



Universidad Autónoma de Querétaro  
Facultad de Ingeniería  
Maestría en Ciencias en Inteligencia Artificial

Ensamble de Modelos de Aprendizaje Profundo Para  
Clasificación de Riesgo de Coledocolitiasis

Tesis

Que como parte de los requisitos para obtener el Grado de  
Maestro en Ciencias en Inteligencia Artificial

Presenta  
Ing. Enrique Mena Camilo

---

Dirigido por:  
Dr. Sebastián Salazar Colores

---

Co-dirigido por:  
Dr. Edgar Efrén Lozada Hernández

---

Dr. Sebastián Salazar Colores  
Presidente

Dr. Edgar Efrén Lozada Hernández  
Secretario

Dr. Juan Manuel Ramos Arreguín  
Vocal

Dr. Marco Antonio Aceves Fernández  
Suplente

Dr. Saúl Tovar Arriaga  
Suplente

Centro Universitario, Querétaro, Qro.  
Noviembre 2024  
México

La presente obra está bajo la licencia:  
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>



CC BY-NC-ND 4.0 DEED

Atribución-NoComercial-SinDerivadas 4.0 Internacional

### Usted es libre de:

**Compartir** — copiar y redistribuir el material en cualquier medio o formato

La licenciante no puede revocar estas libertades en tanto usted siga los términos de la licencia

### Bajo los siguientes términos:



**Atribución** — Usted debe dar [crédito de manera adecuada](#), brindar un enlace a la licencia, e [indicar si se han realizado cambios](#). Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.



**NoComercial** — Usted no puede hacer uso del material con [propósitos comerciales](#).



**SinDerivadas** — Si [remezcla, transforma o crea a partir](#) del material, no podrá distribuir el material modificado.

**No hay restricciones adicionales** — No puede aplicar términos legales ni [medidas tecnológicas](#) que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.

### Avisos:

No tiene que cumplir con la licencia para elementos del material en el dominio público o cuando su uso esté permitido por una [excepción o limitación](#) aplicable.

No se dan garantías. La licencia podría no darle todos los permisos que necesita para el uso que tenga previsto. Por ejemplo, otros derechos como [publicidad, privacidad, o derechos morales](#) pueden limitar la forma en que utilice el material.



# Dedicatorias

Dedico este trabajo, de manera especial, a mis padres, Mtra. Laura Elena Camilo García y Mtro. Enrique Mena Salgado. Desde temprana edad, ustedes me acercaron al mundo de la tecnología, me inculcaron el valor del autoaprendizaje y me enseñaron a apreciar la investigación. Gracias por motivarme a salir de mi zona de confort y explorar nuevos horizontes. Este logro es el reflejo de la educación y los valores que siempre me han transmitido, y por eso, es para ustedes.

Dedico también este esfuerzo a mis abuelos, Alfonso Camilo y Bertha García. Desde mi infancia, su apoyo y su sabiduría me han inspirado a superarme y a seguir aprendiendo cada día. Su presencia y enseñanzas forman una parte esencial de este trabajo.

Finalmente, dedico este escrito a mis hermanas, Lic. Vania y Betsy, quienes ven en mí una fuente de inspiración, cuando en realidad son ellas la verdadera inspiración. Son mi motivo de existir y mi mayor alegría, y en cada paso de este camino, han sido mi razón para continuar.



# Agradecimientos

A todos aquellos que contribuyeron con sus valiosos conocimientos y orientación para el desarrollo exitoso de este proyecto: Dr. Sebastián Salazar Colores, Dr. Edgar Efrén Lozada Hernández, y Dr. Marco Antonio Aceves Fernández. Su guía y apoyo a lo largo de estos años han sido esenciales para la culminación de este trabajo, y sin su dedicación y compromiso, este logro no hubiera sido posible.

A mi familia, por su confianza incondicional y el apoyo constante que me han brindado no solo durante estos años, sino a lo largo de toda mi vida. Sin ustedes, mis logros carecerían de sentido y nunca habría llegado hasta aquí.

A mis amigos y colegas, M. Sc. Gerardo Hernández Nava, M. Sc. Sheila Leyva López, M. Sc. Cecilia Gabriela Rodríguez Flores (Neuro Geek). Gracias por su aporte en ideas, recomendaciones y por las risas compartidas, que hicieron que nunca pensara en rendirme.

Al Dr. Omar Piña Ramírez, quien sembró en mí la pasión por la ciencia desde 2018 y me mostró que la investigación debe trascender del laboratorio.

Al M. Sc. Daniel Martínez Aguilar y al M. Sc. José Iván Esquivel Mendoza, por su apoyo, flexibilidad y colaboración durante estos años.

Al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT), por el financiamiento de este proyecto mediante la beca de investigación de posgrado, y al Centro de Investigaciones en Óptica (CIO), por su respaldo para que los productos académicos de

VI

este trabajo pudieran ver la luz.

A todos ustedes, que confiaron en mí y me brindaron su apoyo, tiempo, conocimientos y recursos: gracias.



# Índice general

<b>1</b>	<b>Introducción</b>	<b>1</b>
<b>2</b>	<b>Fundamentación Teórica</b>	<b>5</b>
2.1	Inteligencia artificial aplicada a la medicina . . . . .	5
2.2	Aprendizaje automático . . . . .	5
2.2.1	Aprendizaje supervisado . . . . .	6
2.3	Aprendizaje profundo . . . . .	6
2.4	Redes neuronales artificiales . . . . .	6
2.4.1	Redes neuronales convolucionales 1D . . . . .	7
2.5	Métodos de ensamble . . . . .	9
2.5.1	Ensamblados bagging . . . . .	9
2.5.2	Ensamblados stacking . . . . .	11
2.5.3	Ensamblados boosting . . . . .	12
2.6	Cifras de mérito . . . . .	13
<b>3</b>	<b>Antecedentes</b>	<b>17</b>
<b>4</b>	<b>Hipótesis</b>	<b>21</b>
<b>5</b>	<b>Objetivos</b>	<b>23</b>
5.1	Objetivo general . . . . .	23
5.2	Objetivos específicos . . . . .	23
<b>6</b>	<b>Material y Métodos</b>	<b>25</b>

6.1	Características del sistema utilizado para experimentación . . . . .	25
6.2	Conjunto de datos utilizado . . . . .	26
6.3	Procedimiento general . . . . .	27
6.4	Preprocesamiento de los datos . . . . .	28
6.5	Diseño y selección de modelos base . . . . .	29
6.6	Diseño y selección de modelos de ensamble . . . . .	31
6.7	Criterios de evaluación de modelos . . . . .	32
6.8	Diseño de aplicación web . . . . .	33
<b>7</b>	<b>Resultados y Discusión</b>	<b>35</b>
7.1	Preprocesamiento de los datos . . . . .	35
7.2	Modelos base . . . . .	38
7.3	Modelo de ensamble . . . . .	41
7.4	Comparación con estado del arte . . . . .	43
7.5	Aplicación web . . . . .	45
<b>8</b>	<b>Conclusiones</b>	<b>49</b>
	<b>Referencias bibliográficas</b>	<b>51</b>
<b>A</b>	<b>Requisitos para obtención de grado</b>	<b>57</b>
<b>B</b>	<b>Productos académicos</b>	<b>63</b>
<b>C</b>	<b>Descripción de conjunto de datos original</b>	<b>67</b>
<b>D</b>	<b>Distribuciones de los atributos utilizados</b>	<b>73</b>



# Índice de cuadros

3.1	Métodos utilizados para el diagnóstico de coledocolitiasis mediante datos clínicos multivariados. . . . .	19
6.1	Características de hardware del sistema utilizado para el desarrollo del proyecto.	26
6.2	Características de software del sistema utilizado para el desarrollo del proyecto.	26
7.1	Atributos del conjunto de datos agrupados por la variable objetivo. . . . .	36
7.2	Comparativa del desempeño de los diversos algoritmos desarrollados contra un modelo entrenado con el mismo conjunto de datos (Parte 1). . . . .	43
7.3	Comparativa del desempeño de los diversos algoritmos desarrollados contra un modelo entrenado con el mismo conjunto de datos (Parte 2). . . . .	44
7.4	Comparativa del desempeño del algoritmo de ensamble, CPRE y guías clínicas.	45



# Índice de figuras

2.1	Diagrama de RNA multicapa . . . . .	7
2.2	Diagrama de ensamble de modelos por Bagging . . . . .	10
2.3	Diagrama de ensamble de modelos por Stacking . . . . .	12
2.4	Diagrama de ensamble de modelos por Boosting . . . . .	13
6.1	Diagrama general del desarrollo del proyecto. . . . .	28
6.2	Diagrama del preprocesamiento aplicado al conjunto de datos original. . . . .	29
6.3	Diagrama general para la selección de modelos base. . . . .	31
6.4	Diagrama general para la selección de métodos de ensamble. . . . .	32
6.5	Diagrama general de la interacción del usuario y la aplicación web. . . . .	34
7.1	Análisis del efecto del preprocesamiento en la distribución del atributo Índice de Masa Corporal. . . . .	37
7.2	Distribución de las métricas de desempeño de los 100 experimentos de Regresión Logística. . . . .	39
7.3	Distribución de las métricas de desempeño de los 100 experimentos de Análisis de Discriminante Lineal. . . . .	40
7.4	Distribución de las métricas de desempeño de los 100 experimentos de CNN 1D. . . . .	41
7.5	Diagrama general del ensamble resultante del proyecto. . . . .	42
7.6	Captura de pantalla de la aplicación web mostrando un resultado positivo. . . . .	46
7.7	Código QR para acceder a la aplicación web. . . . .	47

XIV

A.1 Registro del protocolo de investigación ante la Facultad de Ingeniería de la Universidad Autónoma de Querétaro. . . . . 58

A.2 Aprobación de protocolo de investigación por el comité de ética de Facultad de Ingeniería de la Universidad Autónoma de Querétaro. . . . . 59

A.3 Constancia de acreditación de Examen de Comprensión de Textos en Inglés emitida en Agosto 2023. . . . . 60

A.4 Constancia de acreditación de Examen de Manejo de la Lengua (Inglés) emitida en Febrero 2023. . . . . 61

B.1 Porta de artículo publicado en Congreso Nacional de Ingeniería Biomédica en Octubre 2023. DOI: 10.1007/978-3-031-46933-6\_12. . . . . 64

B.2 Portada de artículo publicado en revista MDPI-Diagnostics en Junio 2024. DOI: 10.3390/diagnostics14121278. . . . . 65

D.1 Análisis del efecto del preprocesamiento en la distribución del atributo Edad. 73

D.2 Análisis del efecto del preprocesamiento en la distribución del atributo Género. 74

D.3 Análisis del efecto del preprocesamiento en la distribución del atributo Origen Étnico. . . . . 74

D.4 Análisis del efecto del preprocesamiento en la distribución del atributo Índice de Masa Corporal. . . . . 75

D.5 Análisis del efecto del preprocesamiento en la distribución del atributo Diabetes. 75

D.6 Análisis del efecto del preprocesamiento en la distribución del atributo Cirrosis. 76

D.7 Análisis del efecto del preprocesamiento en la distribución del atributo Bili-rrubina Máxima. . . . . 76

D.8 Análisis del efecto del preprocesamiento en la distribución del atributo Cálculos en Conducto Biliar en Imagenología. . . . . 77

D.9 Análisis del efecto del preprocesamiento en la distribución del atributo Diámetro de Colédoco en Ultrasonido. . . . . 77

D.10 Análisis del efecto del preprocesamiento en la distribución del atributo Vesícula Biliar. . . . . 78





# Resumen

Este proyecto presenta el desarrollo y validación de un modelo de inteligencia artificial para la predicción de riesgo de coledocolitiasis, utilizando un enfoque de ensamble que combina técnicas de *stacking* y *boosting*. Con una sensibilidad de 96.81 %, especificidad de 93.16 % y un puntaje F1 de 94.87 %, el modelo demuestra una alta precisión en la detección de casos positivos y una reducción significativa de falsos positivos. Este enfoque de ensamble integra modelos lineales (Regresión Logística y Análisis de Discriminante Lineal) y no lineales (CNN 1D), aprovechando las fortalezas de cada uno para maximizar el rendimiento.

Además, se implementó una aplicación web en el dominio <https://cbds-risk-app-1035249958862.us-central1.run.app/>, permitiendo que el personal de salud acceda fácilmente al modelo sin necesidad de conocimientos técnicos en IA. Esta herramienta facilita la evaluación preliminar del riesgo de coledocolitiasis, optimizando el flujo de trabajo clínico y reduciendo la dependencia de procedimientos invasivos en pacientes con bajo riesgo.

Los resultados del modelo de ensamble posicionan esta herramienta como una alternativa complementaria al CPRE y las guías clínicas de la ASGE, acercándose a la precisión de los métodos invasivos pero de manera no invasiva. Este trabajo no solo contribuye al campo de la inteligencia artificial aplicada a la medicina, sino que también establece una base sólida para el uso práctico de IA en la detección temprana de enfermedades gastrointestinales.

Aunque el modelo mostró resultados prometedores, futuras investigaciones podrían en-

## XVIII

focarse en ampliar el conjunto de datos clínicos y optimizar la aplicación web. Este estudio evidencia que la inteligencia artificial tiene el potencial de transformar el diagnóstico médico, mejorando la precisión y accesibilidad de las herramientas diagnósticas en el ámbito clínico.

**Palabras clave:** Modelos de ensamble, Coledocolitiasis, Red neuronal convolucional 1D, Clasificación de riesgo

# Abstract

This project presents the development and validation of an artificial intelligence model for predicting the risk of choledocholithiasis, using an ensemble approach that combines stacking and boosting techniques. With a sensitivity of 96.81 %, specificity of 93.16 %, and an F1 score of 94.87 %, the model demonstrates high accuracy in detecting positive cases and significantly reduces false positives. This ensemble approach integrates both linear (Logistic Regression and Linear Discriminant Analysis) and non-linear models (1D CNN), leveraging each one's strengths to maximize performance.

Additionally, a web application was implemented at the domain <https://cbds-risk-app-1035249958862.us-central1.run.app/>, enabling healthcare personnel to access the model easily without requiring technical expertise in AI. This tool facilitates preliminary risk assessment for choledocholithiasis, streamlining the clinical workflow and reducing reliance on invasive procedures for low-risk patients.

The ensemble model's results position this tool as a complementary alternative to ERCP and ASGE clinical guidelines, approaching the accuracy of invasive methods in a non-invasive manner. This work not only contributes to the field of AI applied to medicine but also establishes a solid foundation for the practical use of AI in the early detection of gastrointestinal diseases.

While the model showed promising results, future research could focus on expanding the clinical dataset and optimizing the web application. This study demonstrates that artificial

XX

intelligence has the potential to transform medical diagnostics, enhancing the precision and accessibility of diagnostic tools within the clinical setting.

**Keywords:** Ensemble models, Choledocholithiasis, 1D convolutional neural network, Risk classification

# Capítulo 1

## Introducción

La coledocolitiasis se caracteriza por la obstrucción parcial o completa del conducto colédoco debido a la presencia de cálculos biliares. Esta condición, que a menudo surge como una complicación de la colelitiasis (formación de cálculos biliares en la vesícula biliar), representa un desafío significativo para la salud. Se estima que entre el 10 % y el 20 % de los casos de cálculos biliares presentan cálculos en el conducto colédoco (CBDS, por sus siglas en inglés) [1, 2]. Diversos estudios realizados a nivel mundial han determinado que la prevalencia de esta condición varía entre el 9 % y el 21 %, dependiendo de la región de estudio [3, 4, 5].

Existen diferentes métodos para diagnosticar CBDS, entre los más utilizados se encuentran las imágenes por ultrasonido, resonancia magnética y tomografía computarizada. Sin embargo, el estándar de oro para esta tarea es un procedimiento clínico llamado colangio-pancreatografía retrógrada endoscópica (CPRE), el cual presenta una sensibilidad que oscila entre el 80 % y el 93 %, y una especificidad del 100 % [6, 7, 8]. No obstante, existen diversas complicaciones asociadas a la CPRE, desde variaciones en la disponibilidad de equipo y personal capacitado entre instituciones, hasta complicaciones clínicas como pancreatitis, colangitis, perforaciones del duodeno o del conducto biliar y sangrado interno [7, 8, 9]. Se estima que el 15 % de los pacientes sometidos a la CPRE presentarán alguna de estas compli-

caciones, de los cuales entre el 0.2 % y el 0.5 % sufrirán complicaciones mortales [7, 10, 11].

Instituciones reconocidas como la Sociedad Americana de Endoscopia Gastrointestinal (ASGE) han sugerido criterios basados en indicadores clínicos para categorizar los niveles de sospecha de coledocolitiasis. Estos criterios consideran diversos factores, como la presencia de cálculos en el conducto colédoco, la ocurrencia de colangitis, los niveles de bilirrubina total, la dilatación del conducto colédoco, la pancreatitis y la edad del paciente. Usando estos factores, el riesgo de coledocolitiasis se clasifica en categorías baja, media y alta. Aunque estos criterios han sido evaluados en diversos estudios, mostrando una sensibilidad de hasta el 90.9 % y una especificidad de hasta el 26.7 % [12, 13, 14], son considerados una opción efectiva y no invasiva para el diagnóstico de coledocolitiasis, aunque no alcanzan el nivel de desempeño del método CPRE.

En este contexto, el aprendizaje automático surge como un aliado poderoso, ofreciendo contribuciones significativas al campo médico al aprovechar sus capacidades de reconocimiento de patrones para analizar datos clínicos. Sin embargo, muchos de los enfoques actuales se centran en el uso de métodos tradicionales de aprendizaje automático, como los bosques aleatorios y las máquinas de soporte vectorial, mientras que otros se han enfocado exclusivamente en el diagnóstico basado en imágenes médicas.

Desde hace varios años, se han presentado diversas contribuciones a los métodos de ensamble en el aprendizaje automático. Estos métodos permiten generar un clasificador que combina los resultados de varios clasificadores individuales, estableciendo criterios de votación para producir una nueva clasificación [15]. Los métodos de ensamble buscan aprovechar las fortalezas de cada clasificador individual y fusionarlas para obtener un clasificador compuesto que sea más robusto y estable que los clasificadores por separado.

A pesar de que la CPRE es el estándar de oro para el diagnóstico, su uso está asociado a diversas complicaciones y requiere equipamiento especializado y personal capacitado, lo que limita su disponibilidad. Los métodos basados en criterios clínicos, aunque no invasivos, no alcanzan el mismo desempeño. Por lo tanto, esta tesis propone el desarrollo de un clasi-

ficador de riesgo de coledocolitiasis utilizando técnicas de aprendizaje profundo y métodos de ensamble, con el objetivo de crear una herramienta accesible y de alto rendimiento que pueda integrarse en la práctica clínica diaria. Con esto, se espera mejorar significativamente el diagnóstico y manejo de esta condición, proporcionando una alternativa viable y efectiva que pueda ser utilizada ampliamente en el ámbito médico.



# Capítulo 2

## Fundamentación Teórica

### 2.1. Inteligencia artificial aplicada a la medicina

Con el surgimiento de las áreas de aprendizaje automático y aprendizaje profundo se ha logrado desarrollar diversas aplicaciones enfocadas al área médica, aplicaciones que abarcan sistemas de clasificación, diagnóstico oportuno, segmentación de imágenes médicas, sistemas auxiliares en procedimientos quirúrgicos, entre otras. En específico, en el área de clasificadores, se ha logrado diseñar diversos sistemas capaces de identificar las diversas etapas de una enfermedad, con lo cual el médico puede generar diagnósticos y tratamientos acordes a la necesidad del paciente [16].

### 2.2. Aprendizaje automático

Uno de los campos de la inteligencia artificial es el aprendizaje automático, cuyo objetivo es el desarrollo de técnicas que permitan a los sistemas informáticos adquirir noción sobre los patrones existentes en la información proporcionada [17], [18]. Dichas técnicas se dividen a su vez en tres grupos: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo.

### 2.2.1. Aprendizaje supervisado

Este grupo de técnicas de aprendizaje automático basan su funcionamiento en un conjunto de datos previamente etiquetados. Dichos datos y sus etiquetas se proporcionan al algoritmo, el cual realiza una serie de operaciones matemáticas hasta conseguir un conjunto de reglas que representa los patrones detectados por el algoritmo [19].

## 2.3. Aprendizaje profundo

El aprendizaje profundo es un subcampo del aprendizaje automático, el cual busca obtener un entendimiento profundo de los datos proporcionados. Este tipo de aprendizaje basa su funcionamiento en redes neuronales artificiales, las cuales están inspiradas en nuestro entendimiento del funcionamiento del cerebro [18].

## 2.4. Redes neuronales artificiales

Las redes neuronales artificiales (RNA), usualmente llamadas redes neuronales, es un modelo computacional que busca asemejar el funcionamiento del cerebro, ya que consta de un conjunto de unidades simples de procesamiento (neuronas) que se encuentran conectadas entre sí y trabajan de forma simultánea [17]. Estos modelos suelen conformarse de tres grandes secciones (Figura 2.1):

- **Capa de entrada:** Sección en la cuál se reciben los datos que se desea procesar.
- **Capas ocultas:** Conjunto de capas encargadas de realizar el procesamiento de los datos, así como los cálculos necesarios para obtener el resultado solicitado.
- **Capa de salida:** Sección cuya tarea consiste en realizar conclusiones del procesamiento de los datos.

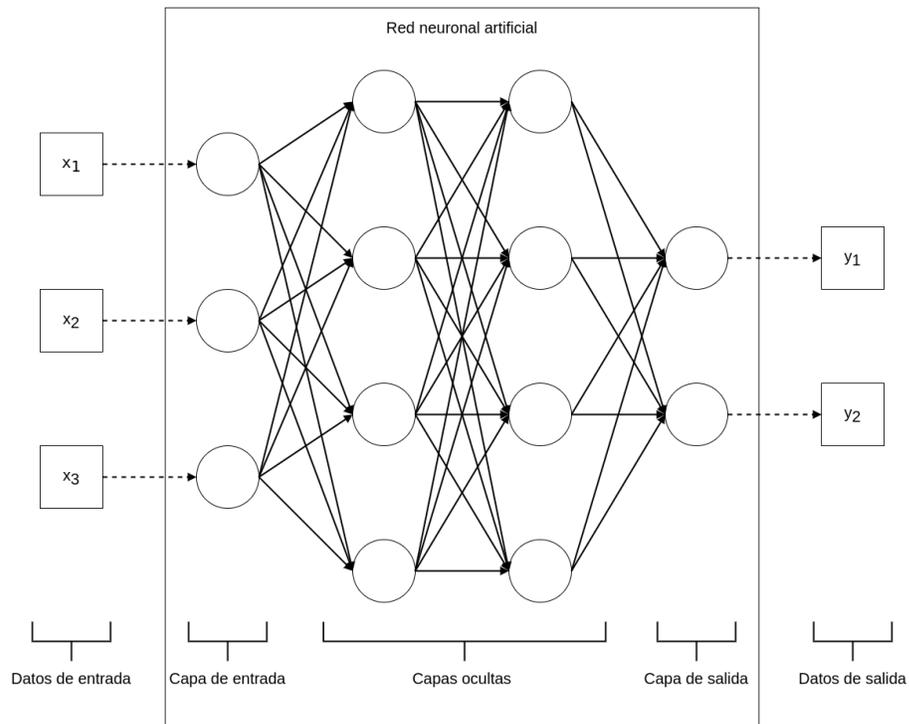


Figura 2.1: Diagrama de RNA multicapa

### 2.4.1. Redes neuronales convolucionales 1D

Las Redes Neuronales Convolucionales 1D (CNN 1D) son un tipo de arquitectura de red neuronal diseñada para procesar datos secuenciales o series temporales, donde los datos tienen una estructura lineal y un orden inherente. A diferencia de las CNN 2D, que operan en datos bidimensionales como imágenes, las CNN 1D aplican convoluciones a lo largo de una sola dimensión.

#### Funcionamiento General

El funcionamiento general de una CNN 1D implica varios pasos clave:

- Convolución:** Se aplica un filtro de tamaño fijo que se desliza a lo largo de la dimensión temporal o secuencial de los datos, produciendo mapas de características. Este proceso captura patrones locales dentro de la secuencia.

- **Activación:** Se utiliza una función de activación, como ReLU (Unidad Lineal Rectificada), para introducir no linealidad en el modelo y permitir que aprenda representaciones más complejas.
- **Agrupamiento (Pooling):** Se aplica una operación de agrupamiento, como *max-pooling*, para reducir la dimensionalidad de los mapas de características y aumentar la invariancia a la traslación.
- **Capas Completamente Conectadas:** Las salidas de las capas convolucionales y de agrupamiento se aplanan y se pasan a través de una o más capas completamente conectadas para realizar la tarea final, como clasificación o regresión.

### Principales Aplicaciones

Las CNN 1D se utilizan en diversas aplicaciones, especialmente en aquellas que involucran datos secuenciales o series temporales. Algunas de las principales aplicaciones incluyen:

- **Análisis de Señales:** Procesamiento y análisis de señales de audio, ECG (electrocardiograma), EMG (electromiograma) y otras señales biomédicas [20].
- **Procesamiento de Texto:** Análisis de sentimientos, clasificación de texto y otras tareas de procesamiento del lenguaje natural cuando los datos están en formato de secuencias de caracteres o palabras [21].
- **Predicción de Series Temporales:** Predicción de precios de acciones, demanda de energía y otros tipos de datos financieros y económicos [22].

### Uso en Datos Tipo Tabla

Si bien las CNN 1D han demostrado una utilidad significativa en aplicaciones de series temporales, su efectividad también se extiende a datos estructurados tipo tabla, incluyendo datos médicos para la extracción de características mediante diversos filtros de CNN 1D [23, 24, 25, 26]. La efectividad en el análisis de datos estructurados tipo tabla se basa en la

operación de convolución, que busca identificar patrones no solo en las variables objetivo, sino también en las características analizadas. Esto permite que los patrones presentes en las variables sean explotados como información útil en el proceso de clasificación.

Las CNN 1D se destacan por su capacidad de generalización de datos, su idoneidad para el aprendizaje supervisado y su robustez frente al ruido y la interferencia, lo que las hace altamente efectivas en escenarios con patrones complejos.

## 2.5. Métodos de ensamble

Los métodos de ensamble son técnicas utilizadas en el aprendizaje automático que combinan múltiples modelos base para mejorar la precisión y robustez de las predicciones. En lugar de confiar en un único modelo, los métodos de ensamble buscan aprovechar las fortalezas individuales de varios modelos y mitigar sus debilidades, lo que resulta en un clasificador compuesto más robusto y estable [15]. Los principales beneficios de los métodos de ensamble incluyen la reducción del sobreajuste, la mejora en la generalización y la capacidad de manejar mejor los datos ruidosos.

### 2.5.1. Ensamblados bagging

Bagging, o *Bootstrap Aggregating*, es una técnica de ensamble que implica entrenar múltiples modelos base independientes en diferentes subconjuntos de los datos de entrenamiento, generados mediante el método de *bootstrap*. Los resultados de estos modelos se combinan típicamente mediante votación mayoritaria para problemas de clasificación o promedio para problemas de regresión [27].

#### Funcionamiento general

1. Se generan múltiples subconjuntos de datos mediante muestreo con reemplazo (*bootstrap*).

2. Se entrena un modelo base en cada subconjunto de datos.
3. Se combinan las predicciones de los modelos mediante votación mayoritaria (clasificación) o promedio (regresión).

### Ventajas

- Reduce la varianza del modelo, mitigando el sobreajuste.
- Funciona bien con modelos base inestables (como árboles de decisión).

### Desventajas

- No necesariamente mejora el sesgo del modelo base.
- Puede ser computacionalmente costoso debido al entrenamiento de múltiples modelos [27].

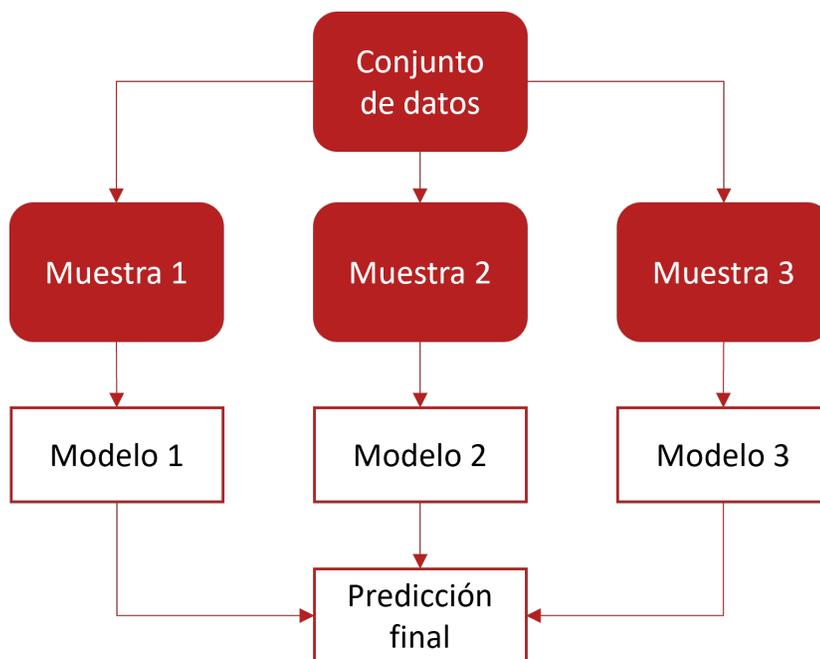


Figura 2.2: Diagrama de ensamblado de modelos por Bagging

### 2.5.2. Ensembles stacking

Stacking, o *Stacked Generalization*, es una técnica de ensamble que combina diferentes tipos de modelos base y utiliza un modelo de nivel superior, o meta-modelo, para hacer las predicciones finales basadas en las predicciones de los modelos base [28].

#### Funcionamiento general

1. Se entrenan varios modelos base diferentes en el conjunto de datos de entrenamiento.
2. Las predicciones de estos modelos base se utilizan como características de entrada para un meta-modelo.
3. El meta-modelo se entrena para combinar las predicciones de los modelos base y generar la predicción final.

#### Ventajas

- Aprovecha la diversidad de diferentes tipos de modelos base.
- Puede capturar relaciones complejas que los modelos base individuales no pueden.

#### Desventajas

- Puede ser difícil de interpretar debido a la complejidad del modelo.
- Requiere un cuidado especial para evitar el sobreajuste [28].

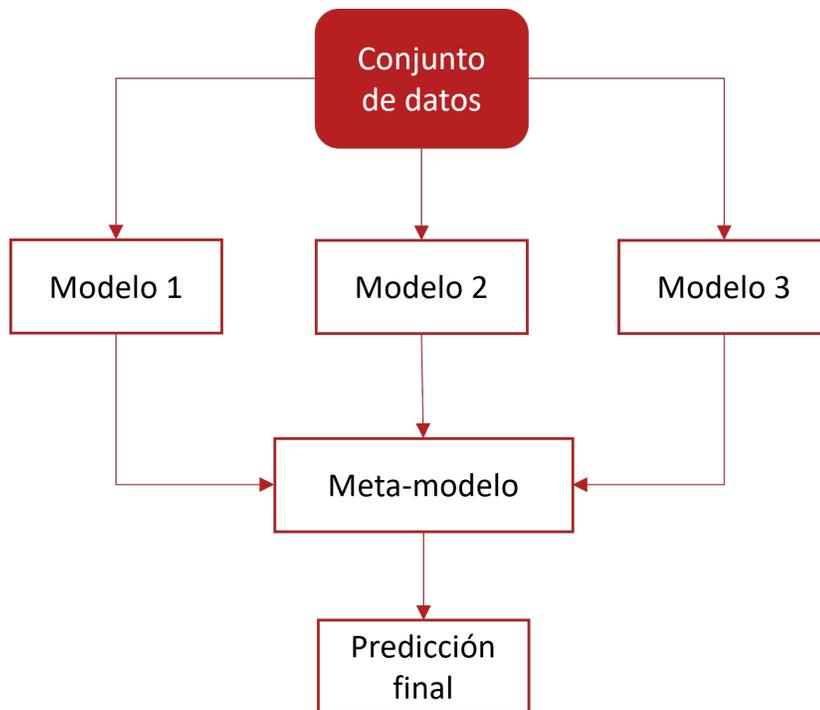


Figura 2.3: Diagrama de ensamble de modelos por Stacking

### 2.5.3. Ensamblés boosting

Boosting es una técnica de ensamble que construye modelos secuencialmente, donde cada nuevo modelo intenta corregir los errores cometidos por los modelos anteriores. A diferencia de bagging, boosting se enfoca en reducir el sesgo del modelo [29].

#### Funcionamiento general

1. Se entrena un modelo base y se evalúan sus errores.
2. Se ajustan los pesos de las instancias mal clasificadas para que los modelos subsiguientes se enfoquen en ellas.
3. Se agrega un nuevo modelo que se entrena con las instancias ajustadas.
4. Las predicciones finales se combinan ponderadamente.

### Ventajas

- Reduce tanto el sesgo como la varianza del modelo.
- Puede transformar modelos débiles en modelos fuertes.

### Desventajas

- Puede ser más susceptible al sobreajuste, especialmente con datos ruidosos.
- Más complejo y computacionalmente intensivo que bagging [29].

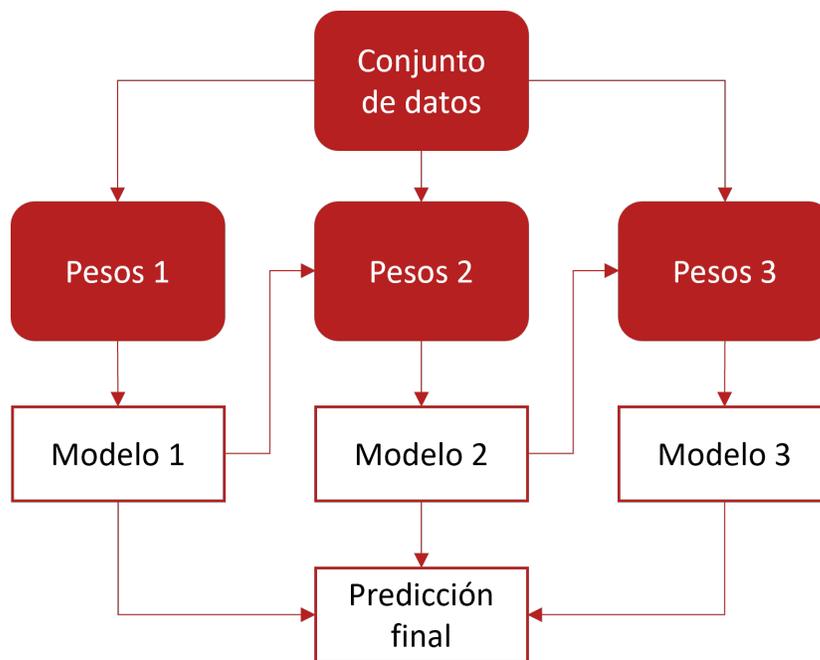


Figura 2.4: Diagrama de ensamble de modelos por Boosting

## 2.6. Cifras de mérito

Para evaluar el desempeño de todo modelo de aprendizaje automático es necesario obtener valores cuantitativos que representen dicho desempeño. Estos valores llevan el nombre de cifras de mérito, y la forma de obtenerlas es mediante el uso de los datos presentes en la

matriz de confusión del modelo de aprendizaje automático.

Las métricas más usadas en modelos desarrollados para objetivos similares a este son:

- **Exactitud** (Ecuación 2.1 [30]). Se define como la proporción de clasificaciones correctas realizadas por el modelo [31] En el ámbito médico, una alta exactitud implica que el modelo es generalmente confiable en sus predicciones, lo cual es crucial para el diagnóstico y tratamiento adecuados de los pacientes [32].

$$\frac{\textit{VerdaderosPositivos} + \textit{VerdaderosNegativos}}{\textit{VerdaderosPositivos} + \textit{FalsosPositivos} + \textit{FalsosNegativos} + \textit{VerdaderosPositivos}} \quad (2.1)$$

- **Sensibilidad** (Ecuación 2.2 [30]). Se define como la proporción de instancias positivas que el modelo clasifica correctamente [31]. En medicina, una alta sensibilidad es vital para asegurar que los pacientes con una enfermedad (verdaderos positivos) sean identificados correctamente, reduciendo el riesgo de pasar por alto casos que necesitan tratamiento [32].

$$\frac{\textit{VerdaderosPositivos}}{\textit{VerdaderosPositivos} + \textit{FalsosNegativos}} \quad (2.2)$$

- **Especificidad** (Ecuación 2.3 [30]). Se define como la proporción de instancias negativas que el modelo clasifica correctamente [31]. En el ámbito médico, una alta especificidad significa que el modelo es eficaz en identificar correctamente a los individuos sanos (verdaderos negativos), minimizando falsos positivos que podrían llevar a tratamientos innecesarios y ansiedad en los pacientes [32].

$$\frac{\textit{VerdaderosNegativos}}{\textit{VerdaderosNegativos} + \textit{FalsosPositivos}} \quad (2.3)$$

- **Valor predictivo positivo** (Ecuación 2.4 [33]). Se define como la probabilidad de que una instancia clasificada como positiva por el modelo sea realmente positiva [32]. En

medicina, un alto valor predictivo positivo indica que cuando el modelo predice una enfermedad, es muy probable que el paciente realmente la tenga, lo cual es esencial para la confianza en el diagnóstico y en la toma de decisiones clínicas [32].

$$\frac{\textit{VerdaderosPositivos}}{\textit{VerdaderosPositivos} + \textit{FalsosPositivos}} \quad (2.4)$$

- **Valor predictivo negativo** (Ecuación 2.5 [33]). Se define como la probabilidad de que una instancia clasificada como negativa por el modelo sea realmente negativa [32]. En el ámbito médico, un alto valor predictivo negativo asegura que cuando el modelo predice la ausencia de una enfermedad, el paciente realmente no la tiene, lo que es crucial para evitar preocupaciones innecesarias y pruebas adicionales [32].

$$\frac{\textit{VerdaderosNegativos}}{\textit{VerdaderosNegativos} + \textit{FalsosNegativos}} \quad (2.5)$$



# Capítulo 3

## Antecedentes

En los últimos años, los avances en aprendizaje automático han realizado grandes contribuciones a la medicina, ya que estos algoritmos han demostrado un excelente desempeño en la extracción de patrones de datos clínicos. Estas contribuciones abarcan desde la optimización de métodos de diagnóstico de enfermedades hasta el desarrollo de nuevos medicamentos.

En relación con las aportaciones al campo de estudio de este proyecto, se han documentado contribuciones en el área del aprendizaje automático desde 1998 [34]. No obstante, para obtener una visión actual del estado del arte en este campo, se ha llevado a cabo una revisión bibliográfica que abarca los últimos 10 años, donde se ha logrado recolectar información de 5 proyectos altamente relevantes en el campo de estudio de este proyecto.

- **2014.** P. Jovanovic *et al.* [35] presentaron dos modelos de aprendizaje automático cuyo objetivo es determinar si hay presencia de cálculos biliares en el conducto colédoco. El primer modelo consta de una regresión logística que toma como parámetros tres variables clínicas; las cifras de mérito obtenidas en dicho modelo fueron: sensibilidad 69.66 %, especificidad 78.95 %, valor predictivo positivo 93.14 % y valor predictivo negativo 39.79 %. Mientras que, el segundo modelo, un perceptrón multicapa con 9 parámetros de entrada obtuvo las siguientes cifras de mérito: sensibilidad 92.74 %,

especificidad 68.42 %, valor predictivo positivo 92.34 % y valor predictivo negativo 69.64 %.

- **2016.** A. M. Vukicevic *et al.* [36] presentan una evaluación de diversos modelos de aprendizaje automático y una propuesta para el diseño de una red neuronal artificial (ANN) con el objetivo de generar un sistema capaz de diagnosticar la coledocolitiasis mediante diversos parámetros clínicos. El método propuesto implementa un algoritmo genético para determinar los parámetros óptimos para lograr el mejor desempeño de la ANN. Dicho método propuesto reporta las siguientes cifras de mérito: exactitud 94.69 %, sensibilidad 88.20 %, especificidad 95.80 %, valor predictivo positivo 78.90 % y valor predictivo negativo 97.80 %.
- **2019.** V. S. Akshintala *et al.* [37] muestran el desempeño de un modelo de bosque aleatorio basado en los métodos *gradient boosting machine* y *XG boost*, logrando un modelo capaz de manejar 12 parámetros clínicos y obteniendo las siguientes cifras de mérito: exactitud 77.90 %, sensibilidad 82.40 % y especificidad 63.30 %.
- **2021.** R. Z. Cohen *et al.* [38] presentan un modelo de regresión logística para diagnóstico de coledocolitiasis en pacientes pediátricos. Dicho modelo toma 4 variables clínicas como parámetros de entrada, además de reforzarse mediante el uso de un algoritmo de vecino más cercano para imputación de datos. Dicha mezcla de algoritmos consigue las siguientes cifras de mérito: exactitud 71.50 %, sensibilidad 40.80 %, especificidad 90.30 %, valor predictivo positivo 71.40 %, valor predictivo negativo 72.10 %. Ese mismo año, C. Dalai *et al.* [12] presentaron una comparación de desempeño de diversos modelos de aprendizaje automático con los criterios propuestos por la *American Society for Gastrointestinal Endoscopy* (ASGE) para el diagnóstico de coledocolitiasis. El mejor modelo resultante de esta investigación fue un bosque aleatorio con 8 variables clínicas como parámetros de entrada, logrando conseguir las siguientes cifras de mérito: exactitud 76.90 %, sensibilidad 77.70 %, especificidad 75.00 %, valor predictivo positivo 94.40 % y valor predictivo negativo 37.50 %.

Complementando la información presentada anteriormente, se presenta el Cuadro 3.1, el cual muestra las cifras de mérito de todos los modelos de aprendizaje automático elaborados en los trabajos previamente mencionados.

Cuadro 3.1: Métodos utilizados para el diagnóstico de coledocolitiasis mediante datos clínicos multivariados.

<b>Método</b>	<b>Exactitud [ %]</b>	<b>Sensibilidad [ %]</b>	<b>Especificidad [ %]</b>	<b>PPV [ %]</b>	<b>NPV [ %]</b>
LR [35]	-	69.66	78.95	93.14	38.79
LR [36]	88.49	76.40	90.60	59.00	95.60
LR [38]	71.50	40.80	90.30	71.40	72.10
SVM [36]	91.15	76.40	93.70	68.40	95.70
SVM-L [12]	19.20	4.60	100.00	100.00	16.00
SVM-RBF [12]	67.30	68.20	62.50	90.90	26.30
DT [36]	92.92	70.50	96.80	80.00	94.48
RF [37]	77.90	82.40	63.30	-	-
RF [12]	76.90	77.70	75.00	94.40	37.50
KNN [36]	90.26	70.58	93.70	66.60	94.70
GLM [12]	71.20	70.50	75.00	93.90	31.60
NB [36]	91.15	76.40	93.70	68.40	95.70
ANN [35]	-	92.74	68.42	92.34	69.64
EAANN [36]	92.92	82.30	94.70	73.60	96.80

PPV: valor predictivo positivo. NPV: valor predictivo negativo. LR: regresión logística. SVM: máquina de soporte vectorial. SVM-L: máquina de soporte vectorial con kernel lineal. SVM-RBF: máquina de soporte vectorial con kernel de función de base radial. DT: árbol de decisión. RF: bosque aleatorio. KNN: vecino más cercano. GLM: modelo generalizado lineal. NB: clasificador bayesiano. ANN: red neuronal artificial. EAANN: red neuronal artificial de ensamble evolutivo.

Analizando el Cuadro 3.1, se pueden observar dos hechos bastante importantes. En primer lugar, aunque existen varias propuestas que logran igualar o superar el desempeño de las guías clínicas propuestas por la ASGE, solo dos de ellas se acercan al rendimiento del método

CPRE. Por otro lado, es importante destacar que los métodos que obtienen resultados cercanos al del método CPRE son algoritmos novedosos del campo del aprendizaje profundo, una rama de la inteligencia artificial que ha demostrado una alta capacidad para extraer patrones en los datos proporcionados.

Estos hallazgos subrayan la necesidad de continuar explorando y desarrollando modelos de aprendizaje profundo y técnicas de ensamble para crear herramientas de diagnóstico más robustas y accesibles.

# Capítulo 4

## Hipótesis

Utilizar métodos de ensamble para combinar las fortalezas de diversos algoritmos de aprendizaje profundo en la clasificación del riesgo de coledocolitiasis permitirá obtener un modelo de clasificación con un desempeño comparable al método actual de colangiopancreatografía retrógrada endoscópica (CPRE), logrando valores de sensibilidad y especificidad de por lo menos 85 %.



# Capítulo 5

## Objetivos

### 5.1. Objetivo general

Diseñar e implementar un clasificador de riesgo de coledocolitiasis utilizando técnicas de aprendizaje profundo y datos clínicos multivariados, con un desempeño comparable al método de colangiopancreatografía retrógrada endoscópica (CPRE), que sea accesible al público en general.

### 5.2. Objetivos específicos

- Desarrollar una serie de modelos de aprendizaje profundo para clasificar el riesgo de coledocolitiasis en dos categorías: bajo y alto.
- Crear un modelo de inteligencia colectiva mediante el ensamble de los mejores modelos desarrollados, con el fin de detectar patrones tanto lineales como no lineales en la información clínica utilizada.
- Implementar un servicio web utilizando tecnologías en la nube para que el clasificador resultante sea accesible a todo el público.



# Capítulo 6

## Material y Métodos

### 6.1. Características del sistema utilizado para experimentación

Para el desarrollo de este proyecto se utilizó un equipo de cómputo personal con hardware y software especializado para proyectos de inteligencia artificial. Sin embargo, es importante mencionar que no es necesario contar con un equipo con las mismas especificaciones de hardware para replicar este proyecto. En el Cuadro 6.1 se muestran las características de hardware utilizadas, mientras que el Cuadro 6.2 presenta las versiones de software instaladas en el sistema.

Cuadro 6.1: Características de hardware del sistema utilizado para el desarrollo del proyecto.

<b>Característica</b>	<b>Valor</b>
Unidad central de procesamiento (CPU)	Intel Core i7-12700F
Unidad de procesamiento gráfico (GPU)	NVIDIA RTX 3060 12 GB
Memoria de acceso aleatorio (RAM)	64 GB
Unidad de almacenamiento	1 TB NVMe

Cuadro 6.2: Características de software del sistema utilizado para el desarrollo del proyecto.

<b>Característica</b>	<b>Versión</b>
Sistema operativo	Ubuntu 20.04
Python	3.10
Numpy	1.26.4
Pandas	1.5.3
Tensorflow	2.10.1
Keras	2.10.0
Scikit-learn	1.4.1
SciKeras	0.12.0
Joblib	1.2.0
Flask	3.0.3

## 6.2. Conjunto de datos utilizado

El conjunto de datos utilizado proviene del *Olive View - UCLA Medical Center* y fue proporcionado de manera voluntaria y gratuita por los autores de [12]. Este conjunto incluye información sobre todas las CPRE realizadas entre noviembre de 2015 y diciembre de 2019.

Los datos fueron recolectados de pacientes hospitalizados mayores de edad, con diversas características demográficas.

El conjunto de datos consta de un total de 551 instancias, acompañadas de 26 atributos y 1 variable de decisión. La variable de decisión es un valor numérico que indica si el paciente presenta CBDS confirmada por CPRE. El desglose completo de los nombres de todos los atributos disponibles se encuentra en el Anexo C. A continuación, se presentan algunos ejemplos de estos atributos:

- **Datos demográficos:** edad, género y raza.
- **Variables clínicas:** índice de masa corporal, diagnóstico de diabetes mellitus, diagnóstico de cirrosis, valor de bilirrubina máxima, presencia de cálculos biliares en pruebas de imagen no invasivas, diámetro del conducto biliar común y existencia de la vesícula biliar.
- **Anotaciones/comentarios:** textos mixtos elaborados por especialistas en el campo.

### 6.3. Procedimiento general

El desarrollo de este trabajo de tesis se llevó a cabo en 4 etapas lineales, cada una de ellas con procesos iterativos y en bucle, que se detallan en las secciones correspondientes. La Figura 6.1 muestra el flujo de estas etapas, las cuales son:

1. **Preprocesamiento de los datos:** En esta etapa se realizó un análisis exploratorio de los datos, seguido de la selección de características y la limpieza de los datos.
2. **Diseño y selección de modelos base:** En esta etapa se realizaron experimentos para identificar los mejores modelos que se utilizarían en los métodos de ensamble.
3. **Diseño y selección de modelos de ensamble:** En esta fase se probaron diferentes configuraciones de ensamble de los modelos base, buscando aquella que proporcionara el mejor desempeño.

4. **Desarrollo de la aplicación web:** Durante esta etapa se implementaron el frontend y backend de la aplicación web, permitiendo que el modelo final estuviera disponible a través de internet.

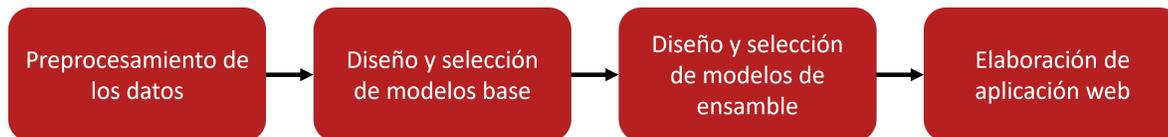


Figura 6.1: Diagrama general del desarrollo del proyecto.

## 6.4. Preprocesamiento de los datos

Para garantizar una alta calidad de los datos, se llevó a cabo un riguroso proceso de preprocesamiento que incluyó etapas de limpieza, codificación, imputación y estandarización. Este preprocesamiento se describe de manera general en la Figura 6.2, y cada uno de sus elementos se detalla en los párrafos siguientes.

Se realizó un análisis exploratorio inicial del conjunto de datos, determinándose que los atributos relacionados con anotaciones y comentarios eran irrelevantes para el estudio. Por lo tanto, se eliminaron los siguientes atributos: id, la fecha del procedimiento quirúrgico, los motivos de la intervención quirúrgica y los comentarios del especialista. También se identificó y eliminó información nula en la variable objetivo.

Durante la exploración, se detectaron notaciones no estándar, como los valores "x", "0x" y variantes similares, que fueron tratadas como datos nulos. Además, los datos categóricos en formato de texto se convirtieron en valores enteros.

Posteriormente, se encontró que seis atributos contenían más del 25 % de datos faltantes, por lo que se decidió eliminarlos. Para los atributos restantes, se aplicaron métodos tradicionales de imputación: los atributos categóricos se imputaron utilizando la moda, y los atributos continuos se imputaron utilizando la media.

Con el conjunto de datos completo, se aplicaron técnicas de normalización a los atributos continuos. Para este propósito, se eligió el método del z-score, que estandariza los datos continuos para que tengan una media final de 0 y una desviación estándar de 1.

El conjunto de datos resultante tras el preprocesamiento consta de 292 instancias, con 10 atributos y una variable de decisión. La distribución de esta variable de decisión indica 251 casos positivos (86 %) y 41 casos negativos (14 %), lo que evidencia un notable desbalance de clases. Este desbalance presenta un desafío significativo, cuyas posibles soluciones se abordarán en secciones posteriores.

Finalmente, se crearon dos copias del conjunto de datos preprocesado. La primera copia se utilizó para los procesos de validación cruzada, mientras que la segunda se reservó para los procesos de validación final y exportación del modelo.

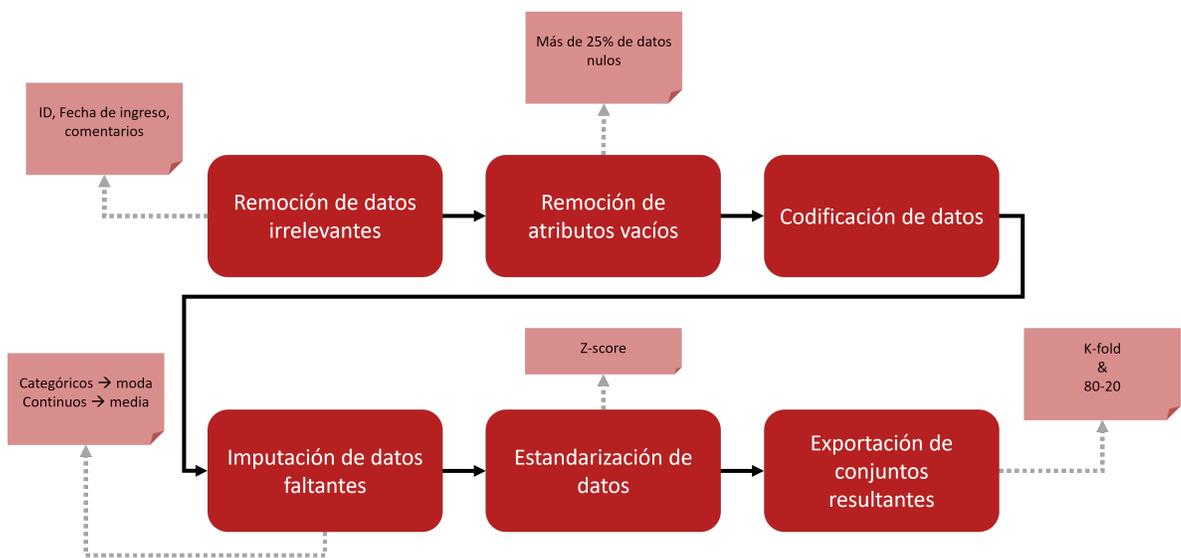


Figura 6.2: Diagrama del preprocesamiento aplicado al conjunto de datos original.

## 6.5. Diseño y selección de modelos base

La primera etapa para desarrollar los modelos de ensamble consistió en definir los modelos que serían utilizados como base para los ensambles. Para ello, se siguió de manera

iterativa el procedimiento mostrado en la Figura 6.3, que detalla las etapas involucradas en la selección de los modelos base.

Como punto de partida, se evaluaron 15 algoritmos de aprendizaje automático tradicionales (Regresión Logística, Bosque Aleatorio, Máquina de Soporte Vectorial, entre otros) y 5 configuraciones de CNN 1D.

Utilizando el conjunto de datos segmentado en 80 % para entrenamiento y 20 % para pruebas, se realizó un entrenamiento inicial de cada uno de los modelos propuestos con el 80 % de los datos, evaluando su rendimiento en el 20 % restante. Aquellos modelos que presentaron un desempeño superior al 85 % en sensibilidad y especificidad avanzaron a una etapa de optimización de hiperparámetros. Durante esta fase, se aplicaron algoritmos de *grid search* y *random search* para identificar los mejores hiperparámetros para cada modelo.

Con los modelos optimizados, se realizaron 100 experimentos de validación cruzada utilizando la técnica *k-fold* estratificada con 10 particiones. Este proceso permitió obtener las métricas de desempeño, con sus valores promedio y desviación estándar.

Al finalizar la optimización, se seleccionaron los tres modelos con mayor sensibilidad y especificidad. Estos modelos fueron exportados en formato `pkl` para su posterior uso, tanto de manera individual como en configuraciones de ensamble.

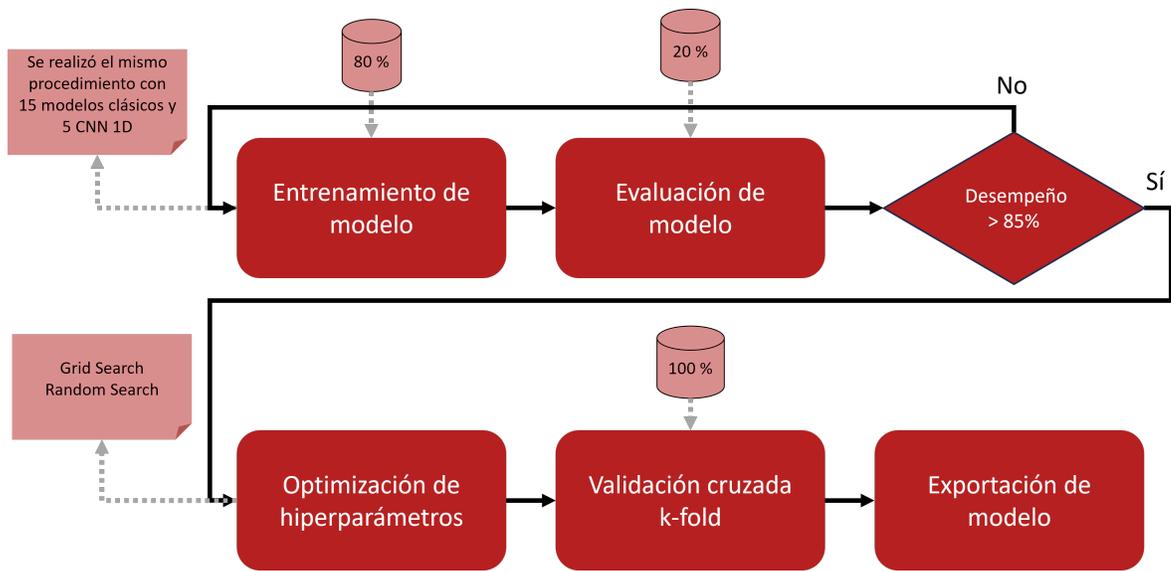


Figura 6.3: Diagrama general para la selección de modelos base.

## 6.6. Diseño y selección de modelos de ensemble

Con los mejores modelos base optimizados, se optó por probar cuatro configuraciones distintas de ensemble utilizando dichos modelos. La Figura 6.4 muestra de manera general los pasos seguidos para la selección del mejor ensemble.

Las configuraciones de ensemble evaluadas fueron:

- **Bagging:** Implementado mediante un algoritmo *derandom forest* con 100 estimadores.
- **Stacking:** Configurado utilizando los tres mejores modelos base obtenidos en etapas previas.
- **Boosting:** Utilizando la técnica XGBoost con 100 estimadores.
- **Stacking + Boosting:** Configuración de dos etapas. En la primera etapa se implementa el *stacking* descrito anteriormente, y en la segunda etapa se utiliza un modelo de *boosting*, que toma como entrada la salida del *stacking*.

Cada una de estas configuraciones fue sometida a una validación cruzada *k-fold* estratificada con 10 particiones, realizando 100 experimentos para cada validación. Tras evaluar el desempeño de cada configuración, se seleccionó el ensamble con mejores resultados, el cual fue exportado en formato `pkl` para su uso futuro en la aplicación web.

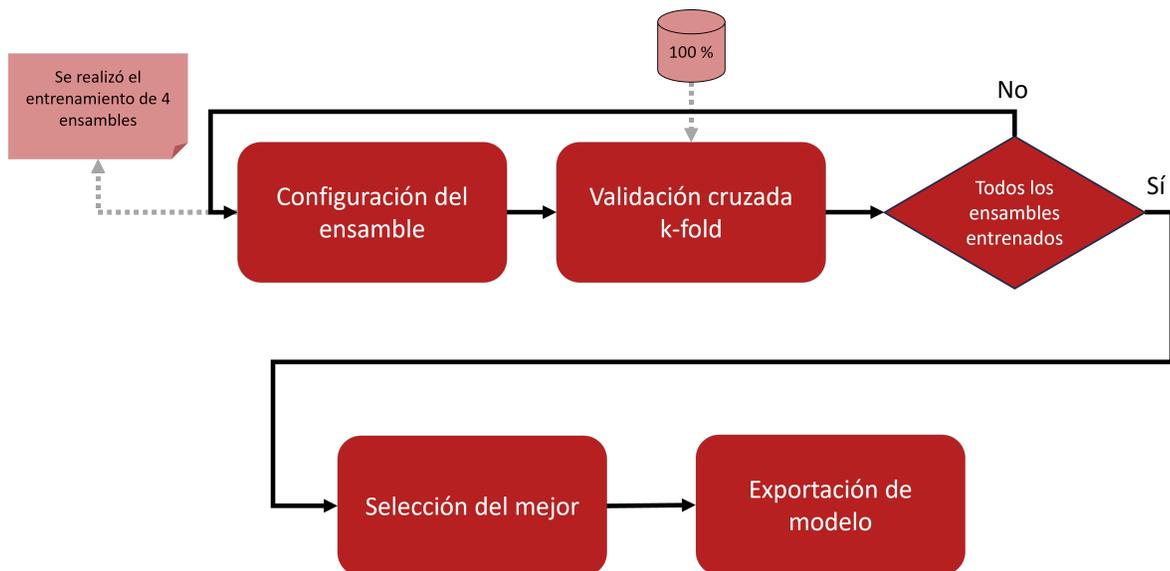


Figura 6.4: Diagrama general para la selección de métodos de ensamble.

## 6.7. Criterios de evaluación de modelos

Los criterios de evaluación de los modelos base y de ensamble se basaron en el desempeño medido a través de las métricas de Sensibilidad y Especificidad. Estas métricas fueron evaluadas mediante dos enfoques: una evaluación inicial con el estándar 80-20 y una evaluación más rigurosa a través de validación cruzada.

En la evaluación inicial utilizando el estándar 80-20, se empleó el 80 % del conjunto de datos para entrenar los modelos, mientras que el 20 % restante se utilizó para evaluar el rendimiento general del algoritmo. Esta etapa permitió tomar decisiones rápidas sobre si el algoritmo era un candidato ideal para el problema planteado.

Para la evaluación mediante validación cruzada, dividimos el conjunto de datos en 10

subconjuntos utilizando la técnica de *k-fold* estratificada. Esta metodología implica que 9 subconjuntos se utilizan para entrenar el modelo, mientras que el subconjunto restante se reserva como conjunto de prueba. Una vez completado el entrenamiento, evaluamos el modelo con el conjunto de prueba y almacenamos temporalmente los resultados. Luego, se intercambiaba el conjunto de prueba por uno de los subconjuntos de entrenamiento no utilizado previamente como prueba, se reiniciaban los pesos del modelo y se volvía a entrenar. Este proceso se repitió hasta que todos los subconjuntos fueron utilizados como conjunto de prueba.

Posteriormente, promediamos los resultados de cada experimento para obtener el rendimiento promedio del modelo. Además, este proceso de validación cruzada se repitió 100 veces, lo que permitió que el mecanismo de *k-fold* seleccionara diferentes elementos en cada ejecución, proporcionando así una estimación más precisa del rendimiento del modelo.

## 6.8. Diseño de aplicación web

Con el fin de proporcionar un acceso fácil y rápido al modelo de ensamble resultante, se desarrolló una aplicación web gratuita que permite a todo el personal de salud hacer uso de esta herramienta.

A nivel técnico, la aplicación se ejecuta en la infraestructura de Google Cloud Platform, específicamente en el servicio Google Cloud Run. La lógica de la aplicación está implementada con Flask, mientras que la interfaz de usuario utiliza JavaScript, HTML y CSS.

La interacción entre el usuario, la aplicación web y el modelo de ensamble se describe en la Figura 6.5, donde se destacan los siguientes pasos:

1. **Lectura de los datos del usuario:** A través de campos de entrada predefinidos, JavaScript se encarga de capturar los datos proporcionados por el usuario y enviarlos a la aplicación web para la inferencia.

2. **Acondicionamiento de los datos:** En la aplicación web, los datos se procesan a través de un pipeline definido en Scikit-learn, que se encarga de codificar y normalizar los datos utilizando los normalizadores y codificadores generados durante el desarrollo del proyecto.
3. **Carga del modelo:** El modelo de ensamble se carga mediante la lectura del archivo `pk1` donde fue almacenado, dejando todo listo para la inferencia.
4. **Inferencia:** Una vez que los datos han sido acondicionados y el modelo de ensamble está preparado, se procede a realizar la inferencia con los datos proporcionados por el usuario.
5. **Entrega de resultados al usuario:** Tras obtener el resultado numérico de la inferencia, éste se devuelve a JavaScript para que se renderice un mensaje adecuado en la interfaz de la aplicación web, mostrando la predicción correspondiente.

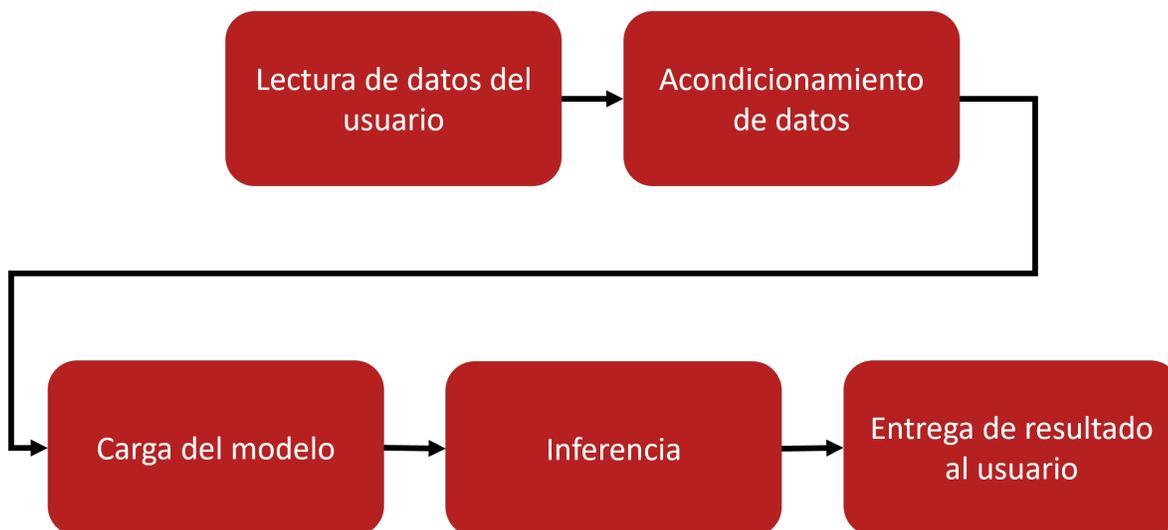


Figura 6.5: Diagrama general de la interacción del usuario y la aplicación web.

# Capítulo 7

## Resultados y Discusión

### 7.1. Preprocesamiento de los datos

Como resultado del exhaustivo preprocesamiento realizado, se obtuvo un conjunto de datos con las siguientes características destacables:

- Cantidad de instancias: 292
- Número de atributos: 10
- Balance de la variable de decisión:
  - Coledocolitiasis positiva: 251 instancias (86 %)
  - Coledocolitiasis negativa: 41 instancias (41 %)

En el Cuadro 7.1 se presenta una descripción detallada de todos los atributos disponibles en el conjunto de datos final, agrupados según los valores de la variable de decisión.

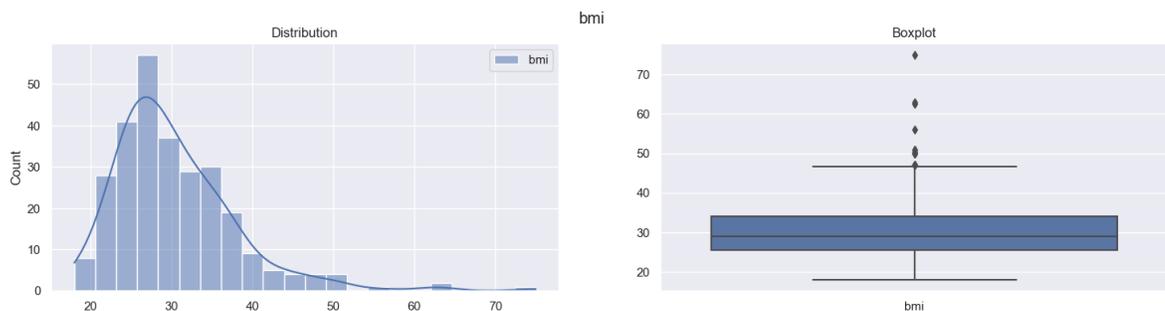
Cuadro 7.1: Atributos del conjunto de datos agrupados por la variable objetivo.

Atributo	Coledocolitiasis	Coledocolitiasis
	positiva (n = 251)	negativa (n = 41)
<b>Edad (años), mediana (IQR)</b>	46 (33-56)	47 (32-62)
<b>Género, n ( %)</b>		
Femenino	166 (66.1)	20 (48.8)
Masculino	85 (33.9)	24 (51.2)
<b>Origen étnico, n ( %)</b>		
Caucásico/Blanco	14 (5.6)	4 (9.7)
Hispano/Latino	188 (74.9)	28 (68.3)
Otro	49 (19.5)	9 (22.0)
<b>IMC, mediana (IQR)</b>	29.0 (25.6-33.6)	30 (26.0-36.7)
<b>Diabetes, n ( %)</b>		
Sí	35 (13.9)	11 (26.8)
No	216 (86.1)	30 (73.2)
<b>Cirrosis, n ( %)</b>		
Sí	8 (3.2)	2 (4.9)
No	243 (96.8)	39 (95.1)
<b>Bilirrubina máxima (mg/dL), mediana (IQR)</b>	3.6 (1.8-5.5)	3.5 (1.9-5.8)
<b>Obstrucción en el tracto biliar, n ( %)</b>		
Sí	135 (53.8)	19 (46.3)
No	116 (46.2)	22 (53.7)
<b>Diámetro de colédoco (mm), mediana (IQR)</b>	8.2 (6-10)	6.3 (4.2-9)
<b>Vesícula biliar, n ( %)</b>		
Sí	175 (69.7)	36 (87.8)
No	76 (30.3)	5 (12.2)

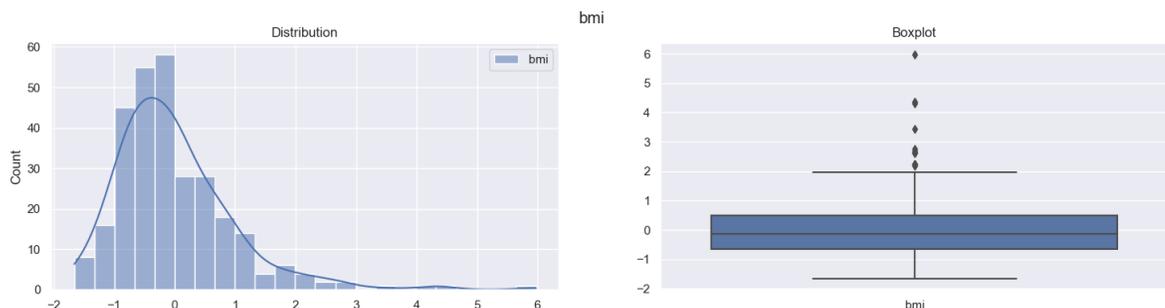
Abreviaturas: IQR, rango intercuartílico; IMC, índice de masa corporal.

Al analizar el Cuadro 7.1, se observan diferencias significativas entre los pacientes con coledocolitiasis positiva y negativa, especialmente en las variables clínicas. Los pacientes con coledocolitiasis positiva presentan un menor porcentaje de diabetes (13.9 %) y un diámetro del colédoco notablemente mayor (8.2 mm frente a 6.3 mm en el grupo negativo), lo que resalta la importancia de este último atributo como un indicador clínico relevante. Estos resultados refuerzan la validez del estudio, demostrando que existen diferencias claras que pueden ser capturadas eficazmente por los modelos de aprendizaje automático, aumentando su capacidad para realizar predicciones más precisas y confiables.

Adicionalmente, se realizó un análisis de la distribución de los atributos. La Figura 7.1 muestra la distribución del atributo *Índice de Masa Corporal* antes y después del preprocesamiento. Las distribuciones completas de todos los atributos del conjunto de datos se encuentran disponibles en el Anexo D.



(a) Distribución del atributo Índice de Masa Corporal previo al preprocesamiento.



(b) Distribución del atributo Índice de Masa Corporal posterior al preprocesamiento.

Figura 7.1: Análisis del efecto del preprocesamiento en la distribución del atributo Índice de Masa Corporal.

El preprocesamiento aplicado a los atributos del conjunto de datos, como se ilustra en la Figura 7.1, permitió normalizar las distribuciones y reducir el impacto de los valores atípicos. Inicialmente, varios atributos presentaban distribuciones asimétricas y valores extremos que podrían afectar negativamente el rendimiento de los modelos de aprendizaje automático. Tras la aplicación de la etapa de preprocesamiento, se logró que los atributos tuvieran medias cercanas a 0 y desviaciones estándar cercanas a 1, lo que garantiza que los modelos procesen los datos de manera más eficiente y consistente. Este proceso fue clave para preparar los datos y mejorar la capacidad de los modelos para identificar patrones de manera más robusta y precisa, asegurando un mejor desempeño general en las etapas posteriores.

## 7.2. Modelos base

En cuanto a los modelos base seleccionados, se determinó que, de los 15 enfoques clásicos evaluados, los algoritmos de regresión logística (LR) y análisis de discriminante lineal (LDA) eran los más adecuados para este problema. Además, se logró entrenar una CNN 1D capaz de reconocer patrones complejos en los datos de manera eficiente, convirtiéndose en una opción robusta para el análisis de coledocolitiasis.

Cada modelo fue evaluado principalmente en función de su sensibilidad, especificidad y puntaje F1, utilizando una validación cruzada *k-fold* con 10 particiones, repetida 100 veces para asegurar una estimación robusta. A continuación, se presentan los resultados de cada modelo, junto con sus ventajas y características principales.

El modelo de regresión logística, conocido por su simplicidad e interpretabilidad, alcanzó una sensibilidad de 94.30 %, una especificidad de 83.98 %, y un puntaje F1 de 95.73 %, como se muestra en la Figura 7.2. Aunque la especificidad de este modelo indica una posibilidad de falsos positivos, lo que podría llevar a pruebas adicionales en algunos pacientes sanos, su alta sensibilidad lo convierte en una herramienta confiable para detectar casos de coledocolitiasis. La ecuación del modelo resultante, que se presenta en las Ecuaciones 7.1 y 7.2, permite interpretar los coeficientes asociados a cada variable, ayudando a los profesionales de la salud

a comprender los factores más relevantes que contribuyen al diagnóstico.

$$y = \frac{1}{1 + e^{-t}} \quad (7.1)$$

$$t = -1,7832 + 0,0160x_1 - 0,2718x_2 + 0,2439x_3 + 0,1170x_4 - 0,6438x_5 - 0,1492x_6 + 0,6289x_7 - 0,0755x_8 + 0,0217x_9 + 0,1447x_{10} \quad (7.2)$$

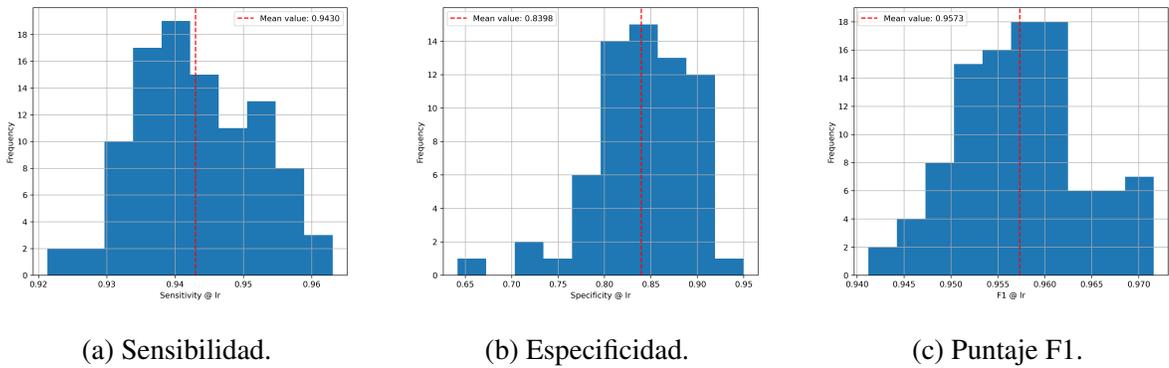


Figura 7.2: Distribución de las métricas de desempeño de los 100 experimentos de Regresión Logística.

El análisis de discriminante lineal obtuvo una sensibilidad de 95.46 %, una especificidad de 77.75 %, y un puntaje F1 de 95.49 %, como se observa en la Figura 7.3. Este modelo, aunque presenta una especificidad relativamente baja, sigue siendo una opción viable cuando se busca un enfoque rápido y fácil de desplegar. La menor especificidad sugiere una mayor probabilidad de falsos positivos, pero su alta sensibilidad y puntaje F1 lo hacen adecuado para entornos en los que es crucial priorizar la identificación de casos positivos. La ecuación que define el modelo se encuentra en la Ecuación 7.3, lo que facilita su interpretación para los profesionales de la salud.

$$f(x) = -3,5391 + 0,1828x_1 - 0,8358x_2 + 0,3553x_3 + 0,1245x_4 - 1,3415x_5 - 1,2359x_6 + 0,7679x_7 + 0,2481x_8 + 0,1693x_9 + 0,6648x_{10} \quad (7.3)$$

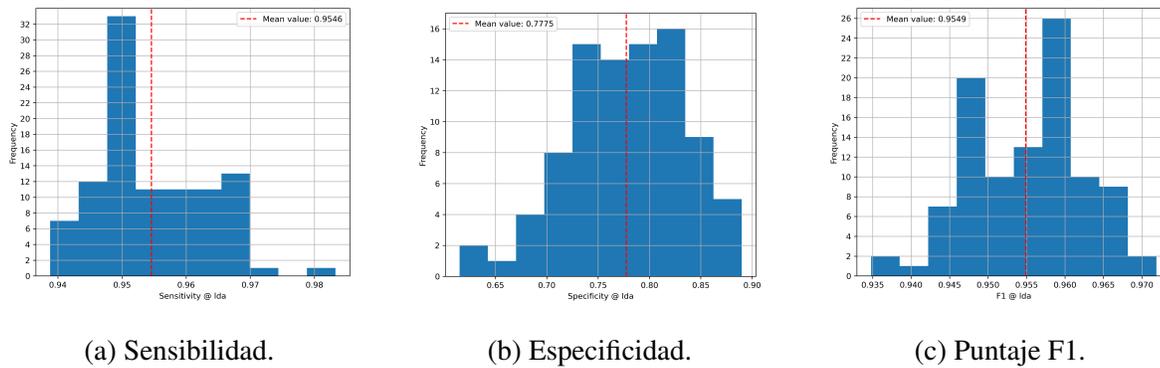


Figura 7.3: Distribución de las métricas de desempeño de los 100 experimentos de Análisis de Discriminante Lineal.

La CNN 1D representa una opción más avanzada y compleja, con una sensibilidad de 96.77 %, especificidad de 92.87 %, y un puntaje F1 de 94.77 %, como se muestra en la Figura 7.4. A diferencia de los modelos lineales, la CNN 1D tiene la capacidad de aprender relaciones no lineales en los datos, permitiéndole identificar patrones complejos que podrían pasar desapercibidos en enfoques más simples. Este excelente equilibrio entre la detección de casos positivos y la reducción de falsos positivos hace que la CNN 1D sea especialmente adecuada para aplicaciones médicas, donde es esencial minimizar tanto los errores en la detección como los procedimientos innecesarios para los pacientes.

La arquitectura de la CNN 1D utilizada en este estudio se detalla a continuación:

- Capa de entrada: (1, 10, 1)
- Capa oculta 1: (128, (1, 10))
- Capa oculta 2: (512, (1, 1))
- Capa de salida: (1, 1)

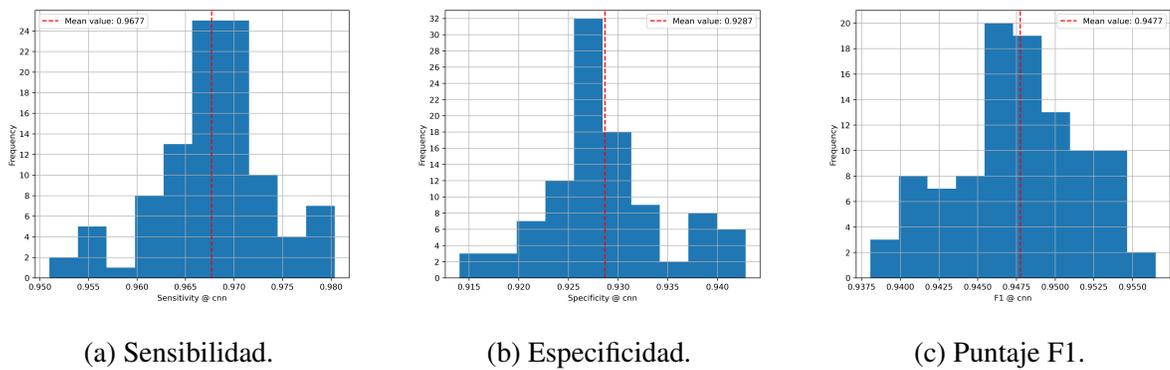


Figura 7.4: Distribución de las métricas de desempeño de los 100 experimentos de CNN 1D.

### 7.3. Modelo de ensemble

El modelo de ensemble final implementado para este proyecto combina las fortalezas de los tres modelos base previamente mencionados mediante una estrategia de dos etapas: *stacking* y *boosting*. Esta combinación permite aprovechar la capacidad de cada modelo base para capturar diferentes aspectos de los datos, al tiempo que maximiza la precisión y robustez de las predicciones. La Figura 7.5 muestra el diagrama general de este ensemble.

En la primera etapa, *stacking*, cada uno de los modelos base es entrenado individualmente con el conjunto de datos y realiza predicciones de manera independiente. Los resultados individuales de cada modelo se recopilan y forman un nuevo conjunto de características, que representa las predicciones de cada modelo base para cada instancia de los datos.

El objetivo de esta etapa de *stacking* es permitir que el modelo de ensemble aprenda patrones complementarios en los datos. Cada modelo base aporta una perspectiva única: LR y LDA, como modelos lineales, destacan en la interpretación y generalización de patrones lineales, mientras que la CNN 1D tiene la capacidad de capturar relaciones no lineales y patrones complejos en los datos. La combinación de estas salidas permite crear una representación más completa y detallada del problema.

La segunda etapa del ensemble implementa el enfoque de *boosting*, específicamente utilizando XGBoost. Los resultados del *stacking* (las predicciones de cada modelo base) se

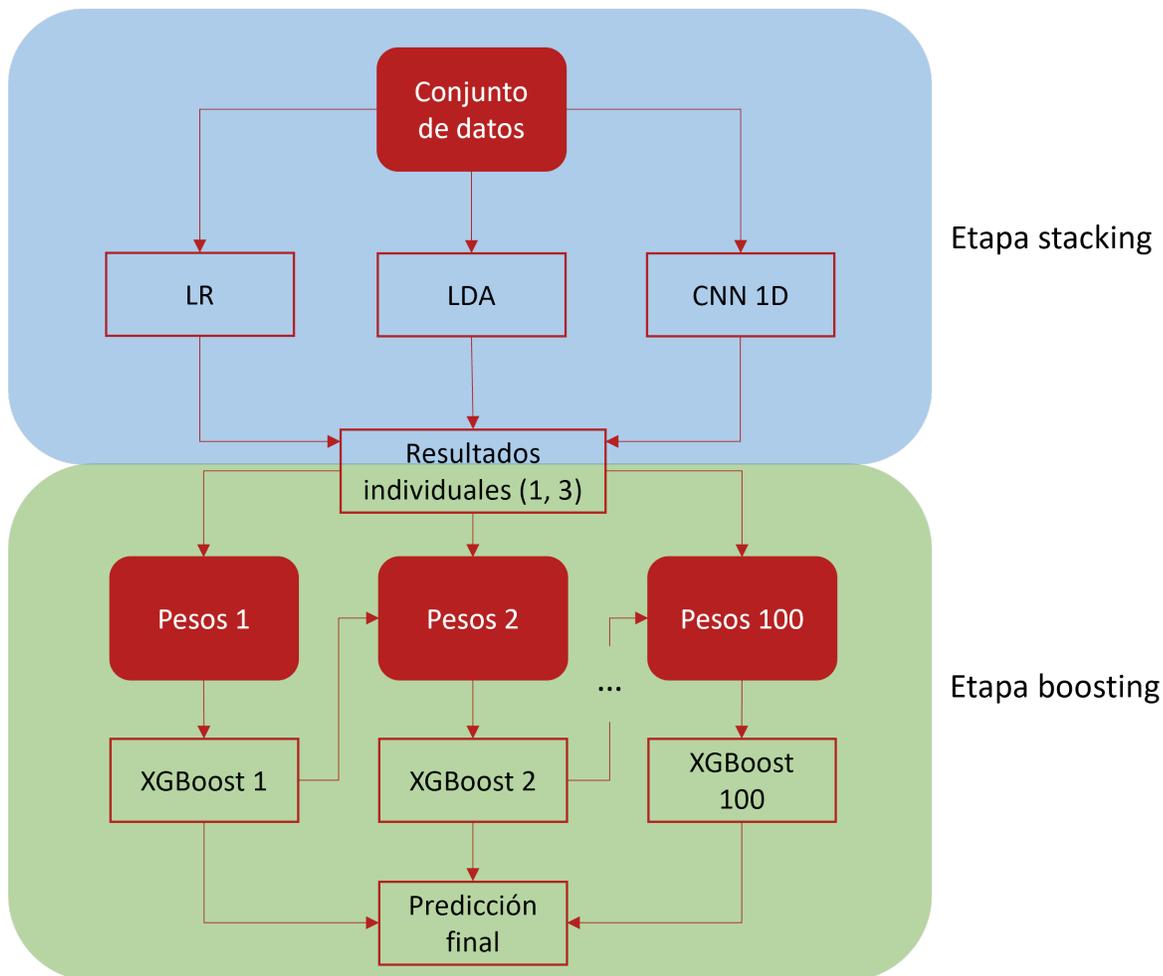


Figura 7.5: Diagrama general del ensamble resultante del proyecto.

ingresan en una serie de clasificadores XGBoost, cada uno de los cuales es ajustado con diferentes pesos. Este proceso de *boosting* se realiza en 100 iteraciones, donde cada clasificador XGBoost se ajusta para corregir los errores de los clasificadores previos, con el objetivo de mejorar el rendimiento general del ensamble.

Este uso de *boosting* permite al modelo de ensamble reducir los errores en las predicciones al dar mayor peso a las instancias en las que los modelos base pueden haber cometido errores, logrando así una mayor precisión y robustez. En el contexto clínico, esto es particularmente valioso, ya que minimiza tanto los falsos negativos (evitando omitir casos positivos) como los falsos positivos (reduciendo pruebas innecesarias).

En cuanto al desempeño de este modelo, los resultados de la validación cruzada muestran una sensibilidad de 96.81 %, una especificidad de 93.16 % y un puntaje F1 de 94.87 %. Estos valores aseguran que el modelo es capaz de detectar la mayoría de los casos positivos, al tiempo que reduce los falsos positivos, lo cual se traduce en una disminución de pruebas clínicas adicionales y procedimientos innecesarios en pacientes sanos. Además, el alto valor del puntaje F1 indica un equilibrio sólido entre sensibilidad y precisión, lo que respalda la eficacia general del modelo en un contexto clínico.

## 7.4. Comparación con estado del arte

Los Cuadros 7.2, 7.3 y 7.4 muestran un análisis comparativo del rendimiento de los modelos seleccionados y el ensamble final frente a métodos de diagnóstico establecidos, como el CPRE y las guías clínicas de la ASGE.

Cuadro 7.2: Comparativa del desempeño de los diversos algoritmos desarrollados contra un modelo entrenado con el mismo conjunto de datos (Parte 1).

Modelo	Sensibilidad ( %)		Especificidad ( %)		Puntaje F1 ( %)	
	Promedio	Desv. Est.	Promedio	Desv. Est.	Promedio	Desv. Est.
LR	94.30	0.0087	83.98	0.0534	<b>95.73</b>	0.0062
LDA	95.46	0.0084	77.75	0.0563	95.49	0.0074
CNN 1D	96.77	0.0005	92.87	0.0062	94.77	0.0040
Ensamble	<b>96.81</b>	0.0491	<b>93.16</b>	0.0276	94.87	0.0284
RF UCLA*	77.30	-	75.00	-	-	-

Abreviaturas: LR, Regresión Logística; LDA, Análisis de Discriminante Lineal; CNN 1D, Red Neuronal Convocución de 1 Dimensión; RF UCLA, Bosque Aleatorio reportado por Olive View - UCLA Medical Center. \*Los resultados mostrados se reportan en [12].

Cuadro 7.3: Comparativa del desempeño de los diversos algoritmos desarrollados contra un modelo entrenado con el mismo conjunto de datos (Parte 2).

Modelo	Exactitud ( %)		VPP ( %)		VPN ( %)	
	Promedio	Desv. Est.	Promedio	Desv. Est.	Promedio	Desv. Est.
LR	<b>92.53</b>	0.0110	94.29	0.0087	<b>83.97</b>	0.0534
LDA	92.22	0.0124	<b>95.45</b>	0.0084	77.74	0.0563
CNN 1D	90.77	0.0071	92.86	0.0062	72.35	0.0366
Ensamble	91.06	0.0468	93.16	0.0276	82.50	0.2203
RF UCLA*	76.90	-	94.40	-	37.50	-

Abreviaturas: LR, Regresión Logística; LDA, Análisis de Discriminante Lineal; CNN 1D, Red Neuronal Convolución de 1 Dimensión; RF UCLA, Bosque Aleatorio reportado por Olive View - UCLA Medical Center. \*Los resultados mostrados se reportan en [12].

En los Cuadros 7.2 y 7.3, se observa que el modelo de ensamble alcanza una sensibilidad de 96.81 % y una especificidad de 93.16 %, superando los valores de especificidad de los modelos base de Regresión Logística (83.98 %), Análisis de Discriminante Lineal (77.75 %), y CNN 1D (92.87 %). Esto indica que el modelo de ensamble no solo mantiene una alta capacidad para identificar casos positivos de coledocolitiasis (alta sensibilidad), sino que también es efectivo en la reducción de falsos positivos (alta especificidad). En cuanto al puntaje F1, el modelo de ensamble (94.87 %) demuestra un balance adecuado entre precisión y sensibilidad, posicionándose como el modelo más robusto entre los evaluados.

Por otro lado, el Cuadro 7.4 permite una comparación directa entre el modelo de ensamble y los métodos de referencia clínica. Si bien el CPRE presenta una especificidad del 100 %, su sensibilidad se reporta en un rango de 80-93 %, inferior a la sensibilidad del modelo de ensamble. Esto sugiere que el modelo de ensamble podría ser útil en escenarios donde la detección temprana y precisa de coledocolitiasis es crítica, aunque no reemplaza la precisión absoluta del CPRE en la reducción de falsos positivos. Comparado con las guías de la ASGE, el modelo de ensamble también muestra una mejora significativa, especialmente

Cuadro 7.4: Comparativa del desempeño del algoritmo de ensamble, CPRE y guías clínicas.

Modelo	Sensibilidad ( %)		Especificidad ( %)		Puntaje F1 ( %)	
	Promedio	Desv. Est.	Promedio	Desv. Est.	Promedio	Desv. Est.
Ensamble	<b>96.81</b>	0.0491	93.16	0.0276	<b>94.87</b>	0.0284
CPRE <sup>1</sup>	80 - 93	-	<b>100</b>	-	-	-
ASGE <sup>2</sup>	90.90	-	25.00	-	-	-

<sup>1</sup>Los resultados mostrados se reportan en [6, 7, 8]. <sup>2</sup>Los resultados mostrados se reportan en [12]

en especificidad (93.16 % frente al 25 % de la ASGE), lo que respalda su aplicabilidad como una herramienta de apoyo para el diagnóstico en escenarios donde se busque minimizar intervenciones innecesarias.

Estos resultados sugieren que el modelo de ensamble tiene el potencial de ser una herramienta diagnóstica complementaria, que puede reducir la carga de pruebas invasivas para los pacientes al ofrecer una precisión de diagnóstico superior a otros modelos de IA y alineada con los métodos clínicos de referencia. Este enfoque híbrido podría tener un impacto positivo en la práctica clínica, optimizando los recursos y mejorando la experiencia del paciente.

## 7.5. Aplicación web

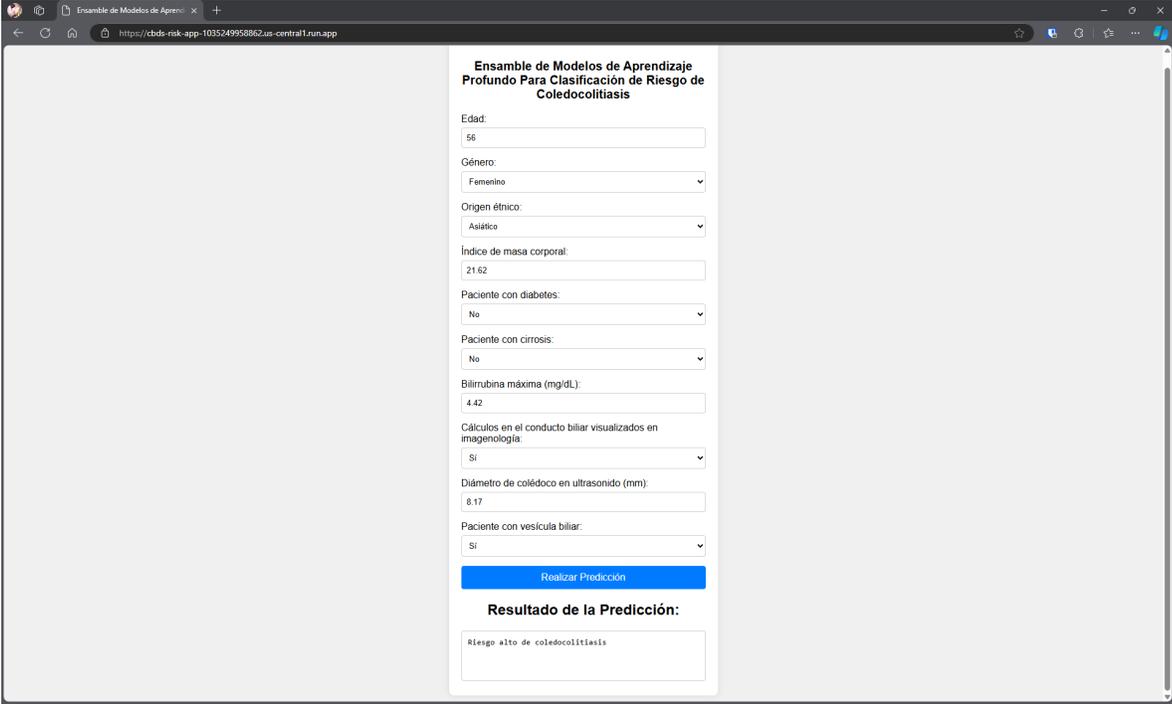
Alineado con el objetivo de desarrollar una aplicación web que permita el acceso fácil y rápido al modelo de IA, se implementó una plataforma accesible para el personal de salud en el dominio `https://cbds-risk-app-1035249958862.us-central1.run.app/`, facilitando así la evaluación del riesgo de coledocolitiasis en pacientes. La aplicación fue diseñada para ser intuitiva, proporcionando una interfaz simple en la que el usuario introduce información clínica relevante para obtener una predicción basada en el modelo de IA.

El flujo de datos en la aplicación comienza con la captura de datos clínicos proporciona-

dos por el usuario, los cuales se procesan automáticamente para asegurar su compatibilidad con el modelo de IA. Posteriormente, el modelo genera una predicción que es presentada en un formato interpretable, indicando si el riesgo de coledocolitiasis es alto o bajo. Este resultado permite al usuario obtener una evaluación rápida que apoya la toma de decisiones en un entorno clínico.

La Figura 7.6 muestra una captura de pantalla de la aplicación web, ilustrando su interfaz y funcionalidad. Adicionalmente, se incluye un código QR en la Figura 7.7, que facilita el acceso directo a la aplicación para los lectores interesados en explorar la herramienta.

La aplicación web representa un valor significativo para la práctica médica, ya que permite al personal de salud utilizar un modelo avanzado de IA sin necesidad de conocimientos técnicos en esta área. Al reducir el tiempo necesario para obtener una evaluación preliminar, la herramienta optimiza el flujo de trabajo clínico y potencialmente disminuye la necesidad de procedimientos invasivos en pacientes con bajo riesgo.



The screenshot shows a web browser window with the URL <https://cbds-risk-app-1035249958862.us-central1.run.app>. The page title is "Ensamble de Modelos de Aprendizaje Profundo Para Clasificación de Riesgo de Coledocolitiasis". The form contains the following fields and values:

Field	Value
Edad	56
Género	Femenino
Origen étnico	Asiático
Índice de masa corporal	21.62
Paciente con diabetes	No
Paciente con cirrosis	No
Bilirrubina máxima (mg/dL)	4.42
Cálculos en el conducto biliar visualizados en imagenología	Si
Diámetro de colédoco en ultrasonido (mm)	8.17
Paciente con vesícula biliar	Si

Below the form is a blue button labeled "Realizar Predicción". Underneath, the "Resultado de la Predicción:" section displays "Riesgo alto de coledocolitiasis".

Figura 7.6: Captura de pantalla de la aplicación web mostrando un resultado positivo.

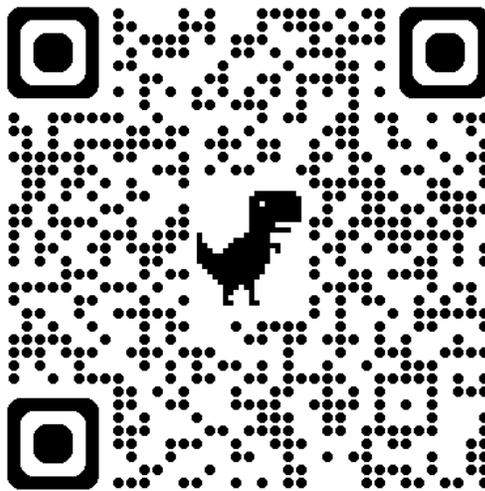


Figura 7.7: Código QR para acceder a la aplicación web.



# Capítulo 8

## Conclusiones

Este trabajo de tesis ha logrado desarrollar y validar un modelo de inteligencia artificial para la predicción de riesgo de coledocolitiasis, utilizando un enfoque de ensamble de modelos que combina técnicas de *stacking* y *boosting*. Los resultados obtenidos, con una sensibilidad de 96.81 %, especificidad de 93.16 % y puntaje F1 de 94.87 %, demuestran que el modelo de ensamble no solo es eficaz en la detección de casos positivos, sino también en la reducción de falsos positivos. Esto representa una mejora significativa en comparación con otros modelos de aprendizaje automático y guías clínicas existentes, lo que confirma la aplicabilidad y el impacto de la IA en el diagnóstico de enfermedades gastrointestinales.

Una de las principales contribuciones de este proyecto ha sido la implementación de un enfoque híbrido de ensamble, que aprovecha las fortalezas de modelos lineales (Regresión Logística y Análisis de Discriminante Lineal) y no lineales (CNN 1D), maximizando la precisión y robustez del diagnóstico. Este enfoque de combinación de modelos permitió desarrollar un sistema de predicción que supera las limitaciones de los modelos individuales, proporcionando un diagnóstico más equilibrado y confiable. Esta metodología no solo contribuye al campo de la inteligencia artificial aplicada a la medicina, sino que también ofrece un modelo replicable para otros estudios en enfermedades de características complejas y multifactoriales.

Además, se desarrolló una aplicación web accesible para el personal de salud, permitiendo la implementación práctica del modelo en un entorno clínico. Esta herramienta reduce la complejidad en el uso de modelos de IA, haciendo que la predicción del riesgo de coledocolitiasis esté al alcance de profesionales de la salud sin experiencia técnica en IA. La aplicación tiene el potencial de optimizar el flujo de trabajo clínico, al facilitar una evaluación rápida y preliminar del riesgo, lo cual podría ayudar en la toma de decisiones y en la reducción de procedimientos invasivos innecesarios en pacientes con bajo riesgo.

El impacto de este trabajo es relevante tanto en el ámbito de la inteligencia artificial como en la práctica clínica. En IA, se ha demostrado la efectividad de los modelos de ensamble en el análisis de datos clínicos, destacando la capacidad de estas técnicas para mejorar la precisión en diagnósticos complejos. En el ámbito médico, el modelo desarrollado ofrece un apoyo valioso para los profesionales de la salud en la detección temprana de coledocolitiasis, una condición que, si no se trata oportunamente, puede tener consecuencias graves. Este modelo tiene el potencial de ser una herramienta complementaria al CPRE y las guías clínicas de la ASGE, ofreciendo una precisión de diagnóstico que se acerca a la de métodos invasivos, pero de manera no invasiva.

No obstante, existen limitaciones y áreas de mejora que deben ser abordadas en futuros trabajos. La dependencia de la calidad y la cantidad de los datos clínicos es un factor crucial para el rendimiento del modelo. La inclusión de datos adicionales y la validación en entornos clínicos más amplios podrían fortalecer la robustez del modelo y su generalización. Además, aunque la aplicación web es funcional, futuros desarrollos podrían centrarse en optimizar su interfaz y funcionalidad para mejorar la experiencia del usuario y garantizar una integración más fluida en los sistemas de salud.

## Referencias bibliográficas

- [1] N. Méndez-Sánchez, J. Jessurun, G. Ponciano-Rodríguez, P. Alonso-De-Ruiz, M. Uribe, and M. Hernández-Avila, “Prevalence of gallstone disease in Mexico - a necropsy study,” *Digestive Diseases and Sciences*, vol. 38, no. 4, pp. 680–683, 1993. [En línea]. Disponible: <https://doi.org/10.1007/BF01316800>
- [2] L. G. González-Pérez, F. R. Zaldívar-Ramírez, B. R. Tapia-Contla, C. M. Díaz-Contreras-Piedras, P. R. Arellano-López, and L. M. Hurtado-López, “Factores de riesgo de la coledocolitiasis asintomática; experiencia en el hospital general de México,” *Cirujano General*, vol. 40, pp. 164–168, 2018. [En línea]. Disponible: <https://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=83961>
- [3] J. F. Flores-Mendoza, “Eficacia de los criterios predictores de coledocolitiasis de la asge con hallazgos en cpre,” *Endoscopia*, vol. 2, pp. 633–636, 6 2022. [En línea]. Disponible: [https://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S2444-64832020000600633](https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2444-64832020000600633)
- [4] S. Li, L. Guizzetti, C. Ma, A. A. Shaheen, E. Dixon, C. Ball, S. Wani, and N. Forbes, “Epidemiology and outcomes of choledocholithiasis and cholangitis in the United States: trends and urban-rural variations,” *BMC Gastroenterology*, vol. 23, no. 1, 7 2023. [En línea]. Disponible: <https://doi.org/10.1186/s12876-023-02868-3>
- [5] X. Gao, L. Zhang, S. Wang, Y. Xiao, D. Song, D. Zhou, and X. Wang, “Prevalence, risk factors, and complications of cholelithiasis in adults with short bowel syndrome:

- A longitudinal cohort study,” *Frontiers in Nutrition*, vol. 8, 11 2021. [En línea]. Disponible: <https://doi.org/10.3389/fnut.2021.762240>
- [6] L. González, F. Zaldívar, B. Tapia, C. Díaz, P. Arellano, and L. Hurtado, “Factores de riesgo de la coledocolitiasis asintomática; experiencia en el hospital general de máximo,” *Cirujano general*, vol. 40, no. 3, pp. 164–168, 2018. [En línea]. Disponible: <https://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=83961>
- [7] M. L. Freitas, R. L. Bell, and A. J. Duffy, “Choledocholithiasis: Evolving standards for diagnosis and management,” *World Journal of Gastroenterology*, vol. 12, no. 20, pp. 1007–9327, 2006. [En línea]. Disponible: <https://doi.org/10.3748/wjg.v12.i20.3162>
- [8] A. Copelan and B. S. Kapoor, “Choledocholithiasis: Diagnosis and management,” *Techniques in Vascular and Interventional Radiology*, vol. 18, no. 4, pp. 244–255, 2015. [En línea]. Disponible: <https://doi.org/10.1053/j.tvir.2015.07.008>
- [9] J. García-Cano Lizcano, J. A. González Martín, J. Morillas Ariño, and A. Pérez Sola, “Complications of endoscopic retrograde cholangiopancreatography: A study in a small ercp unit,” *Revista Española de Enfermedades Digestivas*, vol. 96, no. 3, pp. 163–173, 2004. [En línea]. Disponible: [https://scielo.isciii.es/scielo.php?script=sci\\_arttext&pid=S1130-01082004000300002](https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1130-01082004000300002)
- [10] M. Rhodes, L. Sussman, L. Cohen, and M. Lewis, “Randomised trial of laparoscopic exploration of common bile duct versus postoperative endoscopic retrograde cholangiography for common bile duct stones,” *The Lancet*, vol. 351, no. 9097, pp. 159–161, 1998. [En línea]. Disponible: [https://doi.org/10.1016/S0140-6736\(97\)09175-7](https://doi.org/10.1016/S0140-6736(97)09175-7)
- [11] T. Z. Ong, J. L. Khor, D. S. Selamat, K. G. Yeoh, and K. Y. Ho, “Complications of endoscopic retrograde cholangiography in the post-mrcp era: A tertiary center experience,” *World Journal of Gastroenterology*, vol. 11, no. 33, pp. 5209–5212, 2005. [En línea]. Disponible: <https://www.wjgnet.com/1007-9327/full/v11/i33/5209.htm>

- [12] C. Dalai, J. Azizian, H. Trieu, A. Rajan, F. Chen, T. Dong, S. Beaven, and J. H. Tabibian, “Machine learning models compared to existing criteria for noninvasive prediction of endoscopic retrograde cholangiopancreatography-confirmed choledocholithiasis,” *Liver Research*, vol. 5, no. 4, pp. 224–231, 2021. [En línea]. Disponible: <https://doi.org/10.1016/j.livres.2021.10.001>
- [13] H. E. B. Goñi, F. V. P. Salas, J. L. A. CusiHuallpa, R. A. Morocco, and N. S. S. Valle, “Rendimiento de los criterios predictivos de la asge en el diagnóstico de coledocolitiasis en el hospital edgardo rebagliati martins,” *Revista de Gastroenterología del Perú*, vol. 37, pp. 111–119, 2017. [En línea]. Disponible: [http://www.scielo.org.pe/scielo.php?script=sci\\_arttext&pid=S1022-51292017000200002](http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1022-51292017000200002)
- [14] C. Ovalle-Chao, D. Guajardo-Nieto, and R. Elizondo-Pereo, “Rendimiento de los criterios predictivos de la sociedad americana de endoscopia gastrointestinal en el diagnóstico de coledocolitiasis en un hospital público de segundo nivel del estado de nuevo león, méxico,” *Revista de Gastroenterología de México*, 5 2022.
- [15] T. G. Dietterich, “Ensemble methods in machine learning,” in *Multiple Classifier Systems. MCS 2000. Lecture Notes in Computer Science*, vol. 1857. Springer Berlin Heidelberg, 2000, pp. 1–15. [En línea]. Disponible: [https://doi.org/10.1007/3-540-45014-9\\_1](https://doi.org/10.1007/3-540-45014-9_1)
- [16] V. Kaul, S. Enslin, and S. A. Gross, “History of artificial intelligence in medicine,” *Gastrointestinal Endoscopy*, vol. 92, no. 4, pp. 807–812, 2020. [En línea]. Disponible: <https://doi.org/10.1016/j.gie.2020.06.040>
- [17] J. Patterson and A. Gibson, *Deep Learning*. O’Reilly Media, Inc., 2017, ISBN: 9781491914250.
- [18] F. Chollet, *Deep Learning with Python*. Manning Publications, 2017, ISBN: 9781617294433.

- [19] J. P. Mueller and L. Massaron, *Machine Learning For Dummies*. For Dummies, 2016, ISBN: 9781119245513.
- [20] S. Kiranyaz, T. Ince, and M. Gabbouj, “Real-time patient-specific ecg classification by 1-d convolutional neural networks,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 664–675, 2015. [En línea]. Disponible: <https://doi.org/10.1109/TBME.2015.2468589>
- [21] Y. Kim, “Convolutional neural networks for sentence classification,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1746–1751. [En línea]. Disponible: <https://doi.org/10.3115/v1/D14-1181>
- [22] A. Borovykh, S. Bohte, and C. W. Oosterlee, “Conditional time series forecasting with convolutional neural networks,” *arXiv preprint*, 2017. [En línea]. Disponible: <https://doi.org/10.48550/arXiv.1703.04691>
- [23] K. Surendro, M. I. C. Rachmatullah, and J. Santoso, “Improving 1d convolutional neural network (1d cnn) performance in processing tabular datasets using principal component analysis,” *PREPRINT*, oct 2022. [En línea]. Disponible: <http://dx.doi.org/10.21203/rs.3.rs-2149472/v1>
- [24] S. Kiranyaz, A. Gastli, L. Ben-Brahim, N. Al-Emadi, and M. Gabbouj, “Real-time fault detection and identification for mmc using 1-d convolutional neural networks,” *IEEE Transactions on Industrial Electronics*, vol. 66, no. 11, p. 8760–8771, nov 2019. [En línea]. Disponible: <http://dx.doi.org/10.1109/TIE.2018.2833045>
- [25] A. Salamatian and A. Khadem, “Automatic sleep stage classification using 1d convolutional neural network,” *Frontiers in Biomedical Technologies*, vol. 7, no. 3, pp. 142–150, nov 2020. [En línea]. Disponible: <http://dx.doi.org/10.18502/fbt.v7i3.4616>
- [26] F. Mattioli, C. Porcaro, and G. Baldassarre, “A 1d cnn for high accuracy classification and transfer learning in motor imagery eeg-based brain-computer interface,” *Journal*

- of Neural Engineering*, vol. 18, no. 6, p. 066053, dec 2021. [En línea]. Disponible: <http://dx.doi.org/10.1088/1741-2552/ac4430>
- [27] L. Breiman, “Bagging predictors,” *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996. [En línea]. Disponible: <https://doi.org/10.1007/BF00058655>
- [28] D. H. Wolpert, “Stacked generalization,” *Neural Networks*, vol. 5, pp. 241–259, 1992. [En línea]. Disponible: [https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1)
- [29] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997. [En línea]. Disponible: <https://doi.org/10.1006/jcss.1997.1504>
- [30] M. A. Aceves Fernández, *Inteligencia artificial para programadores con prisa*. Universo de Letras, 2021, ISBN: 9788418854613.
- [31] “Glossary of terms,” *Machine Learning*, vol. 30, pp. 271–274, 1998. [En línea]. Disponible: <https://doi.org/10.1023/A:1017181826899>
- [32] R. H. Fletcher, S. W. Fletcher, and G. S. Fletcher, *Clinical epidemiology: the essentials*. Lippincott Williams & Wilkins, 2012, ISBN: 9781451144475.
- [33] S. Safari, A. Baratloo, E. M., and N. A., “Evidence based emergency medicine part 2: Positive and negative predictive values of diagnostic tests,” *Emergency (Tehran, Iran)*, vol. 3, no. 3, pp. 87–88, 2015. [En línea]. Disponible: <https://pubmed.ncbi.nlm.nih.gov/26495390/>
- [34] R. Golub, R. J. Cantu, and M. Tan, “The prediction of common bile duct stones using a neural network,” *Journal of the American College of Surgeons*, vol. 187, no. 6, pp. 584–590, 1998. [En línea]. Disponible: [https://doi.org/10.1016/s1072-7515\(98\)00241-5](https://doi.org/10.1016/s1072-7515(98)00241-5)
- [35] P. Jovanovic, N. N. Salkic, and E. Zerem, “Artificial neural network predicts the need for therapeutic ercp in patients with suspected choledocholithiasis,”

*Gastrointestinal Endoscopy*, vol. 80, no. 2, pp. 260–268, 2014. [En línea]. Disponible: <https://doi.org/10.1016/j.gie.2014.01.023>

- [36] A. M. Vukicevic, M. Stojadinovic, M. Radovic, M. Djordjevic, B. A. Cirkovic, T. Pejovic, G. Jovicic, and N. Filipovic, “Automated development of artificial neural networks for clinical purposes: Application for predicting the outcome of choledocholithiasis surgery,” *Computers in Biology and Medicine*, vol. 75, pp. 80–89, 2016. [En línea]. Disponible: <https://doi.org/10.1016/j.combiomed.2016.05.016>
- [37] V. S. Akshintala, B. Tang, A. Kamal, J. L. Buxbaum, B. J. Elmunzer, S. B. Wani, C. Y. Yu, A. H. Tieu, A. N. Kalloo, V. Singh, B. Caffo, and M. A. Khashab, “Sa1470 risk estimation, machine learning based ercp decision-making tool for suspected choledocholithiasis,” *Gastrointestinal Endoscopy*, vol. 89, no. 6, Supplement, pp. AB246–AB247, 2019. [En línea]. Disponible: <https://doi.org/10.1016/j.gie.2019.03.264>
- [38] R. Z. Cohen, H. Tian, C. G. Sauer, F. F. Willingham, M. T. Santore, Y. Mei, and A. J. Freeman, “Creation of a pediatric choledocholithiasis prediction model,” *Journal of Pediatric Gastroenterology and Nutrition*, vol. 73, no. 5, pp. 636–641, 2021. [En línea]. Disponible: <https://doi.org/10.1097/MPG.0000000000003219>

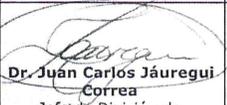
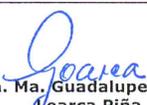
## **Apéndice A**

### **Requisitos para obtención de grado**



UNIVERSIDAD AUTÓNOMA DE QUERÉTARO  
FACULTAD DE INGENIERÍA  
DIVISIÓN DE INVESTIGACIÓN Y POSGRADO

REGISTRO DEL PROTOCOLO DE INVESTIGACIÓN  
DEL ESTUDIANTE DE POSGRADO

Los 2 Espacios oscuros exclusivos para la Dirección	No. Registro de Proyecto*: 74027	Fecha de Registro*: 23/06/2023
	Fecha de inicio de proyecto: 23/01/2023	Fecha de término de proyecto: 31/07/2024
<b>1. DATOS DEL SOLICITANTE</b>		
No. de expediente:	309238	
Apellido Paterno	Apellido Materno	Nombre(s)
MENA	CAMILO	ENRIQUE
Dirección:		
Calle y número	Colonia	C.P.
Herreros #4	Peñuelas	76148
Estado	Teléfono (incluir lada)	Correo Electrónico
22	7331187123	enriquec92@gmail.com
<b>2. DATOS DEL PROYECTO</b>		
Facultad:	INGENIERÍA	
Programa:	MAESTRIA EN CIENCIAS EN INTELIGENCIA ARTIFICIAL	
Tema específico del proyecto:	Ensamble de modelos de aprendizaje profundo para clasificación de riesgo de coledocolitiasis	
 SALAZAR COLORES SEBASTIÁN Director de tesis	 TOVAR ARRIAGA SAUL Coordinador de programa	 MENA CAMILO ENRIQUE Alumno
 Dr. Juan Carlos Jáuregui Correa Jefe de División de Investigación y Posgrado de la Fac. de Ing.	 Dr. Manuel Toledano Ayala Director de Fac. Ing.	 Dra. Ma. Guadalupe Flavia Joarca Piña Directora de Investigación y Posgrado UAQ



Anexo 3 Manual de procedimientos Administrativos de Posgrado de la Facultad de Ingeniería

Figura A.1: Registro del protocolo de investigación ante la Facultad de Ingeniería de la Universidad Autónoma de Querétaro.



C. U., 10 de junio de 2023

**Enrique Mena Camilo**  
**Estudiante de Maestría en Ciencias en Inteligencia Artificial**  
**Expediente 309238**

**Presente**

El Comité de Ética Aplicada a la Investigación de la Facultad de Ingeniería ha revisado el protocolo del trabajo de tesis:

CEAIFI-062-2023-TP

**Ensamble de modelos de aprendizaje profundo  
para clasificación de riesgo de coledocolitiasis**

Con apego a los lineamientos éticos de beneficencia, no maleficencia, justicia y autonomía, este comité ha dado el siguiente dictamen:

**Aprobado éticamente**

El presente dictamen tiene vigencia de un año a partir de su fecha de emisión.

Sirva esta carta para los fines académicos que al interesado convengan.

Atentamente  
"El ingenio para crear, no para destruir"

*Aurora Femat*  
Dra. Aurora Femat Díaz  
Presidente del CEAIFI  
afemat@uaq.mx

Figura A.2: Aprobación de protocolo de investigación por el comité de ética de Facultad de Ingeniería de la Universidad Autónoma de Querétaro.



UNIVERSIDAD AUTÓNOMA DE QUERÉTARO  
FACULTAD DE LENGUAS Y LETRAS



**A QUIEN CORRESPONDA:**

La que suscribe, Directora de la Facultad de Lenguas y Letras, hace **C O N S T A R** que

**MENA CAMILO ENRIQUE**

Presentó y acreditó el **Examen de Comprensión de Textos en Inglés** efectuado el día diecinueve de junio de dos mil veintitrés.

Se extiende la presente a petición de la parte interesada, para los fines escolares y legales que le convengan, en el Campus Aeropuerto de la Universidad Autónoma de Querétaro, el día nueve de agosto de dos mil veintitrés.



Atentamente,  
"Enlazar Culturas por la Palabra"



**DRA. ADELINA VELÁZQUEZ HERRERA**



**AVH/thb\*CL\*FLL-C.-1683**

---

**SOMOS UAQ®**  
SERVIR CONSTRUIR TRANSFORMAR

Campus Aeropuerto, Anillo Vial Fray Junípero Serra S/N, Querétaro, Qro. C.P. 76140  
Tel. 442 192 12 00 Dirección Ext. 61010, Secretaría Administrativa Ext.61300, Posgrado Ext. 61140,  
Licenciatura Ext.61070, Centro de Lenguas Ext.61050, Secretaría Académica Ext.61100 y Planeación Ext.61110

Figura A.3: Constancia de acreditación de Examen de Comprensión de Textos en Inglés emitida en Agosto 2023.




**UNIVERSIDAD AUTÓNOMA DE QUERÉTARO**  
**FACULTAD DE LENGUAS Y LETRAS**



**A QUIEN CORRESPONDA:**

La que suscribe, Directora de la Facultad de Lenguas y Letras, hace **C O N S T A R** que

**MENA CAMILO ENRIQUE**

Presentó el **Examen de Manejo de la Lengua** efectuado el día ocho de febrero de dos mil veintitres, en el cual obtuvo la siguiente calificación:

**8**

Se extiende la presente a petición de la parte interesada, para los fines escolares y legales que le convengan, en el Campus Aeropuerto de la Universidad Autónoma de Querétaro, el día veintiuno de febrero de dos mil veintitres.



Atentamente,  
"Enlazar Culturas por la Palabra"



**DRA. ADELINA VELÁZQUEZ HERRERA**



**AVH/daa\*CL\*FLL-C.-434**

---

**SOMOS UAQ**  
SERVIR CONSTRUIR TRANSFORMAR

Campus Aeropuerto, Anillo Vial Fray Junípero Serra S/N, Querétaro, Qro. C.P. 76140  
Tel. 442 192 12 00 Dirección Ext. 61010, Secretaría Administrativa Ext.61300, Posgrado Ext. 61140,  
Licenciatura Ext.61070, Centro de Lenguas Ext.61050, Secretaría Académica Ext.61100 y Planeación Ext.61110

Figura A.4: Constancia de acreditación de Examen de Manejo de la Lengua (Inglés) emitida en Febrero 2023.



## **Apéndice B**

### **Productos académicos**



## Classification of COVID-19 Mortality Risk: A Neural Network-Based Approach Using Mexican Healthcare Sector Data

Enrique Mena-Camilo<sup>1</sup> , Gerardo Hernández-Nava<sup>1</sup> ,  
Sheila Leyva-López<sup>1</sup> , and Sebastián Salazar-Colores<sup>2</sup> 

<sup>1</sup> Facultad de Ingeniería, Universidad Autónoma de Querétaro,  
Santiago de Querétaro, 76010 Querétaro, Mexico  
enriquece97@gmail.com

<sup>2</sup> Centro de Investigaciones en Óptica, 37150 León, Guanajuato, Mexico

**Abstract.** This study focuses on the use of neural networks and clinical data collected by the Mexican Ministry of Health to classify the risk of death from COVID-19. A multi-layer perceptron neural network model was designed, achieving remarkable results with an accuracy of 96.28%, sensitivity of 99.23%, and an F1 score of 0.9773. The model was optimized through meticulous exploration of various network configurations and performance enhancement techniques. The results showcase the efficacy of neural networks in predicting the risk of death, allowing healthcare professionals to prioritize treatment and allocate resources more efficiently. The value of artificial intelligence in the fight against the pandemic is emphasized, along with its potential application in diverse geographical and healthcare contexts. This work contributes to the advancement of predictive models and encourages further research in the fields of epidemiology and artificial intelligence to combat COVID-19.

**Keywords:** COVID-19 · Artificial Neural Network · Multi-Layer Perceptron · Mortality Risk

### 1 Introduction

The COVID-19 pandemic, has posed an unprecedented global health challenge. The World Health Organization (WHO) reports that as of June 2023, the virus has caused the death of over 6.9 million people worldwide, with the actual number possibly exceeding 20 million [13, 15]. Despite the declaration of the national emergency as concluded, the number of infections and fatalities continues to rise. In just June 2023, WHO reports 1 million new cases and 5,700 deaths [14], while the Mexican Government during the same period reports 14,820 new cases and 143 deaths [4].

The pandemic has highlighted the need for more sophisticated data processing techniques to handle the vast amount of generated information. Artificial

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024  
J. de J. A. Flores Cuautle et al. (Eds.): CNIB 2023, IFMBE Proceedings 96, pp. 112–119, 2024.  
[https://doi.org/10.1007/978-3-031-46933-6\\_12](https://doi.org/10.1007/978-3-031-46933-6_12)

Figura B.1: Porta de artículo publicado en Congreso Nacional de Ingeniería Biomédica en Octubre 2023. DOI: 10.1007/978-3-031-46933-6\_12.

Article

## Non-Invasive Prediction of Choledocholithiasis Using 1D Convolutional Neural Networks and Clinical Data

Enrique Mena-Camilo <sup>1</sup>, Sebastián Salazar-Colores <sup>2,\*</sup>, Marco Antonio Aceves-Fernández <sup>1</sup>,  
Edgard Efrén Lozada-Hernández <sup>3</sup> and Juan Manuel Ramos-Arreguín <sup>1</sup>

<sup>1</sup> Facultad de Ingeniería, Universidad Autónoma de Querétaro, Querétaro 76010, Mexico; emena05@alumnos.uaq.mx (E.M.-C.); marco.aceves@uaq.mx (M.A.A.-F.); jramos@mecamex.net (J.M.R.-A.)  
<sup>2</sup> Centro de Investigaciones en Óptica, León 37150, Mexico  
<sup>3</sup> Hospital Regional de Alta Especialidad del Bajío, León 37660, Mexico; edgardlozada2@gmail.com  
 \* Correspondence: sebastian.salazar@cio.mx

**Abstract:** This paper introduces a novel one-dimensional convolutional neural network that utilizes clinical data to accurately detect choledocholithiasis, where gallstones obstruct the common bile duct. Swift and precise detection of this condition is critical to preventing severe complications, such as biliary colic, jaundice, and pancreatitis. This cutting-edge model was rigorously compared with other machine learning methods commonly used in similar problems, such as logistic regression, linear discriminant analysis, and a state-of-the-art random forest, using a dataset derived from endoscopic retrograde cholangiopancreatography scans performed at Olive View—University of California, Los Angeles Medical Center. The one-dimensional convolutional neural network model demonstrated exceptional performance, achieving 90.77% accuracy and 92.86% specificity, with an area under the curve of 0.9270. While the paper acknowledges potential areas for improvement, it emphasizes the effectiveness of the one-dimensional convolutional neural network architecture. The results suggest that this one-dimensional convolutional neural network approach could serve as a plausible alternative to endoscopic retrograde cholangiopancreatography, considering its disadvantages, such as the need for specialized equipment and skilled personnel and the risk of postoperative complications. The potential of the one-dimensional convolutional neural network model to significantly advance the clinical diagnosis of this gallstone-related condition is notable, offering a less invasive, potentially safer, and more accessible alternative.

**Keywords:** choledocholithiasis; convolutional neural network; endoscopic retrograde cholangiopancreatography; risk prediction



**Citation:** Mena-Camilo, E.; Salazar-Colores, S.; Aceves-Fernández, M.A.; Lozada-Hernández, E.E.; Ramos-Arreguín, J.M. Non-Invasive Prediction of Choledocholithiasis Using 1D Convolutional Neural Networks and Clinical Data. *Diagnostics* **2024**, *14*, 1278. <https://doi.org/10.3390/diagnostics14121278>

Academic Editors: Mukul Vij and Julien Calderaro

Received: 9 May 2024

Revised: 30 May 2024

Accepted: 4 June 2024

Published: 17 June 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

### 1. Introduction

Choledocholithiasis is characterized by the obstruction of the common bile duct, which can be either partial or complete, owing to the presence of gallstones. Often emerging as a complication of cholelithiasis—gallstone formation in the gallbladder—this condition represents a significant health challenge. It is estimated that between 10% and 20% of gallstone cases will present common bile duct stones (CBDs) [1,2]. Various studies conducted worldwide have determined that the prevalence of this condition ranges from 9% to 21%, varying by region of study [3–5].

At present, endoscopic retrograde cholangiopancreatography (ERCP) stands as the diagnostic method with the highest accuracy in identifying choledocholithiasis. Various studies support that ERCP has a sensitivity that fluctuates between 80% and 93%, and it boasts a specificity reaching 100% [2,6,7]. While the efficacy of ERCP is high, the procedure is not without its challenges. These include the requirement for both specialized equipment and highly trained personnel. Moreover, there are risks of postoperative complications associated with ERCP. Such complications may range from pancreatitis and perforations of the duodenum to internal bleeding. In the most severe cases, these complications can have

Figura B.2: Portada de artículo publicado en revista MDPI-Diagnostics en Junio 2024. DOI: 10.3390/diagnostics14121278.



# Apéndice C

## Descripción de conjunto de datos original

- ercp\_indication: Breve descripción del motivo por el cual se solicita la CPRE
- indication\_code: Código de indicación estandarizado
  - 1 - Cálculos o lodos del conducto biliar (coledoco)
  - 2 - Estenosis del conducto biliar
  - 3 - Bloqueo 2/2 CA
  - 4 - Extracción del stent
  - 5 - CPRE ambulatoria para cálculos en el CBD
  - 6 - Otra
- ercp\_date: Fecha en la que se realizó la CPRE
- age\_at\_ercp: Edad del paciente en el momento de la CPRE
- gender: Género del paciente
  - 0 - Femenino

- 1 - Masculino
- race: Raza del paciente
  - 0 - Blanco
  - 1 - Hispano
  - 2 - Negro
  - 3 - Asiático
  - 4 - Otra/desconocida
- bmi: IMC del paciente
  - x - No evaluado
- parity: Número de veces que ha dado a luz a un feto de 24 semanas o más
  - 0 - Nulípara: no ha dado a luz previamente
  - M - Masculino
  - x - No evaluado
- dm: Diabetes Mellitus
  - 0 - No
  - 1 - Sí
  - (A1C > 6.5 %)
- ibd: Enfermedad inflamatoria intestinal (EII)
  - 0 - No
  - 1 - Sí

- cirrhosis:
  - 0 - No
  - 1 - Sí
- peak\_bili: Nivel máximo de bilirrubina antes de la CPRE
- stones\_on\_bd: Cálculos biliares visualizados en imágenes antes de la CPRE
  - 0 - No
  - 1 - Sí en TC
  - 2 - Sí en US
  - 3 - Sí en MRCP
- cbd\_diameter\_us: Diámetro máximo del conducto biliar común en US (mm)
  - x - US no realizado
- cbd\_diameter\_mrcp: Diámetro máximo del conducto biliar común en MRCP (mm)
  - x - MRCP no realizado
- cbd\_diameter\_ercp: Diámetro máximo del conducto biliar común en CPRE (mm)
  - x - CPRE no realizado
- intraductal\_filling: Defecto de llenado intraductal en CPRE
  - 0 - No
  - 1 - Lodo
  - 2 - Piedra o escombros de piedra
  - 3 - Lodo y piedra

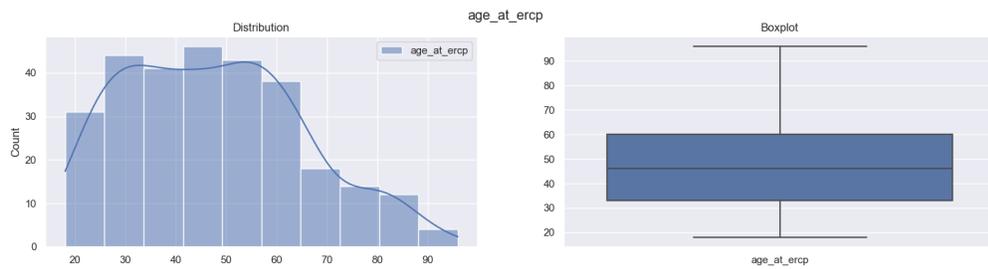
- x - No mencionado, desconocido
  
- cystic\_duct\_filling: Llenado del conducto cístico en CPRE
  - 0 - No
  - 1 - Sí
  - x - Desconocido
  
- stone\_sludge\_ercp: Piedra/lodo visualizado en CPRE
  - 0 - No
  - 1 - Sí
  
- stone\_shape\_ercp: Forma de la piedra en CPRE
  - 0 - Esferoidal
  - 1 - Irregular
  - x - No comentado en CPRE o patología
  
- stone\_color\_ercp: Color de la piedra en CPRE
  - 0 - Amarillo/ocre
  - 1 - Negro
  - 2 - Multicolor
  - x - Desconocido
  
- pyobilia\_ercp: ¿Pióbilis observada en CPRE?
  - 1 - Sí
  - 0 - No

- x - No mencionado en el informe
- stones\_number: Número de piedras o fragmentos observados en CPRE
  - M - Múltiples
  - S - Lodo
- gallbladder: Presencia de vesícula antes de CPRE
  - 0 - Sí
  - 1 - No
- cholecystectomy\_date: Fecha de colecistectomía
- ercp\_cholecystectomy\_delta: Tiempo desde la CPRE hasta la colecistectomía en días
  - x - Ya extraída

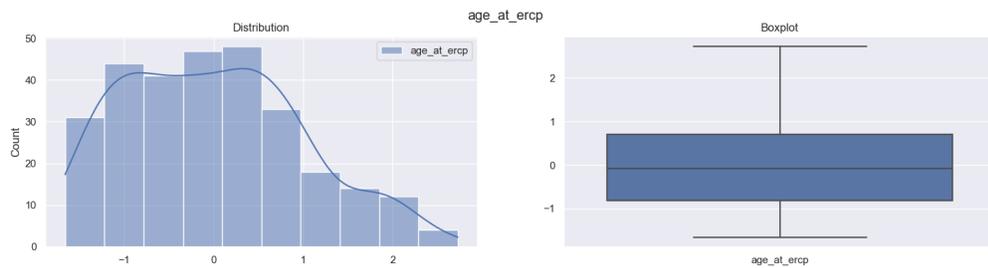


# Apéndice D

## Distribuciones de los atributos utilizados

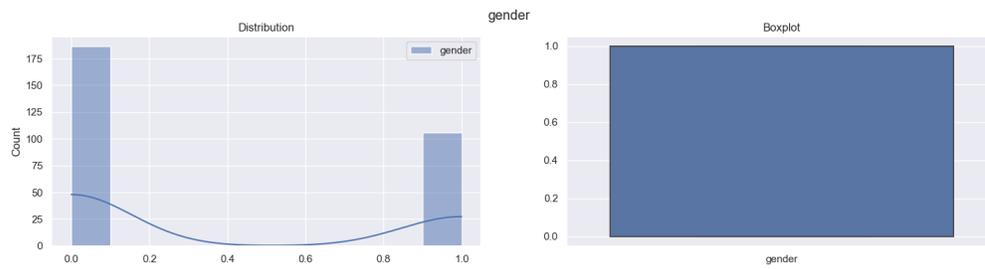


(a) Distribución previo al preprocesamiento.

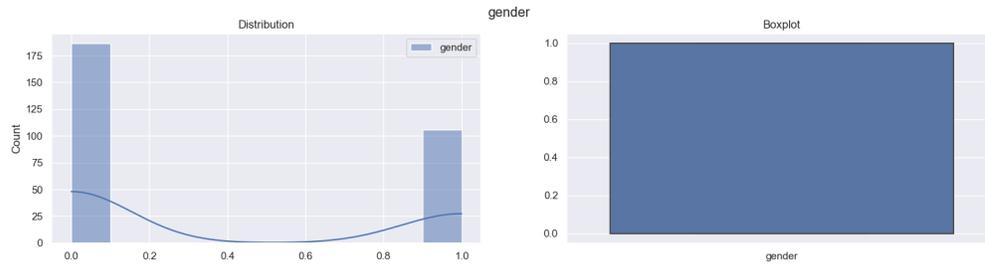


(b) Distribución posterior al preprocesamiento.

Figura D.1: Análisis del efecto del preprocesamiento en la distribución del atributo Edad.

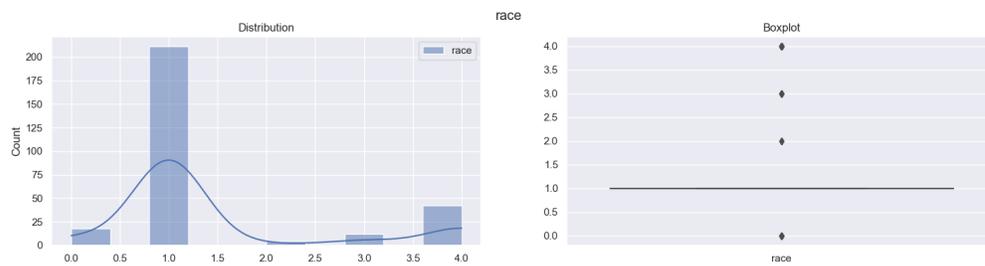


(a) Distribución previo al preprocesamiento.

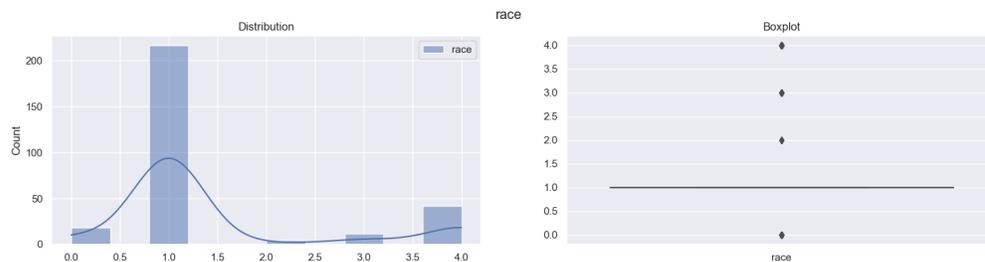


(b) Distribución posterior al preprocesamiento.

Figura D.2: Análisis del efecto del preprocesamiento en la distribución del atributo Género.

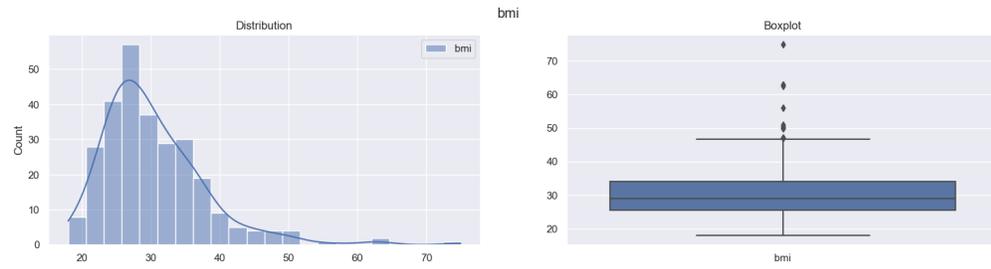


(a) Distribución previo al preprocesamiento.

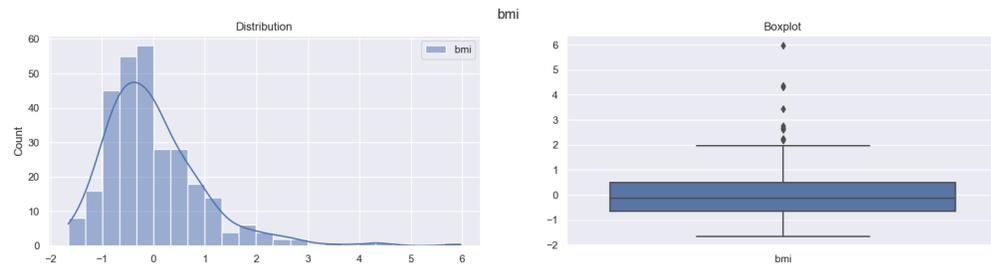


(b) Distribución posterior al preprocesamiento.

Figura D.3: Análisis del efecto del preprocesamiento en la distribución del atributo Origen Étnico.

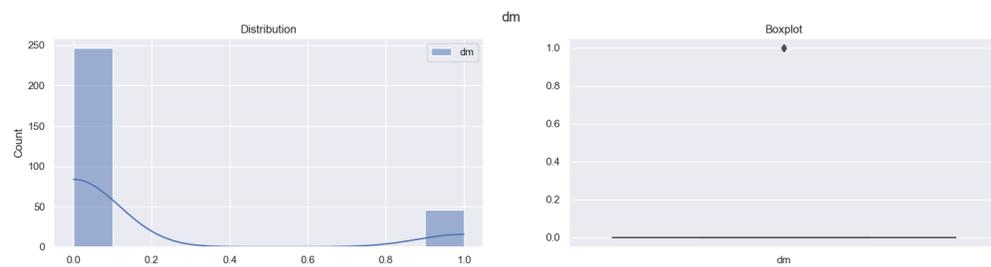


(a) Distribución previo al preprocesamiento.

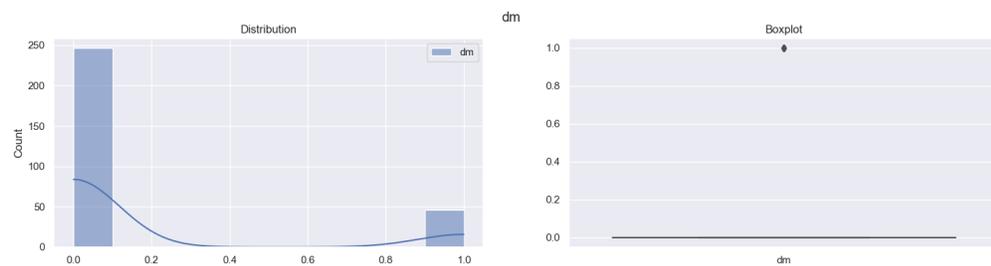


(b) Distribución posterior al preprocesamiento.

Figura D.4: Análisis del efecto del preprocesamiento en la distribución del atributo Índice de Masa Corporal.

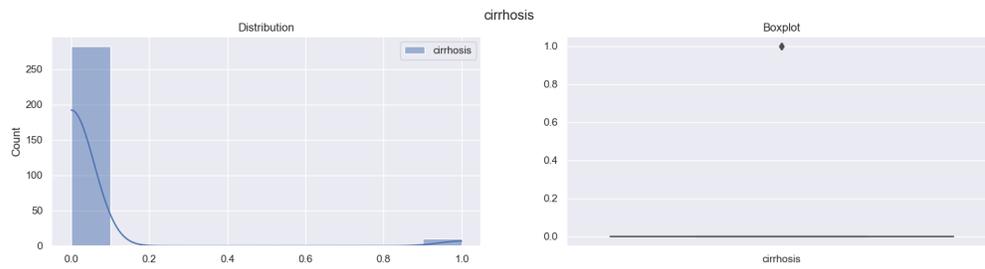


(a) Distribución previo al preprocesamiento.

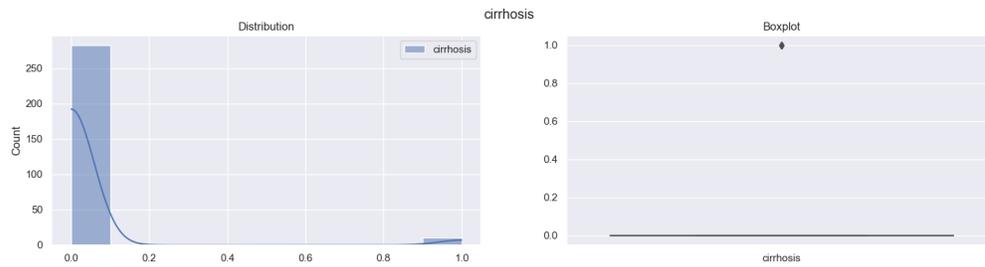


(b) Distribución posterior al preprocesamiento.

Figura D.5: Análisis del efecto del preprocesamiento en la distribución del atributo Diabetes.

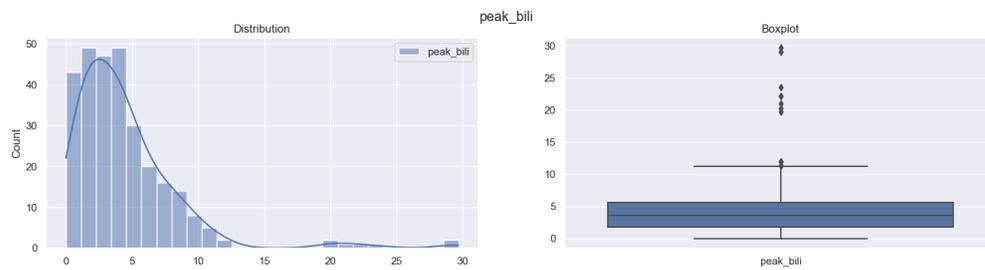


(a) Distribución previo al preprocesamiento.

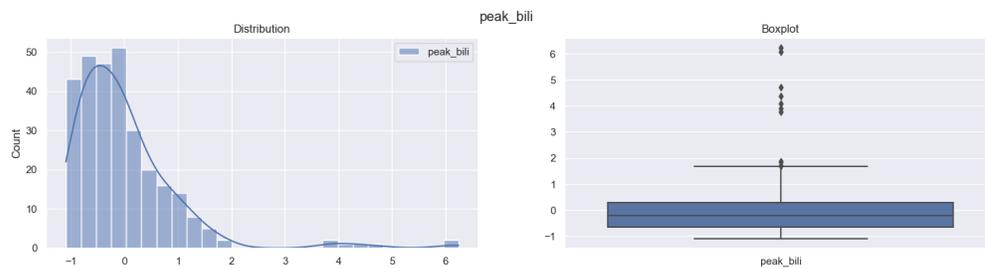


(b) Distribución posterior al preprocesamiento.

Figura D.6: Análisis del efecto del preprocesamiento en la distribución del atributo Cirrosis.

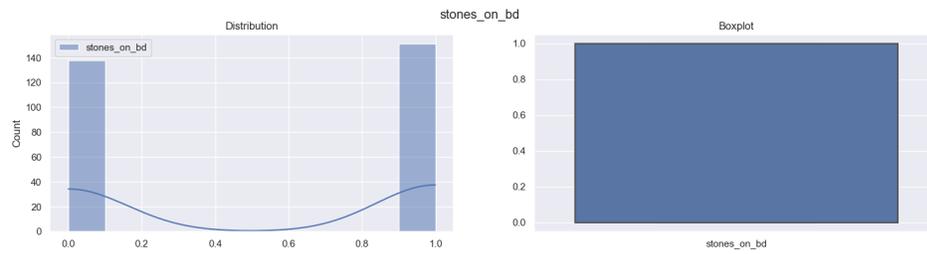


(a) Distribución previo al preprocesamiento.

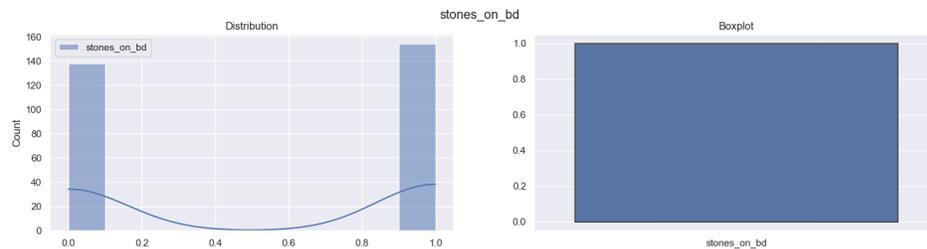


(b) Distribución posterior al preprocesamiento.

Figura D.7: Análisis del efecto del preprocesamiento en la distribución del atributo Bilirrubina Máxima.

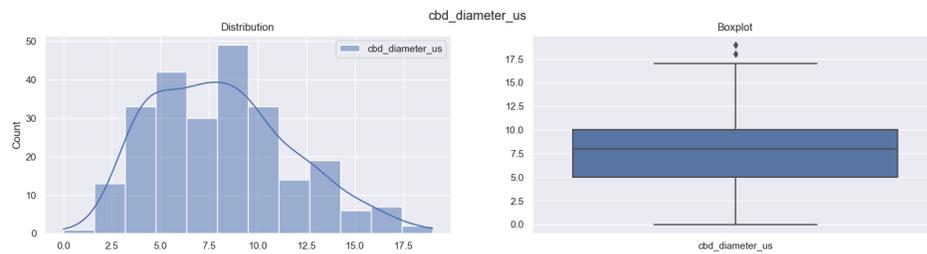


(a) Distribución previo al preprocesamiento.

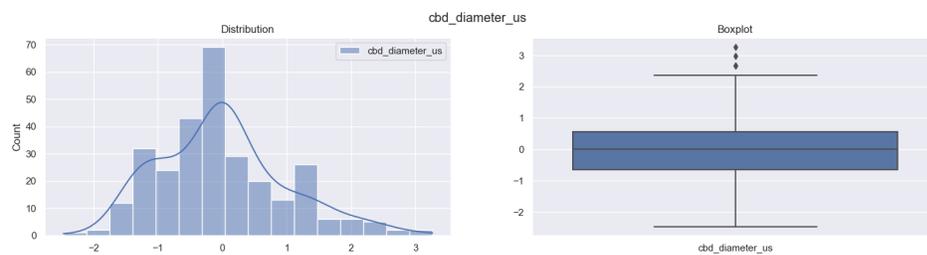


(b) Distribución posterior al preprocesamiento.

Figura D.8: Análisis del efecto del preprocesamiento en la distribución del atributo Cálculos en Conducto Biliar en Imagenología.

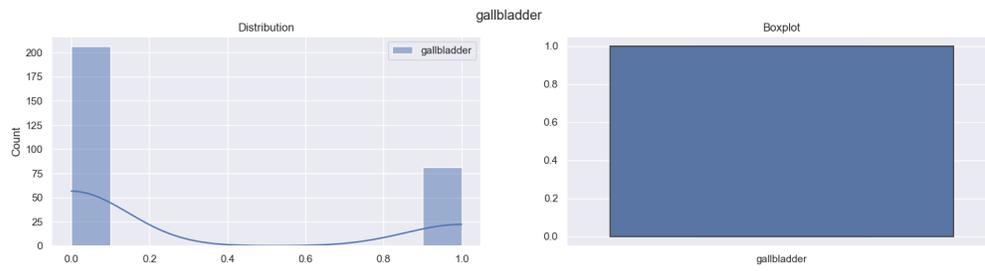


(a) Distribución previo al preprocesamiento.

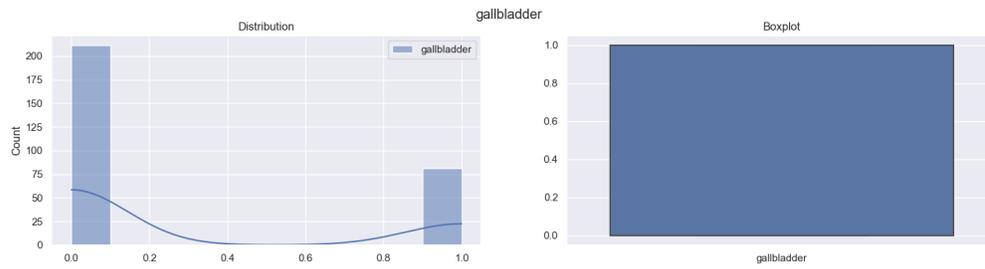


(b) Distribución posterior al preprocesamiento.

Figura D.9: Análisis del efecto del preprocesamiento en la distribución del atributo Diámetro de Colédoco en Ultrasonido.



(a) Distribución previo al preprocesamiento.



(b) Distribución posterior al preprocesamiento.

Figura D.10: Análisis del efecto del preprocesamiento en la distribución del atributo Vesícula Biliar.