rivera, Paulo Andrés Cabrera-Rivera, Leonardo Briceño-Ayala
- Revisión crítica del contenido intelectual: Nicolás Lozano-Suárez, Paulo Andrés Cabrera-Rivera, Leonardo Briceño-Ayala, Carlos J. Pérez-Rivera.

Referencias

1. Bath M, Bashford T, Fitzgerald JE. What is 'global surgery'? Defining the multidisciplinary interface between surgery, anaesthesia and public health. *BMJ Glob Health*. 2019;4:e001808. <https://doi.org/10.1136/bmjgh-2019-001808>
2. Meara JG, Leather AJM, Hagander L, Alkire BC, Alonso N, Ameh EA, et al. Global Surgery 2030: Evidence and solutions for achieving health, welfare, and economic development. *Lancet*. 2015;386:569-624. [https://doi.org/10.1016/S0140-6736\(15\)60160-X](https://doi.org/10.1016/S0140-6736(15)60160-X)
3. Biccard BM, Madiba TE, Kluyts HL, Munlemvo DM, Mazimbamuto FD, Basenero A, et al. Perioperative patient outcomes in the African surgical outcomes study: A 7-day prospective observational cohort study. *Lancet*. 2018;391:1589-98. [https://doi.org/10.1016/S0140-6736\(18\)30001-1](https://doi.org/10.1016/S0140-6736(18)30001-1)
4. Jawad M, Baigi A, Oldner A, Pearse RM, Rhodes A, Seeman-Lodding H, et al. Swedish surgical outcomes study (SweSOS). *Eur J Anaesthesiol*. 2016;33:317-25. <https://doi.org/10.1097/EJA.0000000000000352>
5. Osinaike B, Ayandipo O, Onyeka T, Alagbe-Briggs O, Mohammed A, Oyedepo O, et al. Nigerian surgical outcomes - Report of a 7-day prospective cohort study and external validation of the African surgical outcomes study surgical risk calculator. *Int J Surg*. 2019;68:148-56. <https://doi.org/10.1016/j.ijssu.2019.06.003>
6. Hewitt-Smith A, Bulamba F, Olupot C, Musana F, Ochieng JP, Lipnick MS, et al. Surgical outcomes in eastern Uganda: a one-year cohort study. *South Afr J Anaesth Analg*. 2018;24:122-7. <https://doi.org/10.1080/22201181.2018.1517476>
7. Gómez-Samper AF, Herrera-Almarío GE, Tulloch D, Blanco D, Cardoso LL, Nunez-Rocha RE, et al. A granular analysis of service delivery for surgical system strengthening: Application of the Lancet indicators for policy development in Colombia. *Lancet Reg Health Am*. 2022;10:100217. <https://doi.org/10.1016/j.lana.2022.100217>
8. Hanna JS, Herrera-Almarío GE, Pinilla-Roncancio M, Tulloch D, Valencia SA, Sabatino ME, et al. Use of the six core surgical indicators from the Lancet Commission on Global Surgery in Colombia: a situational analysis. *Lancet Glob Health*. 2020;8:e699-e710. [https://doi.org/10.1016/S2214-109X\(20\)30090-5](https://doi.org/10.1016/S2214-109X(20)30090-5)
9. Lozano-Suárez N, Polanía-Sandoval CA, García-Méndez JP, Velandia-Sánchez CA, Saavedra JD, Corso-Ramírez JM, et al. Mortalidad perioperatoria en Tolima, perspectivas del cuarto indicador de The Lancet Commission on Global Surgery: Un análisis preliminar de la cohorte prospectiva del estudio Colombian Surgical Outcomes Study (ColSOS). *Rev Colomb Cir*. 2023;38:501-11. <https://doi.org/10.30944/20117582.2289>
10. Bruno E, White MC, Baxter LS, Ravelojaona VA, Rakotoarison HN, Andriamanjato HH, et al. An evaluation of preparedness, delivery and impact of surgical and anesthesia care in Madagascar: A framework for a National Surgical Plan. *World J Surg*. 2017;41:1218-24. <https://doi.org/10.1007/s00268-016-3847-9>
11. Burssa D, Teshome A, Iverson K, Ahearn O, Ashengo T, Barash D, et al. Safe surgery for all: Early lessons from implementing a national government-driven surgical plan in Ethiopia. *World J Surg*. 2017;41:3038-45. <https://doi.org/10.1007/s00268-017-4271-5>
12. Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmiecik TE, Ko CY, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: A decision aid and informed consent tool for patients and surgeons. *J Am Coll Surg*. 2013;217:833-42.e1-3. <https://doi.org/10.1016/j.jamcollsurg.2013.07.385>
13. Liu Y, Ko CY, Hall BL, Cohen ME. American College of Surgeons NSQIP risk calculator accuracy using a machine learning algorithm compared with regression. *J Am Coll Surg*. 2023;236:1024-30. <https://doi.org/10.1097/XCS.0000000000000556>
14. Cohen ME, Liu Y, Ko CY, Hall BL. An examination of American College of Surgeons NSQIP surgical risk calculator accuracy. *J Am Coll Surg*. 2017;224:787-795.e1. <https://doi.org/10.1016/j.jamcollsurg.2016.12.057>
15. Pérez-Rivera CJ, Lozano-Suárez N, Velandia-Sánchez A, Polanía-Sandoval CA, García-Méndez JP, Idarraga-Ayala SV, et al. Perioperative mortality in Colombia: Perspectives of the fourth indicator in the Lancet Commission on Global Surgery - Colombian Surgical Outcomes Study (ColSOS) - A protocol for a multicentre prospective cohort study. *BMJ Open*. 2022;12:e063182. <https://doi.org/10.1136/bmjopen-2022-063182>
16. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med*. 2015;13:1. <https://doi.org/10.1186/s12916-014-0241-z>
17. Meara JG, Greenberg SLM. The Lancet Commission on Global Surgery. Global surgery 2030: Evidence and solutions for achieving health, welfare and economic development. *Surgery*. 2015;157:834-5. <https://doi.org/10.1016/j.surg.2015.02.009>
18. Meara JG, Leather AJM, Hagander L, Alkire BC, Alonso N, Ameh EA, et al. Global Surgery 2030: evidence and solutions for achieving health, welfare, and economic development. *Int J Obstet Anesth*. 2016;25:75-8. <https://doi.org/10.1016/j.ijoa.2015.09.006>
19. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models. A framework for traditional and novel measures. *Epidemiology*. 2010;21:128-38. <https://doi.org/10.1097/EDE.0b013e3181c30fb2>
20. Dankers FJWM, Traverso A, Wee L, van Kuijk SMJ. Prediction modeling methodology. En: Kubben P, Dumontier M, Dekker A, eds. *Fundamentals of clinical data science*. Cham: Springer International Publishing; 2019. p. 101-20. https://doi.org/10.1007/978-3-319-99713-1_8

21. Brodersen KH, Ong CS, Stephan KE, Buhmann JM. The balanced accuracy and its posterior distribution. En: 2010 20th International Conference on Pattern Recognition. Istanbul, Turkey: IEEE; 2010. p. 3121-4. <https://doi.org/10.1109/ICPR.2010.764>
22. Dobson GP. Trauma of major surgery: A global problem that is not going away. *Int J Surg.* 2020;81:47-54. <https://doi.org/10.1016/j.ijssu.2020.07.017>
23. Mock CN, Donkor P, Gawande A, Jamison DT, Kruk ME, Debas HT, et al; DCP3 Essential Surgery Author Group. Essential surgery: Key messages from Disease Control Priorities, 3rd edition. *Lancet.* 2015;385:2209-19. [https://doi.org/10.1016/S0140-6736\(15\)60091-5](https://doi.org/10.1016/S0140-6736(15)60091-5)
24. Toma M, Wei OC. Predictive modeling in medicine. *Encyclopedia.* 2023;3:590-601. <https://doi.org/10.3390/encyclopedia3020042>
25. Zhou Y, Gould D, Choong P, Dowsey M, Schilling C. Implementing predictive tools in surgery: A narrative review in the context of orthopaedic surgery. *ANZ J Surg.* 2022;92:3162-9. <https://doi.org/10.1111/ans.18044>
26. Horstmannshoff HFJ, Stol MF, van Tilburg CR. Magic and rationality in ancient Near Eastern and Graeco-Roman medicine. 1st ed. Leiden, Boston: Medicine; 2004. https://doi.org/10.1163/9789047414315_002
27. Ramkumar T, Ng V, Fowler L, Farouk R. A Comparison of POSSUM, P-POSSUM and Colorectal POSSUM for the prediction of postoperative mortality in patients undergoing colorectal resection. *Dis Colon Rectum.* 2006;49:330-5. <https://doi.org/10.1007/s10350-005-0290-7>
28. Mathlouthi A, Abdelkarim A, Elsayed N, Ramakrishnan G, Naazie I, Malas MB. Novel risk score calculator for perioperative mortality after EVAR with incorporation of anatomical factors. *Ann Vasc Surg.* 2023;94:289-95. <https://doi.org/10.1016/j.avsg.2023.02.020>
29. Mahmud N, Fricker Z, Hubbard RA, Ioannou GN, Lewis JD, Taddei TH, et al. Risk prediction models for post-operative mortality in patients with cirrhosis. *Hepatology.* 2021;73:204-18. <https://doi.org/10.1002/hep.31558>
30. Bruns H, Lozanovski VJ, Schultze D, Hillebrand N, Hinz U, Büchler MW, et al. Prediction of postoperative mortality in liver transplantation in the era of MELD-based liver allocation: A multivariate analysis. *PLoS One.* 2014;9:e98782. <https://doi.org/10.1371/journal.pone.0098782>
31. Eslami MH, Rybin DV, Doros G, Siracuse JJ, Farber A. External validation of Vascular Study Group of New England risk predictive model of mortality after elective abdominal aorta aneurysm repair in the Vascular Quality Initiative and comparison against established models. *J Vasc Surg.* 2018;67:143-50. <https://doi.org/10.1016/j.jvs.2017.05.087>
32. Fan Y, Dong J, Wu Y, Shen M, Zhu S, He X, et al. Development of machine learning models for mortality risk prediction after cardiac surgery. *Cardiovasc Diagn Ther.* 2022;12:12-23. <https://doi.org/10.21037/cdt-21-648>
33. Vaid S, Bell T, Grim R, Ahuja V. Predicting risk of death in general surgery patients on the basis of preoperative variables using American College of Surgeons National Surgical Quality Improvement Program data. *Perm J.* 2012;16:10-7. <https://doi.org/10.7812/TPP/12-019>
34. Bleeker SE, Moll HA, Steyerberg EW, Donders ART, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: A clinical example. *J Clin Epidemiol.* 2003;56:826-32. [https://doi.org/10.1016/S0895-4356\(03\)00207-5](https://doi.org/10.1016/S0895-4356(03)00207-5)