



**UNIVERSIDAD DE CONCEPCIÓN
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA ELECTRICA**



**PREDICCIÓN DE MORTALIDAD EN PACIENTES CON ECV
HOSPITALIZADOS EN UCI MEDIANTE MACHINE LEARNING Y LA
BASE DE DATOS MIMIC-III**

POR

Bastian Eduardo Beytía Hormazábal

Informe de Memoria de Título presentado a la Facultad de Ingeniería de la Universidad de
Concepción para optar al título profesional de Ingeniero Civil Biomédico

Profesora Guía
Rosa Liliana Figueroa Iturrieta

Marzo 2026
Concepción (Chile)

© 2026 Bastian Eduardo Beytía Hormazábal

© 2026 Bastian Eduardo Beytía Hormazábal

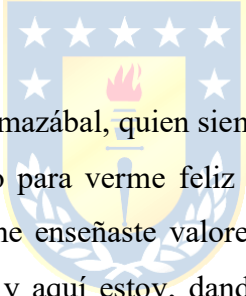
Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento.



Agradecimientos

Deseo expresar mi más profundo agradecimiento a todas las personas que formaron parte de mi proceso universitario, en especial a mi madre, Laura, por apoyo constante en cada uno de mis pasos. Su amor incondicional fue calma en días oscuros, fuerza en momentos de felicidad y compañía en la soledad. Por ello y mil razones más, todos mis logros son por y para ella, su felicidad es la mía, y mi gratitud hacia ella es tan profunda como inefable.

Por otro parte, agradecer a mi padre Evangelo, un hombre gentil y trabajador, cuyo ejemplo y amor han sido fundamentales en mi vida. Gracias por ocupar ese lugar tan importante en nuestro hogar, gracias por enseñarme con paciencia y cariño a ser una buena persona y un buen hombre, por corregirme cuando fue necesario y por entregarme todo lo necesario sin ser tu obligación. Me aceptaste como tu hijo, y yo te amo como mi padre, gracias por cuidar de mis hermanas y de mi madre.



A mi difunto abuelo Manuel Hormazábal, quien siempre confió en que lograría hacer grandes cosas. Siempre pusiste todo tu esfuerzo para verme feliz y motivarme a cumplir mis sueños; me entregaste una infancia maravillosa y me enseñaste valores que aún conservo. Personalmente, me prometí terminar la universidad por ti, y aquí estoy, dando los últimos pasos de este recorrido y recordándote. Este logro también es por y para ti tata, descansa en paz.

A mi pareja, por entregarme siempre el apoyo necesario para poder seguir adelante, por hacer mucho más amenas las horas de estudio, por levantarme la cabeza cada vez que sentía que no podía, y por acompañar mi desarrollo como estudiante y como persona durante una etapa tan importante como la universidad.

Por último, quisiera agradecer a mis amigos, que siempre estuvieron para alegrar las horas de estudio y de clases, y también para apoyarnos y motivarnos mutuamente en lo académico, espero siempre estemos en contacto, y les deseo mucho éxito en la vida.

Resumen

En este proyecto se aborda la predicción de mortalidad en pacientes con enfermedades cardiovasculares (ECV) hospitalizados en UCI, utilizando la base de datos MIMIC-III v1.4 como fuente de información. Se definieron objetivos y alcances enfocados en una cohorte de pacientes UCI adultos con los principales diagnósticos ECV registrados en la base de datos (insuficiencia cardíaca, fibrilaciones auriculares, cardiopatías isquémicas, infartos agudos al miocardio, entre otras.).

La metodología de trabajo incluyó una revisión de literatura inicial, la exploración de MIMIC-III mediante consultas SQL y la construcción de cohortes bajo criterios de integridad (edad y tiempo mínimo en UCI). Posteriormente, se seleccionaron las variables clínicas más relevantes y se diseñó un pipeline en Python para el preprocesamiento, construcción del dataset y entrenamiento de los modelos de machine learning, donde luego, se utilizaron métricas estándar de desempeño para evaluar los modelos mediante un esquema que permitió analizar tanto su rendimiento predictivo como la estabilidad de los resultados frente a las particiones de los datos. Debido a que MIMIC-III es una base de datos masiva, tanto los análisis exploratorios de la base de datos, como el desarrollo y evaluación de los modelos se realizaron utilizando un servidor local de la Universidad.

Finalmente, los modelos basados en gradient boosting (XGBoost, LightGBM y Catboost) obtuvieron el mejor rendimiento, alcanzando valores de AUC-ROC cercanos a 0,80 y superando en gran medida al modelo baseline Random Forest. El análisis de variables más influyentes en el rendimiento de los modelos mostró patrones coherentes con la fisiopatología cardiovascular, destacando parámetros bioquímicos, hematológicos y demográficos tales como los electrolitos principales, hemogramas o la edad. Esto refuerza el potencial del aprendizaje automático como herramienta de apoyo en la predicción de mortalidad en pacientes con ECV ingresados a UCI,

Abstract

This project addresses the prediction of mortality in patients with cardiovascular diseases (CVD) admitted to the intensive care unit (ICU), using the MIMIC-III v1.4 database as the primary data source. The objectives and scope of the study were defined focusing on a cohort of adult ICU patients with the main cardiovascular diagnoses recorded in the database, including heart failure, atrial fibrillation, ischemic heart disease, acute myocardial infarction, among others.

The methodology comprised an initial literature review, exploration of the MIMIC-III database through SQL queries, and cohort construction under integrity criteria such as age and minimum ICU length of stay. Subsequently, the most relevant clinical variables were selected, and a Python-based pipeline was designed for data preprocessing, dataset construction, and machine learning model training. Model performance was evaluated using standard evaluation metrics under a framework that enabled the assessment of both predictive performance and result stability across different data partitions. Given the large-scale nature of the MIMIC-III database, exploratory data analysis as well as model development and evaluation were conducted using a local server provided by the University.

Finally, gradient boosting-based models (XGBoost, LightGBM, and CatBoost) achieved the best performance, reaching AUC-ROC values close to 0,80 and substantially outperforming the Random Forest baseline model. The analysis of the most influential variables revealed patterns consistent with cardiovascular pathophysiology, highlighting biochemical, hematological, and demographic parameters such as key electrolytes, blood count indicators, and age. These findings reinforce the potential of machine learning as a supportive tool for mortality prediction in ICU patients with cardiovascular diseases.

Tabla de Contenidos

CAPÍTULO 1. INTRODUCCIÓN.....	10
1.1. INTRODUCCIÓN GENERAL	10
1.2. TRABAJOS PREVIOS	11
1.2.1 Escalas tradicionales para la medición del riesgo de mortalidad	11
1.2.2 Importancia de las nuevas herramientas.....	11
1.2.3 Aplicaciones anteriores de las nuevas herramientas.....	12
1.2.4 Planteamiento del Problema.....	16
1.3. OBJETIVOS.....	17
1.3.1 Objetivo general.....	17
1.3.2 Objetivos específicos.....	17
1.4. ALCANCES Y LIMITACIONES	17
1.5. TEMARIO	18
CAPÍTULO 2. MARCO TEÓRICO	19
2.1. INTRODUCCIÓN.....	19
2.2. ENFERMEDADES CARDIOVASCULARES EN EL CONTEXTO CLÍNICO	19
2.2.1 Contexto epidemiológico de las principales ECV en Chile	20
2.3. PRINCIPALES ENFERMEDADES CARDIOVASCULARES	21
2.3.1 Insuficiencia cardíaca.....	21
2.3.2 Infarto agudo de miocardio	22
2.3.3 Arritmias	22
2.3.4 Cardiopatía isquémica.....	23
2.3.5 Hipertensión arterial severa	24
2.4. VARIABLES FISIOLÓGICAS INTEGRADAS EN ESTUDIOS PREVIOS.....	24
2.5. BASE DE DATOS	25
CAPÍTULO 3. DESARROLLO TÉCNICO.....	26
3.1. ESTRUCTURA Y EXPLORACIÓN DE MIMIC-III v1.4	26
3.2. METODOLOGÍA DE TRABAJO.....	28
3.2.1 Variables clínicas seleccionadas para el proyecto.....	29
3.3. DESARROLLO DEL PIPELINE EN PYTHON	31
3.3.1 Estructura General	31
3.3.2 Carga y procesamiento de datos.....	32
3.3.3 Construcción del dataset final	33
3.3.4 Partición de datos y manejo del desbalance.....	33
3.3.5 Entrenamiento y evaluación de los modelos de ML	34
CAPÍTULO 4. RESULTADOS.....	34
4.1. INTRODUCCIÓN.....	34
4.2. DESCRIPCIÓN DE LA COHORTE.....	35
4.3. EVALUACIÓN DE LOS MODELOS.....	35
4.3.1 Consideraciones importantes.....	36
4.3.2 Métricas globales por modelo.....	37
4.3.3 Importancia de variables	39
4.4. ANÁLISIS COMPLEMENTARIO	40
4.4.1 Descripción del proceso.....	41
4.4.2 Resultados e interpretación	42
4.5. DISCUSIÓN.....	44
CAPÍTULO 5. CONCLUSIONES.....	46
5.1. CONCLUSIONES	46
5.2. TRABAJO FUTURO	46
CAPÍTULO 6. GLOSARIO.....	48

CAPÍTULO 7. REFERENCIAS.....	50
ANEXO A. DESCRIPCIÓN DE ESCALAS TRADICIONALES DE MORTALIDAD	53
ANEXO B. VARIABLES FISIOLÓGICAS INTEGRADAS EN EL PROYECTO	54



Lista de Tablas

Tabla 1.1. Comparación de estudios previos sobre la predicción de mortalidad en pacientes con enfermedades cardiovasculares en general, mediante el uso de ML y un conjunto de datos.	15
Tabla 2.1 Variables fisiológicas integradas en los modelos de Machine Learning de estudios previos.	25
Tabla 3.1. Diagnósticos cardiovasculares (ICD-9) más predominantes en la cohorte de pacientes con ECV de MIMIC-III, ordenadas según su frecuencia.....	27
Tabla 3.2. Resume los conteos de mortalidad obtenidos de MIMIC-III v1.4, incluyendo el contexto global por admisión, el contexto de UCI sin ECV y el contexto de pacientes UCI con ECV. Lo que permite comparar las cohortes.....	27
Tabla 4.1. Tabla resumen con los resultados obtenidos con los modelos de ML desarrollados en el proyecto.....	38
Tabla 4.2. Se muestran los resultados desglosados obtenidos por los modelos en cada seed utilizada.	42
Tabla 4.3. Se muestra el promedio de los resultados por seed, para cada modelo desarrollado.	42
Tabla 0.1. Variables fisiológicas integradas en los modelos de Machine Learning en este proyecto.	54



Lista de Figuras

Figura 1.1 Comparación de Modelos Basado en AUC, Intervalo de Confianza (CI) y Accuracy (Test Dataset) [4].	13
Figura 1.2 Comparación por pares de la eficacia de predicción del modelo XGBoost, el modelo LASSO y el modo de puntuación de riesgo GWTG-HF, utilizada en el estudio [8].	14
Figura 1.3 AUROC para la predicción de mortalidad a 3 años utilizando el modelo de ML CoxBoost desarrollado por Park et al. (2022). Se destaca el buen desempeño del modelo frente a scores clínicos como BIOSTAT-CHF y AHEAD [7].	14
Figura 3.1 Diagrama de flujo de la metodología de trabajo para la implementación de los modelos de ML en la predicción de la mortalidad mediante datos extraídos de MIMIC-III v1.4.	29
Figura 4.1. Se muestran las matrices de confusión de cada modelo de ML desarrollado en el proyecto.	38
Figura 4.2. Se muestran las 20 variables más importantes para los 3 modelos basados en boosting.	40
Figura 4.3. Se muestran las 20 variables más importantes para los 3 modelos basados en boosting, utilizando el enfoque multi-seed con undersampling mediante NearMiss.	43
Figura 4.4. Se muestran las matrices de confusión obtenidas de los 4 modelos desarrollados con el enfoque multi-seed, utilizando la seed 10.	44



Capítulo 1. Introducción

1.1. Introducción General

Las enfermedades cardiovasculares (ECV) constituyen la principal causa de muerte a nivel mundial. Estas abarcan un gran espectro de condiciones o enfermedades que incluyen patologías coronarias, insuficiencias cardíacas, arritmias, valvulopatías, miocardiopatías, entre otras. Según la Organización Mundial de la Salud (OMS), estas patologías son responsables de aproximadamente un tercio de las muertes globales [1]. Para el caso de aquellos pacientes que fueron ingresados en Unidad de Cuidados Intensivos (UCI), las patologías cardiovasculares representan una de las principales causas de mortalidad y morbilidad [2][3].

Estas condiciones cardíacas suelen compartir mecanismos fisiopatológicos que producen un deterioro multiorgánico progresivo, que en muchos casos puede conllevar a la muerte del paciente si no es detectado y tratado de forma temprana. La mortalidad asociada a la insuficiencia cardíaca, infartos agudos de miocardio, arritmias ventriculares malignas y otras condiciones cardiovasculares críticas puede variar ampliamente, llegando a superar el 30% en algunos casos [3][4].

Ante este escenario, la posibilidad y capacidad de predecir la mortalidad de forma temprana y precisa se torna en una herramienta que puede ser fundamental para orientar la toma de decisiones médicas, optimizar el uso de recursos y así poder reducir la mortalidad en los pacientes. Sin embargo, históricamente se han implementado ciertos instrumentos o escalas para estimar la gravedad y el riesgo de mortalidad de los pacientes críticos. Por ello, y dado al potencial que se observa en las nuevas técnicas de aprendizaje automático en predicciones de mortalidad, se han desarrollado modelos de predicción flexibles y precisos, lo que convierte este tema en un campo de investigación relevante y con potencial suficiente para ser explorado con mayor profundidad.

1.2. Trabajos previos

1.2.1 Escalas tradicionales para la medición del riesgo de mortalidad

Sistemas de puntuación clínica como APACHE II (Acute Physiology and Chronic Health Evaluation II), SAPS II (Simplified Acute Physiology Score II) y SOFA (Sequential Organ Failure Assessment) han sido herramientas relevantes en el campo de la medicina intensiva. Estas están diseñadas para proporcionar una estimación de la mortalidad del paciente basándose en variables fisiológicas, antecedentes clínicos y resultados de laboratorio, recolectados principalmente durante las primeras 24 horas de hospitalización en UCI.

APACHE II, por ejemplo, considera 12 parámetros fisiológicos, edad y condición médica crónica, SAPS II incorpora 17 variables entre fisiológicas y de comorbilidad, mientras que SOFA evalúa el estado funcional de seis sistemas orgánicos principales. Sin embargo, estos sistemas presentan limitaciones en la forma en que integran la información, utilizando ponderaciones fijas y lineales, lo que no permite captar relaciones complejas entre variables fisiológicas [5].

Adicionalmente, estas herramientas se desarrollaron sin una focalización específica en las múltiples formas de enfermedad cardiovascular, lo que motiva la necesidad de explorar nuevas formas de predicción de mortalidad en los pacientes hospitalizados [3]. Una descripción más detallada de cada una de estas escalas, incluyendo su estructura, cálculo y variables específicas, se presenta en el anexo “A”.

1.2.2 Importancia de las nuevas herramientas

Durante los últimos años han emergido técnicas poderosas y versátiles para abordar desafíos complejos de predicción en medicina. Una de estas técnicas es el Machine Learning (ML) que a diferencia de las escalas tradicionales (APACHE II, SAPS II y SOFA), pueden utilizarse para procesar grandes volúmenes de datos en el área clínica, ya sean signos vitales, resultados de laboratorio, registros fisiológicos, historiales médicos, biomarcadores, entre otros. Esto permite identificar ciertos patrones no lineales y complejos en conjuntos de datos heterogéneos, ahorrando así la necesidad de definir o establecer las relaciones que existen entre las variables [5][2].

Un actor fundamental en el desarrollo de investigaciones sobre este tipo de técnicas ha sido la base de datos MIMIC-III, la cual incluye información detallada y estructurada de más de 40.000 pacientes ingresados a la UCI [6]. Esto ha permitido el entrenamiento y la validación de diversos algoritmos en una población diversa y clínicamente relevante [4][7][8].

Lo anterior permite avanzar a una medicina más personalizada y respaldada con datos, lo que mejorar la toma de decisiones médicas y la distribución eficiente de recursos, pudiendo impactar de forma positiva en los resultados clínicos.

1.2.3 Aplicaciones anteriores de las nuevas herramientas

El estudio realizado por Huang *et al.* (2025) se centró en pacientes adultos de la UCI y se utilizaron datos de MIMIC-III para poder evaluar el impacto predictivo de la Variabilidad de la Frecuencia Cardíaca (HRV) en la mortalidad hospitalaria. Se evaluaron algoritmos como Regresión Logística (RL), Random Forest (RF), XGBoost (XGB), LightGBM (LGB), Ensemble y Redes Neuronales (NN). Los datos se representaron de tres formas:

- COB (Clinical-Only Baseline): Sólo incluye variables clínicas.
- HOB (HRV-Only Baseline): Incluye solo características de HRV.
- Integrated: Combina variables clínicas y de HRV.

El mejor rendimiento se obtuvo con el modelo Ensemble (integrated) con un accuracy de 0,91 y el estudio demuestra que la integración de datos clínicos con HRV mejora la capacidad predictiva de los modelos de ML [5].

Por otro lado, Ashrafi *et al.* (2024) desarrolló un modelo enfocado en pacientes con insuficiencia cardíaca, donde el algoritmo XGBoost con ajuste de hiperparámetros y validación cruzada obtuvo un AUROC de 0,92, superando a modelos de ML como Regresión Logística, Random Forest, LightGBM, entre otros, lo cual se puede observar en la Figura 1.1. Además, se utilizó el método SHAP para analizar el impacto de cada variable en las predicciones del modelo [4].

Models	AUROC	AUROC 95% CI	Accuracy
XGBoost	0.9228	[0.8748 - 0.9613]	0.8966
Logistic Regression	0.8739	[0.7859 - 0.9583]	0.8079
Random Forest	0.8727	[0.7633 - 0.9605]	0.9212
LightGBM	0.8471	[0.7489 - 0.9331]	0.9015
SVM	0.8682	[0.7720 - 0.9484]	0.8966
Lasso	0.8739	[0.7861 - 0.9592]	0.8079
KNN	0.8330	[0.7307 - 0.9192]	0.8818

Figura 1.1 Comparación de Modelos Basado en AUC, Intervalo de Confianza (CI) y Accuracy (Test Dataset) [4].

El estudio de Srinivasan *et al.* se enfoca en predecir ECV a partir del Heart Disease Dataset disponible en el UCI Machine Learning Repository. Se aplicaron múltiples modelos de machine learning, tales como: Random Forest, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Decision Tree (DT), XGBoost, Naive Bayes (NB), Redes Neuronales con Función de Base Radial (RBF) y un modelo propuesto de Learning Vector Quantization (LVQ). Este último utiliza vectores prototipo para representar las clases, alcanzando un accuracy de 98,78%, una sensibilidad de 97,91% y una especificidad de 97,1%, superando a los modelos tradicionales. Entre estos, los que más destacaron fueron RBF, con un accuracy de 90,78%, y DT con un 89%. El estudio evaluó el desempeño de cada algoritmo mediante métricas como precisión, F1-Score y curvas ROC, lo que permitió demostrar el rendimiento sobresaliente del modelo LVQ frente a los demás enfoques analizados [9].

El trabajo de Li *et al.* (2021) está enfocado en pacientes ingresados en UCI con diagnóstico de insuficiencia cardíaca, obtenidos de la base de datos MIMIC-III. El objetivo fue desarrollar y validar un modelo predictivo de ML comparando el rendimiento de tres modelos principales: XGBoost, Regresión LASSO y GWTG-HF Risk. El modelo XGBoost mostró un AUROC de 0,83, siendo el modelo con mejor rendimiento, por lo que los autores concluyen que los modelos de ML pueden superar a los scores clínicos tradicionales al ser abordados correctamente [8].

Models	XGBoost model vs LASSO regression model			LASSO regression model vs GWTG-HF risk score model			XGBoost model vs GWTG-HF risk score model		
	XGBoost model	LASSO regression model	P (compare)	LASSO regression model	GWTG-HF risk score model	P (compare)	XGBoost model	GWTG-HF risk score model	P (compare)
ROC area (AUC)	0.8498	0.8378	0.1961	0.8518	0.7743	0.0299	0.8378	0.7834	0.0471
95% CI lower	0.7993	0.7772		0.8011	0.7097		0.7826	0.7239	
95% CI upper	0.8998	0.8921		0.8998	0.8274		0.8847	0.8352	
Specificity	0.7273	0.6537		0.7308	0.6239		0.6537	0.6234	
Sensitivity	0.8696	0.8841		0.8714	0.8000		0.8841	0.8116	
Accuracy	0.7600	0.7067		0.7632	0.6645		0.7067	0.6667	
Positive-LR	3.1884	2.5527		3.2367	2.1273		2.5527	2.1549	
Negative-LR	0.1793	0.1774		0.1759	0.3205		0.1774	0.3022	
Diagnose-OR	17.7778	14.3922		18.3968	6.6364		14.3922	7.1300	
Postive-pv	0.4878	0.4326		0.4919	0.3889		0.4326	0.3916	
Negative-pv	0.9492	0.9497		0.9500	0.9125		0.9497	0.9172	

Figura 1.2 Comparación por pares de la eficacia de predicción del modelo XGBoost, el modelo LASSO y el modo de puntuación de riesgo GWTG-HF, utilizada en el estudio [8].

El estudio realizado por Park et al. (2022) se enfocó en el desarrollo de un modelo predictivo de mortalidad a 3 años en pacientes hospitalizados por insuficiencia cardiaca aguda (AHF), se usaron datos de la cohorte STRATS-AHF la cual incluye 4.312 pacientes. El modelo se desarrolló usando un algoritmo de regresión de Cox potenciado (CoxBoost) e integró 27 variables (19 clínicas y 8 ecocardiográficas). Se obtuvo una AUROC de 0,76 (ver Figura 1.2), logrando una buena capacidad de predicción de mortalidad a largo plazo al combinar datos clínicos y ecocardiográficos [7].

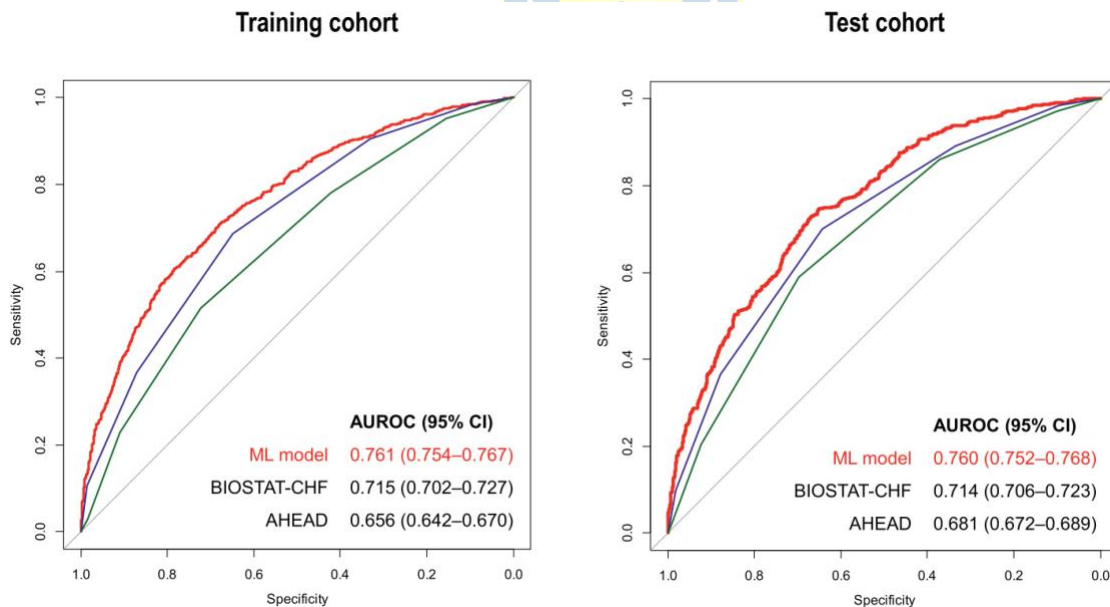


Figura 1.3 AUROC para la predicción de mortalidad a 3 años utilizando el modelo de ML CoxBoost desarrollado por Park et al. (2022). Se destaca el buen desempeño del modelo frente a scores clínicos como BIostat-CHF y AHEAD [7].

En el estudio desarrollado por Jang *et al.* (2023), el objetivo fue evaluar el desempeño y la capacidad del modelo MARKER-HF para predecir la mortalidad a un año en pacientes con y sin insuficiencia cardíaca. Para ello, se utilizó una cohorte hospitalaria de 41.749 pacientes, los cuales fueron divididos en subgrupos según la presencia de IC, enfermedades cardiovasculares u otras comorbilidades como síndrome coronario agudo (SCA), fibrilación auricular (FA), enfermedad pulmonar obstructiva crónica (EPOC), enfermedad renal crónica (ERC), diabetes mellitus, hipertensión o cáncer. El modelo MARKER-HF se basa en árboles de decisión potenciados y fue entrenado con sólo ocho variables clínicas: presión diastólica, creatinina, BUN, hemoglobina, leucocitos, plaquetas, albúmina y RDW. En el subgrupo con insuficiencia cardíaca, se obtuvo un área bajo la curva (AUC) de 0,73, concluyendo que MARKER-HF es un modelo robusto y aplicable para predecir mortalidad en este tipo de pacientes [3].

Tabla 1.1. Comparación de estudios previos sobre la predicción de mortalidad en pacientes con enfermedades cardiovasculares en general, mediante el uso de ML y un conjunto de datos.

Autor (año)	Base de datos	Modelos de ML Evaluados	Métrica destacada
Huan <i>et al.</i> (2025)	MIMIC-III	RL, RF, XGB, LGB, Ensemble, NN	Accuracy: 0,91 (Ensemble)
Ashrafi <i>et al.</i> (2024)	MIMIC-III	XGBoost, RL, RF, LightGBM	AUROC: 0,92 (XGBoost)
Srinivasan <i>et al.</i> (2023)	UCI Heart Disease	RF, KNN, SVM, DT, XGB, NB, RBF, LVQ	Accuracy: 0,99 (LVQ)
Li <i>et al.</i> (2021)	MIMIC-III	XGBoost, LASSO, GWTG-HF Score	AUROC: 0,83 (XGBoost)
Jang <i>et al.</i> (2023)	Cohorte Hospitalaria	MARKER-HF	AUC: 0,73 (MARKER-HF)
Park <i>et al.</i> (2022)	Cohorte Hospitalaria	CoxBoost	AUROC: 0,76 (CoxBoost)

Fuente: Elaboración propia. Información obtenida de [3] [4] [5] [7] [8] [9].

En la Tabla 1.1, se resumen seis estudios previos en los que se aplican modelos de ML para predecir mortalidad en el contexto de ECV. Cabe destacar que los modelos en los que se utilizaron cohortes hospitalarias reales como MARKER-HF y CoxBoost obtuvieron un rendimiento menor que los que usaron datos más estructurados como MIMIC-III o UCI Heart Disease, lo que podría estar relacionado con la calidad y el grado de estandarización de las variables disponibles en cada tipo de base de datos.

1.2.4 Planteamiento del Problema

En el mundo contemporáneo, la predicción de mortalidad en pacientes hospitalizados por ECV continúa representando un desafío relevante. Si bien las escalas tradicionales como APACHE II, SAPS II y SOFA han sido utilizadas para obtener estimaciones de riesgo, éstas presentan limitaciones asociadas a su estructura y a la forma en que integran la información clínica. Estas escalas se basan en modelos con ponderaciones fijas y suposiciones lineales, restringiendo su capacidad para captar interacciones complejas entre variables fisiológicas, adaptarse a poblaciones clínicas específicas tal como pacientes con ECV, y también adaptarse a la heterogeneidad de las cohortes hospitalarias[5].

En el contexto específico de las enfermedades cardíacas, que incluyen una variedad de condiciones tales como insuficiencia cardíaca aguda o crónica, arritmias complejas, síndromes coronarios y enfermedades estructurales en general, la rapidez en la progresión del deterioro hemodinámico, la diversidad de casos clínicos y la presencia de comorbilidades hacen que las actuales herramientas de predicción pierdan precisión y posibles aplicaciones [3]. Además, en la práctica clínica, estas escalas suelen tener dificultades para integrar de forma simultánea distintas fuentes de información, no consideran de forma correcta el manejo de los datos faltantes y requieren en muchos casos de una recolección de información de forma manual, por lo que, en contextos hospitalarios como la UCI, pueden no resultar ser lo más adecuado o eficiente.

Dicho esto, los avances tecnológicos en inteligencia artificial, y particularmente en técnicas de ML, han demostrado un gran potencial para mejorar la capacidad de predecir mortalidad analizando volúmenes de datos más grandes y heterogéneos. Aun así, existen ciertas brechas en este ámbito, como la capacidad de aplicación en contextos clínicos reales, lo que puede significar un problema si es que no se está preparado para aplicar estas técnicas innovadoras. Por otro lado, también se debe considerar la necesidad de definir qué modelos de ML son los más adecuados para realizar las predicciones de mortalidad.

Por lo tanto, en esta investigación se propone el desarrollo, implementación y comparación de modelos de ML para la predicción de mortalidad hospitalaria en pacientes UCI con ECV, esto con el uso de la base de datos MIMIC-III v1.4. Se espera que este enfoque, basado en variables clínicas registradas durante las primeras 24 horas de estancia en UCI, permita identificar patrones clínicos más

complejos, superar algunas limitaciones de las escalas tradicionales y contribuir al desarrollo de técnicas de apoyo a la medicina moderna, haciéndola más eficiente y sustentada en datos cuantitativos.

1.3. Objetivos

1.3.1 Objetivo general

Desarrollar modelos de machine learning para la predicción de la mortalidad hospitalaria en pacientes adultos ingresados en UCI por patologías cardiovasculares, utilizando la base de datos MIMIC-III v1.4.

1.3.2 Objetivos específicos

- Seleccionar desde la base de datos MIMIC-III, un conjunto de variables clínicas relevantes al problema, basándose en la literatura previa y en la caracterización de la base de datos disponible.
- Realizar el preprocesamiento a los datos clínicos para luego seleccionar y extraer las variables más relevantes y relacionadas con el riesgo de mortalidad en pacientes con patologías cardíacas, con el fin de integrarlos a los modelos de ML desarrollados.
- Desarrollar y aplicar distintos modelos de ML para predecir la mortalidad hospitalaria en pacientes con ECV.
- Evaluar y comparar el desempeño de los modelos de machine learning implementados.

1.4. Alcances y limitaciones

- i. Solo se consideraron pacientes hospitalizados en la UCI, por lo que los resultados no aplican a unidades médicas menos críticas.
- ii. No se consideran sujetos menores de 18 o que hayan estado menos de 24 horas en UCI.
- iii. Para efectos de esta memoria de título solo son consideradas las siguientes afecciones: insuficiencia cardíaca, fibrilación auricular, cardiopatías isquémicas, infartos agudos del miocardio, valvulopatías, pericarditis y cardiopatías congénitas. Lo anterior por ser las enfermedades con mayor prevalencia dentro de los pacientes de UCI en MIMIC-III.
- iv. Se utilizó dBeaver Community v25.2.3 para explorar la base de datos MIMIC-III versión 1.4 y se trabajó con Python versión 3.12.7 en Anaconda Spyder 6.0.7.

- v. La investigación y aplicación se realizó con hardware de gama media. GPU: Nvidia GTX 1660S, 16GB RAM, CPU: i3-10100F, SSD 256GB.
- vi. Para el acceso completo a la base de datos MIMIC-III se utilizó un servidor donde ésta fue cargada previamente, esto debido al tamaño elevado de la base de datos.

1.5. Temario

- (i) Capítulo 1: En este capítulo se presenta el contexto general del estudio, abordando la importancia del tema, ciertos trabajos previos, los objetivos tanto generales como específicos, los alcances y limitaciones, además de la metodología propuesta.
- (ii) Capítulo 2: En este capítulo se desarrolla el marco teórico, donde se profundiza en los fundamentos necesarios para el trabajo, se entrega contexto más específico sobre las enfermedades cardiovasculares y técnicas empleadas, por último, se discuten los antecedentes más relevantes.
- (iii) Capítulo 3: En este capítulo se realiza el desarrollo técnico del proyecto, describiendo la estructura de MIMIC-III, la metodología de trabajo, selección de variables clínicas y la implementación del pipeline en Python para entrenar los modelos de ML.
- (iv) Capítulo 4: Acá se presentan los resultados obtenidos del entrenamiento y evaluación de los modelos, incluyendo la descripción de la cohorte, métricas de rendimiento y la importancia de las variables más relevantes.
- (v) Capítulo 5: En este capítulo se exponen las conclusiones principales derivadas del estudio y se propone trabajo futuro para mejoras de la metodología aplicada.
- (vi) Capítulo 6: Se presenta el glosario con los principales términos y signos utilizados en el desarrollo del proyecto.
- (vii) Capítulo 7: Se listan las referencias bibliográficas utilizadas en el proyecto.
- (viii) Anexos: Se incluyen anexos con información complementaria, como la descripción de las escalas tradicionales de mortalidad y el detalle de las variables fisiológicas integradas en los modelos de ML.

Capítulo 2. Marco Teórico

2.1. Introducción

Este capítulo tiene como finalidad aclarar y establecer los fundamentos teóricos y metodológicos necesarios para fundamentar el desarrollo de los modelos predictivos analizados en esta memoria de título. Se busca contextualizar una problemática clínica, como lo es la mortalidad en pacientes hospitalizados que fueron ingresados por alguna enfermedad cardíaca, además de describir las herramientas y técnicas que se han utilizado antes, junto con las tecnologías más emergentes y contemporáneas en esta área.

2.2. Enfermedades cardiovasculares en el contexto clínico

Como se mencionó previamente, las enfermedades cardiovasculares (ECV) constituyen la principal causa muerte a nivel mundial, según datos de la Organización Mundial de la Salud (OMS), estas patologías son las responsables del 32% de todas las muertes en el mundo, lo que significa cerca de 18 millones de fallecidos por año [1]. Estos datos reflejan no solo una mortalidad elevada, sino también una importante carga de morbilidad, hospitalizaciones y deterioros progresivos en salud.

Dentro de los factores de riesgo más recurrentes para el desarrollo de ECV, se encuentran condiciones que aumentan el riesgo de padecer alguna de estas enfermedades, como la hipertensión arterial, diabetes mellitus tipo 2, la dislipidemia, el tabaquismo, la obesidad y el sedentarismo. Estos factores de riesgo son compartidos por diversas ECV, como la cardiopatía isquémica, la insuficiencia cardíaca, arritmias, valvulopatías, miocardiopatías, entre otras [10].

En el contexto hospitalario, y sobre todo en áreas complejas como la Unidad de Cuidados Intensivos (UCI), estas enfermedades se vuelven muy relevantes para la sobrevivencia del paciente, ya que las personas que padecen de patologías cardiovasculares presentan una mayor probabilidad de sufrir complicaciones en su estadía, aumentando así el riesgo de eventos adversos y de mortalidad intrahospitalaria [8] [5].

Existen estudios diversos que han utilizado datos de pacientes críticos para desarrollar modelos predictivos de mortalidad basándose en variables fisiológicas y en antecedentes cardíacos. Trabajos como los de Li et al. [4], Jang et al. [3] y Ashrafi et al. [8], analizan sujetos con insuficiencia cardíaca

a partir de bases de datos como MIMIC-III y de diversas cohortes hospitalarias mediante técnicas de aprendizaje automático. Estos estudios muestran que las características clínicas pueden variar de forma considerable incluso dentro de la categoría de ECV.

2.2.1 Contexto epidemiológico de las principales ECV en Chile

En el contexto nacional, en 2021 se registraron 31.729 muertes relacionadas con ECV, lo que representa un 23,05% del total de fallecidos de ese año [11]. Entre los años 2000 y 2020, el número de pacientes fallecidos por ECV aumento de 21.956 a 29.019, sin embargo, la tasa de mortalidad ajustada tuvo un descenso de 159,5 a 94,6 muertes por cada 100.000 habitantes [12].

La principal causa de muerte directa por ECV en Chile, se da por el infarto agudo de miocardio (IAM), ya que es responsable del 31,2% de las muertes en personas entre los 30 y 69 años, y un 20,4% en personas mayores de 60 años [13].

Para el caso de la IC, esta representa aproximadamente un 10% del total de registros por ECV en Chile, presentando una alta tasa de rehospitalización. Además, constituye una causa importante de mortalidad, con tasas de sobrevida a 1 año de entre 67 y 87% y a los 5 años de 50 y 60%, siendo una de las mayores causas de hospitalización en mayores de 65 años [14].

La arritmia más frecuente en Chile es la Fibrilación Auricular (FA), la cual es una causa frecuente de hospitalización en la población adulta. Si bien su tasa de mortalidad se mantuvo en 7,4 muertes por cada 100.000 habitantes entre 2019 y 2022, la presencia de FA se asocia a un aumento de entre 1,5 y 1,9 veces en el riesgo de mortalidad frente a pacientes que no la padecen [15].

Las cardiopatías isquémicas representan una de las principales ECV en Chile, ubicándose como la primera causa de muerte en hombres y la segunda en mujeres. En el periodo entre el 2000 y el 2021 se reportaron 175.973 muertes por esta condición, sin embargo, la proporción de muertes asociadas a cardiopatías isquémicas se redujo de un 9,7% en el año 2000 a un 6,5% en el 2021. Esta tendencia se puede relacionar con mejoras en la prevención, el diagnóstico y el tratamiento cardiovascular. No obstante, la pandemia de COVID-19 alteró la distribución de causas de muerte, al concentrar una parte importante de estas en 2020 [16].

En el caso de la hipertensión arterial en Chile, su prevalencia en la población mayor de 15 años fue de un 27,6%, cifra que aumenta significativamente con la edad, afectando a un 43,8% de adultos de 45 a 64 años y llegando a un 73,3% en mayores de 65 años [13].

2.3. Principales enfermedades cardiovasculares

Luego de realizar el análisis de los registros clínicos disponibles en MIMIC-III, se identificó un conjunto recurrente de ECV asociadas a un mayor riesgo de mortalidad intrahospitalaria. Estas patologías constituyen el grupo de condiciones frecuentes y relevantes en pacientes ingresados a UCI.

2.3.1 Insuficiencia cardíaca

La insuficiencia cardíaca (IC) se caracteriza por la incapacidad del corazón para bombear sangre de forma correcta, lo que conlleva a una mala perfusión tisular y a congestión venosa sistémica. Los pacientes pueden presentar signos de congestión pulmonar como dificultad para respirar (disnea), que empeora en posición horizontal (ortopnea); además, pueden predominar síntomas de congestión sistémica, como acumulación de líquidos en las extremidades (edema), hepatomegalia o distensión venosa yugular [17].

En el contexto de este proyecto se consideran múltiples variables fisiológicas y de laboratorio que se consideran relevantes dado que reflejan el estado clínico del paciente, entre ellas:

- Presión arterial media (PAM).
- Frecuencia cardíaca.
- Creatinina y urea sérica.
- Sodio y potasio.
- Glasgow Coma Scale (GCS)

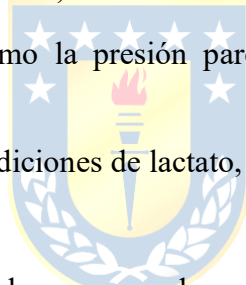
Estas variables permiten evaluar la perfusión tisular, función renal, estado hemodinámico y el estado neurológico, por lo que se consideran marcadores clave en la predicción de mortalidad en pacientes con IC ingresados en UCI.

2.3.2 Infarto agudo de miocardio

El infarto agudo de miocardio (IAM), se define como la necrosis del tejido cardíaco a causa de la interrupción del flujo coronario. Entre sus principales factores de riesgo, destacan la obesidad, el tabaquismo, el sedentarismo, la dislipidemia, el alcoholismo, la diabetes y la hipertensión arterial. Se estima que cerca de un 47% de los IAM son atribuibles a la hipertensión a nivel mundial [13].

Respecto a la sintomatología, el paciente puede presentar dolor torácico opresivo con irradiación a cuello, mandíbula, o extremidades superiores [13]. También se observan alteraciones en el ECG y elevaciones en marcadores cardíacos como la troponina. Entre los parámetros fisiológicos y de laboratorio que son empleados usualmente para caracterizar la gravedad del IAM, entre ellas se incluyen:

- Frecuencia cardíaca y presión arterial media.
- Marcadores del equilibrio ácido-base, como el bicarbonato y el anion gap.
- Parámetros de oxigenación, como la presión parcial de oxígeno (PaO_2) y la saturación periférica de oxígeno (SpO_2).
- Elevación de lactato sérico o mediciones de lactato, como indicador de hipoperfusión.



La mortalidad de esta patología suele elevarse cuando coexiste con otras enfermedades como IC, disfunción renal y arritmias [4][5]. Por ello, el uso de tanto de variables fisiológicas como de laboratorio permite capturar de mejor forma el riesgo de muerte de los pacientes.

2.3.3 Arritmias

Las arritmias, especialmente la fibrilación auricular (FA) y la taquicardia ventricular (TV) se consideran patologías relevantes que afectan al pronóstico de mortalidad y evidencian una correlación significativa con el riesgo de muerte en pacientes hospitalizados [5]. La FA es la arritmia más frecuente, afectando a cerca de 50 millones de personas en el mundo el año 2020, y se caracteriza por alterar la actividad eléctrica y mecánica de las aurículas. Los pacientes pueden presentar síntomas como fatiga, mareos o palpitaciones [15].

En la TV el corazón late de forma rápida y sostenida, lo que puede provocar un deterioro hemodinámico significativo e incluso evolucionar a arritmias ventriculares más graves, como la

fibrilación ventricular. La FA por su lado, se asocia a la formación de trombos y a un mayor riesgo de eventos embólicos como accidentes cerebrovasculares (ACV) [2][3]. Entre las variables más relevantes asociadas a las arritmias se encuentran:

- Frecuencia ventricular.
- Uso de fármacos antiarrítmicos o betabloqueadores.
- Variabilidad de la frecuencia cardíaca (HRV).

En el contexto de modelos predictivos, se pueden combinar variables derivadas del ECG con parámetros clínicos como la frecuencia cardíaca, la presión arterial y el uso de medicamentos, lo que contribuye a una mejora en la sensibilidad y especificidad en la detección de riesgo.

2.3.4 Cardiopatía isquémica

La cardiopatía isquémica corresponde a un conjunto de enfermedades originadas por una obstrucción de las arterias coronarias por aterosclerosis o por déficit crónico o agudo de la irrigación miocárdica. Su diagnóstico es útil para identificar daños miocárdicos previos, angina crónica e incluso revascularización coronaria en los pacientes. La cardiopatía isquémica es la principal causa de muerte a escala mundial, siendo responsable del 13% de todas las muertes, y ha presentado un aumento significativo del número de decesos entre el año 2000 y 2021, pasando de 2,7 a 9,1 millones, respectivamente [18].

Algunos factores de riesgo de esta patología pueden ser la edad, el tabaquismo, la hipertensión arterial, la diabetes, la obesidad y los niveles altos de colesterol. Esta condición puede causar dolor torácico, fatiga, o incluso infartos al miocardio, además existen pacientes con isquemia silenciosa, por lo que pueden no presentar síntomas evidentes [19]. En el contexto clínico, entre algunas de las características fisiológicas relevantes se encuentran:

- Presión arterial sistólica y frecuencia respiratoria.
- Parámetros de oxigenación y del equilibrio ácido-base, como la relación $\text{PaO}_2/\text{FiO}_2$ y el pH arterial.
- Niveles de lactato y bicarbonato sérico.

Estas variables permiten detectar la disfunción hemodinámica, el compromiso respiratorio y la presencia de hipoperfusión tisular, lo que se relaciona con el riesgo de mortalidad o complicaciones en pacientes con cardiopatía isquémica hospitalizados [4].

2.3.5 Hipertensión arterial severa

La hipertensión arterial (HTA) se caracteriza por elevar los niveles de presión arterial, donde la presión arterial sistólica es mayor o igual a 140 mmHg y la presión arterial diastólica es mayor o igual a 90 mmHg [20]. Las formas más severas de HTA se asocian a un mayor riesgo de eventos cardiovasculares o de falla multiorgánica. Esta enfermedad es uno de los factores de riesgo modificables más importantes para las ECV, y se estima que a está presente en 1 de cada 3 enfermedades cardiovasculares [13].

Además, la HTA representa una de las mayores cargas de morbilidad y mortalidad a nivel mundial. Este factor de riesgo es responsable de aproximadamente 10,4 millones de muertes al año y algunos factores de riesgo pueden ser: Obesidad, sedentarismo, ingesta excesiva de sodio, estrés crónico, el parto prematuro o el bajo peso al nacer, e incluso la exposición a la polución ambiental y contaminación acústica [20].

En la base de datos MIMIC-III la HTA se asocia a mayores niveles de creatinina, mayor riesgo de hemorragia intracraneal, edema pulmonar y uso de fármacos intravenosos para controlar la presión arterial [2][3][5]. Entre las variables representativas en este contexto se encuentran:

- Presión arterial media y sistólica.
- Niveles séricos de creatinina y sodio.
- Frecuencia cardíaca y necesidad de fármacos antihipertensivos.

2.4. Variables Fisiológicas Integradas en estudios previos

En los modelos de ML usados para la predicción de mortalidad en pacientes con ECV hospitalizados en UCI, las variables fisiológicas representan mediciones cuantificables del estado clínico del paciente, lo que permite a algoritmos de ML detectar patrones complejos que no siempre son evidentes para el humano. Estas variables no solo sirven para caracterizar la gravedad de la enfermedad en un momento dado, sino que también para anticipar eventos como paros cardíacos,

fallas multiorgánicas o incluso la muerte. A continuación, se presenta la Tabla 2.1 con las variables fisiológicas claves a considerar:

Tabla 2.1 Variables fisiológicas integradas en los modelos de Machine Learning de estudios previos.

Clasificación	Interpretación Clínica
Presión Arterial Sistólica (SAS)	Indicador de perfusión sistémica.
Frecuencia Cardíaca (HR)	Refleja compensación hemodinámica o arritmias.
Creatinina sérica	Marcador de función renal.
Urea	Indica hipoperfusión o catabolismo, complementa función renal.
Sodio sérico	Hiponatremia causa activación neurohormonal o uso de diuréticos.
Potasio sérico	Riesgo de arritmias, indica uso de diuréticos o IECA.
Lactato sérico	Hipoperfusión y metabolismo anaeróbico. Predictor de mortalidad.
Anión GAP	Indicador de acidosis metabólica. Se eleva en sepsis o hipoxia.
Temperatura corporal	Hipotermia puede causar falla autonómica y la fiebre infección.
Saturación de oxígeno (SpO ₂)	Hipoxia, congestión pulmonar y edemas agudos.
Diuresis	Perfusión renal y signo de disfunción multiorgánica.
Glasgow Coma Scale (GCS)	★ Evalúa el estado neurológico.

Fuente: Elaboración propia. Información obtenida de [4] [3] [8].

2.5. Base de Datos

La base de datos Medical Information Mart for Intensive Care III o MIMIC III, es un repositorio clínico usado para la investigación médica, el desarrollo de inteligencia artificial aplicada al campo de la salud y el análisis de datos clínicos. Esta base de datos fue desarrollada por el MIT Laboratory for Computational Physiology junto con el BIDMC (Beth Israel Deaconess Medical Center), y su principal objetivo es otorgar un acceso libre a datos clínicos anonimizados para realizar investigaciones que puedan tener un impacto positivo en el área de la salud [6][21]. Sin embargo, para obtener un acceso completo se requiere una certificación de ética y un registro del investigador, que deben ser aprobados por la plataforma PhysioNet [22].

MIMIC-III es una base de datos libre que contiene información estructurada y detallada de 46.520 pacientes hospitalizados que hayan ingresado a la Unidad de Cuidados Intensivos (UCI) entre el año 2001 y 2012. La base de datos está compuesta por 26 tablas interconectadas, que integran datos demográficos, administrativos, de laboratorio, diagnósticos, procedimientos, prescripciones y registros de pacientes en la UCI [6][22].

En total, la base de datos registra 651.047 diagnósticos, de los cuales 139.423 corresponden a ECV, lo que respalda y confirma la relevancia de estas patologías en la población clínica incluida en el proyecto. Para el manejo del repositorio existen herramientas como dBeaver, que permiten la exploración visual de las tablas y la obtención de cohortes de datos estructurados, lo que facilita el uso para modelos de predicción al optimizar procesos de extracción y organización de datos.

Todo esto permite un mejor desarrollo de modelos para la predicción de mortalidad y facilita la comparación y el análisis de patrones en pacientes críticos, resultando esencial en muchos proyectos de investigación que buscan aplicar herramientas de inteligencia artificial en contextos clínicos reales [2][3][6].

Capítulo 3. Desarrollo técnico

Este capítulo detalla los procedimientos, las herramientas y las estrategias utilizadas para implementar y validar los modelos de predicción de mortalidad en pacientes hospitalizados por enfermedades cardíacas. Se abordan decisiones técnicas tomadas tanto para el procesamiento de datos como para la configuración del entorno de trabajo y el diseño de los modelos de ML, estableciendo el flujo completo desde la extracción de datos, hasta la evaluación final de los modelos.

3.1. Estructura y exploración de MIMIC-III v1.4

La información de la base de datos MIMIC-III v1.4 registra 46.520 pacientes, de los cuales 26.121 son hombres y 20.399 son mujeres, los cuales tienen en promedio 53 años. A nivel de diagnósticos, la base contiene 651.047 registros ICD-9, de los cuales 139.423 (21,42%) corresponden a alguna ECV, lo que refleja la alta frecuencia de estas patologías en pacientes críticos. Además, del total de pacientes, 32.457 presentó al menos un diagnóstico ECV, lo que representa un 69,77% del total de pacientes.

Durante la etapa de exploración se realizaron consultas en SQL que permitieron caracterizar la población y definir distintas cohortes analíticas basadas en el contexto hospitalario y de UCI. Estos conteos de exploración son coherentes con lo reportado en la literatura, donde las enfermedades cardiovasculares constituyen una de las principales causas de hospitalización y mortalidad en unidades críticas. Además, estos resultados exploratorios sirven como base para la construcción de las cohortes

específicas empleadas en el pipeline de Python. El resumen de conteos de diagnósticos obtenidos se puede observar en la Tabla 3.1 y los de mortalidad en la Tabla 3.2, presentadas a continuación.

Tabla 3.1. Diagnósticos cardiovasculares (ICD-9) más predominantes en la cohorte de pacientes con ECV de MIMIC-III, ordenadas según su frecuencia.

Diagnóstico ECV	Frecuencia del diagnóstico
Otros trastornos cardiovasculares	50.254
Hipertensión esencial (primaria)	21.329
Otras arritmias	20.998
Insuficiencia cardíaca	20.676
Enfermedad renal hipertensiva	6.440
Infarto agudo de miocardio (IAM)	5.496
Enfermedad cerebrovascular	4.895
Infarto antiguo (IAM previo)	3.278
Aterosclerosis	2.135
Trombosis venosa profunda	1.725
Trombosis cerebral / Oclusión	1.480
Embolia arterial	1.528
Cardiopatía hipertensiva	1.107
Cardiopatía + enfermedad renal hipertensiva	82

Fuente: Elaboración propia.

Tabla 3.2. Resume los conteos de mortalidad obtenidos de MIMIC-III v1.4, incluyendo el contexto global por admisión, el contexto de UCI sin ECV y el contexto de pacientes UCI con ECV. Lo que permite comparar las cohortes.

Contexto	Criterio de mortalidad	Número pacientes	Muertes	Sobrevivientes
Sin contexto de UCI ni de ECV.	Global por admisión	46.520	5.813	40.707
	En la primera admisión	46.520	4.393	42.127
Contexto UCI sin ECV	Cualquier estadía en UCI	15.052	551	14.501
Contexto UCI por ECV	Primera estadía UCI	32.401	2.734	29.667
	Cualquier estadía en UCI	32.401	3.961	28.440

Fuente: Elaboración propia.

3.2. Metodología de trabajo

Para el desarrollo e implementación del proyecto se utilizó una metodología que consiste en dos etapas principales. La primera corresponde a una etapa de exploración y preparación en SQL, donde se utilizó una herramienta de interfaz gráfica llamada dBeaver para conectarse a la base de datos. En este software se validó la estructura original de las tablas, se construyeron queries de conteo y exploración y también se establecieron distintas cohortes con ciertos criterios de inclusión como la edad y el tiempo de estadía en UCI. Además, a partir de los códigos ICD-9 se definió la cohorte de diagnósticos cardiovasculares de interés, lo que permitió generar tablas intermedias, filtrar ECV y agrupar variables clínicas que posteriormente se usarían en el desarrollo del pipeline en Python.

La segunda etapa consistió en el desarrollo del código usando Python 3.12.7 en el entorno de desarrollo Spyder 6.0.7 de Anaconda. Se utilizaron librerías como sqlalchemy y pandas para extraer las tablas donde se aplicó el preprocesamiento correspondiente, seguido de la construcción del dataset analítico y la selección de las variables clínicas más relevantes para el problema. Sobre este dataset se implementaron los modelos de ML, junto con el proceso de entrenamiento y evaluación mediante las métricas estándar.

Como complemento al enfoque original, el pipeline fue extendido mediante una estructura multi-seed, en la cual el proceso de partición, entrenamiento y evaluación fue repetido con distintas semillas. Esto permitió analizar la estabilidad del rendimiento de los modelos frente a la variabilidad generada por el muestreo de los datos, lo que permitió obtener un modelo más estable. Adicionalmente, se realizó undersampling parcial controlado en el set de datos para poder evaluar la respuesta de los modelos frente a una distribución de clases más equilibrada, por lo que se refuerza la robustez e interpretación de los resultados obtenidos.

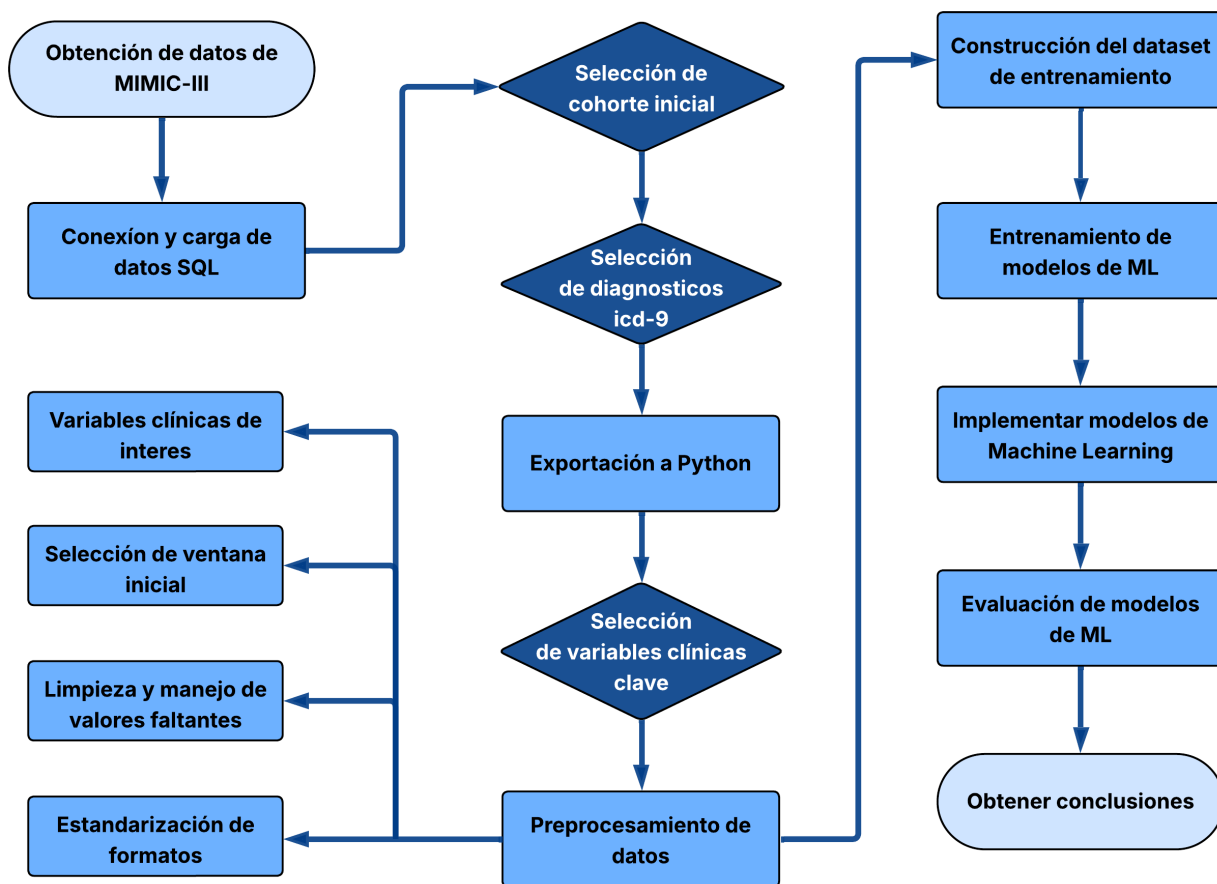


Figura 3.1 Diagrama de flujo de la metodología de trabajo para la implementación de los modelos de ML en la predicción de la mortalidad mediante datos extraídos de MIMIC-III v1.4.

3.2.1 Variables clínicas seleccionadas para el proyecto

Las variables clínicas consideradas en este estudio fueron seleccionadas en base al estudio bibliográfico de investigaciones previas [4][3][8] y de la disponibilidad de estos parámetros en la base de datos MIMIC-III. Todas ellas se extrajeron de las tablas originales considerando las primeras 24 horas de estadía en UCI de cada paciente. El conjunto incluye datos demográficos, signos vitales, resultados de laboratorio y algunas variables relacionadas con intervenciones o medidas clínicas, priorizando aquellas con mayor respaldo fisiopatológico en relación con las ECV incluidas en la cohorte.

A. Datos demográficos y antecedentes

Las variables demográficas constituyen el primer paso para evaluar el riesgo de mortalidad en UCI. En este proyecto se consideran la edad, el sexo y también el tipo de admisión hospitalaria (urgencia, consulta, etc.).

La edad se asocia directamente con mayor mortalidad y menor reserva fisiológica [23]. El sexo introduce diferencias fisiopatológicas, y en la respuesta a terapias; los hombres presentan un mayor riesgo de eventos coronarios, mientras que en las mujeres se observa una mayor prevalencia de IC [24]. Por su parte, el tipo de admisión, en especial las admisiones de urgencia suelen reflejar cuadros más agudos, por lo que se relaciona con un mayor riesgo de eventos adversos durante la hospitalización.

Cabe destacar, que otras variables demográficas y antecedentes como el índice de masa corporal o comorbilidades también influyen en el riesgo de cada paciente. No obstante, no fueron integradas de forma explícita en el desarrollo de los modelos dado a su poca disponibilidad en MIMIC-III, por lo que su rol se considera a nivel conceptual.

B. Signos vitales al ingresar a UCI

Los signos vitales son los predictores directos del riesgo de muerte en pacientes críticos con ECV, dentro de estos se incluyen parámetros como la presión arterial media (PAM), frecuencia cardíaca, frecuencia respiratoria, saturación de oxígeno periférica y temperatura durante las primeras 24 horas.

Por un lado, la PAM es un marcador hemodinámico central, la hipotensión sostenida se asocia con una mala perfusión, y por ende, a una mayor mortalidad [25]; las alteraciones del ritmo cardíaco como la taquicardia, son reflejo de inestabilidad hemodinámica frecuente en ECV [26]. Por otro lado, una frecuencia respiratoria elevada junto con la hipoxemia (SpO_2 disminuida) se relacionan con insuficiencia respiratoria, congestión pulmonar y mayor riesgo de eventos cardiovasculares adversos. Las alteraciones de la temperatura normal, tanto la hipotermia como la hipertermia, se vinculan con infección o disfunción sistémica, por lo que suelen indicar un mal pronóstico [27].

C. Resultados de laboratorio

Los biomarcadores de laboratorio obtenidos durante las primeras 24 horas entregan información directa sobre la perfusión, función orgánica y también del estado inflamatorio del paciente, estos aspectos se asocian estrechamente con la mortalidad por ECV. Entre los principales parámetros incluidos se encuentra el lactato sérico, marcador de hipoperfusión e indicador de mortalidad [8]; los electrolitos como el sodio y el potasio, cuyas alteraciones se relacionan con arritmias y descompensaciones hemodinámicas [4][3]; y los parámetros del hemograma de cada paciente, que han sido vinculados con síndromes coronarios agudos y con un mayor riesgo de muerte en general [8][3].

También se integraron indicadores de función renal como la urea y la creatinina, que representan tanto comorbilidades previas o deterioro del estado de salud del paciente; por último, también se incluyen perfiles de bioquímica hepática como la bilirrubina, que pueden sugerir congestión venosa o hipoperfusión sistémica [4][8]. Estas variables son respaldadas en múltiples estudios basados en MIMIC-III y en cohortes cardiovasculares, y se pueden observar sus interpretaciones clínicas en la Tabla 0-1 del Anexo B del proyecto.

3.3. Desarrollo del pipeline en Python

En esta sección se presenta el diseño y la implementación del pipeline de Python utilizado para realizar las predicciones de mortalidad en los pacientes a partir de MIMIC-III. El enfoque del desarrollo apunta a estructurar un código que permita procesar los datos clínicos, para luego entrenar modelos clasificadores de ML.

3.3.1 Estructura General

El pipeline fue desarrollado completamente en Python, donde se estructuró su flujo en etapas secuenciales y bloques que permiten mantener claridad en el proceso. El entorno de trabajo se conectó directamente al servidor donde se alojó la base de datos mediante scripts que permiten la integración de nuevas funciones según sea necesario.

El flujo general del pipeline consiste en cinco grandes bloques principales:

1. Conexión y extracción de datos mediante consultas SQL ejecutadas sobre PostgreSQL.
2. Preprocesamiento y limpieza de los registros clínicos obtenidos.

3. Integración de las variables clínicas seleccionadas para entrenar los modelos, que corresponden a las primeras 24 horas de estancia en UCI.
4. Ingeniería de características, balanceo de clases y partición del dataset.
5. Entrenamiento y evaluación de los modelos de ML.

3.3.2 Carga y procesamiento de datos

La construcción del dataset se realizó a partir de la base de datos MIMIC-III mediante consultas SQL ejecutadas con Python utilizando SQLAlchemy. Para esto se estableció una conexión con el servidor de PostgreSQL y se configuró la ruta según los esquemas utilizados en dBeaver, lo que permitió hacer las consultas necesarias en las tablas estructurales de la base de datos.

La cohorte final se construyó a partir de una serie de Common Table Expressions (CTE) que fueron encadenándose. En primer lugar, se identificaron las hospitalizaciones con diagnósticos cardiovasculares según los códigos ICD-9 de MIMIC. Luego, estas admisiones fueron vinculadas con las estancias en UCI, restringiendo el análisis a pacientes adultos (>18 años) y a una ventana de observación de 24 horas a partir del ingreso a UCI. A partir de esta ventana se extrajeron las mediciones clínicas necesarias para luego guardar los resultados en dataframes mediante la librería pandas, donde cada fila representó una estancia única en UCI.

Posteriormente, se integraron las mediciones de laboratorio y los parámetros fisiológicos recopilados durante la ventana de 24 horas. Para cada variable se obtuvo información relevante, como el primer valor, o el valor máximo o mínimo según la variable. Para el procesamiento de los datos, estos fueron estandarizados en tipo y formato, y se incluyeron indicadores de presencia para las variables clínicas cuya ausencia puede reflejar decisiones médicas o severidad del paciente. Estos indicadores se usaron para analizar valores faltantes o poco relevantes, y también como variables predictoras adicionales, ya que mediciones como el lactato, de gases o del equilibrio ácido-base, no son exámenes que se requieran en casos menos severos.

Como resultado, se obtuvo un dataframe final que contiene las variables demográficas, administrativas y clínicas esenciales asociadas a cada estancia en UCI, todas obtenidas durante las primeras 24 horas. El tratamiento a detalle de los valores faltantes, selección final de variables y construcción del dataset definitivo se describen en secciones siguientes.

3.3.3 Construcción del dataset final

Tras la integración de los datos, se procedió a la construcción del dataset utilizado en el modelado. Primero, se calculó el porcentaje de valores faltantes por variable, lo que permitió identificar aquellas con altos niveles de ausencia de datos, o bien, con baja relevancia clínica en el contexto de ECV. Ciertas variables como la diuresis, albuminas, bilirrubinas y creatina quinasa fueron excluidas dado su escasa disponibilidad y relevancia. Luego, para el resto de las variables numéricas, se aplicó imputación por mediana, la cual fue calculada únicamente con el conjunto de entrenamiento, mientras que las variables categóricas (sexo, tipo de ingreso, etc.) y los indicadores de presencia se mantuvieron sin imputar.

Posteriormente, se definió la variable objetivo (`icu_expire_flag`) para representar la mortalidad durante la estancia en UCI y se eliminaron los identificadores administrativos. Finalmente, el dataframe definitivo quedó compuesto por 7.817 estancias de pacientes adultos con ECV, de las cuales 829 corresponden a fallecimientos, lo que refleja un problema de clasificación binaria con desbalance de clases. Este dataset sirvió como base para las etapas posteriores de partición estratificada, entrenamiento y optimización de los modelos de ML a desarrollar.

3.3.4 Partición de datos y manejo del desbalance

Con el dataset ya definido, se realizó la partición de los datos en conjuntos de entrenamiento y prueba mediante un train-test split estratificado con proporción 80/20, lo que deja 6.253 casos para entrenamiento y 1564 para prueba. La estratificación permitió mantener la proporción de la variable objetivo en ambos conjuntos, permitiendo aliviar en cierta medida el efecto del desbalance de la cohorte (10,6% fallecidos). Estas medidas permiten evaluaciones más estables y comparables entre modelos.

Para manejar el desbalance de clases presente en la cohorte, se utilizaron los mecanismos internos de ponderación de clases disponibles en los modelos implementados. Estas funciones ajustan automáticamente la importancia asignada a la clase minoritaria, reforzando el peso de los casos de muerte, lo que compensa su baja frecuencia y evita que el modelo favorezca en exceso a la clase mayoritaria, lo que permite mejorar los resultados sin necesidad de modificar la distribución de datos.

Por otro lado, el conjunto de entrenamiento se utilizó tanto para entrenar los modelos como para optimizarlos, mientras que el conjunto de prueba se empleó solamente para la evaluación final de los modelos, evitando fugas de información durante el ajuste de hiperparámetros.

3.3.5 Entrenamiento y evaluación de los modelos de ML

Luego de realizar la partición estratificada, se procedió con el entrenamiento de los modelos de ML seleccionados: Random Forest, XGBoost, LightGBM y CatBoost. Estos modelos son ampliamente usados en problemas del ámbito clínico debido a su capacidad de manejar relaciones no lineales, variables heterogéneas y estructuras de datos complejas. Los modelos fueron configurados según nuestro dataset final para manejar de forma adecuada los valores faltantes, las variables heterogéneas, y en general para obtener un mejor rendimiento.

El proceso de entrenamiento se llevó a cabo usando exclusivamente el conjunto de entrenamiento y los modelos utilizados fueron evaluados primero con sus configuraciones iniciales. Posteriormente, se optimizaron mediante la búsqueda de hiperparámetros de la librería Optuna, con el fin de maximizar su rendimiento y adaptarlo al proyecto. La optimización se realizó con validación cruzada estratificada, evaluando múltiples configuraciones y seleccionando aquellas que maximizaban el rendimiento de cada algoritmo, permitiendo reducir riesgos como el sobreajuste o la sensibilidad al ruido.

Luego de encontrar los hiperparámetros óptimos, los modelos fueron reentrenados para ser evaluados de nuevo. La evaluación se realizó utilizando métricas estándar de clasificación binaria tales como: AUC ROC, precisión, sensibilidad y F1-Score. Además, se realizó un análisis específico para poder encontrar el umbral de decisión óptimo para cada modelo, priorizando el equilibrio entre sensibilidad y precisión. Los resultados obtenidos se detallan en la sección siguiente.

Capítulo 4. Resultados

4.1. Introducción

En este capítulo se presentan los resultados obtenidos tras los procesos de entrenamiento y evaluación de los modelos de ML desarrollados para el proyecto. Se incluyen métricas de discriminación, análisis de importancia de variables y el desempeño según distintos umbrales de

decisión. Esto con la finalidad de caracterizar el rendimiento de los modelos y evaluar su potencial para aplicarlo en contextos clínicos reales.

4.2. Descripción de la cohorte

Tras aplicar una serie de filtros definidos para el proyecto, se obtuvo una cohorte final constituida por 7.817 estancias en UCI de pacientes adultos con al menos un diagnóstico cardiovascular identificado con los códigos ICD-9, y que cumplan con tener la información mínima necesaria para poder hacer el análisis. Para cada paciente se consideró exclusivamente la información obtenida durante las primeras 24 horas posteriores al ingreso en UCI, con tal de obtener el estado clínico inicial del paciente y mejorar la toma de decisiones rápida. Dentro de la cohorte, 829 pacientes (10,6%) fallecieron durante su estadía en UCI, mientras que 6.988 (89,4%) sobrevivieron, lo que deja un problema de clasificación binaria con desbalance entre las clases.

El dataset final incluyó también información demográfica, datos administrativos, y un conjunto amplio de parámetros clínicos, incluyendo perfiles hematológicos, electrolitos, biomarcadores metabólicos, entre otros (ver Tabla 0.1 en Anexo B). Para cada medición se extrajeron los valores clínicamente más relevantes dentro de la ventana de análisis, tales como valor inicial y valores máximos o mínimos, según la naturaleza de la variable. Finalmente, tras la eliminación de columnas con alta proporción de valores faltantes, el conjunto final para el modelado quedó conformado por un total de 51 variables predictoras, que constituyen la base para el análisis comparativo del rendimiento presentado por los modelos de ML.

4.3. Evaluación de los modelos

El rendimiento de los modelos entrenados se evaluó utilizando exclusivamente el conjunto de prueba. En términos generales, los cuatro algoritmos tuvieron desempeños similares en cuando a AUC ROC, con valores cercanos a 0,80, lo que indica una buena capacidad para discriminar de forma temprana aquellos pacientes que fallecen y los que sobreviven. Sin embargo, al analizar las otras métricas, como sensibilidad y precisión, se observaron diferencias importantes entre el desempeño de los modelos.

Como referencia metodológica o baseline, se incluyó Random Forest sin hacerle un proceso de optimización intensivo como a los otros modelos, esto con el fin de establecer un punto de

comparación base. Pese a esto, alcanzó un AUC-ROC bastante comparable con los modelos basados en boosting, sin embargo, obtuvo una sensibilidad muy baja, no logrando identificar los pacientes que fallecieron. Esto lo posicionó como el modelo con menor rendimiento en un contexto clínico donde omitir casos de alto riesgo es más crítico que obtener falsos positivos.

Para el caso de los modelos basados en boosting como LightGBM, CatBoost y XGBoost, mostraron un rendimiento evidentemente superior a RF. Por un lado, XGBoost fue el modelo con el mejor equilibrio entre recall y precisión, y además también obtuvo valores de AUC-ROC cercanos al 80%, sosteniendo una buena capacidad para discriminar, lo que lo convierte en la opción más estable de los modelos evaluados.

Por otro lado, LightGBM y CatBoost fueron los más destacados en sensibilidad, logrando identificar una mayor proporción de pacientes que fallecieron durante su estancia en UCI. Este rendimiento se obtuvo tras la optimización de hiperparámetros y el ajuste de los umbrales de decisión, permitiendo hacer énfasis en la detección de los fallecidos sin sacrificar de gran manera la precisión del modelo. Pese a esto, el aumento en la sensibilidad implicó un incremento de los falsos positivos, reduciendo los valores de precisión. No obstante, debemos recordar que para el contexto clínico del proyecto, los falsos negativos pueden ser considerados más costosos que los falsos positivos, por lo que se asume esa pequeña pérdida de precisión.

4.3.1 Consideraciones importantes

Para la interpretación del desempeño de los modelos no solo se deben considerar las métricas globales, sino también la distribución de los errores y el efecto del umbral de decisión, el cual fue definido de manera específica para cada modelo. Debido a la baja prevalencia de la variable objetivo, el valor estándar del umbral de decisión de 0,50 no resulta adecuado para obtener un óptimo desempeño del modelo en la detección de pacientes con alto riesgo de muerte. Por esto, se evaluó el comportamiento de los modelos basados en boosting frente a distintos umbrales de decisión, seleccionando aquellos que mejoraran la sensibilidad sin afectar en exceso a la precisión. Esto permitió definir un umbral de 0,40 para XGBoost y CatBoost, y de 0,45 para LightGBM. Para el caso de RF, y al ser el modelo baseline, se mantuvo el valor estándar del threshold debido a su capacidad limitada para obtener buenos resultados en sensibilidad.

Para complementar la idea de la interpretación del desempeño, también se obtuvieron matrices de confusión, las que permiten visualizar el tipo de errores que cometen los modelos y su relevancia clínica. Estos resultados mostraron que XGBoost logró un equilibrio razonable entre los verdaderos positivos y los falsos positivos. Por otro lado, LightGBM y CatBoost alcanzaron sensibilidades más altas, logrando identificar en mayor grado a los pacientes que fallecieron durante su estancia en UCI. A raíz de este aumento en sensibilidad, se notó un leve incremento en los falsos positivos, pero como se mencionó antes, en un contexto clínico, un falso negativo representa un riesgo bastante crítico en comparación con los falsos positivos, especialmente si se trata de decisiones que se deben tomar de forma temprana.

Estas consideraciones, tanto el umbral de decisión como las matrices de confusión, nos entregan una visión más completa del rendimiento de los modelos, lo que permitió seleccionar un enfoque más adecuado y personalizado a las condiciones del problema, complementando así el análisis del rendimiento basado en las métricas globales.

4.3.2 Métricas globales por modelo

Las métricas de evaluación calculadas fueron: AUC-ROC, exactitud, precisión, sensibilidad y F1-Score. Adicional a esto, se incluyeron las matrices de confusión correspondientes para poder caracterizar la distribución de los errores y aciertos de cada modelo.

En términos generales, los modelos basados en boosting presentaron desempeños muy superiores al baseline (RF), obteniendo valores de AUC-ROC cercanos a 0,80 y métricas más coherentes con el balance entre falsos positivos y falsos negativos. Por un lado, XGBoost alcanzó el F1-score más alto, lo que refleja un equilibrio moderado entre precisión y sensibilidad, mientras que LightGBM y Catboost lograron las sensibilidades más altas, con valores cercanos a 0,73, esto indica que identificaron la mayor proporción de pacientes fallecidos en el conjunto de pruebas. Por último, RF fue el que obtuvo el accuracy más alto de todas, pero esto con una sensibilidad muy reducida, lo que refleja una mala capacidad para detectar eventos de mortalidad.

Las matrices de confusión (Figura 4.1) permiten visualizar el comportamiento de los modelos de forma más detallada, permitiendo diferenciar tanto los verdaderos negativos y verdaderos positivos, como los falsos positivos y falsos negativos. Esto permite una evaluación más estructurada que refleja

el tipo de error que predomina en cada modelo. A continuación, se presenta la Tabla 4.1, la cual resume todas las métricas calculadas para poder caracterizar el desempeño de los modelos desarrollados, lo que sirve como base para las interpretaciones clínicas.

Tabla 4.1. Tabla resumen con los resultados obtenidos con los modelos de ML desarrollados en el proyecto.

Clasificación	AUC-ROC	Accuracy	Recall	Precision	F1-Score
XGBoost	0,804	0,804	0,633	0,300	0,407
CatBoost	0,801	0,753	0,747	0,265	0,391
LightGBM	0,807	0,747	0,729	0,257	0,379
Random Forest	0,792	0,889	0,078	0,382	0,130

Fuente: Elaboración propia.

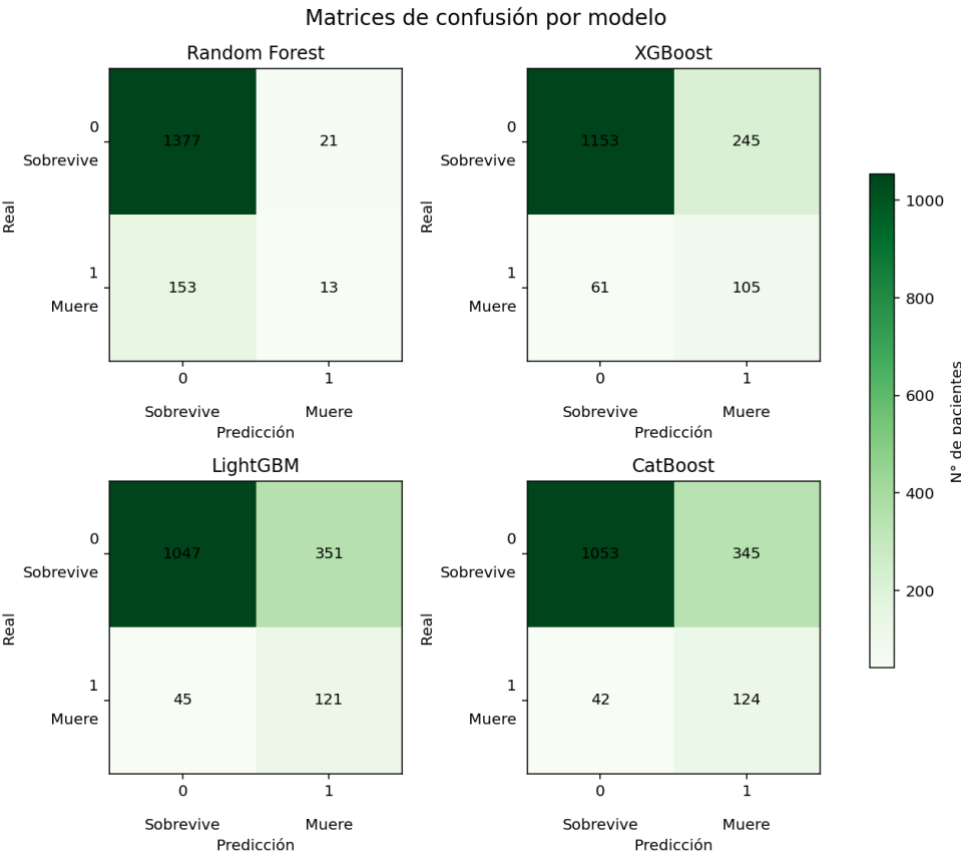


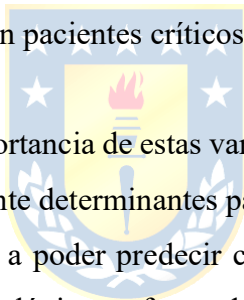
Figura 4.1. Se muestran las matrices de confusión de cada modelo de ML desarrollado en el proyecto.

4.3.3 Importancia de variables

Cabe destacar, que la interpretación de los resultados de los modelos es un aspecto fundamental en tareas clínicas, ya que permite comprender cuales son los parámetros fisiológicos o de laboratorio que más contribuyen a la predicción del riesgo de mortalidad. Para hacer esto, se obtuvo la importancia de variables de los modelos basados en boosting, para luego promediarlas e identificar a los predictores con mayor influencia en el rendimiento general.

Los resultados indican que los predictores más relevantes son aquellos asociados a marcadores de disfunción metabólica, inflamatoria y de alta severidad fisiológica. Entre las variables que más destacan podemos encontrar anion gap, PTT máximo, edad, glucosa máxima, sodio y potasio iniciales, junto con parámetros hematológicos como plaquetas mínimas y los recuentos de WBC inicial y máximo. Estos factores coinciden con la evidencia clínica disponible, donde generalmente, alteraciones en el equilibrio ácido-base, coagulopatías, hiperglicemia y anomalías electrolíticas se han asociado a mayor riesgo de mortalidad en pacientes críticos.

Es importante señalar que la importancia de estas variables debe ser interpretada considerando el contexto clínico, estas no son totalmente determinantes para el fallecimiento del paciente, sino que son las que más contribuyen al modelo a poder predecir correctamente. Sin embargo, los patrones indican una coherencia a nivel fisiopatológico, reforzando su importancia en pacientes con ECV hospitalizados en UCI. En la Figura 4.2 se pueden observar las variables más importantes para la predicción:



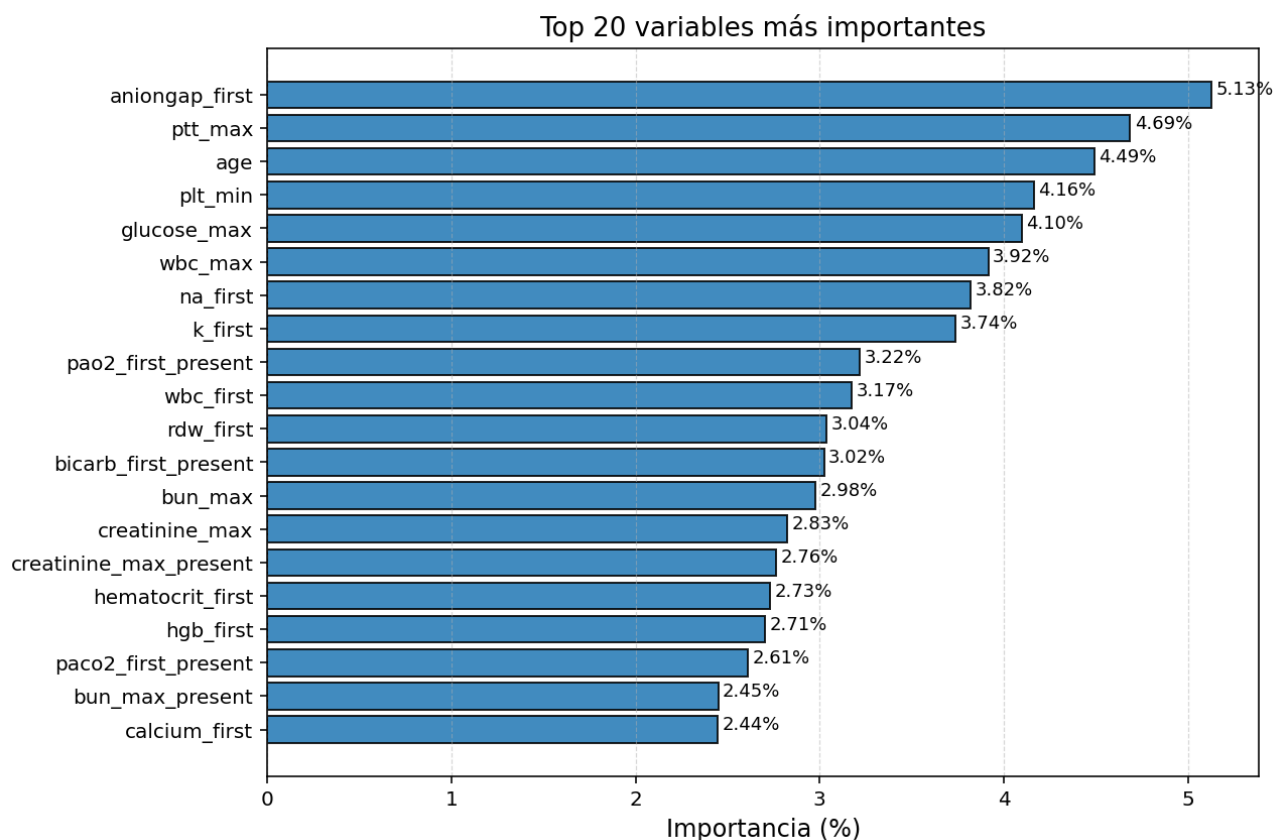


Figura 4.2. Se muestran las 20 variables más importantes para los 3 modelos basados en boosting.

4.4. Análisis complementario

El conjunto de datos utilizado presenta un marcado desbalance entre las clases, con una baja prevalencia de muertes frente a la clase mayoritaria. Esto introduce un problema para el desarrollo de los modelos de ML, ya que los resultados se ven influenciados por la distribución de clases, lo que no permite identificar si las limitaciones son propias de los modelos o son consecuencia de la estructura y la prevalencia en el conjunto de datos. Esta situación motiva la necesidad de complementar el análisis principal con un enfoque que permita evaluar los modelos en distintas condiciones de distribución de clases, con el fin de evaluar la robustez y estabilidad del pipeline.

Además, se complementó el pipeline aplicando un enfoque basado en una estructura multi-seed que permite analizar la estabilidad de los resultados y reducir el factor de aleatoriedad introducido al realizar la partición de los datos. Esto permite tener una visión más robusta del desempeño de los

modelos, y también evaluar la variabilidad de las métricas obtenidas, lo que reduce la dependencia de un único split.

Junto con lo anterior, se aplicó una técnica de undersampling (NearMiss) sobre el conjunto de datos, lo que compensa de forma controlada el desbalance original y disminuye su impacto en el desempeño de los modelos sin modificar la arquitectura de estos. Finalmente, el dataset después del undersampling quedó conformado por 3.316 casos totales, de los cuales 829 (25%) fallecen y 2.487 (75%) sobreviven. Por otro lado, el conjunto de entrenamiento tiene 2.652 casos y el de prueba 664 casos.

Esto permite evaluar si las limitaciones presentadas responden a los modelos como tal o a restricciones estructurales o de prevalencia de clases, permitiendo fortalecer la interpretación del estudio y aportando evidencia adicional sobre la estabilidad.

4.4.1 Descripción del proceso

El proceso descrito se implementó como una extensión del pipeline, manteniendo sin modificaciones la selección de variables, los modelos entrenados y las métricas de rendimiento. En primer lugar, la estructura se realizó con un enfoque multi-seed, por lo que se repitió el proceso para distintas semillas con el fin de evaluar la variabilidad de los resultados debido a la aleatoriedad al seleccionar los datos.

A partir de lo anterior, se realizó un procedimiento de undersampling al dataset mediante el algoritmo NearMiss, el cual reduce el número de casos de la clase mayoritaria, conservando los casos más parecidos con la clase minoritaria, es decir, los más cercanos al umbral de decisión, por lo que resultan más informativos. Esto evita inventar instancias y modificar la clase minoritaria, preservando un poco más la coherencia clínica de los datos.

El undersampling fue configurado utilizando una proporción de 25/75 entre los casos positivos y negativos, lo que permite reducir el impacto del alto desbalance del dataset original conservando una estructura de datos cercana a los escenarios clínicos. Posteriormente, esto se integró al esquema multi-seed descrito antes, asegurando también considerar la variabilidad de resultados.

4.4.2 Resultados e interpretación

Los resultados obtenidos mediante este enfoque multi-seed con undersampling parcial mostraron un comportamiento bastante consistente y estable frente a las variaciones de los datos. Como se observa en la Tabla 4.2, el rendimiento de los modelos no depende en gran medida de una sola partición, donde se obtienen valores de AUC-ROC, F1-Score y recall bastante comparables entre las semillas utilizadas.

Tabla 4.2. Se muestran los resultados desglosados obtenidos por los modelos en cada seed utilizada.

Seed	Model	AUC-ROC	Accuracy	Precision	Recall	F1-score
42	RF	0,91	0,88	0,82	0,67	0,74
42	XGB	0,94	0,89	0,84	0,67	0,75
42	LGBM	0,94	0,87	0,69	0,89	0,78
42	CatBoost	0,95	0,89	0,74	0,86	0,80
10	RF	0,91	0,87	0,75	0,70	0,73
10	XGB	0,93	0,88	0,83	0,65	0,73
10	LGBM	0,93	0,84	0,64	0,88	0,74
10	CatBoost	0,93	0,86	0,67	0,85	0,75
2022	RF	0,89	0,86	0,77	0,67	0,71
2022	XGB	0,91	0,88	0,89	0,61	0,72
2022	LGBM	0,91	0,81	0,59	0,85	0,70
2022	CatBoost	0,92	0,84	0,64	0,83	0,72

Fuente: Elaboración propia

En la Tabla 4.3, se muestra el promedio de los resultados por seed y se observa que las diferencias entre semillas son mínimas, lo que respalda la robustez del pipeline frente a la aleatoriedad de la selección de datos.

Tabla 4.3. Se muestra el promedio de los resultados por seed, para cada modelo desarrollado.

Model	AUC-ROC	Accuracy	Precision	Recall	F1-score
CatBoost	0,93	0,86	0,69	0,85	0,76
LightGBM	0,93	0,84	0,64	0,87	0,73
Random Forest	0,91	0,87	0,78	0,68	0,73
XGBoost	0,93	0,88	0,85	0,65	0,73

Fuente: Elaboración propia

La incorporación del undersampling permitió obtener un mejor equilibrio entre precisión y recall, que era justamente una de las limitaciones del pipeline original, mientras el AUC-ROC se mantuvo en valores altos. Esto se traduce en una excelente capacidad para obtener el riesgo de los pacientes, mientras mejoran otras métricas claves. Por otro lado, el análisis de la importancia de

variables coincide con variables clínicas reconocidas en el contexto de pacientes críticos, tales como gases arteriales, alteraciones ácido-base, marcadores hematológicos y la edad del paciente. Estas variables se pueden observar en la Figura 4.3.

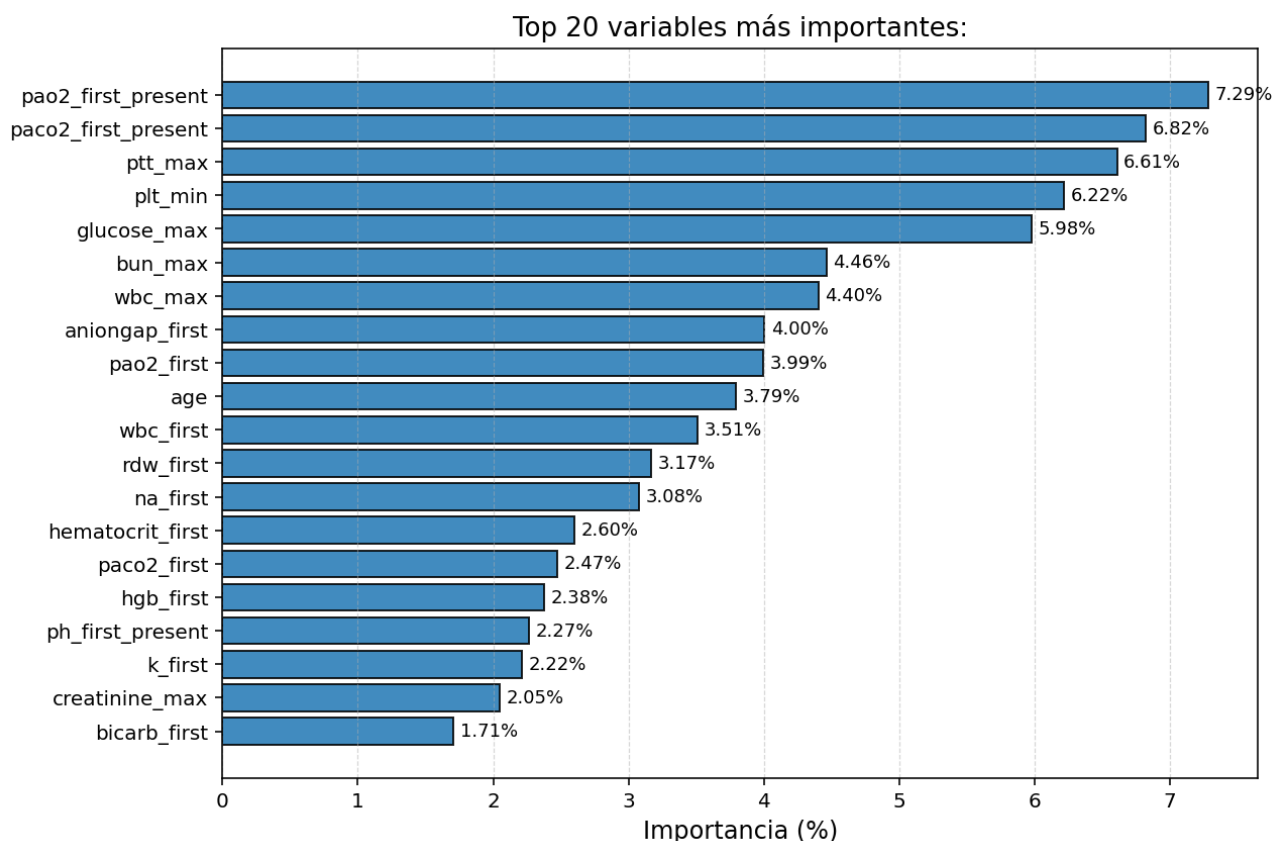


Figura 4.3. Se muestran las 20 variables más importantes para los 3 modelos basados en boosting, utilizando el enfoque multi-seed con undersampling mediante NearMiss.

Los modelos basados en boosting muestran un mejor rendimiento al modelo baseline, donde CatBoost destaca como el mejor desempeño global, alcanzando los valores más altos de AUC-ROC y F1-Score en promedio. Este rendimiento también se puede observar en las matrices de confusión de la semilla con mejor rendimiento (Figura 4.4) donde se observa un menor número de falsos negativos respecto a otros modelos.

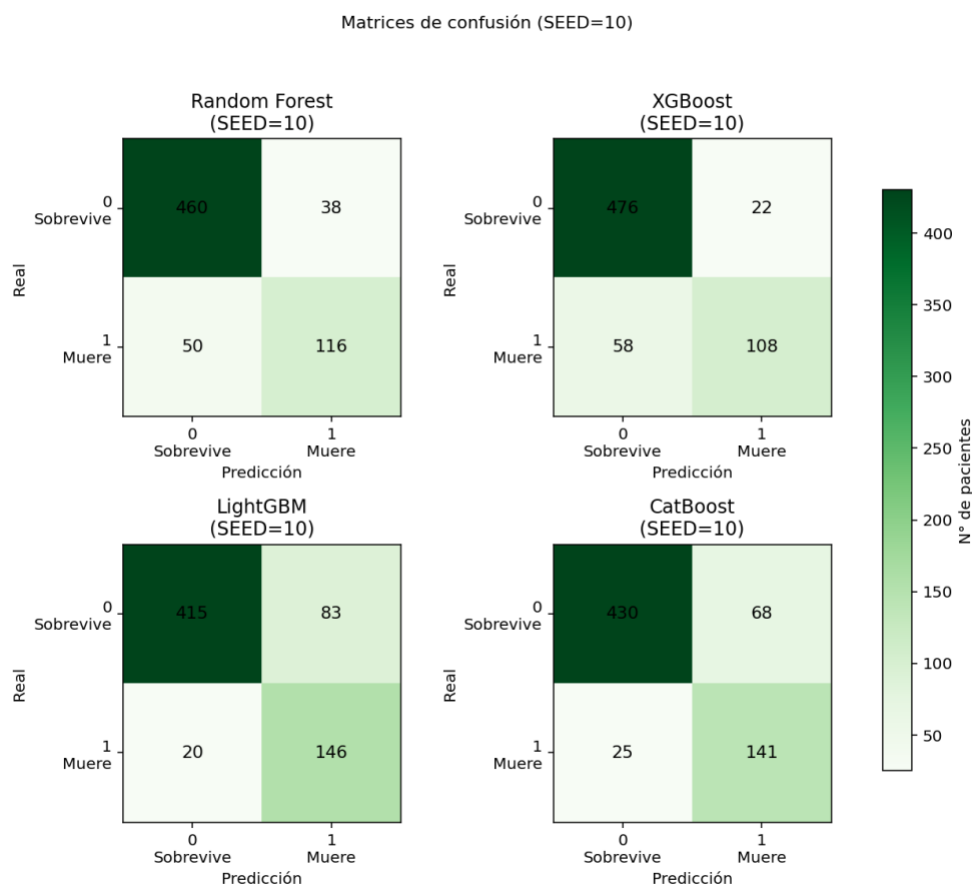


Figura 4.4. Se muestran las matrices de confusión obtenidas de los 4 modelos desarrollados con el enfoque multi-seed, utilizando la seed 10.

Estos resultados indican que una proporción importante de las limitaciones observadas previamente se asocian al desbalance que existía, y al atenuar parcialmente este mediante el undersampling controlado, el pipeline muestra un comportamiento más equilibrado y estable. Lo que refuerza la interpretación de los resultados iniciales y aporta evidencia adicional sobre la robustez del enfoque del proyecto.

4.5. Discusión

Los resultados del estudio indican que los modelos basados en gradient boosting presentan un rendimiento sólido para predecir la mortalidad temprana en pacientes con ECV internados en UCI. Estos resultados son bastante consistentes con los observados en otros estudios basados en MIMIC-III, donde este tipo de modelos ha demostrado una mayor capacidad para trabajar relaciones no lineales y patrones complejos en datos clínicos tan variados. Por otro lado, Random Forest mostró una sensibilidad muy baja, lo que lo posicionó como el modelo menos adecuado en un contexto omitir casos críticos tiene un impacto muy significativo.

En relación con la selección y el preprocesamiento de las variables clínicas, el análisis de importancia de características dejó en evidencia patrones fisiopatológicos coherentes con la literatura, siendo los predictores principales marcadores reconocidos de disfunción orgánica y de gravedad en el estado de los pacientes. La coherencia mostrada entre los resultados de los modelos con la interpretación clínica de las variables refuerza la validez de las variables predictoras obtenidas.

Por otro lado, el ajuste de los umbrales de decisión fue otra modificación clave para optimizar la sensibilidad de los modelos en contextos con desbalance de clases. Esta estrategia permitió mejorar la identificación de los casos de mayor riesgo de muerte, manteniendo un equilibrio aceptable entre sensibilidad y precisión, además de alinearse con el criterio de disminuir falsos negativos.

De forma complementaria, el enfoque de evaluación multi-seed permitió analizar la estabilidad del rendimiento de los modelos frente a diferentes semillas, donde se observó que las métricas de desempeño se mantienen relativamente estables, lo que indica que el rendimiento no es dependiente de una única partición del conjunto de datos, lo que le otorga más robustez al pipeline.

Adicionalmente, la aplicación de undersampling permitió evaluar los modelos bajo escenarios con un desbalance de clases menos severo. Lo que mostró que al reducir este en cierta medida, los modelos logran un desempeño más equilibrado, sugiriendo que parte de las limitaciones observadas se asocian a la estructura del dataset más que a deficiencias de la configuración de los modelos.

Sin embargo, es importante considerar que, si bien los objetivos planteados en esta investigación fueron alcanzados, los resultados obtenidos presentan ciertas limitaciones, ya que estos fueron obtenidos a partir de un único centro hospitalario y considerando solo las primeras 24 horas de estancia en UCI, lo que limita el uso de información. Aun así, los hallazgos confirman la potencial utilidad del aprendizaje automático en contextos clínicos, pudiendo ser una herramienta de apoyo para la obtención temprana del riesgo en pacientes cardiovasculares críticos, lo que permite mejorar la toma de decisiones rápidas en el contexto hospitalario.

Capítulo 5. Conclusiones

5.1. Conclusiones

Este proyecto tuvo como objetivo desarrollar modelos de ML capaces de predecir la mortalidad intrahospitalaria en pacientes adultos con ECV durante su estancia en UCI, esto considerando las variables clínicas obtenidas durante las primeras 24 horas tras su ingreso. Se utilizó la base de datos MIMIC-III y técnicas de aprendizaje automático, donde los resultados demostraron que los modelos basados en gradient boosting, particularmente XGBoost, LightGBM y CatBoost, presentaron un rendimiento superior al modelo baseline Random Forest.

El ajuste de hiperparámetros y la selección de los umbrales de decisión permitieron mejorar en gran medida la sensibilidad de los modelos en un dataset desbalanceado, logrando mejorar su capacidad de identificar los casos con más riesgo de mortalidad. Entre los modelos evaluados, XGBoost fue el que alcanzó el mejor equilibrio entre sensibilidad y precisión, mientras que LightGBM obtuvo la sensibilidad más alta. En términos de AUC-ROC, todos obtuvieron un rendimiento que ronda el 0,80, lo cual respalda su capacidad discriminativa en este contexto clínico.

Los resultados que se observaron en la importancia de variables reflejaron coherencia fisiopatológica de las variables más influyentes, lo que les entrega una confianza adicional a los resultados obtenidos. Adicionalmente, la evaluación bajo el esquema multi-seed y la aplicación de undersampling le otorgó robustez al pipeline y permitió confirmar la estabilidad del rendimiento de los modelos frente a variaciones en la partición de los datos y la distribución de clases.

Finalmente, las conclusiones obtenidas de este proyecto confirman el potencial de los modelos de ML como herramienta de apoyo en la clasificación temprana del riesgo en pacientes críticos con ECV, esto mientras se integren con el juicio médico y los protocolos hospitalarios necesarios.

5.2. Trabajo Futuro

Para finalizar, a partir del proceso completo y de los resultados del proyecto, se logran identificar ciertos enfoques de trabajo que podrían fortalecer o mejorar el alcance de este estudio. Dentro de estos podría incluirse la aplicación de modelos más avanzados diseñados para series

temporales más largas, lo que permitiría capturar la evolución fisiológica completa de los pacientes en lugar de usar una ventana inicial de 24 horas.

Otro enfoque que permitiría un mayor desarrollo del objetivo del proyecto sería el integrar nuevas variables clínicas, tales como tratamientos realizados, comorbilidades, uso de ventilación mecánica, o incluso incluir las escalas fisiológicas típicas como SOFA, SAPS II o APACHE II. El uso combinado de estas herramientas con modelos de ML podría aumentar significativamente la capacidad predictiva de los modelos.

Y como adicional, se podría validar el rendimiento de los modelos aplicándolo a otras cohortes hospitalarias o en distintos centros clínicos para permitir evaluar su robustez y aplicabilidad. Además, podría ampliarse el enfoque hacia más patologías críticas, permitiendo extender aún más su alcance en diversos problemas clínicos.



Capítulo 6. Glosario

APACHE II	: Acute Physiology and Chronic Health Evaluation II
BIDMC	: Beth Israel Deaconess Medical Center
ECV	: Enfermedades Cardiovasculares
FiO ₂	: Fracción Inspirada de Oxígeno
GCS	: Glasgow Coma Scale
HOB	: HRV-Only Baseline
HTA	: Hipertensión Arterial
HR	: Heart Rate (Frecuencia Cardíaca)
HRV	: Heart Rate Variability (Variabilidad de la Frecuencia Cardíaca)
IC	: Insuficiencia Cardíaca
ICD-9	: International Classification of Diseases v9
IAM	: Infarto Agudo de Miocardio
LGB	: Light Gradient Boosting
LASSO	: Least Absolute Shrinkage and Selection Operator
MIMIC-III	: Medical Information Mart for Intensive Care III
ML	: Machine Learning
PaO ₂	: Presión Parcial de Oxígeno
PaO ₂ /FiO ₂	: Relación entre la presión parcial de oxígeno y la fracción inspirada de oxígeno
SAPS II	: Simplified Acute Physiology Score II
SOFA	: Sequential Organ Failure Assessment
SpO ₂	: Saturación Periférica de Oxígeno
UCI	: Unidad de Cuidados Intensivos
AUC-ROC	: Área Bajo la Curva Receiver Operating Characteristic
CTE	: Common Table Expression
F1-Score	: Media armónica entre precisión y sensibilidad
INR	: International Normalized Ratio
PAM	: Presión Arterial Media
OPTUNA	: Framework de optimización de hiperparámetros utilizado en el proyecto
PTT	: Partial Thromboplastin Time
SHAP	: SHapley Additive exPlanations

SQL : Structured Query Language
WBC : White Blood Cell Count
RF : Random Forest
XGBoost : Extreme Gradient Boosting

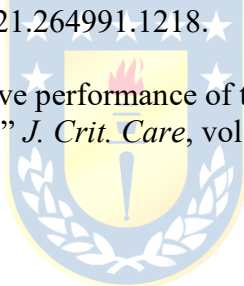


Capítulo 7. Referencias

- [1] OMS, “Enfermedades Cardiovasculares.” Accessed: May 15, 2025. [Online]. Available: <https://www.who.int/es/news-room/fact-sheets/detail/cardiovascular-diseases-%28cvds%29>
- [2] D. Li, J. Fu, J. Zhao, J. Qin, and L. Zhang, “A deep learning system for heart failure mortality prediction,” *PLoS One*, vol. 18, no. 2 February, Feb. 2023, doi: 10.1371/journal.pone.0276835.
- [3] S. Y. Jang *et al.*, “Mortality Prediction in Patients With or Without Heart Failure Using a Machine Learning Model,” *JACC: Advances*, vol. 2, no. 7, Sep. 2023, doi: 10.1016/j.jacadv.2023.100554.
- [4] N. Ashrafi, A. Abdollahi, J. Zhang, and M. Pishgar, “Optimizing Mortality Prediction for ICU Heart Failure Patients: Leveraging XGBoost and Advanced Machine Learning with the MIMIC-III Database,” arXiv preprint arXiv:2409.01685, Sep. 2024.
- [5] L. Huang, Z. Dou, F. Fang, B. Zhou, P. Zhang, and R. Jiang, “Prediction of mortality in intensive care unit with short-term heart rate variability: Machine learning-based analysis of the MIMIC-III database,” *Comput. Biol. Med.*, vol. 186, Mar. 2025, doi: 10.1016/j.combiomed.2024.109635.
- [6] S. Wang, M. B. A. McDermott, G. Chauhan, M. Ghassemi, M. C. Hughes, and T. Naumann, “MIMIC-Extract,” in *ACM CHIL 2020 - Proceedings of the 2020 ACM Conference on Health, Inference, and Learning*, Feb. 2020, pp. 222–235. doi: 10.1145/3368555.3384469.
- [7] J. Park, I.-C. Hwang, Y. Yoon, J.-B. Park, J.-H. Park, and G.-Y. Cho, “Predicting Long-Term Mortality in Patients With Acute Heart Failure by Using Machine Learning,” *J. Card. Fail.*, vol. 28, no. 7, pp. 1078–1087, Jul. 2022, doi: 10.1016/j.cardfail.2022.02.012.
- [8] F. Li, H. Xin, J. Zhang, M. Fu, J. Zhou, and Z. Lian, “Prediction model of in-hospital mortality in intensive care unit patients with heart failure: Machine learning-based, retrospective analysis of the MIMIC-III database,” *BMJ Open*, vol. 11, no. 7, Jul. 2021, doi: 10.1136/bmjopen-2020-044779.
- [9] S. Srinivasan, S. Gunasekaran, S. K. Mathivanan, Benjula Anbu, P. Jayagopal, and G. T. Dalu, “An active learning machine technique based prediction of cardiovascular heart disease from UCI-repository database,” *Sci. Rep.*, vol. 13, no. 1, p. 13588, Dec. 2023, doi: 10.1038/s41598-023-40717-1.
- [10] World Heart Federation, “Cardiovascular Disease (CVD) | World Heart Federation.” Accessed: May 30, 2025. [Online]. Available: <https://world-heart-federation.org/what-is-cvd/>
- [11] B. Moraga-Arias, S. Chamorro, D. Godoy, P. Caorsi, and L. Goitia, “Analysis of the epidemiological distribution of emergency care for cardiovascular causes between 2018-2022 in Chile,” *Revista de Estudiantes de Medicina del Sur*, vol. 12, no. 2, pp. 29–36, May 2025, doi: 10.56754/0718-9958.2025.224.

- [12] A. Soto, T. Balboa-Castillo, O. Andrade-Mayorga, G. Marzuca-Nassr, S. Muñoz, and G. Morales, “Trends in mortality from cardiovascular diseases in Chile, 2000–2020,” *Revista Panamericana de Salud Pública*, pp. 1–8, 2023, doi: 10.26633/rpsp.2023.127.
- [13] Departamento de Estrategia Nacional de Salud, Ministerio de Salud de Chile, “Análisis crítico de temas: Enfermedad cardiovascular. Insumo para la evaluación final de periodo Estrategia Nacional de Salud 2011–2020,” Santiago, Chile, Jan. 2020.: contentReference[oaicite:0]{index=0}.
- [14] G. Palma, F. Verdugo, N. Saldías, C. Morales, and G. Henríquez, “Prevalencia, caracterización y pronóstico de pacientes con sospecha de insuficiencia cardíaca en atención primaria de salud en Chile,” *Revista Chilena de Cardiología*, vol. 43, no. 1, pp. 9–21, 2024.
- [15] C. Zúñiga Pizarro, B. Muñoz Vasquez, and C. Yañez Fernandez, “Análisis sociodemográfico: tasa de mortalidad por fibrilación auricular 2019-2022 en Chile,” *Revista Confluencia*, vol. 7, Oct. 2024, doi: 10.52611/confluencia.2024.1138.
- [16] S. Muñoz-Molina, G. Año-Maturana, M. Galleguillos-Tolosa, L. Barra-Muñoz, M. Hernández-Maturana, and P. Benítez-Penroz, “Análisis epidemiológico de la mortalidad por isquemia miocárdica en Chile entre 2000 y 2021,” *Revista Confluencia*, vol. 8, 2025, 10.52611/confluencia.2025.1326.
- [17] Sociedad Chilena de Cardiología y Cirugía Cardiovascular y Ministerio de Salud, “Guía Clínica de Insuficiencia Cardíaca,” 2015.
- [18] OMS, Organización Mundial de la Salud, “Las diez causas principales de defunción”: Accessed: Jun. 16, 2025. [Online]. Available: <https://www.who.int/es/news-room/fact-sheets/detail/the-top-10-causes-of-death>
- [19] Z. Zhang *et al.*, “Reactive Oxygen Species Scavenging Nanomedicine for the Treatment of Ischemic Heart Disease,” Sep. 01, 2022, *John Wiley and Sons Inc.* doi: 10.1002/adma.202202169.
- [20] S. Brouwers, I. Sudano, Y. Kokubo, and E. M. Sulaica, “Arterial hypertension,” Jul. 17, 2021, *Elsevier B.V.* doi: 10.1016/S0140-6736(21)00221-X.
- [21] N. Miguel and O. Tomás, “Uma interface para análise exploratória da base de dados MIMIC-III,” in *Anais do Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, Sociedade Brasileira de Computação, 2018.
- [22] Z. Dai, S. Liu, J. Wu, M. Li, J. Liu, and K. Li, “Analysis of adult disease characteristics and mortality on MIMIC-III,” *PLoS One*, vol. 15, no. 4, Apr. 2020, doi: 10.1371/journal.pone.0232176.
- [23] K. Akinosoglou, G. Schinas, M. P. Almyroudi, C. Gogos, and G. Dimopoulos, “The impact of age on intensive care,” *Ageing Res. Rev.*, vol. 84, p. 101832, Feb. 2023, doi: 10.1016/j.arr.2022.101832.

- [24] M. L. Bates *et al.*, “Sex differences in cardiovascular disease and dysregulation in Down syndrome,” *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 324, no. 4, pp. H542–H552, Apr. 2023, doi: 10.1152/ajpheart.00544.2022.
- [25] S. Valente *et al.*, “[Circulatory shock: early diagnosis and therapy].,” *G. Ital. Cardiol. (Rome)*, vol. 24, no. 2, pp. 110–121, Feb. 2023, doi: 10.1714/3963.39418.
- [26] Y. Zhao, H. Yu, A. Gong, S. Zhang, and B. Xiao, “Heart rate variability and cardiovascular diseases: A Mendelian randomization study.,” *Eur. J. Clin. Invest.*, vol. 54, no. 1, p. e14085, Jan. 2024, doi: 10.1111/eci.14085.
- [27] W. Ai *et al.*, “Urine Output as a Novel Predictor for In-Hospital Mortality in Acute Pulmonary Embolism Patients: Training With the MIMIC Database and Validation With Independent Cohort.,” *Cardiovasc. Ther.*, vol. 2025, p. 7907049, 2025, doi: 10.1155/cdr/7907049.
- [28] M. R. T. Gouce, A. Ghanbari, T. K. Mohammadi, E. K. Leili, and S. Ashrafi, “Evaluating APACHE II and APACHE IV Scoring Systems in Predicting Mortality Rate and Length of Hospital Stay in Patients with Head Trauma,” *Trauma Mon.*, vol. 26, no. 6, pp. 317–321, Nov. 2021, doi: 10.30491/TM.2021.264991.1218.
- [29] A. Katsounas *et al.*, “The predictive performance of the SAPS II and SAPS 3 scoring systems: A retrospective analysis,” *J. Crit. Care*, vol. 33, pp. 180–185, Jun. 2016, doi: 10.1016/j.jcrc.2016.01.013.



Anexo A. Descripción de Escalas Tradicionales de Mortalidad

APACHE II (Acute Physiology and Chronic Health Evaluation II) es un score utilizado en la UCI para estimar la mortalidad hospitalaria. Este sistema considera:

- 12 variables fisiológicas: Temperatura, presión arterial media, frecuencia cardíaca, frecuencia respiratoria, oxigenación, pH arterial, sodio, potasio, creatinina, hematocrito, leucocitos y Glasgow Coma Scale (GCS)
- La edad del paciente.
- Presencia de enfermedades crónicas severas.
- Rango de puntaje que va del 0 al 71, donde a mayor puntaje, mayor es el riesgo de mortalidad.
- La evaluación se realiza las primeras 24 horas de ingreso a la UCI [28].

SAPS II (Simplified Acute Physiology Score II) fue desarrollado como una alternativa más simple que APACHE. Este sistema:

- Considera las 12 variables fisiológicas: Edad, frecuencia cardíaca, presión arterial sistólica, temperatura corporal, GCS, frecuencia respiratoria, la relación entre presión parcial de oxígeno y fracción inspirada de oxígeno ($\text{PaO}_2/\text{FiO}_2$), pH arterial, bicarbonato plasmático (si no se tiene el pH arterial), bilirrubina sérica total, urea sérica o creatinina (según disponibilidad) y el recuento de leucocitos.
- Incluye 5 variables adicionales: Tipo de admisión (programada o de urgencia), y la presencia de patologías severas como cáncer, SIDA o enfermedades hematológicas.
- Presenta un rango de puntaje entre 0 y 163, donde a mayor puntaje, mayor riesgo de mortalidad.
- Utiliza una fórmula logística para estimar probabilidad de mortalidad en porcentaje.
- Está diseñado para aplicarse durante los primeros días de ingreso a la UCI. [29]

SOFA (Sequential Organ Failure Assessment) es un score que evalúa la disfunción orgánica en base al funcionamiento de distintos sistemas del organismo. En particular, considera lo siguiente:

- Sistema respiratorio: Analiza la relación entre la presión parcial de oxígeno arterial (PaO_2) y la fracción inspirada de oxígeno (FiO_2).

- Coagulación: Evalúa el recuento de plaquetas como indicador de alteraciones de hemostasia.
- Hígado: Niveles de bilirrubina.
- Cardiovascular: Presión arterial y requerimiento de vasopresores.
- Neurológico: Glasgow Coma Scale (GCS)
- Renal: Creatinina y diuresis

Anexo B. Variables fisiológicas integradas en el proyecto

Tabla 0.1. Variables fisiológicas integradas en los modelos de Machine Learning en este proyecto.

Variable	Clasificación	Interpretación Clínica
Sodio (Na^+)	Electrolitos esenciales	Disnatremias reflejan desbalance hídrico y neurohormonal.
Potasio	Electrolitos esenciales	La hipo e hiperpotasemia predisponen a arritmias ventriculares
Cloro	Electrolitos esenciales	Alteraciones producen trastornos ácido-base.
Calcio	Electrolitos esenciales	Afecta contractilidad y estabilidad neuromuscular.
Magnesio	Electrolitos esenciales	Niveles bajos se asocian a arritmia y disfunción miocárdica.
Bicarbonato	Acido-Base	Indicador de equilibrio metabólico, disminuye en acidosis.
Anión Gap	Acido-Base	Indicador de acidosis metabólica, se eleva en sepsis o hipoxia.
pH arterial	Gasometría	Refleja equilibrio ácido-base.
PaO ₂	Gasometría	Mide la oxigenación, hipoxemia indica fallas respiratorias.
PaCO ₂	Gasometría	Evalúa la ventilación y compensación respiratoria.
Lactato	Perfusión	Indicador de hipoperfusión y metabolismo anaeróbico.
Glucosa	Metabolismo	Hiperglucemia indica estrés metabólico y mayores riesgos.
Creatinina	Función renal	Evalúa filtración glomerular, niveles altos indican disfunción renal.
BUN (urea)	Función renal	Indicador de hipoperfusión o catabolismo.
Albúmina	Función hepática	Niveles bajos reflejan mal estado nutricional o congestión hepática.
TP / INR	Coagulación	Evalúa la función hepática y el riesgo de hemorragia.
Hemoglobina	Hemoglobina (Hb)	Anemia reduce el transporte de oxígeno e indica perfusión tisular.
Hematocrito (Hto)	Hematocrito (Hto)	Corroborar la anemia.
Leucocitos	Leucocitos	Leucocitosis indica inflamación o infección.
Plaquetas	Leucocitos	Trombocitopenia aumenta riesgo hemorrágico.
RDW	Hemograma	Niveles altos reflejan inflamación crónica o estrés oxidativo.
Creatina quinasa		Niveles altos indican necrosis muscular o daño cardíaco.

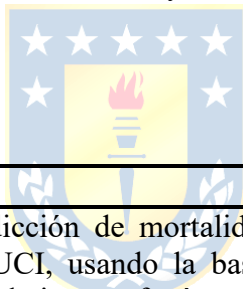
Fuente: Elaboración propia. Información obtenida de [4] [3] [8].

ANEXO D: FORMATO RESUMEN DE MEMORIA DE TÍTULO

UNIVERSIDAD DE CONCEPCION – FACULTAD DE INGENIERIA RESUMEN DE MEMORIA DE TITULO

Departamento : Departamento de Ingeniería Eléctrica
Carrera : Ingeniería Civil Biomédica
Nombre del memorista : Bastian Eduardo Beytía Hormazábal
Título de la memoria : Predicción de Mortalidad en Pacientes con ECV Hospitalizados en UCI
 Mediante Machine Learning y la Base de Datos MIMIC-III.
Fecha de la presentación oral : 27/03/2026

Profesor(es) Guía : Rosa Figueroa Iturrieta
Profesor(es) Revisor(es) : Pamela Guevara A. y Gerónimo Bolado L.
Concepto :
Calificación :



Resumen (máximo 200 palabras)

En este proyecto se estudió la predicción de mortalidad en pacientes con enfermedades cardiovasculares hospitalizados en UCI, usando la base de datos MIMIC-III v1.4 como principal fuente de información. El trabajo se enfocó en una cohorte de pacientes adultos con diagnósticos cardiovasculares relevantes e incluyó revisión bibliográfica, exploración de datos, selección de variables clínicas y la construcción de un proceso de análisis para entrenar y evaluar modelos de machine learning. Como MIMIC-III es una base de datos de gran tamaño, tanto el procesamiento como la evaluación se realizaron en un servidor local de la Universidad. Los resultados mostraron que los modelos basados en gradient boosting tuvieron un mejor desempeño que Random Forest. Además, las variables más influyentes presentaron una relación coherente con el cuadro clínico de los pacientes, destacando parámetros bioquímicos, hematológicos y demográficos. Esto muestra que el aprendizaje automático puede ser un apoyo útil para la evaluación temprana del riesgo en UCI.