



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

PREDICCIÓN DE EVENTOS CLÍNICOS ADVERSOS A PARTIR DE DATOS  
ESTÁTICOS Y DINÁMICOS DEL PACIENTE EN EL SERVICIO DE URGENCIAS  
PARA COVID-19

TESIS PARA OPTAR AL GRADO DE  
MAGÍSTER EN CIENCIAS DE LA INGENIERÍA, MENCIÓN ELÉCTRICA

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO

RICARDO ESTEBAN ISRAEL JARA GONZÁLEZ

PROFESOR GUÍA:  
MARCOS ORCHARD CONCHA  
PROFESOR CO-GUÍA:  
STEFFEN HÄRTEL GRÜNDLER

MIEMBROS DE LA COMISIÓN:  
PABLO ESTÉVEZ VALENCIA  
MATÍAS ZAÑARTU SALAS

Este trabajo ha sido parcialmente financiado por IA-TRAD CHILE: Inteligencia Artificial para Apoyo Diagnóstico y Procesos de Atención en Radiología y Tele-Radiología Clínica, ANID COVID0733, 2020 – 2021

SANTIAGO DE CHILE  
2024

RESUMEN DE LA TESIS PARA OPTAR AL  
GRADO DE MAGÍSTER EN CIENCIAS DE LA INGENIERÍA, MENCIÓN ELÉCTRICA  
Y MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO  
POR: RICARDO ESTEBAN ISRAEL JARA GONZÁLEZ

FECHA: 2024

PROF. GUÍA: MARCOS ORCHARD CONCHA  
PROF. CO-GUÍA: STEFFEN HÄRTEL GRÜNDLER

PREDICCIÓN DE EVENTOS CLÍNICOS ADVERSOS A PARTIR DE DATOS  
ESTÁTICOS Y DINÁMICOS DEL PACIENTE EN EL SERVICIO DE URGENCIAS  
PARA COVID-19

En el contexto de la pandemia, el mayor impacto repercute sobre el personal de salud y la gran responsabilidad puesta sobre ellos. Se propone un sistema de predicción de eventos clínicos adversos para pacientes COVID-19 positivos, con tal de apoyar al personal de salud en la toma de decisiones. Se consideran los eventos indicación de ventilación mecánica invasiva en los próximos dos días, y mortalidad en los próximos cinco días. Modelos en base a XGBoost son diseñados para predecir la probabilidad de ocurrencia de éstos, a partir de datos clínicos del paciente y su evolución temporal. El estudio considera 395 pacientes. En el caso de ventilación mecánica invasiva se encuentran 7 variables clínicas de alto impacto, de las que se obtiene un *recall* o sensibilidad de 0.93 y área bajo la curva ROC de 0.88. Para mortalidad, se encuentran otras 7 variables claves, consiguiendo un *recall* de 0.82 y área bajo la curva de 0.82. Estos resultados son comparables con la literatura, a pesar de la acotada cantidad de datos. Al entregar la probabilidad de que el paciente sufra estos eventos en el futuro cercano, se presenta una herramienta que apoya a los profesionales en la toma de decisiones efectiva, agregando carácter predictivo a sus decisiones reactivas.

*A mi nina, estoy orgulloso de ser tu hijo*

# Agradecimientos

A si es, las gracias. Quiero agradecer a todo ser que a sido parte de mi camino por la u, haciendolo lo maravilloso que es. Sin orden en particular:

A mi familia, en especial a mi padre, a mi madre y a mi hermana, por su apoyo incondicional y darme absolutamente todo lo que soy. No podría haber llegado donde estoy sin ustedes. Gracias su confianza en mí.

Al Dilan, Terry y Remi, porque si.

A mis amigos Ale buena onda, Negro apu rana Méndez, San, Pudú, Simmon, Pelao, Kako y Seba heiter.

A los sabios Diego y Martín.

A mi gente Kebim, Gabriel, Tío P, Benja, Tomás, Néstor y a la tripleta EMP.

A los cabros Loreh, Alonso, Fabi, el hombre, Kim, Javierrr, fla, Diland, Nico Silva, Sacco cgn, lucho trucho, dani, ale mala onda y Manu.

A los k amigo francisco, metralleta y el rubio.

A la gente Marx, Lucho diego, lala, Edu, Lucho malo, Albert, Lucho bueno y Caro.

A los eternos el wardo, el oso, el metal, el Rios, señorito sebastian y el jimmy.

A los pandémicos Henrique y Anthony.

Al equipo IA·TRAD Maca, Cota, Valu, Gonzalo, Alonso, Camilo, Guillermo y Victor.

A Loreto, no hay mejor funcionaria en toda la universidad.

Cada uno a estado presente en algún punto de mi transcurso por la carrera, influenciando de manera positiva, y sabe como a tomado parte en mí.

A mis profes Marcos y Steffen, por su guía, dedicación y experiencia compartida, por su paciencia y comprensión, y por darme la oportunidad de desarrollarme en el área que me gusta. Muchas gracias.

Lamentablemente no me es posible nombrar a absolutamente todos, pero los tengo en el cora igual. Infinitas gracias por haber participado a todos. *Muito Obrigado.*

# Tabla de Contenido

<b>1. Introducción</b>	<b>1</b>
1.1. Motivación . . . . .	1
1.2. Hipótesis . . . . .	3
1.3. Objetivos . . . . .	3
1.3.1. Objetivo General . . . . .	3
1.3.2. Objetivos Específicos . . . . .	3
1.4. Estructura del Documento . . . . .	3
<b>2. Marco Teórico</b>	<b>4</b>
2.1. Trabajo Relacionado . . . . .	4
2.2. Inteligencia Artificial . . . . .	5
2.2.1. UMAP . . . . .	5
2.2.2. Árboles de decisión . . . . .	7
2.2.3. <i>Gradient Boosting</i> . . . . .	8
2.2.4. XGBoost . . . . .	8
2.2.5. Análisis SHAP . . . . .	10
2.2.6. Métricas de evaluación . . . . .	12
<b>3. Propuesta de Sistema para Predicción de Eventos Clínicos Adversos</b>	<b>17</b>
3.1. Presentación del Problema . . . . .	17
3.2. Datos Utilizados . . . . .	19
3.2.1. Adquisición de Datos . . . . .	19
3.2.2. Conjunto de Datos . . . . .	20
3.3. Propuesta de Sistema de Predicción . . . . .	23
3.3.1. Preprocesamiento de Datos . . . . .	23
3.3.2. Extracción de Características . . . . .	25
3.3.3. Diseño de Modelos . . . . .	27
3.3.4. Selección de Variables . . . . .	31
<b>4. Resultados y Discusión</b>	<b>39</b>
4.1. Predicción de Ventilación Mecánica Invasiva . . . . .	39
4.2. Predicción de Mortalidad . . . . .	51
<b>5. Conclusiones</b>	<b>60</b>
5.1. Trabajo Futuro . . . . .	62

<b>Bibliografía</b>	<b>63</b>
<b>A. Acrónimos</b>	<b>67</b>
<b>B. Variables Datos Estáticos</b>	<b>68</b>

# Índice de Tablas

3.1. Cantidad de días que se proponen a predecir los evento clínicos adversos. . . . .	17
3.2. Cantidad de pacientes en cada evento. . . . .	21
3.3. Representación del formato de los datos longitudinales preprocesados . . . . .	25
3.4. Datos subconjunto de entrenamiento y prueba para evento VMI. . . . .	29
3.5. Datos subconjunto de entrenamiento y prueba para Mortalidad. . . . .	31
3.6. Variables basales escogidas para comenzar el análisis de sensibilidad. . . . .	32
3.7. Métricas obtenidas para predicción de VMI utilizando distintas variables de entrada. . . . .	34
3.8. Métricas obtenidas para predicción de Mortalidad utilizando distintas variables de entrada. . . . .	37
4.1. Métricas de desempeño para el modelo de predicción de VMI Variables Seleccionadas (7). . . . .	40
4.2. Métricas de desempeño para el modelo de predicción de VMI Variables Seleccionadas (9). . . . .	43
4.3. Métricas de desempeño para los modelos de predicción de VMI. . . . .	46
4.4. Comparación de modelos de predicción de VMI en el estado del arte y en el presente trabajo. . . . .	49
4.5. Métricas de desempeño para el modelo de predicción de Mortalidad Variables Seleccionadas. . . . .	51
4.6. Comparación de métricas de desempeño para los modelos de predicción de Mortalidad. . . . .	55
4.7. Comparación de modelos de predicción de Mortalidad en el estado del arte y en el presente trabajo. . . . .	57
B.1. Variables incluidas en el estudio a pacientes COVID-19 ingresados al servicio de urgencia durante los primeros 100 días, datos estáticos. . . . .	68

# Índice de Ilustraciones

2.1. Conjuntos difusos asociados a los vecindarios generados a partir de los datos de entrada. . . . .	5
2.2. Ejemplo de UMAP. . . . .	6
2.3. Diagrama representativo de un árbol de decisión . . . . .	7
2.4. Representación de un árbol de decisiones con direcciones por defecto. . . . .	10
2.5. Ejemplo de análisis SHAP de la relación entre mortalidad con las variables clínicas Frecuencia Cardíaca, Edad y PCR. . . . .	11
2.6. Ejemplo de una matriz de confusión de dos clases. . . . .	12
2.7. Ejemplo de curvas ROC. La curva azul representa una ejemplo de un modelo con mejor desempeño que la amarilla. . . . .	13
2.8. Representación gráfica del área bajo la curva ROC . . . . .	14
2.9. Relación entre matriz de confusión y métricas de desempeño. . . . .	15
2.10. Explicación y demostración de Diagramas de Caja . . . . .	16
3.1. Muestras de UMAP para los Datos Estáticos. . . . .	18
3.2. Distribución de los 395 pacientes en los eventos clínicos presentados. . . . .	21
3.3. Descripción gráfica de la extracción de características sobre los datos longitudinales. . . . .	26
3.4. Ejemplo de un paciente y su organización de clases en el contexto del modelo de predicción de Ventilación Mecánica Invasiva. . . . .	27
3.5. Diagrama de flujo del paciente hasta la predicción de eventos clínicos. Este muestra de manera simplificada el camino seguido por los datos para ser ingresado al modelo. . . . .	29
3.6. Importancia de características sobre el modelo de predicción de VMI entrenado con las Variables Basales. . . . .	33
3.7. Importancia de características sobre el modelo de predicción de Mortalidad entrenado con las Variables Basales. . . . .	35
3.8. Análisis SHAP de la variable “Obesidad” sobre el modelo de predicción de Mortalidad entrenado con las Variables Basales. . . . .	36
3.9. Porción de pacientes con Obesidad en el estudio. . . . .	36
3.10. Histograma de edades de los 58 pacientes fallecidos durante el estudio. . . . .	37
4.1. Resultados del modelo de predicción de VMI asociado a Variables Seleccionadas (7). . . . .	40
4.2. Histograma normalizado de probabilidad de ocurrencia predicha para el modelo de predicción de VMI asociado a las Variables Seleccionadas (7). . . . .	41



4.3. Análisis SHAP para el modelo de predicción de VMI asociado a las Variables Seleccionadas (7). . . . .	42
4.4. Resultados del modelo de predicción de VMI asociado a las Variables Seleccionadas (9). . . . .	43
4.5. Histograma de probabilidad predicha para el modelo de predicción de VMI asociado a las Variables Seleccionadas (9). . . . .	44
4.6. Análisis SHAP para el modelo de predicción de VMI asociado a las Variables Seleccionadas (9). . . . .	45
4.7. Curvas ROC de los modelos de predicción de VMI Variables Seleccionadas (7), Variables Seleccionadas (9) y de Variables Basales. . . . .	45
4.8. Probabilidad de ingresar a VMI predicha por el modelo Variables Seleccionadas (7) de pacientes VMI, calculada por día. . . . .	47
4.9. Diagramas de caja generados sobre la predicción de VMI por día. . . . .	48
4.10. Resultados del modelo de predicción de Mortalidad asociado a las variables seleccionadas. . . . .	51
4.11. Histograma de probabilidad predicha para el modelo de predicción de Mortalidad asociado a las Variables Seleccionadas. . . . .	52
4.12. Análisis SHAP para el modelo de predicción de Mortalidad asociado a las Variables Seleccionadas. . . . .	53
4.13. Curvas ROC del modelo de predicción de Mortalidad de Variables Seleccionadas y de Variables Basales. . . . .	54
4.14. Probabilidad de fallecer predicha por el modelo Variables Seleccionadas de pacientes que fallecen, calculada por día. . . . .	55
4.15. Diagramas de caja generados sobre la predicción de Mortalidad por día. . . . .	56

# Capítulo 1

## Introducción

### 1.1. Motivación

En el inicio de la pandemia en Chile (Marzo del 2020) la situación en los centros de salud era abrumadora al punto del colapso [1]. Si bien solo el 14 % de los pacientes requerían cuidados en la Unidad de Pacientes Críticos (UPC) o tratamientos más drásticos [2], la gran cantidad de infectados convertía ese pequeño porcentaje en una cifra significativa. En este contexto, surgió la necesidad de optimizar y agilizar los procesos de atención en los hospitales, especialmente en las unidades de urgencia, donde la toma de decisiones fundamentales es de carácter vital para los pacientes.

El estado del paciente, caracterizado por su información clínica y la experiencia del personal de salud, es clave para la toma de decisiones efectiva. Los procedimientos de admisión y cuidados de pacientes continúan demandando una atención meticulosa lo cual resulta en una labor humana intensiva para la misma cantidad trabajadores del servicio de urgencia. Un gran desafío de este escenario es seleccionar a quién priorizar la atención médica en situaciones críticas, ya que recursos como camas clínicas o ventiladores mecánicos son limitados. Para estas decisiones no solo se considera el estado del paciente, si no también su evolución en el tiempo lo que convierte esto en un proceso reactivo.

Los algoritmos de Inteligencia Artificial (IA) sobresalen en la tarea de obtener información a partir de patrones que a menudo son difíciles de discernir por parte de los seres humanos o que para lograrlo se requiere de especial concentración [3]. Por otro lado la IA tiene la limitación de que no puede alcanzar el desempeño dada por la experiencia de un experto de la salud, referente a comportamientos humanos del paciente que no pueden ser registrados. Bajo este panorama ésta tiene utilidad como un sistema de apoyo a la toma de decisiones, ayudando a agilizar la atención médica, liberando recursos y aliviando la carga al personal de salud.

Durante los primeros 100 días de pandemia, médicos del Hospital Clínico de la Universidad de Chile (HCUCH) llevaron a cabo un estudio prospectivo con 395 pacientes COVID-19 positivos que ingresaron al servicio de urgencias. Este estudio recopiló, entre muchos otros, datos demográficos, signos vitales, resultados de exámenes de laboratorio y eventos clínicos

adversos. Cada paciente cuenta con un seguimiento de traslados, fechas y eventos hasta que este sea dado de alta, fallezca o alcance la fecha de censura del estudio. Para el resto de los datos, debido a que deben ser obtenidos de manera manual lo cual es un trabajo exhaustivo, se opta por registrar solo el peor valor fisiológico en las primeras 24 horas desde el ingreso al servicio de urgencias.

En colaboración con los autores del estudio recién mencionado<sup>1</sup>, personal de la Universidad de Concepción y la Facultad de Medicina de la Universidad de Chile nace el proyecto IA·TRAD-Chile, con la idea de proporcionar al personal de salud del HCUCH herramientas para asistir la toma de decisiones informadas y precisas. Dentro de este proyecto se encuentra el presente trabajo, el cual tiene como objetivo desarrollar un sistema de predicción de eventos clínicos adversos relevantes como lo son la indicación de Ventilación Mecánica Invasiva (VMI) y el posible fallecimiento del paciente.

La importancia de esta investigación radica en su potencial de brindar apoyo al diagnóstico y la toma de decisiones clínicas, y mejorar la calidad de atención a los pacientes. La implementación exitosa de este sistema de predicción puede otorgar a los profesionales de la salud una herramienta de apoyo que agregaría valor predictivo a sus procesos reactivos.

Con intención de mejorar la experiencia de atención a pacientes y al personal de salud, el trabajo está contextualizado en el servicio de urgencias del hospital, el cual es la primera instancia de atención médica desde el ingreso del paciente. Como tal, la toma de decisiones rápida y precisa es crucial para agilizar el proceso hospitalario.

En el presente documento se propone un sistema de predicción de los eventos clínicos adversos VMI y mortalidad para pacientes COVID-19 positivos, utilizando datos clínicos del paciente, a partir de modelos de *Machine Learning* (ML) clásicos, en particular XGBoost [4], el cual es el estado del arte en múltiples problemas de clasificación [5–8].

Estos algoritmos se muestran particularmente interesantes para el tópico a trabajar, pues entregan información concreta sobre qué variables son las que más relevancia tienen en la predicción, entregando interpretabilidad a las decisiones. Así mismo, se puede plantear el cómo podría haber cambiado el escenario del inicio de la pandemia de tener un sistema que apoye la toma de decisiones, como también la información que este tiene para ofrecer en cuanto a variables de relevancia.

---

<sup>1</sup>Equipo de internistas del HCUCH liderado por Dr. Rodrigo Cornejo.

## 1.2. Hipótesis

Un modelo de clasificación basado en árboles de decisión XGBoost es capaz de predecir la ocurrencia de eventos clínicos adversos en el futuro inmediato o cercano sobre pacientes COVID-19 locales en el servicio de urgencias utilizando sus datos clínicos y evolución temporal, con un desempeño comparable al estado del arte representado por un área bajo la curva ROC mayor a 0.85 en el escenario de Ventilación Mecánica Invasiva y 0.80 en el de Mortalidad.

## 1.3. Objetivos

### 1.3.1. Objetivo General

Desarrollar un sistema de apoyo a la toma de decisiones para efecto de asistir al personal de salud a partir de recomendación de opciones para dos eventos clínicos adversos: Indicación de Ventilación Mecánica Invasiva y Mortalidad, en base a modelos de Inteligencia Artificial XGBoost.

### 1.3.2. Objetivos Específicos

**OE1** Garantizar la consistencia y confiabilidad de los datos mediante interpretación clínica experta, estandarización de variables y sistemas de medida, con el fin de mejorar la precisión de los datos y su usabilidad en modelos.

**OE2** Identificar y reducir las variables clínicas a considerar para factibilidad de uso en la práctica clínica en base a conocimiento experto y análisis de sensibilidad, manteniendo o mejorando el rendimiento de predicciones.

**OE3** Desarrollar un sistema que permita predecir eventos clínicos adversos en el futuro cercano con un área bajo la curva ROC comparables con el estado del arte, a partir del diseño y optimización de algoritmos de predicción basados en XGBoost para los eventos clínicos “Indicación de Ventilación Mecánica Invasiva” y “Mortalidad”.

## 1.4. Estructura del Documento

**Capítulo 1:** Se muestra la introducción, en donde se revisa el contexto y motivación del presente trabajo. Se presentan también la hipótesis y objetivos.

**Capítulo 2:** Se expone el marco teórico, en el cual se hace un breve repaso del estado del arte, y la teoría y explicación de las herramientas que se usarán en el documento.

**Capítulo 3:** Se presenta la propuesta del sistema de predicción de eventos adversos, con los pasos seguidos en el diseño de los modelos desde los datos hasta la optimización.

**Capítulo 4:** Se presentan y discuten los resultados de los modelos de predicción. También se evalúan éstos a partir de su comportamiento e importancia de variables.

**Capítulo 5:** Se concluye el trabajo realizado en base a los objetivos y se propone trabajo futuro.

# Capítulo 2

## Marco Teórico

### 2.1. Trabajo Relacionado

Los modelos de clasificación o predicción para datos numéricos han sido ampliamente abordados en el pasado teniendo estudios específicos sobre éstos para prácticamente cada área. Con la aparición del COVID-19 se presentó un escenario inexplorado, motivando la implementación de modelos de Inteligencia Artificial (IA) con el afán de apoyar al personal de salud y generar entendimiento de distintos tópicos sobre el nuevo escenario [9, 10]. Previo, en paralelo y posterior al presente trabajo se desarrollan modelos de predicción de escenarios críticos alrededor del mundo utilizando datos de sus localidades buscando dar solución al problema planteado.

Numerosos esfuerzos son descritos en la literatura con el objetivo de abordar la misión de predecir el índice de mortalidad de los pacientes [9–14]. En su mayor parte éstos utilizan aprendizaje supervisado, en el cual el modelo aprende utilizando datos etiquetados los que pueden ser por ejemplo si el paciente fallece o no. Algoritmos de árboles de decisión basados en *Boosting* como *Gradient Boosting Machine* (GBM) [15] y XGBoost [4] son los que han mostrado tener mejor desempeño sobre bases de datos acotadas [11–14, 16, 17]. El estudio que ha obtenido mejores resultados presenta un *accuracy* del 90% y área bajo la curva ROC de 0.951 para predicción de mortalidad hasta 10 días en el futuro [12]. La base de datos utilizada está limitada a variables de laboratorio, y tal desempeño es alcanzado utilizando LDH, linfocitos y PCR. Otras aproximaciones al problema se han hecho con *deep learning* utilizando imágenes clínicas, en particular el modelo ResNet-50 entrenado con rayos-X de tórax de pacientes COVID-19 positivos, el cual obtiene un área bajo la curva ROC de 0.78~0.82 [18].

Por el lado de ventilación mecánica invasiva, los trabajos que abordan este tema son menos abundantes, y tienen similitudes en el enfoque y métodos utilizados. Otra vez los algoritmos basados en *Boosting* tienden a ser los que destacan en la tarea [10, 19–22]. El modelo de predicción con mayor desempeño en la literatura está basado en GBM, con un área bajo la curva ROC de 0.94 [22]. En éste se utiliza un *dataset* de 472 pacientes COVID-19 positivos con datos, de los cuales 137 presentan eventos. Las variables utilizadas incluyen datos clínicos, signos vitales y exámenes de laboratorio.

## 2.2. Inteligencia Artificial

### 2.2.1. UMAP

*Uniform Manifold Approximation and Projection* (UMAP) [23] es una técnica de reducción de dimensionalidad que destaca como herramienta de visualización y exploración de datos de alta dimensión. Tiene la cualidad de conservar información topológica de los grafos más complejos en sus proyecciones dando directa interpretabilidad de los datos visualizados. El algoritmo asume que los datos de entrada se encuentran uniformemente distribuidos dentro de un espacio topológico infinito, el cual puede ser aproximado a partir de las muestras de datos finitas entregadas y luego proyectado sobre un espacio de menores dimensiones.

Conocer como se posicionan los datos en el espacio de alta dimensionalidad es necesario para graficarlos en el espacio de dimensiones menores. UMAP comienza por encontrar los **vecinos más cercanos** en función de una medida de distancia para cada punto utilizando como hiperparámetro la cantidad máxima de éstos a tener en cuenta. Este paso es esencial para la correcta interpretación del grafo pues controla como el algoritmo balancea la estructura local contra la global en los datos.

Los vecindarios encontrados, formados por los vecinos cercanos previamente conectados, permiten obtener la forma del grafo original. Al asumir que los puntos se encuentran uniformemente distribuidos los vecindarios sugieren que el espacio dentro y entre ellos se encuentra contraído o dilatado según corresponde. Esto permite utilizar una noción de distancia local dada por geometría de Riemann [24], implicando que la medida de distancia utilizada no es universal. Aquí es donde se elige el siguiente hiperparámetro, el de **conectividad local**, en donde se decide a cuantos puntos mínimo debe estar conectado cada dato en el espacio original.

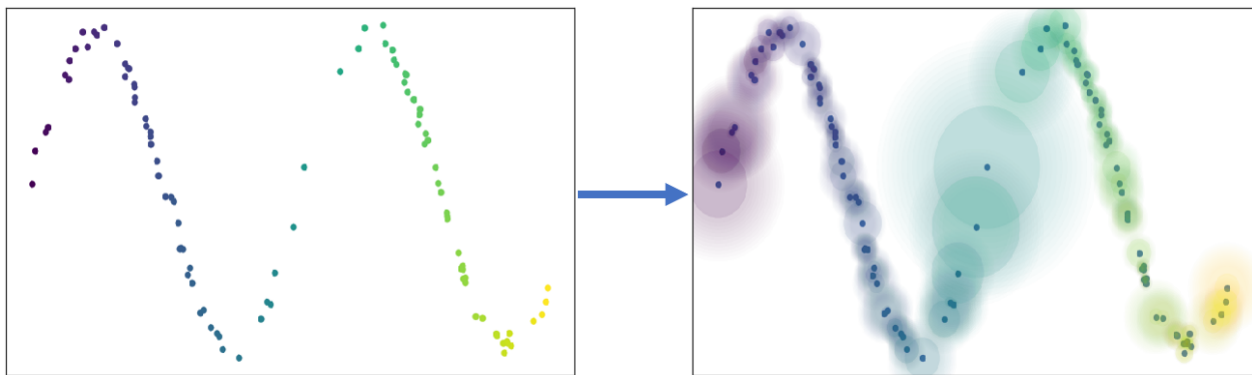


Figura 2.1: Conjuntos difusos (a la derecha) asociados a los vecindarios generados a partir de los datos de entrada (a la izquierda). Los hiperparámetros utilizados para este resultado son 15 como número de vecinos y 1 como conectividad local. (Figura tomada de [25]).

En la Figura 2.1 se muestra un ejemplo de los conjuntos difusos generados en torno a las muestras una vez elegidos los hiperparámetros. Los radios de los vecindarios poseen la misma longitud según la medida de distancia utilizada, mostrándose de diferentes tamaños debido a como varía el espacio. Cada conjunto difuso representa la probabilidad de que la muestra central este conectada con el resto de las muestras. A partir de estos vecindarios locales se

crea un grafo ponderado considerando las distancias entre los puntos y la probabilidad de conexión entre ellos.

Una vez que el algoritmo posee la información sobre el espacio topológico de alta dimensionalidad se puede entrar a calcular la proyección sobre un espacio de menores dimensiones. Teniendo en cuenta que uno de los objetivos del algoritmo es la visualización de datos, se desea que este último espacio presente distancias Euclidianas estándar con respecto a su sistema de referencia. Cambiar entre espacios produce también un cambio entre las distancias de las muestras, lo que puede producir que muchos puntos queden juntos de manera indistinguible o difíciles de visualizar. Para solventar esto se utiliza el tercer hiperparámetro, la distancia mínima, que define el mínimo de separación entre las muestras en la proyección.

UMAP busca minimizar una función de costos para encontrar una representación apropiada en un espacio de menor dimensión. La función de costos a minimizar es *cross entropy* ( $CE$ ) que en este caso se puede escribir como:

$$CE = \sum_{e \in E} w_h(e) \log \left( \frac{w_h(e)}{w_l(e)} \right) + (1 - w_h(e)) \log \left( \frac{1 - w_h(e)}{1 - w_l(e)} \right). \quad (2.1)$$

donde  $E$  es el conjunto de las muestras borde en un conjunto difuso, y  $e \in E$  las muestras.  $w_h(e)$  son los pesos del espacio de alta dimensión determinados por la unión de los conjuntos difusos entre las muestras borde, y  $w_l(e)$  los pesos del espacio de baja dimensión análogos. Para mayor detalle sobre la matemática detrás del algoritmo referirse a la publicación original [23].

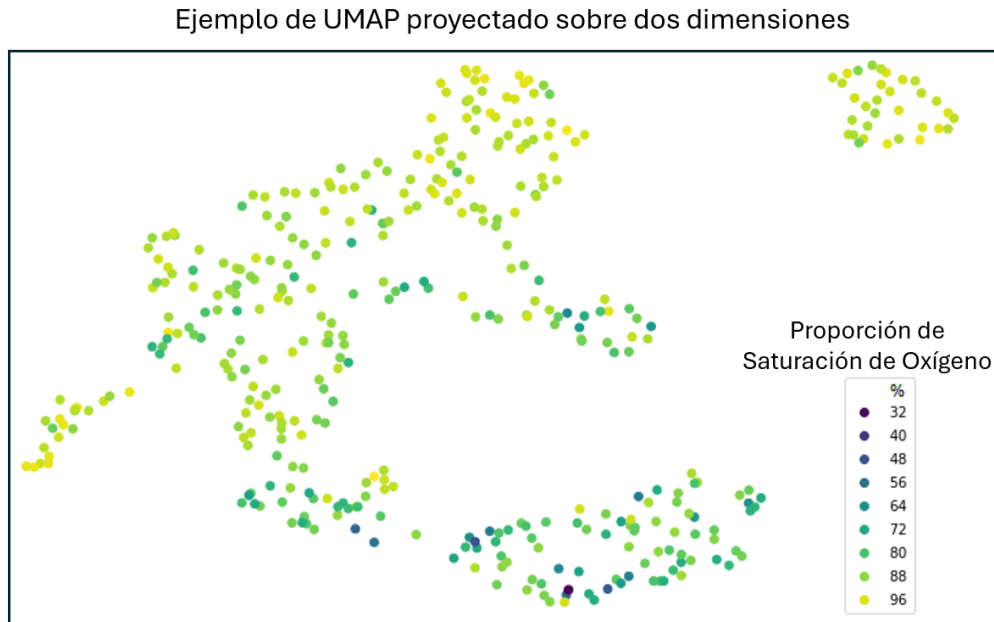


Figura 2.2: Ejemplo de UMAP generado con 395 datos donde cada uno es representado por un punto. La proyección es generada desde un espacio de 39 dimensiones sobre las dos dimensiones mostradas. El código de colores representa el porcentaje de Saturación de Oxígeno presente en cada dato, una de las variables utilizadas por el algoritmo.

El algoritmo itera hasta encontrar los pesos optimizados para el espacio de baja dimensionalidad. Una vez obtenidos, se pueden proyectar los datos, los que se distribuyen de tal manera que reflejan las relaciones y agrupaciones presentes en el espacio de alta dimensión [26]. En la Figura 2.2 se muestra un ejemplo de UMAP generado a partir de un espacio de 39 dimensiones, y proyectado sobre dos dimensiones. Éste permite explorar de manera visual el problema, lo que se hace en el caso del ejemplo destacando los datos a partir de “Saturación de Oxígeno” que es una de sus dimensiones. De aquí se puede obtener información, como por ejemplo, que existen vecindarios generados que se asemejan por su porcentaje de Saturación de Oxígeno, lo que describe una correlación entre el espacio y dicha variable en el problema.

## 2.2.2. Árboles de decisión

Los árboles de decisión son un algoritmo de IA clásico basado en aprendizaje supervisado, utilizados para problemas de clasificación y regresión. Su estructura corresponde a un árbol en donde cada nodo (o bifurcación) representa una decisión tomada en base de el vector de características de entrada.

Este modelo posee una distribución jerárquica (ver Figura 2.3), en donde el primer nodo representa la característica con más importancia para la clasificación, siguiendo hasta el final del árbol con características menos importantes. En las “hojas”, los últimos escalafones del diagrama, se encuentran representados los resultados de la clasificación o clases. En cada nodo se toma una decisión en base a su característica asociada de la cual se bifurca a los siguientes nodos u hojas.

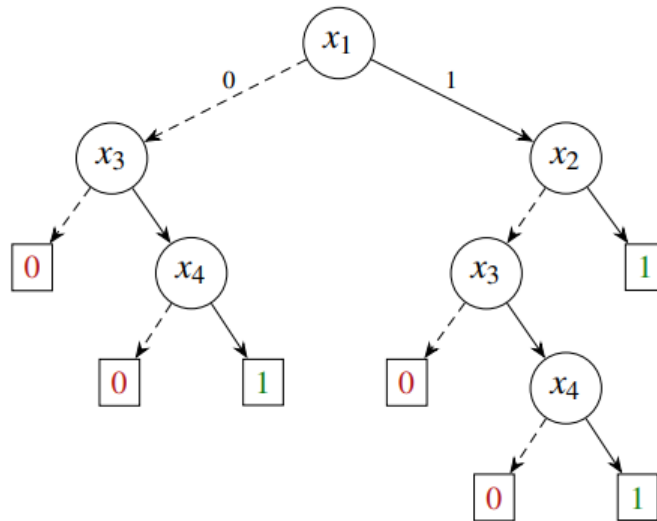


Figura 2.3: Diagrama representativo de un árbol de decisión construido para un vector de características de entrada de 4 elementos  $\{x_1, x_2, x_3, x_4\}$  y 2 clases  $\{0, 1\}$ . (Figura tomada de [27]).

El entrenamiento es llevado a cabo seleccionando los mejores atributos para jerarquizar el árbol basándose en métricas como entropía o impureza de Gini, las cuales cuantifican la cantidad de información o aleatoriedad de cada variable. De manera iterativa se busca maximizar la pureza de los subgrupos que representan las ramas.



El sobreajuste es un problema recurrente en este tipo de algoritmos si es que no se considera al momento del diseño, pues este presenta mejores resultados sobre los datos de entrenamiento pues mejor se adapta a estos mientras más profundidad posea. La forma común de resolver esto es limitando la profundidad del árbol, es decir, el número de niveles o decisiones de contiene de manera consecutiva. [27,28]

### 2.2.3. *Gradient Boosting*

En el campo del aprendizaje de máquinas *Gradient Boosting* [29] es una técnica utilizada en el entrenamiento de modelos de IA mayormente para problemas de regresión y clasificación. La idea detrás del algoritmo es ensamblar *weak learners* para generar un gran modelo robusto.

Un *weak learner* es un modelo simple de clasificación o regresión como lo son los árboles de decisión de poca profundidad. A grandes rasgos se busca entrenar los árboles de decisión de manera secuencial, tal que cada *weak learner* tome en consideración el error del árbol anterior en su predicción. Esta forma en cascada de entrenar se le denomina *Boosting*. La manera de optimizar cada *weak learner* es a partir de descenso de gradiente sobre la función de pérdida, he de aquí el nombre *Gradient Boosting*.

El algoritmo trabaja de manera iterativa comenzando con un árbol de decisión base. En cada hoja de éste se computa el siguiente árbol donde se minimiza la función de pérdida con los datos de entrenamiento y se generan nuevas hojas, obteniendo un puntaje (peso) con la información de cuanto contribuye el camino actual a la reducción de residuos. El proceso se repite un número predefinido de veces o hasta alcanzar alguna métrica umbral.

El resultado de este algoritmo es un robusto modelo como los *Gradient Boosting Machine* (GBM) [15], el cual es un árbol de decisión de alto desempeño capaz de resolver tareas complejas gracias a la generalidad que le entrega la agrupación de *weak learners*, y la particularidad de cada uno por separado. Sin embargo es importante mencionar que los GBMs pueden ser muy costosos computacionalmente y fáciles de sobreajustar si no se es cuidadoso con la elección de parámetros [30].

### 2.2.4. *XGBoost*

*Extreme Gradient Boosting* o XGBoost [4] es un algoritmo de IA basado en *Gradient Boosting* utilizado principalmente en tareas de regresión y clasificación. La principal mejora sobre los GBM es como XGBoost opera la función de pérdida para mitigar las varianzas, reduciendo con esto también la complejidad del modelo y evitando el sobreajuste. El algoritmo busca también optimizar los recursos computacionales lo cual consigue a partir de lo anterior y añadiendo métodos de ejecución en paralelo durante el entrenamiento.

La base de los modelos de ML es minimizar una función objetivo. En el caso de XGBoost la función objetivo está dada por la combinación entre la función de pérdida (*Loss*) y un termino de regularización:

$$\begin{aligned} \text{Función Objetivo}(T) &= \mathcal{L}(T) = \text{Loss} + \text{Regularización}, \\ \mathcal{L}(T) &= \sum l(y_i, \hat{y}_i) + \sum \Omega(f_k). \end{aligned} \quad (2.2)$$

donde  $T$  es el número de hojas en cada árbol y  $l(y_i, \hat{y}_i)$  una función de pérdida diferenciable que mide la diferencia entre la clase real  $y_i$  y la clase predicha  $\hat{y}_i$ .

El término de regularización se define como  $\Omega(f_k) = \gamma T + \frac{1}{2}\lambda\|w\|^2$ , siendo  $w$  los pesos de las hojas correspondientes al árbol actual  $f_k$ , y  $\gamma$  con  $\lambda$  hiperparámetros definidos por el usuario. El término adicional de regularización cumple la función de penalizar la complejidad del modelo. Agregar éste ayuda a mejorar la generalización y con ello evitar también el sobreajuste. Si  $\Omega = 0$  la función objetivo se convierte en la de un *Gradient Tree Boosting*.

XGBoost va aprendiendo la función a minimizar de manera aditiva, generando agrupaciones de árboles de decisión secuenciales durante la iteración del algoritmo. En cada paso se agrega un nuevo árbol que busca corregir el error de los anteriores optimizando la función objetivo al considerar en la predicción actual la predicción anterior. Formalizando, si se considera  $\hat{y}^{(t)}$  la predicción en la iteración número  $t$ , y  $f_t(x)$  la estructura de árbol añadida en la misma iteración se tiene que:

$$\hat{y}^{(t)} = \hat{y}^{(t-1)} + f_t(x). \quad (2.3)$$

Luego, una aproximación de segundo orden es utilizada para optimizar la función objetivo de manera rápida. Esto es calcular el gradiente ( $g$ ) y la hessiana ( $h$ ) (primera y segunda derivada) de la función de pérdida con respecto a la salida predicha en cada instancia.

Aplicando esto desde el principio, XGBoost comienza con un árbol vacío que sucesivamente se divide en sus hojas para minimizar la ecuación (2.4) que corresponde a la ganancia del sistema. Aquí  $I_j$  es el conjunto de instancias presentes para la hoja  $j$ , y  $q$  el director encargado de localizar cada muestra al índice de la hoja correspondiente. Por un tema de costo computacional considerar todas las posibles estructuras de árboles no es viable, por lo que se usa el algoritmo *greedy* [31] construyendo el siguiente árbol a partir de la mejor hoja.

$$\mathcal{L}^{(t)}(q) = -\frac{1}{2} \sum_{j=1}^T \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T. \quad (2.4)$$

La ecuación (2.4) se obtiene al agregar la aproximación de segundo orden sobre la ecuación (2.2), expandir los términos  $f_t$  y  $\Omega$ , y remover las constantes por simplicidad. Para mayor detalle de la matemática detrás del algoritmo se sugiere referirse a la publicación original [4]. Notar que la ecuación de ganancia combina los términos de regularización y la reducción de pérdida, entregando en un solo paso el balance óptimo entre una buena predicción y disminución de complejidad. Esto produce eficiencia en cuanto a tiempo de ejecución y reducción de sobreajuste.

XGBoost toma en cuenta que comúnmente las bases de datos representativas a la realidad poseen valores faltantes en sus muestras. La manera que utiliza para abordar esto es añadir una dirección por defecto en cada nodo como se muestra en la Figura 2.4, la cual es optimizada a partir de los datos de entrenamiento. En esta se aprecia como cada muestra “X” es capaz de llegar al final del árbol aunque no tenga todos los valores en las características. Con esto se obtiene un modelo que no solo puede ser utilizado con anotaciones faltantes, si no que aprende de ellas y las considera al momento de la predicción.

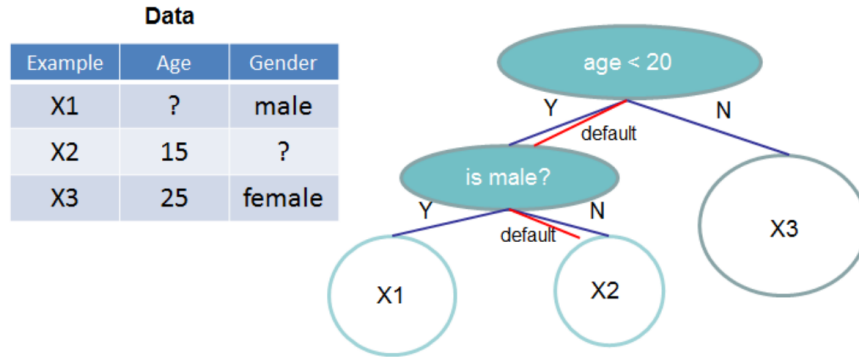


Figura 2.4: Representación de un árbol de decisiones con direcciones por defecto para características faltantes. La hoja a la que llega cada ejemplo X depende del valor de sus características, y de no poseer se toma el camino por defecto marcado por las líneas rojas. (Figura tomada de [4]).

Al momento de entrenar un XGBoost hay que tener en cuenta los hiperparámetros configurables por el usuario. Existe un gran número de estos por lo que se presentan a continuación una selección acotada de los hiperparámetros clave para el entrenamiento, con su método de Python entre paréntesis. El *learning rate*  $\eta$  (eta) supervisa el tamaño del paso al actualizar la función en cada iteración, acotando el peso de cada árbol para prevenir sobreajuste. La **profundidad máxima** de cada árbol (`max_depth`) da control sobre la complejidad de éste. El **número de árboles** (`n_estimators`) equivalente a la cantidad de rondas de *gradient boosting* a computar. Se pueden considerar también los términos de penalización  $\lambda$  (`reg_lambda`) y  $\gamma$  (`gamma`), los que al ajustarlos varían el comportamiento de la regularización [32, 33].

Las ventajas de XGBoost se pueden resumir en su alto desempeño en tareas de clasificación y regresión alcanzando el estado del arte de manera consistente en varias competiciones y problemas de la vida real [4]. Su buen rendimiento va de la mano con su capacidad de escalar grandes bases de datos gracias a su algoritmo basado en *Gradient Boosting* que forma el robusto modelo final. El algoritmo está constantemente pendiente de evitar el sobreajuste al agregar el término de regularización en la optimización. A esto se le añade su capacidad de aprender con valores faltantes en las bases de datos, característica que lo hace particularmente sobresaliente frente a modelos similares. Por otro lado, es importante tener en cuenta su limitación sobre bases de datos desbalanceadas, como es común en los algoritmos de árboles de decisión.

### 2.2.5. Análisis SHAP

*Shapley Additive Explanations* (SHAP) [34] es una herramienta de análisis de modelos cuyo objetivo es explicar la predicción de algoritmos de IA, a partir de la contribución de cada

atributo sobre la predicción. Su funcionamiento se basa en teoría de juegos cooperativa, en donde se calculan los valores *Shapley* [35], que son los que miden el impacto de cada jugador en el resultado final. Aterrizando esto al modelo de IA, se pueden tomar las características de entrada como los jugadores y el resultado final como la predicción del modelo.

Así, el algoritmo entrega un índice llamado “valor SHAP”, cuya única restricción es que su suma total sea 1 dentro de la misma variable. Un valor SHAP positivo implica una contribución positiva al modelo, así como un valor negativo implica una contribución negativa. Es decir, para un modelo de predicción de dos clases, 0 y 1, un SHAP *value* positivo lleva al modelo a clasificar hacia la clase 1, un valor negativo lleva al modelo a clasificar hacia la clase 0 y un valor cero no aporta en la clasificación.

Para entender como leer un análisis SHAP, en la Figura 2.5 se muestra un ejemplo de éste que relaciona mortalidad con las variables clínicas Frecuencia Cardiaca, Edad y Proteína C Reactiva (PCR)<sup>1</sup>. En este las variables se encuentran ordenadas según su contribución al modelo de arriba hacia abajo, siendo Frecuencia Cardiaca la que más aporte tiene. Cada punto en el gráfico representa un dato, y cada variable posee todos los pacientes en su eje horizontal, teniendo así una cantidad de puntos igual en cada fila.

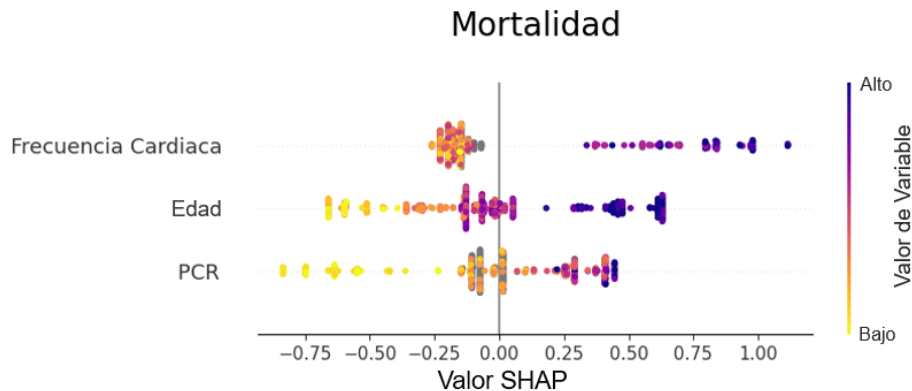


Figura 2.5: Ejemplo de análisis SHAP de la relación entre mortalidad con las variables clínicas Frecuencia Cardiaca, Edad y PCR. Un Valor SHAP positivo se correlaciona con fallecer, y uno negativo con no fallecer. El color de cada punto representa el valor relativo de la variable, según el código de colores mostrado a la derecha de la Figura.

Como se ve en el código de colores a la derecha, el valor de la variable es más alto si el color del dato es más oscuro y en grados de azul, y más bajo si el color es más claro y en grados de amarillo. Estos valores son relativos a cada variable, por lo que por ejemplo una PCR de 9 (mg/L) se marcaría como un color azul muy oscuro, mientras que una edad de 30 años de un amarillo claro. Los puntos grises, representan valores faltantes. Finalmente, en el eje horizontal se distribuyen los datos según sus valores SHAP, los cuales se posicionan más hacia la derecha mientras más positivo sea el valor SHAP y más hacia la izquierda a más negativo sea el mismo. Un valor SHAP positivo está asociado a la clase 1, que en este caso sería probabilidad de fallecer y uno negativo a la clase 0 con mayor probabilidad de no fallecer.

<sup>1</sup>No confundir con el examen de identificación de COVID-19 del mismo nombre.

Del análisis SHAP se da a interpretar que mientras más alta sea la Frecuencia Cardiaca, más se induce a la clase “Fallece”, mientras que una Frecuencia Cardiaca menor es más propensa a la clase “No Fallece”. Esta segunda presenta menos relevancia en el ejemplo, teniendo un gran cúmulo de puntos cerca del -0.25, mostrando que tener esta variable en valor promedio produce un impacto similar en el modelo a tener un valor más bajo. Tomando como ejemplo también la variable “Edad”, el análisis SHAP da a interpretar que una persona con mayor edad tiene más probabilidad de fallecer. Similar ocurre con “PCR”, en donde esta vez se ve el cúmulo de valores bajos más esparcidos, sugiriendo que esta variable tiene un gran aporte para discernir la clase 0.

Es importante tener en cuenta que el análisis SHAP no entrega información sobre que tan eficiente es el modelo, y no debe ser usado como métrica de elección sino como una herramienta de análisis.

## 2.2.6. Métricas de evaluación

### Matriz de Confusión

La matriz de confusión es una representación del desempeño de un clasificador, que permite resumir el resultado de manera visual. Se considera como “positivo” a un dato en el que es detectado el evento buscado, y “negativo” en el que no. Los conceptos a tener en cuenta son:

- **Verdadero Positivo (VP):** El dato positivo es predicho correctamente.
- **Verdadero Negativo (VN):** El dato negativo es predicho correctamente.

Etiqueta Real	Negativo (0)	VN	FP
	Positivo (1)	FN	VP
		Negativo (0)	Positivo (1)
		Etiqueta Predicha	

Figura 2.6: Ejemplo de una matriz de confusión de dos clases. (VN: Verdadero Negativo, FP: Falso Positivo, FN: Falso Negativo, VP: Verdadero Positivo)

- **Falso Positivo (FP):** El dato positivo es predicho como negativo. También conocido como error de tipo 1.
- **False Negativo (FN):** El dato negativo es predicho como positivo. También conocido como error de tipo 2.

Estos 4 conceptos son los que componen la matriz de confusión como se muestra en la Figura 2.6. La diagonal de ésta (en verde) representa los valores correctamente predichos. Se quiere que el modelo posea la mayor cantidad de verdaderos posibles, minimizando los falsos negativos y positivos.

## Curva ROC

La curva ROC (*Receiver Operating Characteristic*) es una forma gráfica de mostrar la eficiencia del modelo dando énfasis a los datos “positivos” correctamente predichos. Para esto se utiliza la tasa de verdaderos positivos (TVP) y la tasa de falsos positivos (TFP) los cuales se definen como:

$$TVP = \frac{\text{Verdaderos Positivos}}{\text{Todos los Positivos}} = \frac{VP}{VP+FN}, \quad (2.5)$$

$$TFP = \frac{\text{Falsos Positivos}}{\text{Todos los Negativos}} = \frac{FP}{FP+VN}. \quad (2.6)$$

Las curvas se grafican sobre las tasas como se ve en la Figura 2.7. Se desea maximizar la tasa de verdaderos positivos y disminuir la de falsos positivos, por lo que un clasificador es mejor mientras más cercana sea su curva al eje vertical. En otra palabras, una curva representa un mejor clasificador con respecto a otro si su curva esta sobre este. En el ejemplo de la Figura, la curva azul muestra un mejor desempeño que la amarilla. En este gráfico, la diagonal simboliza un valor igual para la tasa de verdaderos positivos y de falsos positivos, lo que se traduce en una clasificación aleatoria, implicando que el modelo no logra predecir.

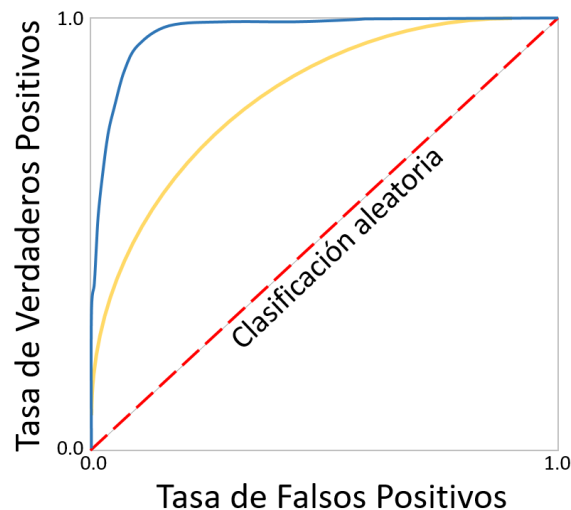


Figura 2.7: Ejemplo de curvas ROC. La curva azul representa una ejemplo de un modelo con mejor desempeño que la amarilla.

## Área bajo la Curva ROC

Dado que un clasificador tiene mejor desempeño que otro si es que su curva ROC esta por encima, se puede utilizar como métrica el área bajo la curva (o “ROC Área” en adelante), la cual se ejemplifica en la Figura 2.8. Esta área puede tomar valores de cero a uno, y un área más grande implica una curva de mejor desempeño.

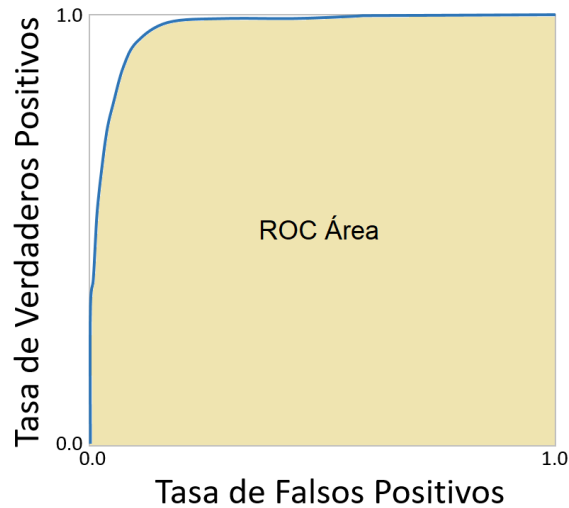


Figura 2.8: Representación gráfica del área bajo la curva ROC.

## *Accuracy* o Exactitud

Es la razón entre los datos correctamente predichos y el total de datos. Se interpreta como el porcentaje de aciertos en la clasificación o predicción.

$$\text{Accuracy} = \frac{\text{Correctamente Predichos}}{\text{Todos los datos}} = \frac{VP + VN}{VP + VN + FP + FN}. \quad (2.7)$$

## *Precision*

Es la razón entre los verdaderos positivos y el total de positivos predichos. Entrega información sobre el acierto en la predicción de verdaderos positivos. Se puede interpretar como el porcentaje correctamente predicho dentro de las predicciones consideradas como positivas por el modelo.

$$\text{Precision} = \frac{VP}{VP + FP}. \quad (2.8)$$

## *Recall* o Sensitividad

Es la razón entre verdaderos positivos y el total de valores positivos. Entrega información sobre el acierto en la predicción de los casos positivos. Se puede interpretar como la probabilidad de que un caso predicho como positivo efectivamente lo sea.

$$\text{Recall} = \text{Sensitividad} = \frac{VP}{VP + FN}. \quad (2.9)$$

## Especificidad

Es la razón entre los verdaderos negativos y el total de valores negativos. Entrega información sobre el acierto en la predicción de los casos negativos. Similar al anterior, se puede interpretar como la probabilidad de que un caso predicho como negativo efectivamente lo sea.

$$\text{Especificidad} = \frac{VN}{VN+FP}. \quad (2.10)$$

A modo de resumen, en la Figura 2.9 se muestra la relación de las métricas con la matriz de confusión. Aquí se representa de manera visual de que habla cada métrica de desempeño.

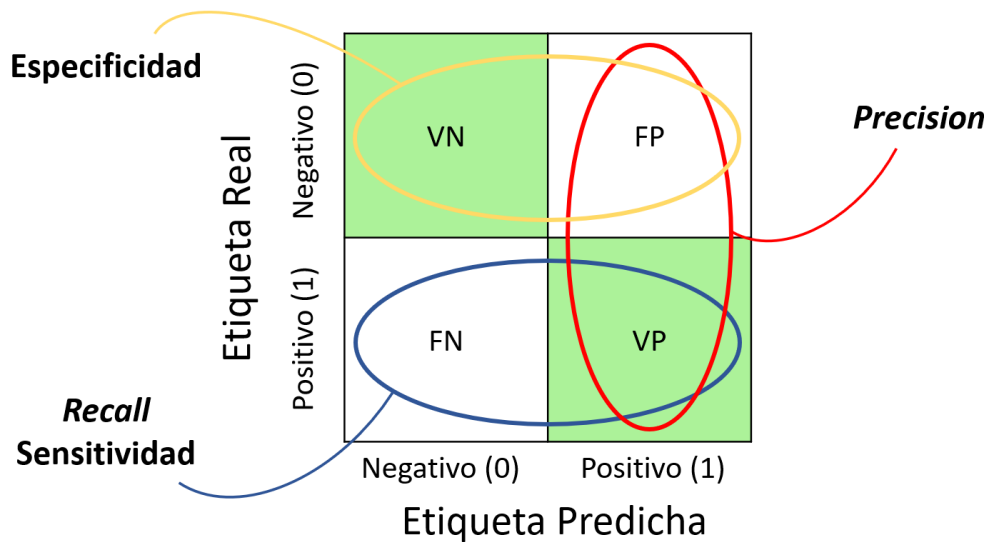


Figura 2.9: Relación entre matriz de confusión y métricas de desempeño.

## Diagrama de Caja

El diagrama de caja o *boxplot* es un método de descripción gráfica de datos numéricos a partir de sus cuartiles. Se caracteriza por ser robusto frente a valores atípicos u *outliers*, a diferencia de métricas estadísticas como el promedio y la desviación estándar. Cada cuartil Q está formado por un 25% de los datos, excluyendo *outliers*.

En la Figura 2.10 (a) se presenta un ejemplo a modo de definición del diagrama. Un *boxplot* está formado por su caja, correspondiente al 50% de los datos pertenecientes a los cuartiles centrales (Q2 y Q3), y los *whiskers* formados por los cuartiles externos Q1 y Q4. La mediana, presente dentro de la caja, es demarcada con una línea naranja, y calza con el valor máximo del segundo cuartil Q2. La mediana permite hacer un análisis de como se distribuyen los datos, ya que siendo ésta la mitad, es posible interpretar la concentración presente en dentro de la caja y la dispersión hacia los *whiskers*.

Los *outliers* no son considerados dentro de los cuartiles, y para definir que valores son atípicos, se calcula el rango de los cuartiles centrales o *Inter Quartil Range* (IQR) que es la diferencia entre el valor máximo y mínimo de la caja. El rango es ponderado usualmente por



1.5, denotando que cualquier valor alejado de la caja 1.5 veces más que su largo se considera *outlier*:

$$outlier < Min(Q2) - 1,5 \cdot IQR, \text{ ó, } outlier > Max(Q3) + 1,5 \cdot IQR. \quad (2.11)$$

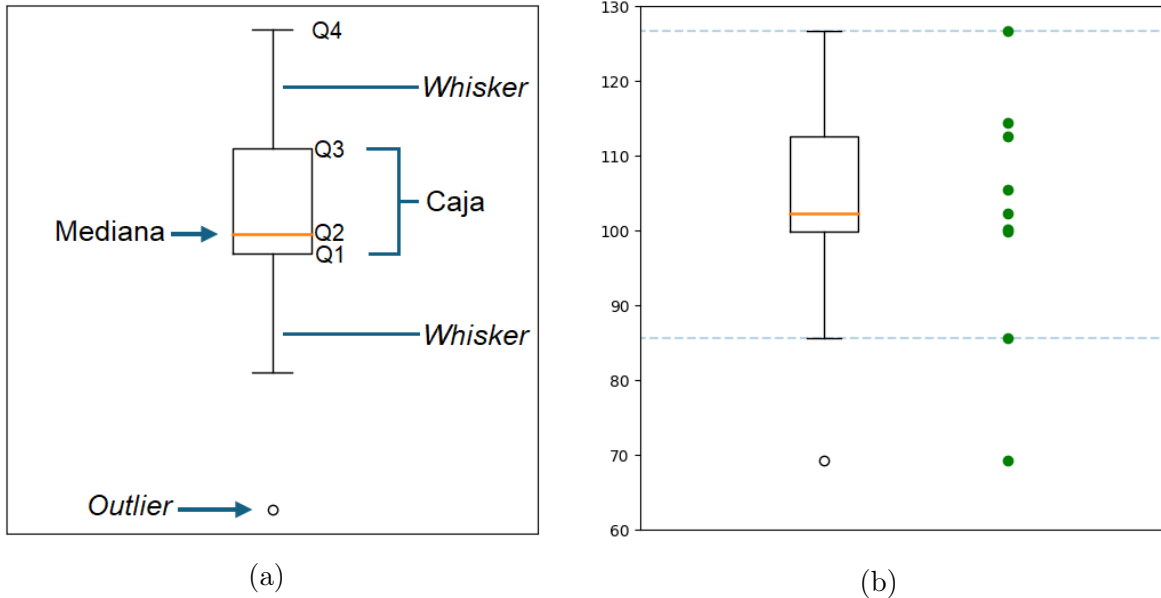


Figura 2.10: Explicación y demostración de Diagramas de Caja. En (a) se muestra un ejemplo descriptivo de lo que corresponde cada parte del diagrama. Q1 hasta Q4 son los cuartiles del 1 al 4. En (b), el mismo *boxplot* junto a los datos que lo formaron (puntos verdes). Las líneas horizontales punteadas son una demostración de que considera el diagrama como máximo y mínimo, excluyendo *outliers*.

En la Figura 2.10 (b) se puede ver el *boxplot* al lado de los datos con los que fue generado. Las líneas punteada horizontales demarcan el valor máximo de Q4, que es el borde del *whisker* superior y el valor máximo de los datos, y el valor mínimo de Q1, que es el borde del *whisker* inferior y el mínimo de los datos sin incluir *outliers*. Éste último es un dato de alta dispersión, marcado en su valor exacto dentro del *boxplot*.

# Capítulo 3

## Propuesta de Sistema para Predicción de Eventos Clínicos Adversos

### 3.1. Presentación del Problema

La Ventilación Mecánica Invasiva (VMI) es un procedimiento de respiración artificial que permite asistir al paciente con su suministro de Oxígeno a partir de intubación endotraqueal. La infección producida por el COVID-19 produce obstrucción pulmonar, la cual de estar muy avanzada, genera complicaciones en la oxigenación del paciente. Como los ventiladores son limitados, elegir quien será asignado para VMI es crucial. Con el afán de generar herramientas para apoyar a los médicos en sus decisiones en el servicio de urgencias, se plantea generar un sistema de Inteligencia Artificial (IA) que permita obtener la probabilidad de que un paciente hospitalizado necesite de VMI o fallezca en el corto plazo.

La predicción de los eventos debe ser acotado por un tiempo máximo, puesto que el estado del paciente ahora no será el mismo que en el futuro, y es dicho estado el que permite discernir si el paciente presentara un evento clínico. A modo de ejemplo, si un paciente requerirá de ventilación mecánica invasiva en dos semanas más, entregar un resultado positivo ahora generaría una alarma sobre un escenario no prioritario. Desde el punto de vista del modelo, considerar al mismo paciente como positivo en su estado actual siendo que requerirá ventilación mecánica en dos semanas, induciría ruido en el entrenamiento, pues estaría etiquetado como que requiere VMI siendo que su estado de salud aún no es tal. Esto confundiría al algoritmo, el que recibiría dos estados de salud muy distintos, el de ahora y el de dos semanas en el futuro cuando si requiera VMI, ambos con la misma etiqueta.

Consultando con médicos del servicio de urgencia colaboradores del proyecto, se concluye que un tiempo adecuado para la predicción, es si el paciente requerirá VMI en las próximas 48 horas, o si fallecerá en los próximos 5 días (Tabla 3.1).

Tabla 3.1: Cantidad de días que se proponen a predecir los eventos clínicos adversos.

	VMI	Mortalidad
Ventana de Predicción	2 días	5 días

Durante los primeros 100 días de pandemia, médicos del Hospital Clínico de la Universidad de Chile (HCUCH) llevaron a cabo un estudio prospectivo con 395 pacientes COVID-19 positivos que ingresaron al servicio de urgencias <sup>1</sup>. Haciendo uso de ésta base de datos se realiza una exploración del problema a partir de UMAP [23], tomando ventaja de que estos permiten analizar los datos sin perder singularidad. Por conocimiento experto se eligen de manera preliminar las siguientes variables:

Edad, Saturación de Oxígeno, Frecuencia Respiratoria, Frecuencia Cardíaca, Presión Arterial Sistólica máxima y mínima, Presión Arterial Media, Compromiso de Conciencia, GCS (*Glasgow Coma Scale*), Fracción Inspirada de Oxígeno, Lactato, Presión Arterial de Oxígeno, Presión Arterial de Dióxido de Carbono, PaFi, uso de Cánula de Alto Flujo, Disnea, Prono Vigil, Comorbilidades (20 en total), cantidad de Comorbilidades y Severidad de Daño Hepático.

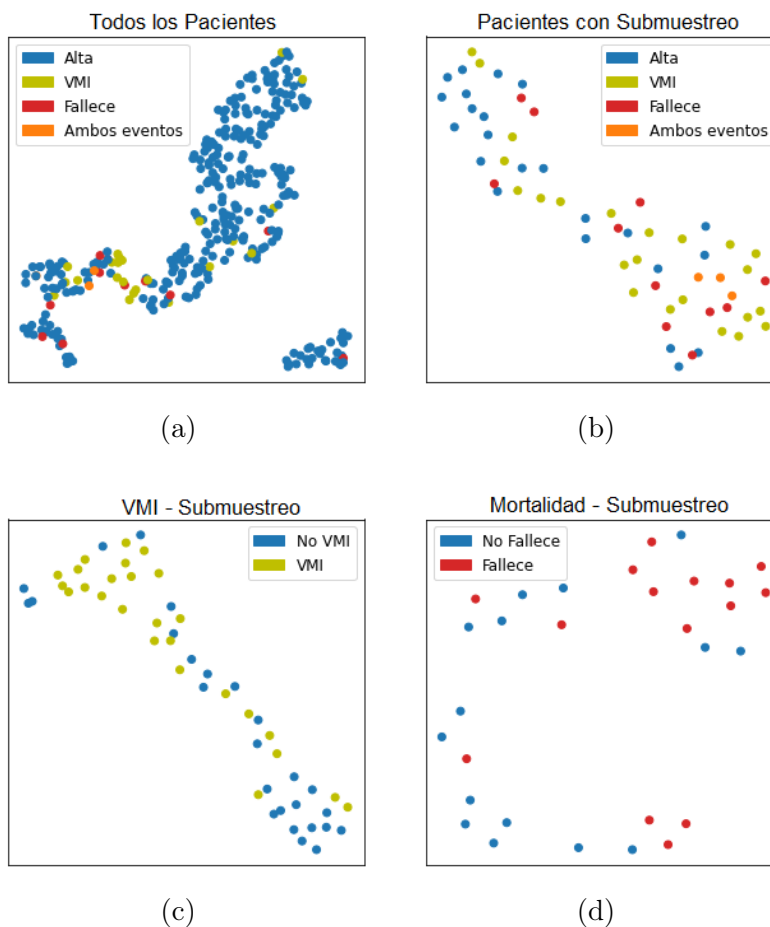


Figura 3.1: Muestras de UMAP para los Datos Estáticos, en los cuales cada punto representa un paciente de los 395. Los pacientes con eventos “Indicación de VMI” y “Fallece” se encuentran resaltados de color amarillo y rojo respectivamente. Si un paciente sufre ambos eventos, se encuentra resaltado en color naranja, y si no sufre ninguno (paciente de “Alta”) de color azul. En (a): Proyección de todos los pacientes. En (b): Proyección luego de submuestreo en torno al evento “Fallece”. En (c): Proyección solo de pacientes “VMI” y “No VMI” con submuestreo en torno al primero. En (d): Proyección solo de pacientes “Fallece” y “No Fallece” con submuestro en torno a “Fallece”.

<sup>1</sup>Las bases de datos se detallan en la siguiente subsección. De ser de interés inmediato, referirse a 3.2.2

En la iteración del algoritmo, mostrado en la Figura 3.1, se logra ver que los diferentes eventos se agrupan de manera más clara dentro de los cúmulos cuando son sometidas a un proceso de balanceo, en este caso submuestreo. En la Figura 3.1 (a) se ve una muestra de UMAP para los 395 datos y en (b) una con submuestreo aplicado a pacientes que no sufren eventos clínicos negativos. En esta última los pacientes sin eventos tienen tendencia a agruparse en el extremo superior izquierdo, mientras los pacientes positivos para “Indicación de VMI” y “Fallece” en el inferior derecho, teniendo por supuesto excepciones en ambos grupos.

Aquí, en la Figura 3.1 (b), se puede notar que separar los pacientes que sufren el evento “Indicación de VMI” de los que sufren “Fallece” dentro del cúmulo de puntos no es directo. Para abordar esto se decide iterar el algoritmo para cada evento clínico por separado, lo que se muestra en la Figura 3.1 (c) y (d), de las cuales en la primera se aprecia como los pacientes que irán a VMI en los próximos dos días tienden a acumularse en el sector superior izquierdo y los que no en el otro extremo. Similar, en (d), los pacientes que fallecen en los próximos cinco días tienden a agruparse en los extremos del cúmulo hacia la derecha, y los que no, en el resto de la curva.

Este análisis del problema sugiere generar dos modelos, uno exclusivo para la predicción de VMI y otro para la predicción de Mortalidad, en vez de un gran modelo que trate de predecir ambas clases. También sugiere que balancear las clases ayuda a mejorar el desempeño de la predicción sobre el escenario. Con esto en mente, la propuesta es diseñar dos modelos de predicción: Un modelo de predicción de VMI con las clases “VMI”, si el paciente sufrirá el evento clínico en los próximos dos días, y “No VMI” si no; Un modelo de predicción de Mortalidad con las clases “Fallece”, si el paciente fallecerá en los próximos cinco días, y “No Fallece” si no.

## 3.2. Datos Utilizados

### 3.2.1. Adquisición de Datos

Debido a que se desean balancear los datos a partir de submuestreo para ambos eventos, se reduce significativamente la cantidad de datos disponibles. Esto ocurre por que los datos tomados corresponden al día de ingreso al servicio de urgencias, por lo que cualquier paciente que va a VMI al tercer día o después no se considera como “VMI” positivo según la definición descrita. Esto se replica para el estudio de mortalidad.

El escenario ideal sería obtener acceso a datos de más pacientes incluso después de los primeros 100 días de la pandemia, pero esto está limitado por el sistema de datos utilizados en el Hospital Clínico de la Universidad de Chile (HCUCH). Obtener la fecha de ingreso, la ocurrencia de eventos, los datos correspondientes y su seguimiento, es un trabajo que se debe hacer de manera manual, además de que muchos de los datos son escritos a mano en papel, lo cual sumado al escenario crítico en el que se encontraba el centro de salud lo convierte en una solución no factible.

Una opción más viable es recolectar los datos de signos vitales y exámenes de laboratorio, pues estos están semi-automatizados por sistema. A partir de éstos se tienen datos clínicos en función del tiempo, o longitudinales, para cada paciente, desde el momento de ingreso hasta la ocurrencia del evento final indicación de VMI o Fallece. En caso de que ninguno de los dos eventos ocurran los datos se recolectan hasta la fecha de alta o censura del estudio. Con esto se tienen datos estáticos correspondientes al estudio realizado los primeros 100 días de la pandemia, y datos longitudinales de signos vitales y exámenes de laboratorio para los mismos pacientes.

### **3.2.2. Conjunto de Datos**

Los datos estáticos, correspondiente a la base de datos obtenida por profesionales del HCUCH, y los datos longitudinales, correspondientes a las bases de datos de signos vitales y exámenes de laboratorio, presentan diferentes formatos e información. Por ello se detallan por separado a continuación.

#### **Datos Estáticos**

Un total de 395 pacientes COVID-19 positivos fueron estudiados a través de entrevista clínica, revisión de ficha clínica electrónica y exámenes de laboratorio en un cohorte prospectivo durante los primeros 100 días de la pandemia, entre el 03-03-2020 y el 10-06-2020 en el HCUCH. La información está contenida en un archivo Excel.

En un comienzo la base de datos cuenta con 493 pacientes ingresados, reduciéndose a los mencionados 395 según los siguientes criterios de inclusión:

- Admisión en el servicio de urgencias del HCUCH.
- COVID-19 positivos.
- Pacientes sintomáticos.
- Mayor de edad.
- Al menos 24 horas de estadía en el hospital previo al alta.
- No presenta embarazo.
- No presenta infección intrahospitalaria.

El estudio se encuentra distribuido en 278 variables las cuales se encuentran agrupadas en:

- I. Identificación (3)
- II. Demográfica (3)
- III. Comorbilidades (23)
- IV. Síntomas (15)
- V. Signos vitales y examen físico (9)

- VI. Laboratorio (24)
- VII. Complicaciones (9)
- VIII. Evaluación clínica (12)
- IX. Terapias farmacológicas (21)
- X. Intervenciones (24)
- XI. Eventos (33)
- XII. Otras (102)

El detalle de éste se muestra en la Tabla Anexo B.1 en donde se presentan todas las variables. Cada una posee una única muestra correspondiente al peor valor fisiológico medido el día de ingreso. De los 395 pacientes, **69** de ellos entran a ventilación mecánica, **58** fallecen, y 26 presentan ambos eventos (éstos están incluidos en los anteriores), lo que se muestra gráficamente en la Figura 3.2. La información se ve resumida en la Tabla 3.2.

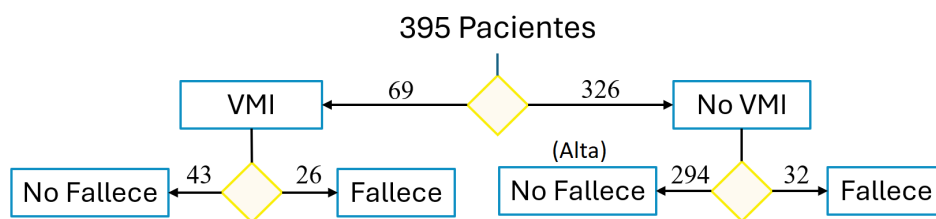


Figura 3.2: Distribución de los 395 pacientes en los eventos clínicos presentados.

Tabla 3.2: Cantidad de pacientes en cada evento. 26 pacientes presentan ambos eventos, VMI y Fallece, los cuales se incluyen en ambas columnas. “Alta” considera los pacientes que no realizan ninguno de los dos eventos anteriores.

Total	VMI	Fallece	Alta
395	69	58	294

## Signos Vitales

Base de datos extraída por el departamento de Tecnología e Información (TI) del HCUCH a petición del proyecto. Contiene datos de signos vitales de 714 pacientes hospitalizados por sospecha de COVID-19 entre el 03-03-2020 y el 03-07-2020, dentro de los cuales se encuentran los 395 pacientes del estudio original. Comprende de datos longitudinales durante las primeras 96 horas desde el ingreso al servicio de urgencias, es decir, cada variable tiene el registro de los valores y el momento (fecha y hora) en que el dato fue tomado y cada vez que fue tomado.

El contenido está distribuido en 7 archivos Excel, en donde cada pestaña representa un paciente distinto. La base de datos consta de las siguientes variables:

1. Temperatura Axilar.
2. Frecuencia Cardiaca.
3. Presión arterial sistólica (Pas), diastólica (Pad), media (Pam).
4. Escala Visual Análoga (EVA).
5. Hemoglucotest (HGT).
6. Uso de Insulina Cristalina.
7. Frecuencia Respiratoria.
8. Saturación de Oxígeno (SatO<sub>2</sub>).
9. Flujo de Oxígeno.
10. Fracción inspirada de Oxígeno (FiO<sub>2</sub>).
11. Dispositivos de Oxígeno.

## Exámenes de Laboratorio

Base de datos extraída por el departamento de TI del HCUCH a petición del proyecto. Contiene datos de exámenes de laboratorio de 714 pacientes hospitalizados por sospecha de COVID-19 entre el 05-01-16 y el 18-12-20, dentro de los cuales se encuentran los 395 pacientes del estudio original. Comprende de datos longitudinales de todos los exámenes de laboratorio hechos al paciente dentro de las fechas.

El contenido se encuentra en un archivo Excel formato tabla, en donde cada pestaña representa un examen diferente. La base de datos consta de las siguientes variables:

- |                                      |   |
|--------------------------------------|---|
| 1. Hemoglobina.                      | 14. Lactato.  |
| 2. Leucocitos.                       | 15. Troponina.  |
| 3. Neutrófilos.                      | 16. Alanina aminotransferasa (GPT).                             |
| 4. Plaquetas.                        | 17. Procalcitonina (PCT).                                       |
| 5. Linfocitos.                       | 18. Proteína C-reactiva (PCR).                                  |
| 6. Virus del Herpes Simple (VHS).    | 19. <i>International Normalized Ratio</i> (INR).                |
| 7. Albumina.                         | 20. Presión Arterial de Oxíneo (paO <sub>2</sub> ).             |
| 8. Nitrógeno Ureico.                 | 21. Presión Arterial de Dióxido de Carbono (pCO <sub>2</sub> ). |
| 9. Aspartato Aminotransferasa (GOT). | 22. Dímero D.   |
| 10. Lactato deshidrogenasa (LDH).    | 23. Tiempo de Tromboplastina Parcial activado (TTPa).           |
| 11. Bilirrubina Total.               | 24. Fibrinógeno.  |
| 12. Creatina plasmática (CREA).      |   |
| 13. Creatina quinasa (CK).           |   |

### 3.3. Propuesta de Sistema de Predicción

La implementación exitosa de un sistema de predicción de eventos clínicos requiere de un meticuloso trabajo interdisciplinario. Desde los datos recibidos que deben ser interpretados, pasando por el desarrollo de los modelos, hasta el análisis de resultados que tengan sentido clínico, son parte importante del trabajo considerando lo sensible que es el trabajar con datos de pacientes. En el presente se ahonda en el desarrollo desde el punto de vista del trabajo ingenieril.

En este contexto se propone realizar un sistema basado en IA de predicción de los eventos clínicos adversos indicación de VMI y mortalidad del paciente, considerando su estado de salud al momento de entrada al servicio de urgencias y la evolución de éste en el tiempo. Es posible utilizar modelos de clasificación para problemas de predicción si la clase a considerar es posterior a los datos asociados a ella. En este caso se opta por utilizar los árboles de decisión XGBoost, los cuales son particularmente adecuados a este caso dada la acotada cantidad de pacientes incluidos en el estudio, la capacidad de entrenar y predecir con datos faltantes y la información que provee sobre la importancia de cada variable utilizada. A continuación se detallan los pasos a seguir con los que se lleva a cabo esta tarea.

#### 3.3.1. Preprocesamiento de Datos

Como se expuso en la Sección 3.2, se tienen 3 grandes bases de datos: datos estáticos, signos vitales y exámenes de laboratorio; cada una de ellas registradas en archivos excel con su propio formato. Estos presentan imperfecciones en cuanto a calidad de datos, debido a que son anotados manualmente induciendo errores humanos comunes.

Las bases de datos poseen inconsistencias en las anotaciones relacionadas por ejemplo con el nombre de las variables, sistemas de medida y duplicación de datos, entre ellas y en cada una de ellas por separado. La primera tarea entonces es limpiar cada una de estas. Para ello se abordaron los siguientes puntos para cada base de datos por separados y en conjunto:

- **Nombre de las Variables:** Una variable clínica puede ser anotada de más de una forma. Tomando el ejemplo de “Fracción Inspirada de Oxígeno” ( $\text{FiO}_2$ ), esta puede ser registrada por su nombre o su abreviación recién mostradas, con lo que tenemos dos maneras de escribirla. Hay que considerar también posibles variaciones que el autor pueda emplear (Fracción Inspirada de  $\text{O}_2$ , FI de Oxígeno, Fracción I.  $\text{O}_2$ , etc.). Por último existen errores humanos producidos al momento de anotar como faltas ortográficas.

Una meticulosa revisión de las variables es llevada a cabo con ayuda de conocimiento experto para identificar cada variable única presente en las bases de datos. Una vez identificadas se procede a dar consistencia a cada una dentro de su propio set, uniendo finalmente todas las variables con el mismo nombre dentro de una sola.

- **Sistemas de medida y formato:** Una misma variable puede ser escrita considerando más de un sistema de medida y/o formato. Para enfrentar esto se identifican las variables cuyos valores no son consistentes y se estandarizan a partir de código en sistema de medida internacional (sistema métrico) o el más utilizado en la práctica clínica. En este punto entran también las fechas y variables temporales.



- **Variables no significativas:** En la práctica clínica existen exámenes y similares que son tomados solo de forma ocasional. Traducido a los datos, esto significa que no existen suficientes pacientes que posean éstas variables como para que sea significativa. Ejemplos de esto son la “Insulina Cristalina” y los datos imagenológicos como los rayos-X. Se identifican las variables no significativas y se desechan de ser correspondiente.
- **Variables derivadas:** Existen variables que se pueden obtener a partir de otras, lo que se toma como una oportunidad de agregar los valores faltantes de éstas. Tomando como ejemplo las variables de presión arterial, éstas están relacionadas a partir de la siguiente expresión:

$$P_{am} = \frac{P_{as} + 2 \cdot P_{ad}}{3}. \quad (3.1)$$

donde  $P_{am}$  es la Presión Arterial Media,  $P_{as}$  la Presión Arterial Sistólica y  $P_{ad}$  la Presión Arterial Diastólica. Dependiendo de quien anote estos datos en cada paciente y momento, puede decidir si anotar las tres variables o cualquier par de estas sin perder información. En el caso de este ejemplo en particular se decide mantener las 3 variables, calculando las que falten.

- **Datos duplicados:** En el manejo de la base de datos, en el momento de hacer las anotaciones o en el traspaso de información entre archivos, es común que ocurran errores humanos como registrar más de una vez a un paciente, una variable, un valor, y más. A partir de código se hace revisión de las bases de datos asegurándose de que los datos únicos ocurran una sola vez, eliminando los datos duplicados.

Con los datos homogeneizados se pueden combinar las bases de datos y darles un formato amigable con las necesidades del problema. Debido a que las variables presentes en la base de datos estáticos solo poseen un valor, es preferible considerar los signos vitales y exámenes de laboratorio según sus propios set de datos con información temporal.

Las bases de datos longitudinales (signos vitales y laboratorio) son unificadas en un único archivo excel. En Tabla 3.3 se muestra una representación de los datos preprocesados y su formato de escritura. En ésta, los símbolos “x” representan donde hay datos en las variables, indicando que hay valores faltantes de manera frecuente en la base de datos, lo cual es normal pues no todos los signos vitales y/o exámenes de laboratorio se toman todo el tiempo ni al mismo tiempo. También se aprecia que cada paciente tiene una cantidad distinta de marcas de tiempo  $t$ , en donde estas describen en que momento se tomó el examen.

Con el preprocesamiento, los datos longitudinales son limitados a **34** variables, 10 de signos vitales y 24 de exámenes de laboratorio. La base de datos estáticos se mantiene de manera independiente, pues con el preprocesamiento previamente mostrado basta para ser utilizada.

Tabla 3.3: Representación del formato de los datos longitudinales preprocesados. En la primera fila  $sv$  de 1 a  $i$  simbolizan los nombres de las variables de los signos vitales, y  $el$  de 1 a  $j$  los exámenes de laboratorio. En las primeras columnas,  $p$  de 1 a 395 simbolizan los IDs únicos de los pacientes y  $t$  las marcas de tiempo con formato “aaaa-mm-dd HH:MM”. Los caracteres “x” representan donde existe un valor en la variable, puestos de manera aleatoria a modo de ejemplo.

ID	Tiempo	$sv_1$	$sv_2$	...	$sv_i$	$el_1$	$el_2$	...	$el_j$
$p_1$	$t_1$	x	x		x				
$p_1$	$t_2$					x			x
	$\vdots$								
$p_1$	$t_{k_1}$	x	x			x			
$p_2$	$t_1$		x						
	$\vdots$								
$p_2$	$t_{k_2}$	x	x		x	x		x	x
$\vdots$	$\vdots$								
$p_{395}$	$t_{k_{395}}$	x	x		x		x		

### 3.3.2. Extracción de Características

Ingresar los datos longitudinales a un modelo de ML no es directo. El principal problema es que, tal como se encuentran registrados los atributos, la cantidad de éstos difiere por paciente, por lo que no es posible utilizarlos como vector de características de entrada. Los nombres de las variables longitudinales son constantes y los mismos para todos los pacientes, por lo que se puede tomar ventaja de éstos para generar características homogéneas.

Se realiza extracción de características a partir de cada variable longitudinal eligiendo cinco estadísticas clave: promedio, desviación estándar, valor máximo, valor mínimo y pendiente. Ésta última se calcula proyectando una regresión lineal, considerando los valores disponibles y el tiempo de adquisición. Cada vez que exista una medición, se hace interpolación de tipo retenedor de orden cero [36] sobre variables con datos faltantes, con tal de tener una mejor proyección. Así, de cada variable, sin importar la cantidad de valores que ésta tenga, se obtendrán 5 características, logrando que cada dato o paciente tenga un extenso vector de características de tamaño estándar para todos, que además contempla la evolución temporal de cada variable.

Considerando el objetivo de predecir eventos clínicos adversos en un plazo cercano, se elige trabajar con valores correspondientes a un día de datos. Si se toman todos los datos que tiene un paciente desde el ingreso hasta el momento de hacer la predicción, puede ocurrir que se consideren días e incluso semanas de valores clínicos, abreviados en solo 5 características. Hay que tener en cuenta que si el paciente presenta algún evento clínico, son los valores cercanos al presente los que representan mejor el estado del paciente al momento de la decisión clínica. Por último, considerar todos los datos del paciente en el pasado en extracción de características, sugiere por generalización la pérdida de la variable temporal de éstos, y con ello la evolución del estado del paciente.

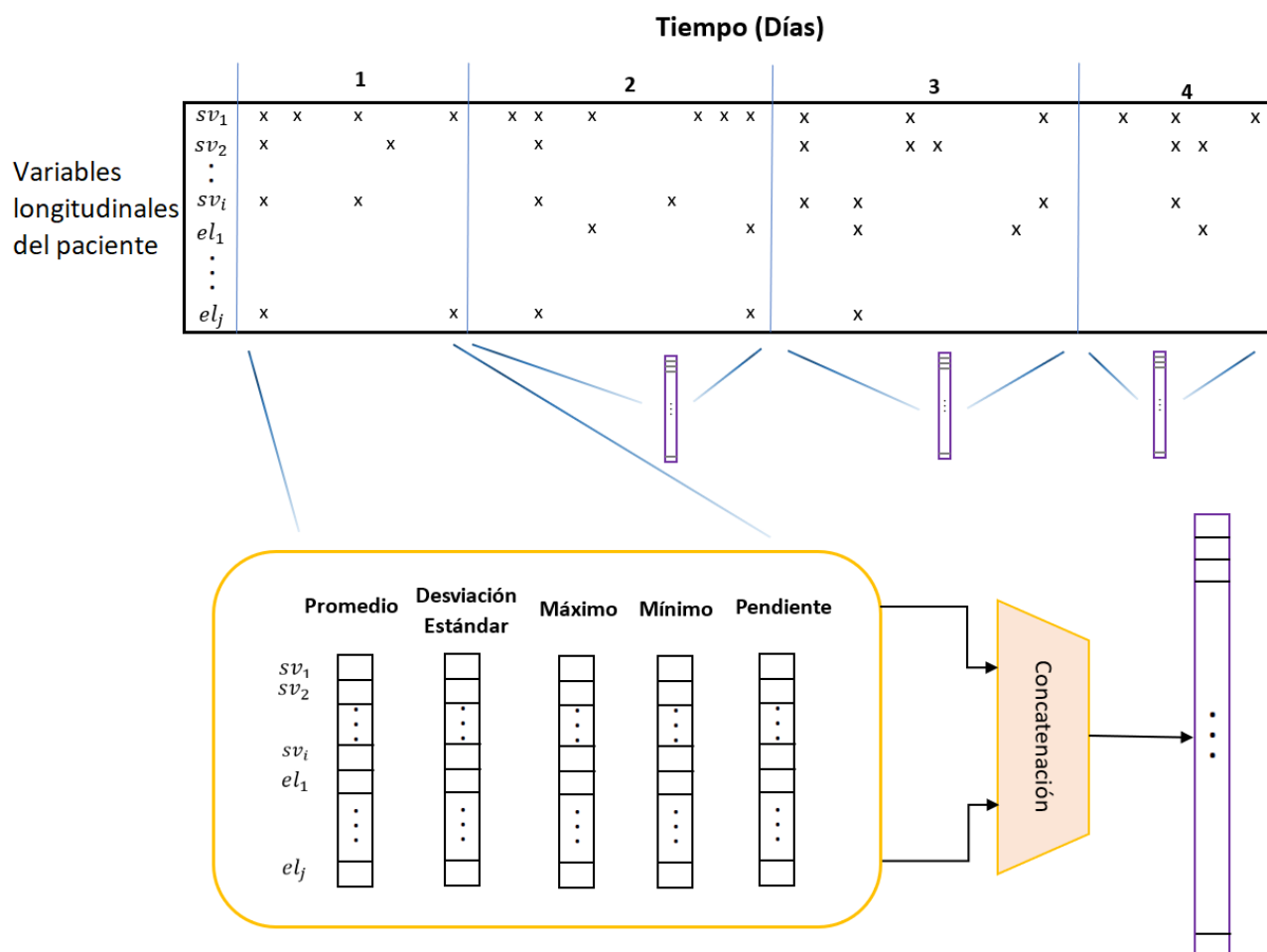


Figura 3.3: Descripción gráfica de la extracción de características sobre los datos longitudinales. En las variables longitudinales del paciente, las líneas verticales separan un día de valores clínicos.  $sv_1$  hasta  $sv_i$  representan las variables de signos vitales y  $el_1$  hasta  $el_j$  los exámenes de laboratorio. Por cada día se considera un vector a partir de extracción de características y concatenación.

Desde las 00:00 hrs hasta las 23:59 hrs se considera un día de datos, ventana de tiempo dentro de la cual se extraen las características. Se puede tener en cuenta un día aunque el paciente no esté las 24 horas en el servicio, basta que posea al menos una medición clínica en ese periodo. Por cada día que el paciente se encuentre hospitalizado y/o posea datos clínicos, es posible calcular un vector de características. Ésto se encuentra representado gráficamente en la Figura 3.3. Una ventaja de este sistema es que cada vez que se ingrese un dato nuevo del paciente, el modelo será capaz de considerar éste para la nueva predicción.

Las 34 variables longitudinales son procesadas utilizando las 5 estadísticas mencionadas obteniendo en total **170** características. Este forma parte del vector de características de entrada de los modelos.

### 3.3.3. Diseño de Modelos

Como se mencionó previamente, se propone utilizar XGBoost como modelo para la tarea de predicción de eventos clínicos adversos, considerando las siguientes razones:

- XGBoost ha estado consistentemente al nivel del estado del arte en diversos problemas de clasificación en cuanto a modelos de IA clásicos [5–8], y en particular para escenarios de medicina con bases de datos similares [4, 11–14, 16, 17].
- Entrega información explícita sobre la importancia de cada variable en la elección de la clase, lo cual da interpretabilidad al modelo, lo que es muy valorado por el personal de salud.
- Gracias a su arquitectura de árbol de decisión, la complejidad del algoritmo es relativamente baja en comparación a otros modelos de clasificación. Esto produce que muestre resultados aún en *datasets* pequeños, lo cual es una ventaja sobre la limitada cantidad de datos en el presente problema.
- Posee la cualidad de poder entrenar y discernir con características faltantes en los datos de entrada. Ésto es particularmente importante en tareas que por su naturaleza presentan datos sensibles, y que por ello se intenta evitar el uso de algoritmos de imputación de datos, como lo es el caso a trabajar.

Los dos eventos a tener en cuenta, Indicación de VMI y Mortalidad, se aproximan por separado cada uno con su propio modelo. En ambos casos se tienen dos clases: “VMI” y “No VMI”; “Fallece” y “No Fallece”. Como tal, se revisan a continuación ambos modelos.

#### Predicción de Indicación de Ventilación Mecánica Invasiva

La ventana de predicción seleccionada de si el paciente necesitará de ventilación mecánica es de dos días. En términos del modelo de predicción, se considera como “clase 0” si el paciente no va a VMI en los próximos dos días y “clase 1” si es que va al día siguiente o subsiguiente. Esta forma de organizar las clases tiene un par de implicancias. La primera es que un paciente puede presentar la clase 0 “No VMI” aunque si vaya a ventilación mecánica en el futuro. En la Figura 3.4 se muestra un ejemplo de paciente que ingresa a VMI en el cuarto día, produciendo que presente las clases 0, 1 y 1 en los tres días de estancia hospitalaria previos al evento clínico.

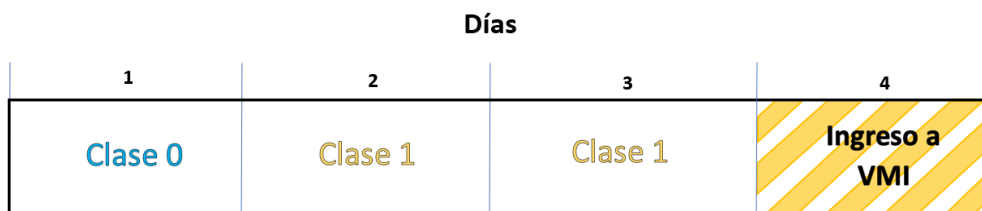


Figura 3.4: Ejemplo de un paciente y su organización de clases en el contexto del modelo de predicción de Ventilación Mecánica Invasiva. En este el paciente ingresa a VMI el cuarto día de hospitalización, por lo que presenta tres datos correspondientes a los días anteriores y sus respectivas clases.

La segunda implicancia es que esto incrementa la cantidad de datos, pues cada paciente está presente al menos un día en el servicio de urgencias<sup>2</sup>, y por cada día que esté hospitalizado después del de ingreso se obtiene un dato más. Así, la cantidad de datos que se puede obtener del *dataset* se describe como:

Cantidad total de datos: **6831**

Datos de pacientes que no realizan evento VMI (Clase 0): **1999**

Datos con evento VMI en las proximas 48 horas (Clase 1): **100**

Datos con evento VMI pero en más de 48 hrs (Clase 0): **111**

Datos que se encuentran en evento VMI o posterior (sin denominación): **4621**

De los 69 pacientes que son derivados a ventilación mecánica se obtienen 100 datos de la clase 1 (VMI) y 111 de clase 0 (No VMI). Se tienen además 1999 datos más para la clase 0 de pacientes que no realizan el evento VMI. Para abordar el desbalance de datos se realiza submuestreo de manera aleatoria sobre la clase 0 para el conjunto de entrenamiento. Debido a que el *dataset* consta de datos sensibles, es preferible evitar otras técnicas como por ejemplo sobremuestreo, que requiere generar datos. Las técnicas de aumentación permiten balancear y obtener bases de datos más grandes, pero en el escenario clínico como el presente es necesario un extenso trabajo que incluye validación experta de los datos generados, lo cual induce carga extra al cuerpo médico, que es lo que se quiere evitar.

Dividir los datos en los subconjuntos de entrenamiento y prueba no es tan directo como elegirlos de manera aleatoria, pues hay que tener en cuenta que los datos de un mismo paciente no deben quedar en subconjuntos distintos. Al momento en que el modelo mira los datos de un paciente al entrenar, éste se ajusta en torno al dato respectivo. Los datos de un mismo paciente tienden a ser similares, en especial si ambos son de la clase “No VMI”, por lo que si encuentra otro dato del mismo paciente al momento de la prueba se corre el riesgo de sesgar los resultados, pudiendo entregar predicciones correctas por sobreajuste. Con esto, los datos de un mismo paciente son exclusivos o para el subconjunto de entrenamiento o para el de prueba.

Se decide utilizar el 70 % para entrenar considerando el submuestreo, y el 30 % para probar en una razón que represente la realidad, dando favor al aprendizaje del modelo debido a la reducida cantidad de datos. Estos porcentajes se ven levemente afectados debido a la restricción sobre la exclusividad de los pacientes en los subconjuntos. De los 395 pacientes, los 69 que van a VMI corresponden a aproximadamente el 18 %. Con la separación aleatoria de los subconjuntos quedan 18 pacientes VMI en prueba, dando que el 82 % de pacientes No VMI en el mismo subconjunto debe ser de 82. Se obtiene así las cantidades de datos presentes en la Tabla 3.4 para cada subconjunto. Con 138 datos para entrenar y 1302 para probar el modelo se preparan los datos para que sean admisibles como características de entrada.

---

<sup>2</sup>Como se mencionó en la Subsección Extracción de Características, un día de datos no implica necesariamente que el paciente estuvo 24 hrs en el servicio, si no que tiene registrado variables clínicas en algún momento desde las 00:00 hrs hasta las 23:59 hrs

Tabla 3.4: Datos subconjunto de entrenamiento y prueba para evento VMI. Se muestra la cantidad de datos y entre paréntesis la de pacientes a los que están asociados, del tipo: #Datos (#Pacientes).

	Entrenamiento	Prueba
Total datos	<b>138</b> (96)	<b>1302</b> (100)
VMI	<b>70</b> (40)	<b>30</b> (18)
No VMI	<b>68</b> (56)	<b>1272</b> (82)

Por cada día que el paciente se encuentra hospitalizado tiene asociado las variables longitudinales del día y sus datos estáticos. Todos estos valores se concatenan de manera que los datos estáticos y los atributos extraídos queden expresados en un gran vector de características. Con éste se puede ingresar el estado del paciente al modelo, según el flujo mostrado en la Figura 3.5.

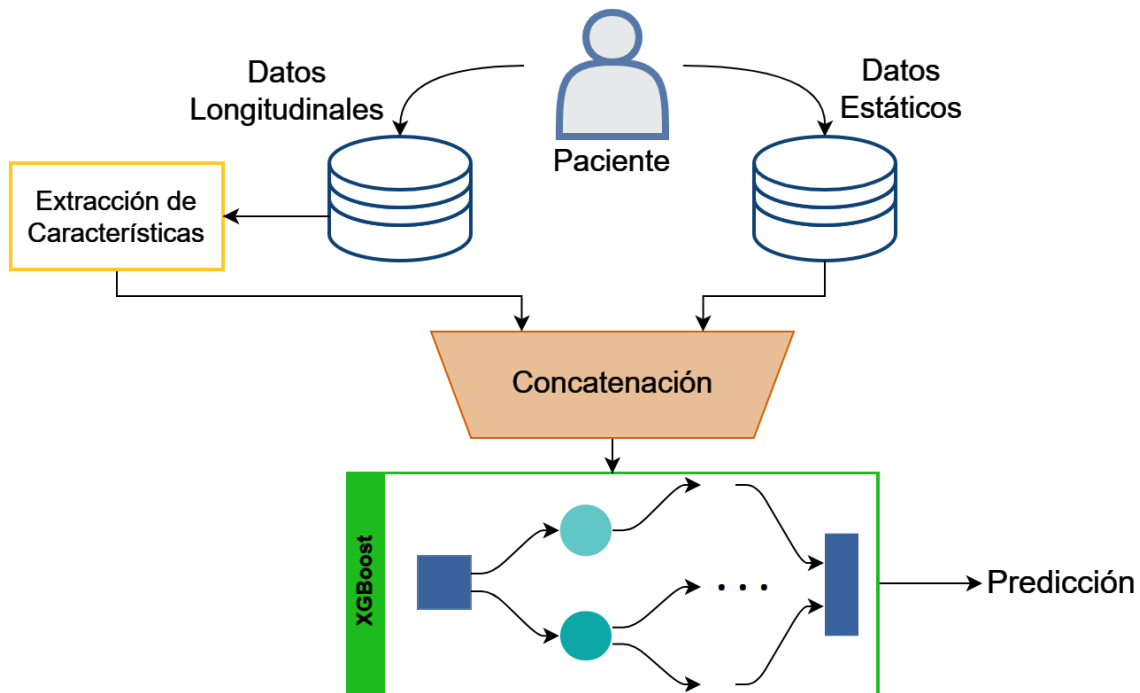


Figura 3.5: Diagrama de flujo del paciente hasta la predicción de eventos clínicos. Este muestra de manera simplificada el camino seguido por los datos para ser ingresado al modelo.

Se presenta en este punto la posibilidad de utilizar todos los datos recolectados en el tiempo del paciente durante su hospitalización, lo que en teoría entregaría al modelo más información sobre la evolución de éste. La forma propuesta de hacer esto es concatenando los datos estáticos, seguido de las características extraídas de los datos longitudinales del día más actual, seguido de las características extraídas por cada día hacia el pasado. Para el caso en que un paciente tenga solo un día de datos, se dejan como datos faltantes (NaN) el resto del vector de características lo cual es manejado por el mismo XGBoost. Este proceso se puede iterar hasta los días que se quieran elegir en el pasado, sin importar la cantidad de días que el paciente se encuentre hospitalizado.

Esto se lleva a cabo probando los datos desde uno hasta tres días en el pasado, donde se encuentra que resultados son muy similares entre sí. Se encuentra además que los atributos

correspondientes al día más actual son a los que el modelo atribuye más importancia por un gran margen. Analizando el escenario completo se concluye que no resulta justificable considerar más de un día de datos como entrada al modelo en este caso. Por un lado tenemos que por cada día se agregan 170 atributos, aumentando considerablemente la complejidad del modelo. Por otro, debido a que el sistema de datos del HCUCH no se encuentra consistentemente automatizado, puede derivar a entregarle la responsabilidad de buscar e ingresar los datos del pasado al personal de salud. Y finalmente, pues los resultados y el análisis de características muestran que el modelo está utilizando básicamente solo los datos del último día para predecir.

En cuanto al ajuste del modelo, los hiperparámetros que definen el comportamiento general del XGBoost son la cantidad de árboles (`n_estimators`), el *learning rate* (`eta`) y la profundidad máxima de cada árbol (`max_depth`)<sup>3</sup>. La selección de estos se realiza a través de *grid search*, un algoritmo iterativo que va variando los parámetros y entregando los resultados. La elección de estos se hace teniendo en cuenta la optimización basada en variables seleccionadas, lo cual se describe en la siguiente Subsección (ver 3.3.4). En ésta se muestra que existen dos propuestas de modelo para predicción de VMI: Variables Seleccionadas (7), cuyos valores elegidos son `n_estimators = 24`, `eta = 0.3`, `max_depth = 2`; y Variables Seleccionadas (9) con `n_estimators = 13`, `eta = 0.3`, `max_depth = 2`.

## Predicción de Mortalidad

Para el caso de mortalidad el diseño del modelo es análogo. La ventana de predicción seleccionada para predecir si el paciente fallecerá es de cinco días. En términos del modelo se considera como “clase 0” si es que éste no fallece en los próximos cinco días o como “clase 1” si el paciente fallece dentro de los siguientes cinco días. Así, considerando cada día hospitalizado como un dato, se obtiene:

Cantidad total de datos: **6831**

Datos con evento Fallece en los próximos 5 días (Clase 1): **262**

Datos con evento Fallece pero en más de 5 días (Clase 0): **572**

Datos de pacientes que no Fallecen (Clase 0): **5997**

De los 58 pacientes que fallecen en el periodo de estudio, se obtienen 262 datos de la clase 1 “Fallece” y 572 de la clase 0 “No Fallece”. Se puede ver que la cantidad de datos generados es tanto mayor que en el caso de predicción de VMI, dado que se consideran cinco días en el futuro en vez de dos. Se tienen además 5997 datos adicionales en la clase 0 pertenecientes a los 337 pacientes que no fallecen. Se aplica submuestreo sobre la clase 0 considerando pacientes que fallecerán en el futuro y los que no.

Teniendo en cuenta la restricción de que cada paciente es exclusivo o para el entrenamiento o para la prueba, se hace la división de subconjuntos en proporción de 70 % y 30 % respectivamente. El porcentaje de pacientes que fallecen en el periodo de estudio es de aproximadamente 15 %. Como 13 de estos pacientes corresponden al 30 % designado al conjunto

---

<sup>3</sup>Entre paréntesis se encuentran los nombres de los hiperparámetros para la librería XGBoost en *Python*

de prueba, la cantidad de No Fallece es de 74. En la Tabla 3.5 se muestran las cantidades de datos para cada clase y subconjunto. Con 388 datos para el entrenamiento y 803 para prueba, se generan los vectores de características, concatenando los valores estáticos y los extraídos de los longitudinales para su ingreso al modelo, al igual que en el caso anterior (Figura 3.5).

Tabla 3.5: Datos subconjunto de entrenamiento y prueba para Mortalidad. Se muestra la cantidad de datos y entre paréntesis la de pacientes a los que están asociados, del tipo: #**Datos** (#Pacientes).

	Entrenamiento	Prueba
Total datos	<b>388</b> (137)	<b>803</b> (87)
Fallece	<b>200</b> (44)	<b>62</b> (13)
No Fallece	<b>188</b> (93)	<b>741</b> (74)

En este escenario tampoco es justificable utilizar más de un día de datos pues los atributos asociados al día más actual son a los que el modelo atribuye más importancia, añadiendo a esto el análisis asociado, mencionado en predicción de VMI.

Finalmente, los parámetros que se configuran para el modelo son `n_estimators = 23`, `eta = 0.2`, `max_depth = 2`. Éstos corresponden a la cantidad de árboles, el *learning rate* y la profundidad máxima de cada árbol respectivamente, y su selección está sujeta a iteraciones de optimización del modelo a partir de *grid search*, considerando las variables seleccionadas, tema que se aborda a continuación.

### 3.3.4. Selección de Variables

Debido a que, además de las variables longitudinales, se tienen cerca de 300 variables estáticas, es esencial llevar a cabo un proceso de selección con tal de disminuir las características de entrada de los modelos. Esto es un proceso crítico en el desarrollo del modelo por motivos de diferente tópicos:

- **Inteligencia Artificial:** Optimizar el rendimiento de los modelos identificando los atributos más importantes para la predicción y eliminando los que no aportan información o introduzcan ruido al modelo.
- **Usuario:** Simplificar el uso del sistema de predicción para el personal de salud, teniendo en cuenta que muchos de los datos deben ingresarse de manera manual. Esto es particularmente importante pues la intención es apoyar al equipo de salud y no darles trabajo extra.

Como tal se realiza el proceso llevando a cabo un análisis de sensibilidad y tomando en cuenta el conocimiento experto. Profesionales de la salud con experiencia en trabajo en el servicio de urgencias, sugieren de primera fuente variables claves para la decisión y otras que pueden ser omitidas. Las variables a considerar bajo su interpretación son las demográficas (Edad, Género), los signos vitales y los exámenes de laboratorio.

Se decide agregar también las comorbilidades y el tiempo que ha estado el paciente hospitalizado en días (Días desde Ingreso) para ambos eventos. Con esto se tiene un total de **57 variables basales**, las cuales se detallan en la Tabla 3.6. Las diferencias en las variables



presentes en la Tabla y las presentadas en la Subsección 3.2.2 Conjunto de Datos es debido al preprocesamiento de datos realizado, mostrado en 3.3.1. A partir de ésto se lleva a cabo el análisis de sensibilidad para cada evento.

Tabla 3.6: Variables Basales (57) escogidas para comenzar el análisis de sensibilidad.

Categoría	Variables
Demográficas (2)	Edad, Género.
Signos Vitales (10)	Temperatura Axilar ( $^{\circ}\text{C}$ ), Frecuencia Cardiaca (U/min), Pas (mmHg), Pad (mmHg), Pam (mmHg), HGT (mg/dL), Frecuencia Respiratoria (U/min), SatO <sub>2</sub> (%), Flujo O <sub>2</sub> (lts/min), FiO <sub>2</sub> (%).
Exámenes de Laboratorio (24)	Hemoglobina (gr/dL), Leucocitos (U/ $\mu\text{L}$ ), Neutrófilos (U/mL), Plaquetas (U/ $\mu\text{L}$ ), Linfocitos (U/ $\mu\text{L}$ ), VHS (mm/hr), Albumina (gr/dL), Nitrógeno Ureico (mg/dL), GOT (U/L), LDH (U/L), Bilirrubina Total (mg/dL), CREA (mg/dL), CK (mg/dL), Lactato (mmol/L), Troponina (ng/mL), GPT (U/L), Procalcitonina (ng/mL), PCR (mg/L), INR, paO <sub>2</sub> (mmHg), pCO <sub>2</sub> (mmHg), Dimero D (ng/mL), TTPa (s), Fibrinógeno (mg/dL).
Comorbilidades (20) (binarias)	Tabaquismo, Hipertensión Arterial, Diabetes Tipo 2, Diabetes Tipo 2 con Daño de Órgano Blanco, Asma, Obesidad, Enfermedad Coronaria, Insuficiencia Cardiaca, EPOC, Malignidad, Enfermedad Renal Crónica, Accidente Cerebrovascular, Hemiplejia, Demencia, Enfermedad Arterial Oclusiva, Ulcera Péptica, Daño Hepático Crónico, Inmunosupresión Celular, Enfermedad de Tejido Conectivo, Taco.
Otros (1)	Días desde Ingreso

**Pas:** Presión Arterial Sistólica. **Pad:** Presión Arterial Diastólica, **Pam:** Presión Arterial Media. **HGT:** Hemoglocutest. **SatO<sub>2</sub>:** Saturación de Oxígeno. **FiO<sub>2</sub>:** Fracción inspirada de Oxígeno. **VHS:** Virus del Herpes Simple. **GOT:** Aspartato Aminotransferasa. **LDH:** Lactato Deshidrogenasa. **CREA:** Creatina Plasmática. **CK:** Creatina Quinasa. **GPT:** Alanina Aminotransferasa. **PCR:** Proteína C Reactiva. **INR:** *International Normalized Ratio*. **paO<sub>2</sub>:** Presión Arterial de Oxígeno. **pCO<sub>2</sub>:** Presión Arterial de Dióxido de Carbono. **TTPa:** Tiempo de Tromboplastina Parcial activado. **EPOC:** Enfermedad Pulmonar Obstructiva Crónica.

## Predicción de Indicación de Ventilación Mecánica Invasiva

Con las 57 variables basales, se tiene un total de **173** características de entrada para el modelo. El análisis de sensibilidad se hace a partir de la información que el mismo XGBoost ofrece una vez entrenado, la importancia de características, la cual es un indicador de la

contribución de cada característica sobre la decisión del modelo. El indicador es un valor relativo calculado en torno a las importancias de todas las demás variables a partir de la pureza y luego ponderados, tal que la suma de éstos de uno [37]. El proceso se lleva a cabo utilizando el conjunto de test balanceado por simplicidad de lectura de las métricas de evaluación.

En la Figura 3.6 se muestra el histograma con los 15 atributos más relevantes para la predicción de VMI según el modelo optimizado. De este se obtiene que las 10 variables clínicas más importantes son:

paO<sub>2</sub>, FiO<sub>2</sub>, PCR, LDH, Frecuencia Cardíaca, Temperatura Axilar, SatO<sub>2</sub>, Frecuencia Respiratoria, Días desde Ingreso, Flujo O<sub>2</sub>.

de las cuales se destaca paO<sub>2</sub> que hace aparición tres veces en el top 15 con su promedio, valor mínimo y desviación estándar. La variable agregada “Días desde Ingreso” hace aparición como una de las más importantes. Por otro lado no aparece ninguna variable de Categoría comorbilidad. Se puede notar que la pendiente y la desviación estándar son las características menos utilizadas en general, y el promedio la más utilizada.

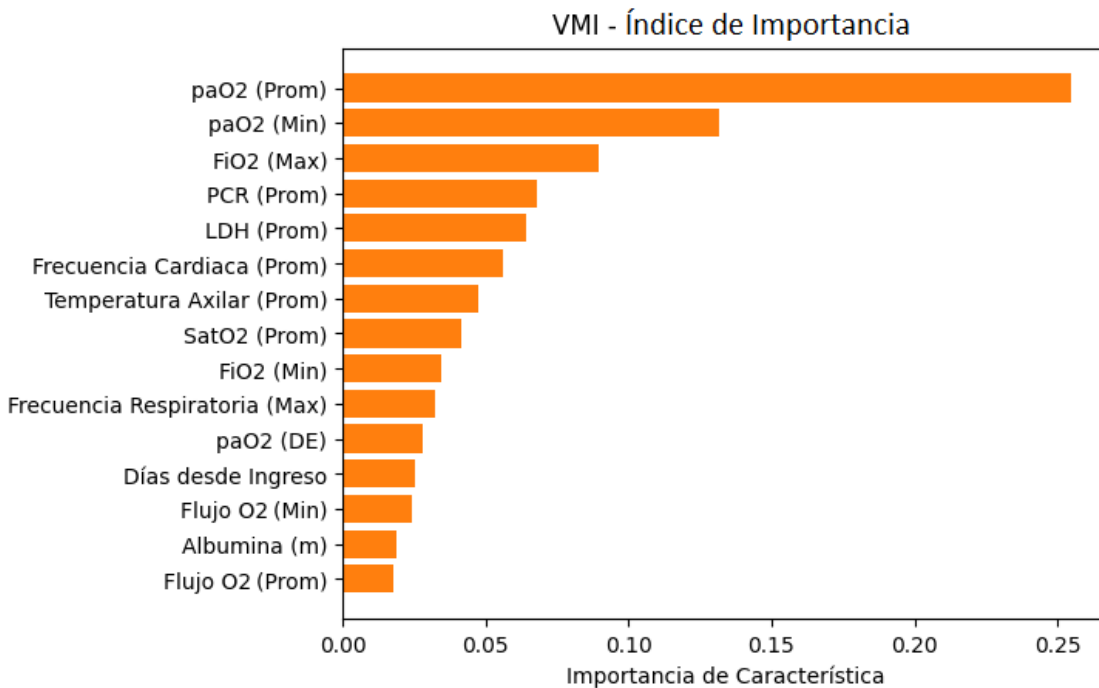


Figura 3.6: Importancia de características sobre el modelo de predicción de VMI entrenado con las Variables Basales. Se visualizan los 15 atributos más importantes según XGBoost. (Prom: Promedio, Min: Mínimo, Max: Máximo, DE: Desviación Estándar, m: Pendiente).

Si bien el análisis de resultados se ahondará en la siguiente Sección, es pertinente mencionar ahora la importancia de la métrica *recall* o sensibilidad sobre este problema. Ésta representa la cantidad de aciertos sobre la clase 1 “VMI” en relación al total de pacientes que efectivamente fueron indicados con VMI. Es decir, el porcentaje de pacientes que requieren de indicación de VMI correctamente predichos, sin considerar los “No VMI”. Si bien el *accuracy* o exactitud es un buen indicador general sobre el desempeño del modelo, el *recall* nos muestra de manera

directa como lo hace sobre la clase que presenta sentido de urgencia. Se desea que la predicción tenga énfasis sobre si el paciente se encuentra en estado crítico y que no pase por alto estos casos.

Para reducir la cantidad de variables de entrada, se prueban modelos que utilicen las 10 variables más importantes y combinaciones entre ellas de manera iterativa. La metodología final se basa en utilizar la variable que más aporte tiene sobre el modelo e ir agregando una a una la siguiente variable más importante según lo encontrado en la iteración.

Se encuentra que existen cuatro variables esenciales para la predicción:  $paO_2$ ,  $FiO_2$ , Frecuencia Respiratoria y Días desde el ingreso. Luego, para obtener resultados similares al de las variables basales, se pueden agregar tres variables de signos vitales: Temperatura Axilar, Frecuencia Cardíaca y Saturación de Oxígeno ( $SatO_2$ ); las cuales valen la pena ser consideradas pues no son complejas de obtener en la práctica clínica. Con esto, el problema se acota a **7 variables** para hacer la predicción, de las cuales 6 serían ingresadas por el personal de salud pues los días desde el ingreso pueden ser calculados de manera automática. En la Tabla 3.7 se ve que al utilizar estas 7 variables se obtiene un *recall* del 93% sin perder desempeño en las demás métricas.

Tabla 3.7: Métricas de desempeño obtenidas para predicción de VMI utilizando distintas variables de entrada. Los resultados de la  $i$ -ésima fila representan el modelo entrenado con  $i$  variables, es decir desde la primera ( $paO_2$ ) hasta la correspondiente a su fila. “(N° variables)” hace referencia a la cantidad de variables utilizadas como entrada del modelo.

(N° variables) Variables	<i>Accuracy</i> o		<i>Recall</i> o		
	Exactitud	<i>Precision</i>	Sensitividad	Especificidad	ROC Área
(1) $paO_2$	0.82	0.88	0.73	0.90	0.82
(2) $FiO_2$	0.85	0.82	0.90	0.81	0.86
(3) Frecuencia Respiratoria	0.87	0.87	0.87	0.88	0.87
(4) Días desde ingreso	0.85	0.82	0.90	0.81	0.86
(5) Temperatura Axilar	0.87	0.82	0.93	0.81	0.87
(6) Frecuencia Cardíaca	0.85	0.82	0.90	0.81	0.86
(7) $SatO_2$	0.89	0.85	<b>0.93</b>	0.84	0.89
(8) PCR	0.89	0.87	0.90	0.88	0.89
(9) LDH	<b>0.92</b>	<b>0.90</b>	<b>0.93</b>	<b>0.91</b>	<b>0.92</b>
(10) Flujo $O_2$	0.90	0.88	0.93	0.89	0.90

Las tres variables de signos vitales mencionadas al último son incluidas gracias a la simplicidad de su medición, pero existen otras dos variables de laboratorio que ayudan a mejorar la predicción: PCR y LDH. En la misma Tabla 3.7 se muestra como el modelo que considera estas **9 variables** obtiene los mejores resultados en todas las métricas. Las variables de exámenes de laboratorio llevan más esfuerzo de obtener, por lo que se plantea este modelo como otra alternativa teniendo en cuenta sus limitaciones.

Se tienen entonces dos propuestas de modelo para predicción de VMI, uno con 7 y otro 9 variables de entrada. Este último posee resultados un tanto mejores que el primero gracias a sus dos variables adicionales, las cuales al ser exámenes de laboratorio son por su naturaleza

más complejas de obtener, justifica el análisis de ambos en profundidad.

## Predicción de Mortalidad

De manera similar para el presente caso, toma en cuenta las importancias de las variables calculadas por el modelo de predicción de mortalidad, de las cuales las 15 más importantes se muestran en la Figura 3.7. Se puede notar como en este caso existen más atributos con alto impacto en el modelo, lo que implica que el algoritmo está haciendo uso de una gran cantidad de variables para obtener su resultado.

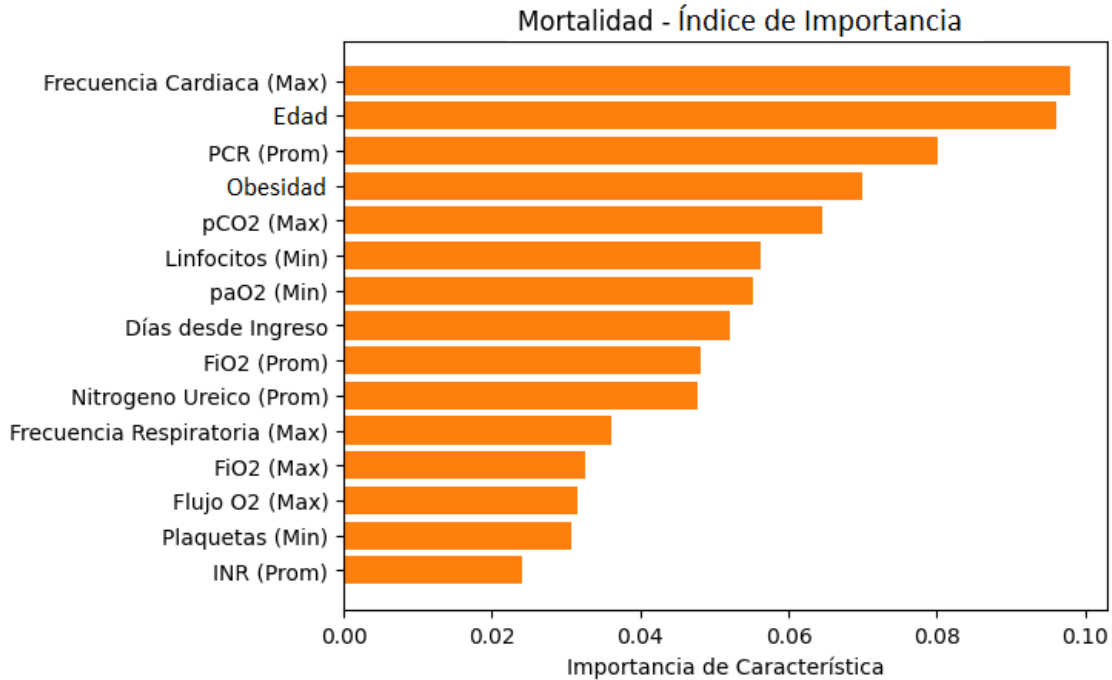


Figura 3.7: Importancia de características sobre el modelo de predicción de Mortalidad entrenado con las Variables Basales. Se visualizan los 15 atributos más importantes según XGBoost. (Prom: Promedio, Min: Mínimo, Max: Máximo, DE: Desviación Estándar, m: Pendiente).

Al explorar las variables y su comportamiento junto con los expertos de la salud se concluye que no es tan directo seleccionar las 10 más importantes a diferencia del caso de predicción de VMI. Existen un par de variables a las cuales prestar especial atención antes de continuar el análisis de sensibilidad. Comenzando por la variable “Obesidad”, la única comorbilidad presente en el top de atributos, ésta tiene un comportamiento opuesto al que uno esperaría. En la Figura 3.8 se muestra el análisis SHAP de ésta dentro del modelo. Como es una variable binaria solo puede tomar dos valores: 1 (en azul) que representa el caso de obesidad positiva, y 0 (en amarillo) obesidad negativa. Los valores SHAP funcionan tal que mientras más positivo sea su valor (hacia la derecha en el eje horizontal de la Figura) más correlación tiene el atributo con la clase 1 (Fallece), y mientras más negativo más correlación con la clase 0 (No Fallece). Aplicando esto al caso, se tiene que los pacientes con obesidad tienden a estar relacionados con la clase “No Fallece”, y los que no tienen obesidad con la clase “Fallece”.

Este fenómeno es conocido como “La paradoja de la obesidad” [38], y es el perfecto ejemplo de que correlación no implica causalidad. Ésta habla sobre como es más común de lo que uno

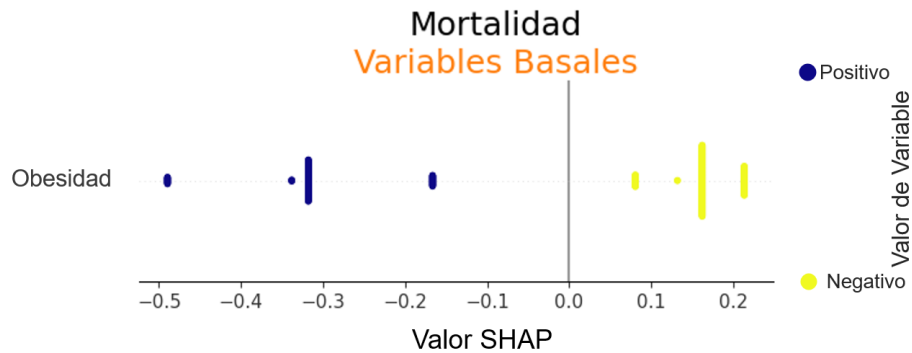


Figura 3.8: Análisis SHAP de la variable “Obesidad” sobre el modelo de predicción de Mortalidad entrenado con las Variables Basales. Un valor SHAP positivo está relacionado con la clase “Fallece”, y uno negativo con “No Fallece”. Un color azul representa Obesidad positiva y Amarillo negativa.

esperaría encontrar una menor tasa de mortalidad en pacientes con obesidad en comparación a pacientes sin obesidad, especialmente en casos como problemas cardiovasculares, obstrucción pulmonar y cáncer [39].

Entre los pacientes involucrados en el estudio, efectivamente sucede que hay un mayor porcentaje de pacientes con obesidad que no fallecen con un 35.3%, que los que fallecen con un 22.4% como se ve en la Figura 3.9. Este caso pone en perspectiva la importancia de la validación de algoritmos basados en IA para escenarios médicos. Gracias al apoyo del conocimiento experto en el análisis, es posible conocer la existencia de este fenómeno y evitar la inclusión de una variable que pareciera tener un elevado impacto en el modelo, pero que en realidad no presenta mayor aporte, dando además sentido clínico al modelo.

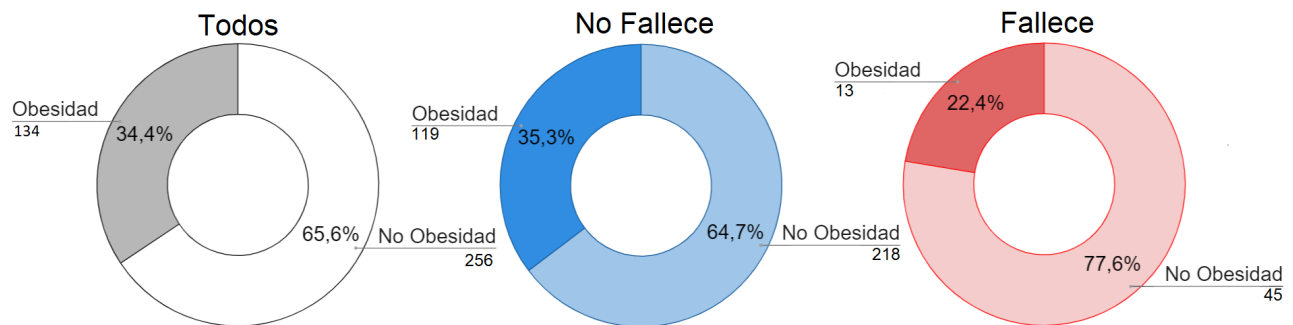


Figura 3.9: Porción de pacientes con Obesidad en el estudio. A la izquierda: Todos los datos; Al centro: Pacientes de la clase No Fallece; A la derecha: Pacientes de la clase Fallece.

Por otro lado está la variable “Edad”, la cual presenta un sesgo dentro del conjunto de datos. Como se muestra en el histograma de la Figura 3.10, de los 58 pacientes que fallecen dentro del periodo de estudio, solo 8 de ellos son menores de 60 años y ninguno menor a 40. Estos umbrales, generan una bifurcación fuerte en el árbol de decisión, clasificando como “No Fallece” a prácticamente todos los pacientes menores de 60 años, y debido al gran desbalance de esta variable esto no es penalizado lo suficiente. Si bien esto es un reflejo del escenario real, se desea que el modelo sea funcional para el caso general y no solo para pacientes de elevada edad. La finalidad es que el modelo logre predecir según el estado del paciente y este

sesgo puede ser contraproducente por lo que se decide no utilizar la variable.

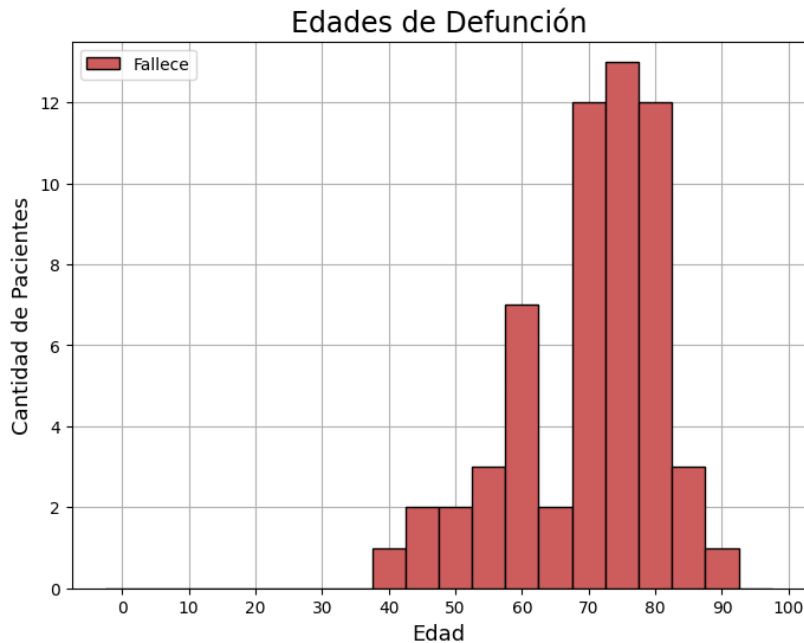


Figura 3.10: Histograma de edades de los 58 pacientes fallecidos durante el estudio, discretizado en múltiplos de cinco.

Tabla 3.8: Métricas de desempeño obtenidas para predicción de Mortalidad utilizando distintas variables de entrada. Los resultados de la  $i$ -ésima fila representan el modelo entrenado con  $i$  variables, es decir desde la primera ( $pCO_2$ ) hasta la correspondiente a su fila. “(N° variables)” hace referencia a la cantidad de variables utilizadas como entrada del modelo.

(N° variables) Variables	<i>Accuracy</i> o		<i>Recall</i> o		
	Exactitud	<i>Precision</i>	Sensitividad	Especificidad	ROC Área
(1) $pCO_2$	0.62	0.6	0.5	0.72	0.61
(2) $paO_2$	0.65	0.61	0.61	0.68	0.64
(3) Frecuencia Cardiaca	0.69	0.64	0.74	0.65	0.70
(4) $SatO_2$	0.77	0.71	<b>0.85</b>	0.70	0.78
(5) Nitrogeno Ureico	0.78	0.74	0.81	0.76	0.78
(6) PCR	0.80	0.80	0.76	<b>0.84</b>	0.80
(7) $FiO_2$	<b>0.81</b>	<b>0.78</b>	<b>0.81</b>	<b>0.81</b>	<b>0.81</b>
(8) LDH	0.82	0.80	0.79	0.84	0.81
(9) Linfocitos	0.79	0.77	0.79	0.80	0.79
(10) Plaquetas	0.80	0.80	0.76	0.84	0.80

Con todo esto en mente se procede a optimizar el modelo considerando el análisis de sensibilidad para disminuir la cantidad de atributos. Se encuentra que existen cuatro variables claves para la predicción de mortalidad:  $pCO_2$ ,  $paO_2$ , Frecuencia Cardiaca y  $SatO_2$ . Notar que  $paO_2$  vuelve a hacer aparición como una de las variables más importantes, mostrando ser un indicador imprescindible en el escenario del COVID-19. Para mejorar el rendimiento del modelo se consideran las siguientes 7 variables:

pCO<sub>2</sub>, paO<sub>2</sub>, Frecuencia Cardiaca, SatO<sub>2</sub>, Nitrógeno Ureico, PCR, FiO<sub>2</sub>

pues como se muestra en la Tabla 3.8, con éstas se alcanzan los mejores resultados. Como se menciono previamente, es de particular interés el *recall* o sensibilidad del modelo, el cual alcanza un 0.85 con solo 4 variables, pero su desempeño sobre la clase “No Fallece” es considerablemente peor, lo que entregaría una exagerada cantidad de falsos positivos. También se puede ver como utilizando 6 variables se alcanza una especificidad de 0.84, siendo esto un valor relativamente alto para los resultados de predicción de mortalidad, pero esto a costa de una considerable disminución en el desempeño basado en *recall*.

A diferencia de los modelos de predicción de VMI, las variables clave para la predicción de mortalidad constan de varias pertenecientes a exámenes de laboratorio. Debido a esto el análisis de la complejidad en conseguir los datos clínicos queda de lado, tomando solo el modelo de 7 variables como propuesta gracias a su buen desempeño general.

# Capítulo 4

## Resultados y Discusión

A continuación se presentan los resultados de los modelos de predicción de VMI y Mortalidad diseñados. Para su calidad de predicción, se eligen métricas de desempeño acordes a problemas de clasificación presentadas en 2.2.6. Para un correcto análisis de dichas métricas es necesario tener en cuenta el contexto del problema, en el cual se propone predecir un evento clínico adverso a un paciente en el servicio de urgencias.

Los modelos desarrollados tienen como salida la probabilidad de que el paciente sufra un evento clínico adverso. En otras palabras, el usuario que utilice la herramienta no verá una predicción del tipo “Alerta de VMI en las próximas 48 hrs”, sino “Existe un 82 % de probabilidad de que el paciente requiera VMI en las próximas 48 hrs”.

Las métricas de evaluación por lo general consideran un umbral en el 50 % de probabilidad de ocurrencia para definir la clase. Esto tiene sentido para una gran variedad de escenarios, pero en medicina no es tan directo. Un ejemplo extremo es el caso de predicción de riesgo de cáncer de pulmón, en donde una probabilidad de a lo menos 1.51 % es suficiente para generar alarmas [40].

Las curvas ROC describen el comportamiento de los modelos para diferentes umbrales lo que les da especial relevancia en este contexto. Se muestran también histogramas normalizados de la probabilidad predicha, que reflejan la creteza con que se aproxima ésta a la clase real (Verdaderos Positivos y Negativos). Y no menos importante, se presenta el análisis SHAP sobre las variables, que ayuda a la interpretabilidad clínica a lo que ocurre dentro del modelo al momento de hacer la predicción.

### 4.1. Predicción de Ventilación Mecánica Invasiva

Como se vio en el Capítulo anterior, se eligen dos alternativas de modelo para la tarea de predicción de VMI. Se referirá a estos como el modelo de “Variables Seleccionadas (7)” para el modelo entrenado con 7 variables, y “Variables Seleccionadas (9)” para el modelo entrenado con las dos variables de laboratorio adicionales. Para tener en mente, las variables utilizadas en los modelos son:



**VARIABLES SELECCIONADAS (7):**  $paO_2$ ,  $FiO_2$ , Frecuencia Respiratoria, Días desde ingreso, Temperatura Axilar, Frecuencia Cardíaca,  $SatO_2$ .

**VARIABLES SELECCIONADAS (9):**  $paO_2$ ,  $FiO_2$ , Frecuencia Respiratoria, Días desde ingreso, Temperatura Axilar, Frecuencia Cardíaca,  $SatO_2$ , PCR, LDH.

En la Figura 4.1 (a) se muestra la matriz de confusión para el modelo de predicción de VMI con Variables Seleccionadas (7), la cual representa un *accuracy* o exactitud de 0.82 como se ve en la Tabla 4.1. La predicción sobre los pacientes que requieren VMI tiene un acierto del 93%, a diferencia de los pacientes No VMI con un 82%. Tomando en cuenta *precision* se tiene que un 11% de los pacientes advertidos como necesidad de VMI por el modelo realmente irán a VMI. Este bajo porcentaje es dado por el desbalance de datos en test, donde el 18% de pacientes falsos positivos son 229 datos, y el 93% de verdaderos positivos son 28.

Tabla 4.1: Métricas de desempeño para el modelo de predicción de VMI Variables Seleccionadas (7).

<i>Accuracy</i> o Exactitud		<i>Recall</i> o Sensitividad		ROC Área
	<i>Precision</i>		Especificidad	
0.82	0.11	0.93	0.82	0.88

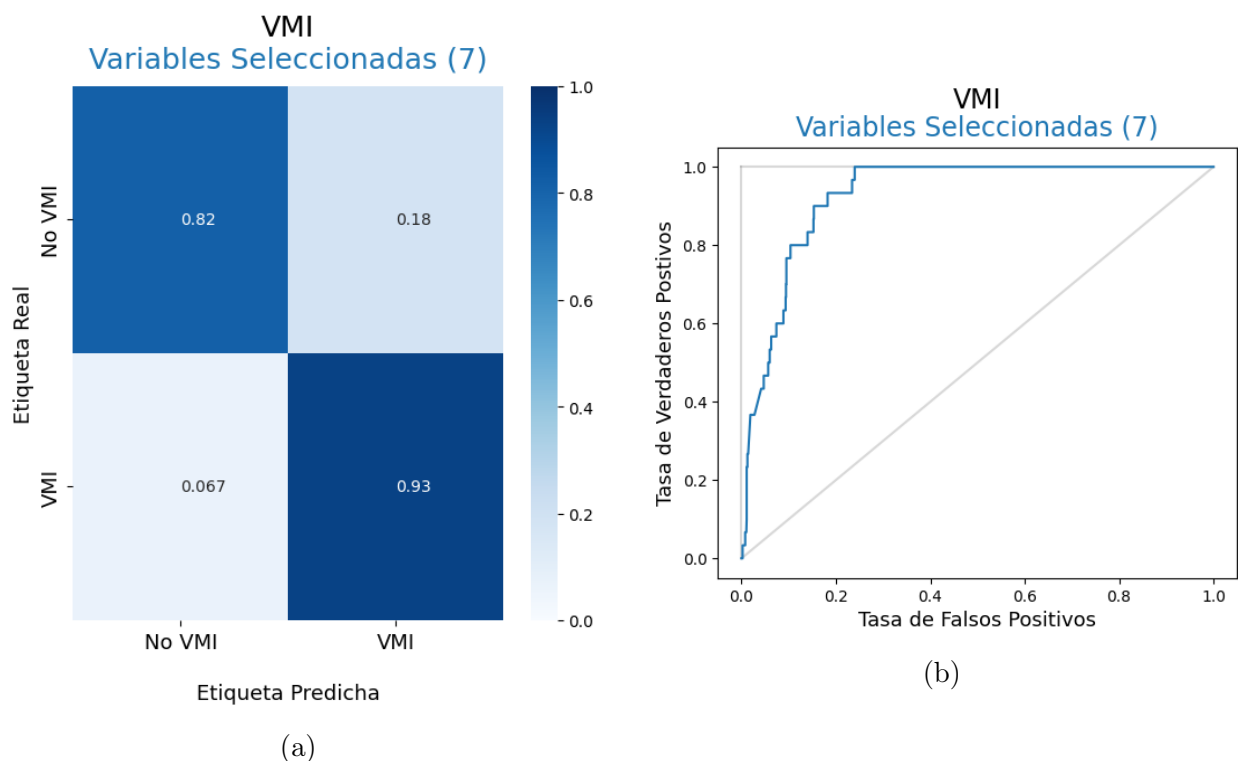


Figura 4.1: Resultados del modelo de predicción de VMI asociado a Variables Seleccionadas (7). (a) Matriz de confusión. (b) Curva ROC.

En la curva ROC de la Figura 4.1 (b), se nota como el modelo da prioridad a la correcta predicción de indicación de VMI gracias a su empujada ascendente, teniendo preferencia por obtener verdaderos positivos sobre disminuir los falsos positivos, que es precisamente el

comportamiento deseado. Ésto se puede ejemplificar en los puntos más altos de la curva antes de converger a 1, en el cual el modelo obtiene un 97 % de aciertos sobre verdaderos pacientes VMI, entregando como consecuencia un 17 % de falsos positivos.

Es de interés también conocer la certeza que entrega la predicción sobre cada clase. En la Figura 4.2 se muestra el histograma normalizado de probabilidad de ocurrencia del evento VMI para el mismo modelo. Cada clase posee una tendencia similar a una curva de distribución de probabilidad, con el máximo en 0 para los pacientes No VMI y en 0.95 para los VMI. De los 30 pacientes del conjunto de prueba de clase VMI, 24 de ellos son predichos con una probabilidad de ocurrencia mayor al 80 %, y solo dos con menos del 50 %.

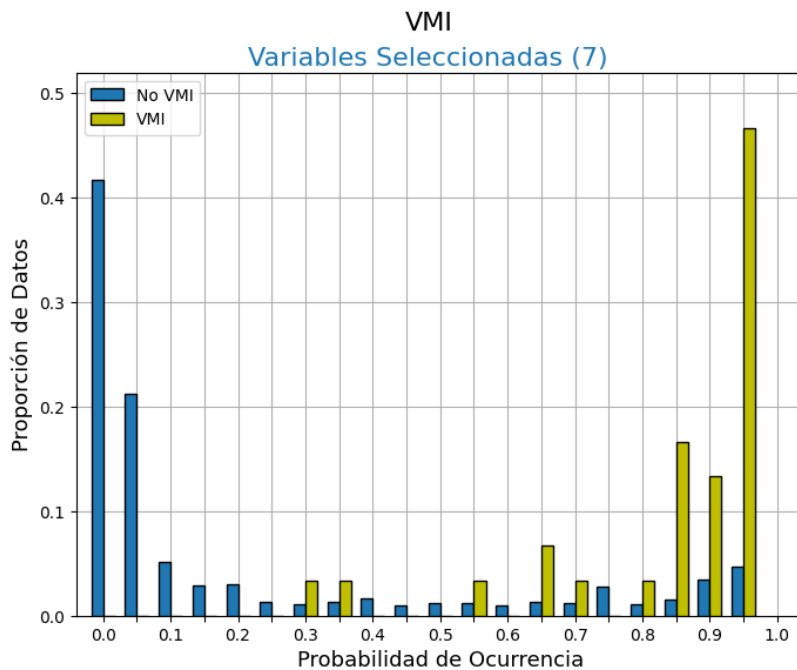


Figura 4.2: Histograma normalizado de probabilidad de ocurrencia predicha para el modelo de predicción de VMI asociado a las Variables Seleccionadas (7). En azul los pacientes de clase 0 “No VMI” y en amarillo los de clase 1 “VMI”.

Para los pacientes de clase No VMI, hay pacientes con una alta probabilidad de tener indicación de ventilación mecánica. En la práctica clínica, tareas médicas como indicar ventilación mecánica invasiva a un paciente pueden no ser tan consistentes. En primer lugar pues es una decisión humana en donde expertos pueden tener opiniones distintas en relación a un mismo paciente. Pero aún asumiendo que los expertos tomen una decisión idéntica existe otro factor de mucho peso, y es que el paciente puede elegir si hacer caso a la indicación o no. El paciente, o incluso su familia, pueden decidir si tomar en cuenta la opinión de ingresar a VMI, la cual puede ser influenciada por la expectativa de vida del paciente, situación financiera, incomodidad, entre otras. Es por esto que el problema de predicción de VMI posee alta incertidumbre de manera sistemática, y queda abierta la posibilidad de que casos como estos pacientes clase 0 con alta probabilidad de requerir ventilación mecánica pertenezcan a éste error.

Se evalúa por último el análisis SHAP del modelo presente en la Figura 4.3. Comenzando

por la Presión Arterial de Oxígeno ( $paO_2$ ), ésta es de manera consistente la variable que más impacto tiene en los modelos de predicción de VMI. Un valor relativamente bajo de  $paO_2$  favorece la clasificación hacia la clase VMI. Es interesante notar que bajo la característica “Promedio”,  $paO_2$  empuja hacia la clase VMI sin importar su valor relativo, pero refuerza la clase No VMI cuando el valor es faltante (NaN). Esto quiere decir que el modelo ha logrado interpretar que si al paciente no se le toma el examen  $paO_2$ , implica que este probablemente no está en riesgo del evento. Si la desviación estándar de esta variable es alta hay mayor impacto sobre la clase VMI, lo que se puede interpretar como que si la  $paO_2$  del paciente varía, mientras que se alcanza un valor relativamente bajo, es más probable que requiera VMI.

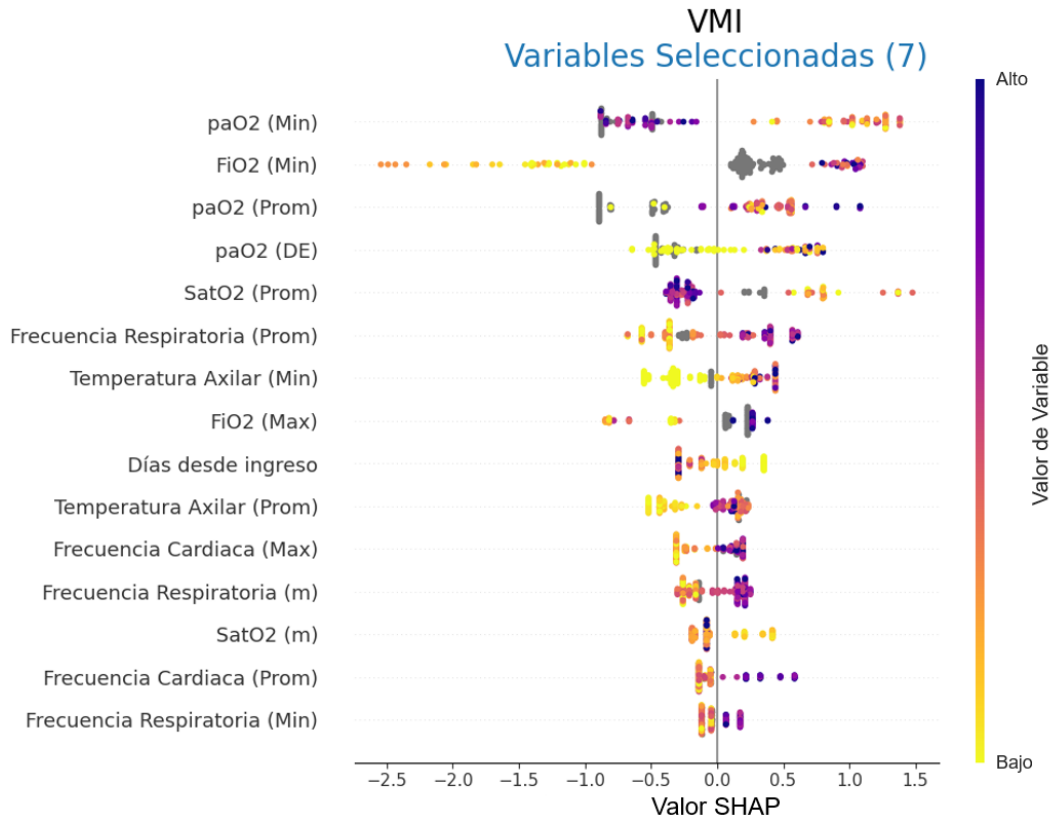


Figura 4.3: Análisis SHAP para el modelo de predicción de VMI asociado a las Variables Seleccionadas (7). Un valor SHAP positivo está relacionado con la clase “VMI”, y uno negativo con “No VMI”. Un color más oscuro representa un valor mayor relativo a la variable y gris un valor faltante. (Prom: Promedio, Min: Mínimo, Max: Máximo, DE: Desviación Estándar, m: Pendiente).

La Fracción Inspirada de Oxígeno ( $FiO_2$ ), es utilizada habitualmente en conjunto con la  $paO_2$  en la práctica como la razón  $PaFi = paO_2/FiO_2$ , por lo que tiene sentido clínico que el modelo encuentre éstas dos variables entre las más importantes y en conjunto. Una  $FiO_2$  alta tiende a empujar hacia la clase VMI, lo cual ocurre pues un paciente con obstrucción en el sistema respiratorio requiere inspirar más Oxígeno para obtener lo que su cuerpo necesita. En casos extremos es requerida la ventilación mecánica invasiva, por lo que la variable se condice con la práctica. Éste análisis es también válido para la variable Frecuencia Respiratoria.

En otras variables, la Temperatura Axilar relativamente alta induce a la clase VMI. Según los datos, los pacientes que van a VMI habitualmente lo hacen en los días tempranos desde

el ingreso. La Saturación de Oxígeno ( $\text{SatO}_2$ ) baja empuja a la clase VMI, y es de las pocas variables en las que el modelo considera la pendiente. Una Frecuencia Cardíaca alta tiende a darse en pacientes más críticos, lo cual tiene sentido con que el análisis SHAP interprete este escenario como un puntaje mayor hacia la clase VMI.

Si al modelo anterior le agregamos las variables de laboratorio PCR y LDH al entrenar, obtenemos el modelo de Variables Seleccionadas (9), cuya matriz de confusión se muestra en la Figura 4.4 (a), representativa de un *accuracy* de 0.82. El desempeño de éste es muy similar al del modelo de Variables Seleccionadas (7) como se puede ver en la Tabla 4.2, manteniendo el mismo *recall*. Éste modelo tiene un comportamiento peor sobre los pacientes No VMI con un acierto del 81%, a diferencia del 82% del modelo anterior, lo que es una diferencia de 14 datos de error sobre los 1272 totales No VMI. La curva ROC de la Figura 4.4 (b), representativa de un Área de 0.87, presenta un comportamiento muy similar a la del modelo anterior, priorizando la correcta detección de los verdaderos positivos que es lo deseado.

Tabla 4.2: Métricas de desempeño para el modelo de predicción de VMI Variables Seleccionadas (9).

<i>Accuracy</i> o Exactitud		<i>Recall</i> o Sensitividad		<i>ROC</i> Área
	<i>Precision</i>		Especificidad	
0.82	0.11	0.93	0.81	0.87

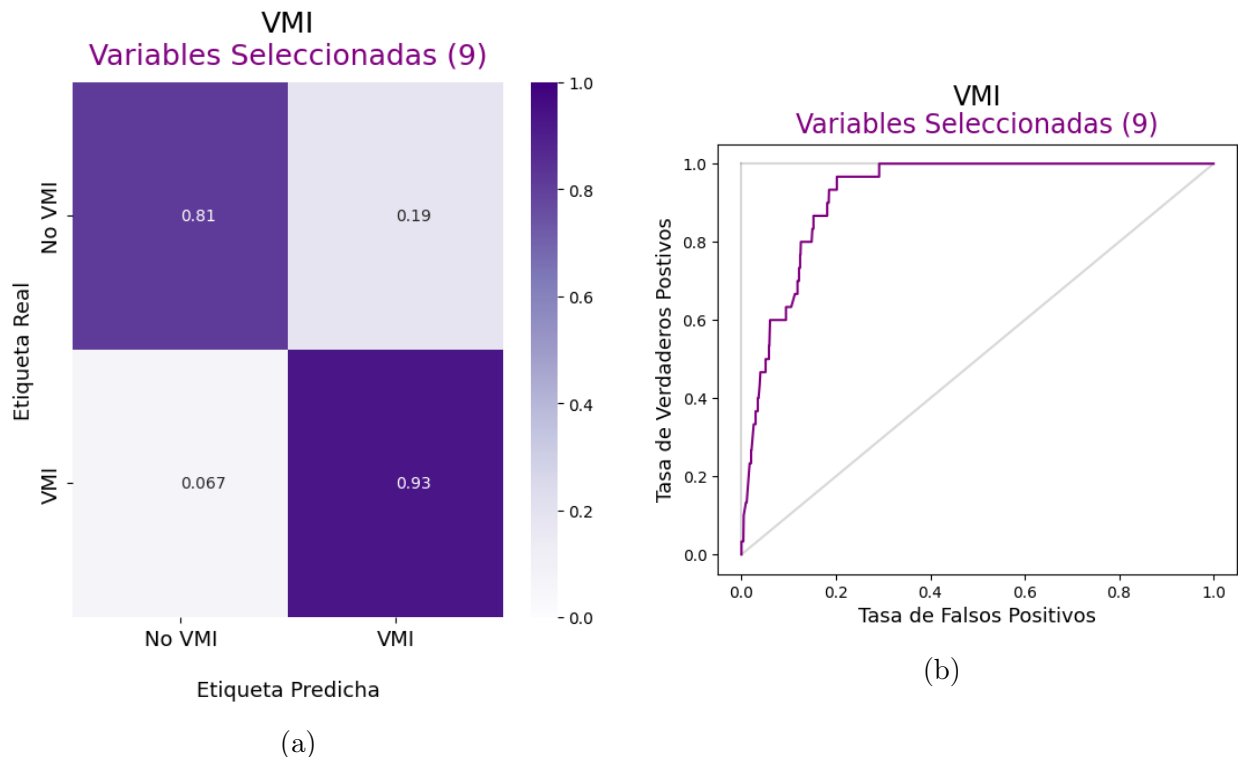


Figura 4.4: Resultados del modelo de predicción de VMI asociado a las Variables Seleccionadas (9). (a) Matriz de confusión. (b) Curva ROC.

En el histograma de probabilidad de ocurrencia del evento VMI, en la Figura 4.5, se ve que la mayor cantidad de pacientes No VMI ahora se encuentran parados sobre la probabilidad 0.05 en vez de 0.0, al igual que los VMI están sobre 0.9 en vez de 0.95 como en el modelo anterior. Esta disminución de certeza implica que ahora hay 19 pacientes VMI con una probabilidad de ocurrencia mayor al 80 %, mientras en el modelo de Variables Seleccionadas (7) hay 18 con probabilidad de ocurrencia mayor a 90 %. Se destaca que a pesar de que el modelo de Variables Seleccionadas (9) tiene una distribución de probabilidad diferente para la clase No VMI, más ancha, debido a disminución de certeza, se ve un comportamiento similar para los falsos positivos de alta probabilidad de VMI, apoyando la idea de que éstos son casos difíciles de predecir.

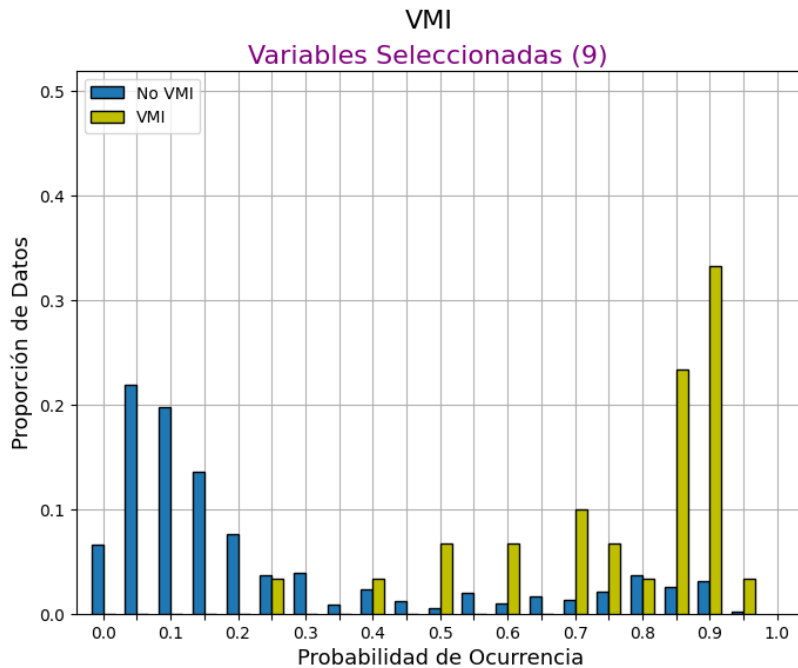


Figura 4.5: Histograma de probabilidad predicha para el modelo de predicción de VMI asociado a las Variables Seleccionadas (9). En azul los pacientes de clase 0 “No VMI” y en amarillo los de clase 1 “VMI”.

El análisis SHAP asociado al modelo de 9 variables (ver Figura 4.6), presenta el mismo comportamiento que el anterior, tomando como variables más importantes la  $paO_2$  y la  $FiO_2$ . Entran en el top 5 variables el LDH y el PCR, ambos relacionados con la clase VMI cuando sus valores relativos son mayores. La variables Días desde ingreso ya no se encuentra entre las 15 características con más impacto en el modelo.

Comparando los modelos, en la Figura 4.7 se muestran las curvas ROC de los dos modelos propuestos y el de Variables Basales. En ésta se ve como los modelos de Variables Seleccionadas (7) y (9) tienen curvas muy similares en la gran mayoría del recorrido, aunque la de 7 variables se posiciona levemente por encima de las demás en algunos tramos. Por otro lado, se puede apreciar como la curva del modelo de Variables Basales no presenta ninguna ventaja particularmente visible, a pesar de estar entrenado con información de 57 variables clínicas.

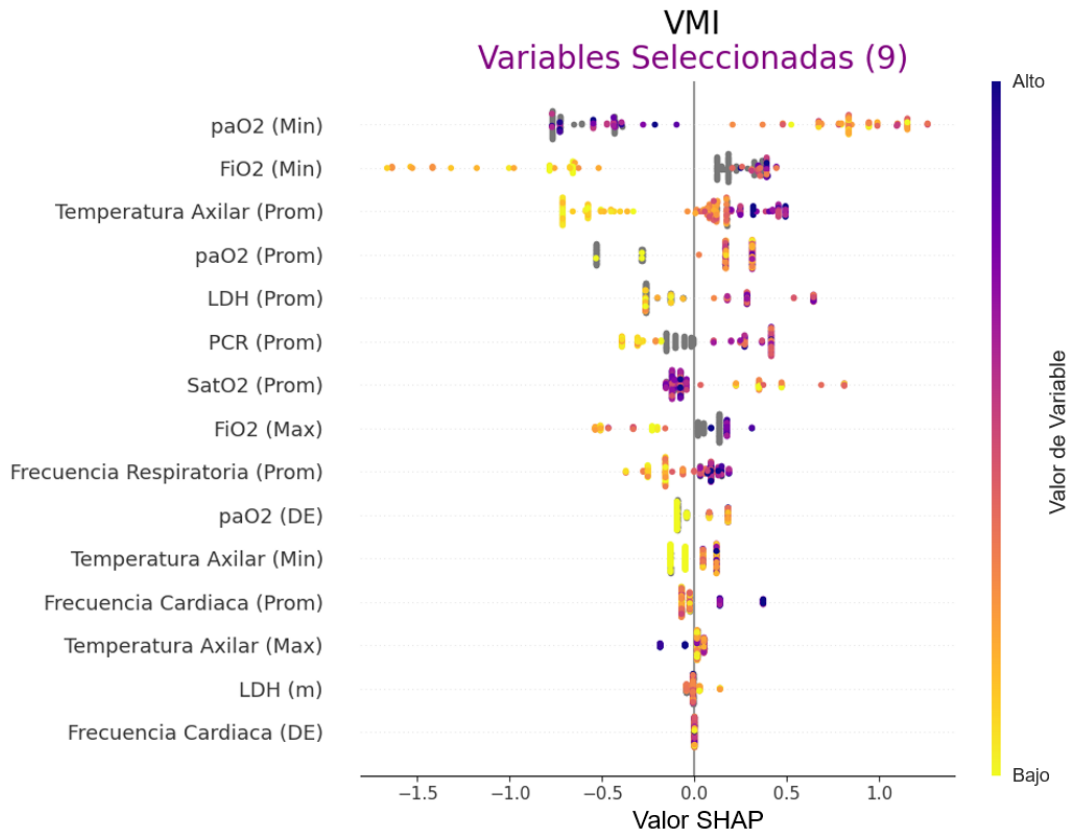


Figura 4.6: Análisis SHAP para el modelo de predicción de VMI asociado a las Variables Seleccionadas (9). Un valor SHAP positivo está relacionado con la clase “VMI”, y uno negativo con “No VMI”. Un color más oscuro representa un valor mayor relativo a la variable y gris un valor faltante. (Prom: Promedio, Min: Mínimo, Max: Máximo, DE: Desviación Estándar, m: Pendiente).

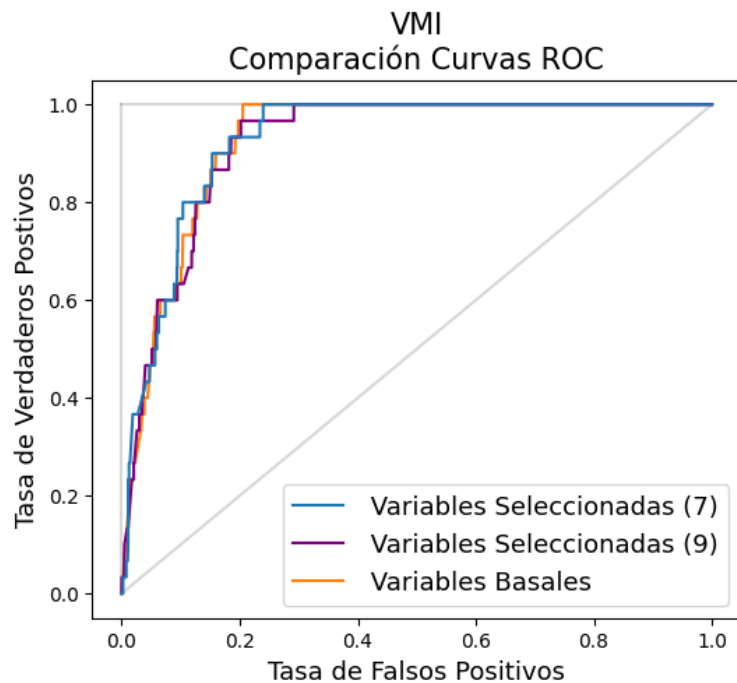


Figura 4.7: Curvas ROC de los modelos de predicción de VMI Variables Seleccionadas (7), Variables Seleccionadas (9) y de Variables Basales.

Para una comparación más amplia, se entrenan modelos basados en *Random Forest* (RF) con las mismas características, procesamientos y datos que los modelos de Variables Seleccionadas, *Random Forest* (7) y *Random Forest* (9). La Tabla 4.3 muestra las métricas de desempeño para los modelos propuestos, los basados en RF y el de Variables Basales, y al igual que en la comparación anterior, en modelo de Variables Basales no presenta ninguna ventaja real. Los modelos basados en XGBoost igualan y superan a los basados en RF, lo que se ve claro en la métrica ROC Área, en donde estos últimos entregan 0.81 y 0.84 para 7 y 9 variables respectivamente, frente a 0.87 y 0.88 para los primeros. En el comportamiento general no es distinto, en donde si los modelos basados en RF logran igualar una métrica de los XGBoost en torno a especificidad, pierden desempeño significativo hacia *recall* y viceversa. Por otro lado, los modelos propuesta Variables Seleccionadas presentan resultados muy similares, donde la única diferencia es en el desempeño sobre los pacientes No VMI, teniendo Variables Seleccionadas (7) una mayor especificidad y por consiguiente una mayor ROC Área que el modelo de 9 variables.

Tabla 4.3: Métricas de desempeño para los modelos de predicción de VMI Variables Seleccionadas (7), Variables Seleccionadas (9), Variables Basales, *Random Forest* (7) y *Random Forest* (9). “VS” representa Variables Seleccionadas, “VB” Variables Basales y “RF” *Random Forest*.

Modelo	<i>Accuracy</i> o		<i>Recall</i> o		ROC Área
	Exactitud	<i>Precision</i>	Sensitividad	Especificidad	
RF(7)	0.70	0.06	<b>0.93</b>	0.70	0.81
RF(9)	0.82	0.10	0.87	<b>0.82</b>	0.84
VB	0.81	0.10	0.90	0.81	0.86
<b>VS(7)</b>	<b>0.82</b>	<b>0.11</b>	<b>0.93</b>	<b>0.82</b>	<b>0.88</b>
VS(9)	<b>0.82</b>	<b>0.11</b>	<b>0.93</b>	0.81	0.87

A esto se agrega que el modelo de Variables Seleccionadas (7) tiene otras ventajas en comparación como: el usuario debe ingresar solo 6 variables al sistema para que éste funcione a máximo desempeño (los días desde el ingreso pueden ser calculados automáticamente); la certeza de las predicciones es relativamente alta como se mostró en la Figura 4.2; exceptuando por la  $paO_2$ , las variables clave del modelo son de signos vitales, que por su naturaleza son más sencillas y rápidas de obtener que los exámenes de laboratorio, impactando positivamente en la práctica. Así, se elige el modelo Variables Seleccionadas (7) como propuesta solución, y se tomara éste para los siguientes análisis.

Como se ha visto hasta ahora, la predicción de VMI toma en consideración una ventana de 48 horas en el futuro, tiempo seleccionado bajo consenso por profesionales de la salud colaboradores del proyecto, teniendo en mente la utilidad práctica. Ésto se traduce en que para el conjunto de prueba, pacientes que si requerirán VMI en tres o más días predichos como positivo por el modelo, son considerados como error. Para probar el comportamiento del modelo en este escenario más amplio, se toman los 18 pacientes que van a VMI del conjunto de prueba y su evolución en el tiempo. En la Figura 4.8 se muestra la probabilidad de requerir VMI de cada paciente durante su hospitalización, predicha por el modelo de Variables Seleccionadas (7). Cada paciente es representado por una línea de color, su primer dato corresponde al día de ingreso al servicio de urgencias y su último dato al día antes de ingresar a VMI.

Tal como se mostró en los resultados del modelo Variables Seleccionadas (7), las predicciones hechas hasta dos días antes del evento presentan una probabilidad relativamente alta de requerimiento de VMI, y existen dos datos que se encuentran bajo el umbral de 0.5. Éstos corresponden a pacientes que según el modelo, están en tendencia a mejorar, pues su probabilidad disminuyó con respecto al día anterior, y el descenso es tal que pueden incluso considerarse como *outliers* dentro del conjunto de prueba. Se ven también otros dos casos que muestran un comportamiento de mejora, disminuyendo su probabilidad de requerir VMI, pero de manera no tan drástica como los anteriores. Excluyendo estos casos, en general los pacientes tienden a mantener o incrementar su probabilidad de requerir VMI a medida que se acercan al evento, en particular si tomamos en cuenta desde tres días antes.

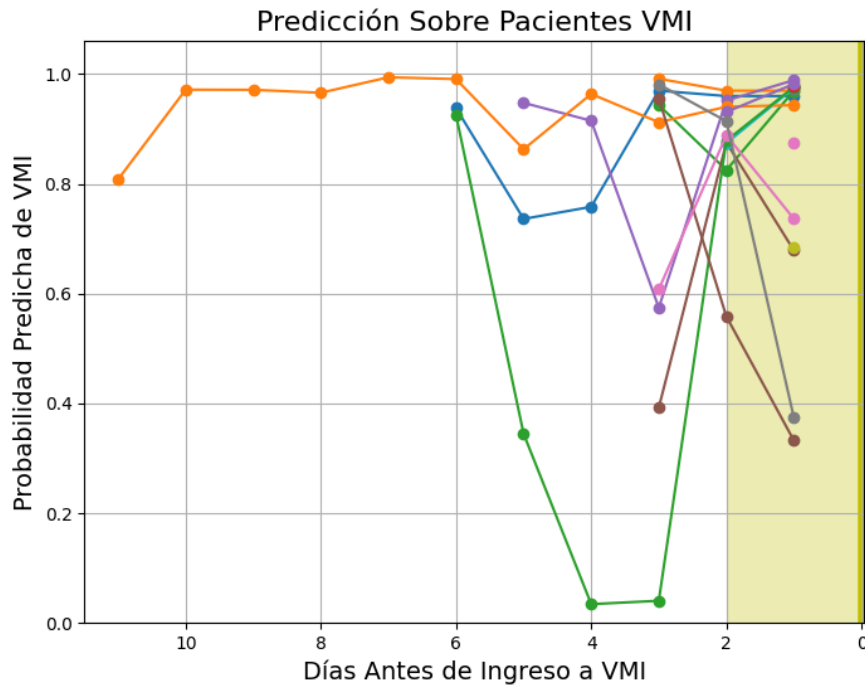


Figura 4.8: Probabilidad de ingresar a VMI predicha por el modelo Variables Seleccionadas (7) de pacientes VMI, calculada por día. En el eje X, Días Antes de Ingreso a VMI, el día 0 demarcado con una línea amarilla a la derecha es el día de ingreso a VMI. Cada punto corresponde a una predicción, y las predicciones de un mismo paciente se encuentran unidas con una línea. La zona destacada en amarillo representa la ventana de predicción de 48 horas elegida.

Del gráfico también se puede ver como los pacientes que entran a VMI, lo hacen mayormente dentro de los primeros tres días desde el ingreso al servicio de urgencias, disminuyendo a solo cuatro pacientes hospitalizados por 4 y 5 días, y a solo uno hospitalizado por 7 días o más. Éste último paciente, hospitalizado por 11 días antes de entrar a ventilación mecánica, es visto por el modelo como que requiere de VMI con más de 0.9 de probabilidad desde el día dos. Las razones por las que no se le asignó un ventilador a este paciente antes pueden ser por falta de camas o ventiladores, el paciente o la familia preferían no hacerlo, el o los profesionales a cargo decidieron que no requería, o sencillamente el modelo está mostrando una alarma donde no la hay.

El caso de este paciente es importante en el análisis, pues por un lado demuestra la incertidumbre propia del problema. Por otro, es un caso atípico, y al ser el único paciente



con más de seis días de hospitalización no logra representar el comportamiento del modelo a esas escalas de tiempo. Así se decide continuar el análisis de este escenario con los días que tengan a lo menos cuatro pacientes, es decir, desde el día cinco en adelante.

En la Figura 4.9 se presentan los diagramas de caja de cada día hasta el quinto antes de ingresar a VMI, en el mismo orden que la Figura anterior. En ésta se ven los *outliers* en el conjunto de prueba, y que los dos pacientes discutidos por tener una probabilidad menor al 50% un día antes de su VMI generan tres de éstos. Observando los *boxplot* en general, se comprueba que las medianas de las probabilidades van disminuyendo mientras más se alejan del día del evento, al igual que las cajas se van posicionando cada vez más abajo.

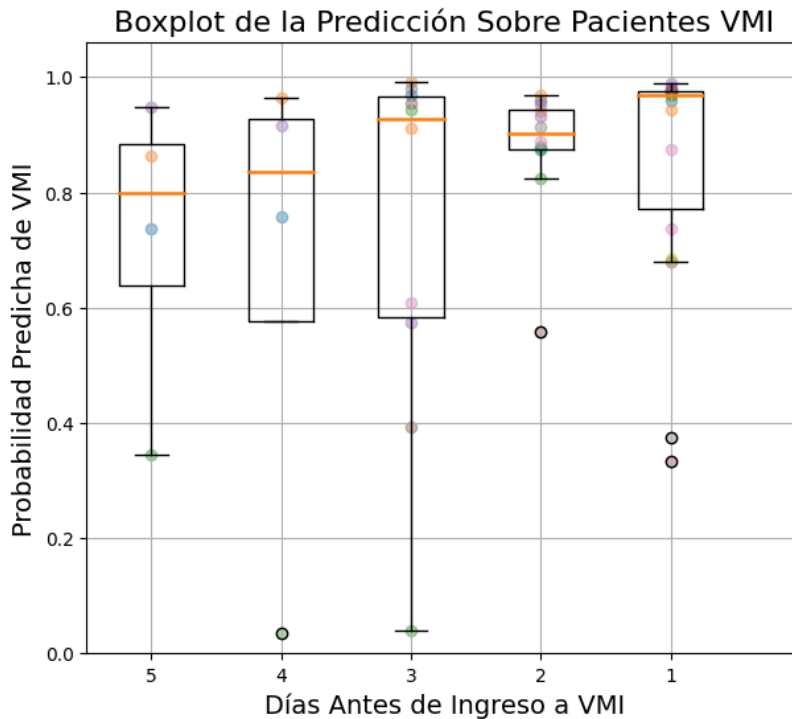


Figura 4.9: Diagramas de caja generados sobre la predicción de VMI por día. La línea naranja dentro de las cajas corresponde a la mediana, y los puntos externos a los *outliers*. Cada punto de color corresponde a una predicción, las cuales son las mismas a las previamente mostradas en la Figura 4.8.

Fuera de los *outliers*, los dos días más cercanos al evento se muestran relativamente compactos en cuanto a la dispersión de las predicciones, mientras que desde el día tres esta aumenta considerablemente. Se puede destacar en el día tres antes del evento, que los pacientes que están sobre la mediana son predichos con más de 0.9 de probabilidad de requerimiento de VMI.

De todo esto, se puede hacer el análisis de que el modelo ha sido diseñado para entregar la probabilidad de que el paciente requiera de VMI en los próximos dos días, pero una descripción más acertada en la práctica es que predice requerimiento de VMI en el “futuro cercano”, siendo futuro cercano una ventana de uno hasta aproximadamente 5 días.

Pasando a la discusión, los resultados obtenidos son comparados con modelos de predicción

de VMI descritos en la literatura, los cuales son mostrados en la Tabla 4.4 y tienen en común entre ellos que predicen en una ventana de 24 horas. Éstos son el estado del arte frente al escenario propuesto, y donde cada uno de ellos ha sido entrenado con diferentes bases de datos. Se encuentran ordenados según el de desempeño sobre la métrica área bajo la curva ROC, que es el que tienen en común todos los trabajos.

Tabla 4.4: Comparación de modelos de predicción de VMI en el estado del arte y en el presente trabajo.

Trabajo	Modelo Base	# Pacientes (VMI)	<i>Acc</i>	<i>Prec</i>	<i>Rec</i>	<i>Esp</i>	ROC Área
Wollenstein, S. <i>et al.</i> [21]	SVM	20,737 (1,809)	0.83				0.77
Marcos, M. <i>et al.</i> [19]	XGBoost	918 (363)					0.86
Burdick, H. <i>et al.</i> [20]	XGBoost	197 (10)			0.90	0.58	0.86
<b>Variables Seleccionadas (7)</b>	<b>XGBoost</b>	<b>395 (69)</b>	<b>0.82</b>	<b>0.11</b>	<b>0.93</b>	<b>0.82</b>	<b>0.88</b>
Shashikumar S., <i>et al.</i> [41]	FNN	3,888 (286)			0.85		0.92
Youssef, A. <i>et al.</i> [22]	GBT	472 (137)	0.70	0.03	0.96	0.70	0.94

**Acc:** Accuracy o Exactitud. **Prec:** Precision. **Rec:** Recall o Sensitividad. **Esp:** Especificidad. **SVM:** Support Vector Machine. **FNN:** Feedforward Neural Network. **GBT:** Gradient Boosting Tree.

Wollenstein, S. *et al.* [21] muestra el modelo de predicción que ha sido entrenado con más datos en la literatura, con 20,737 pacientes COVID-19 positivos de los cuales 1,809 ingresaron a VMI. La base de datos incluye toda persona testada de la enfermedad en México hasta el 1 de Mayo del 2020. En éste se reporta un *accuracy* o exactitud de 0.83, mayor al 0.82 alcanzado por el modelo Variables Seleccionadas (7), y un ROC Área de 0.77, con el menor desempeño dentro del estado del arte. El estudio no detalla como fue trabajado el desbalanceo de datos, ni como fueron conformados los subconjuntos de entrenamiento y prueba. El estudio muestra como *Support Vector Machine* (SVM) obtiene los mismos resultados que RF y XGBoost.

Marcos, M. *et al.* [19] y Burdick, H. *et al.* [20] presentan modelos de predicción de VMI, ambos con un ROC Área de 0.86 y usando XGBoost como base. La similitud en los modelos se contrasta en la cantidad de pacientes con datos disponibles, en particular los que sufren el evento VMI, teniendo el primero 363 y el segundo solo 10 pacientes. Se destaca este último por obtener un *recall* o sensibilidad de 0.90, lo cual implica que tiene buen rendimiento sobre la clase “VMI”, aunque a costa de solo un 58% de aciertos para “No VMI”. Esto hace sentido por el gran desbalance de clases, el cual no es mencionado en el texto.

En Marcos, M. *et al.* [19] se identifican 7 variables importantes para la clasificación: SatO<sub>2</sub>/FiO<sub>2</sub>, Edad, Estimación de la Tasa de Filtración Glomerular, Procalcitonina, PCR, Índice de Comorbilidad de Charlson y Linfocitos. Burdick, H. *et al.* [20] utilizan 12 variables (sin detallar el porqué de la selección): Pad, Pas, Frecuencia Cardiaca, Temperatura, Frecuencia Respiratoria, SatO<sub>2</sub>, Glóbulos Blancos, Plaquetas, Lactato, Nitrógeno Ureico, Creatina, y Bilirrubina. En ambos casos se pueden ver similitudes en las variables clínicas utilizadas con las Variables Basales elegidas en el presente. En particular, las variables que se repiten con las del modelo de Variables Seleccionadas (7) son FiO<sub>2</sub> y SatO<sub>2</sub> para ambos, y Frecuencia Cardiaca, Temperatura Axilar y Frecuencia Respiratoria para el segundo.

Una interesante aproximación al problema es presentada por Shashikumar, S. *et al.* [41], en donde desarrollan una *Feedforward Neural Network* (FNN) de 2 capas con tamaños 40

y 25, entrenada a partir de 3,888 pacientes COVID-19 positivos. Utilizando la red neuronal logran obtener una ROC Área de 0.92, y un *recall* o sensibilidad de 0.85. Una mejor área bajo la curva ROC, pero menor *recall* en comparación al modelo de Variables Seleccionadas (7) habla de que el primero es capaz de predecir mejor sobre la clase “No VMI”, entregando una especificidad relativamente elevada (no es reportada) a costa de su desempeño sobre la clase “VMI”, teniendo el peor *recall* registrado en comparación al resto de trabajos en el estado del arte.

Youssef, A. *et al.* [22] presenta un modelo basado en GBM, cuyo estudio dispone 472 pacientes de los cuales 137 requirieron ventilación mecánica. En cuanto a cantidad de datos, este trabajo es el que más se asemeja al desarrollado en este documento, con la diferencia de que posee alrededor del doble de pacientes de clase “VMI”. La publicación logra los mejores resultados frente al escenario planteado, destacándose en su ROC Área de 0.94 y su *recall* o sensibilidad de 0.96, mostrando un muy buen desempeño relativo frente a la clase “VMI”, aunque con una especificidad de 0.70. Los resultados mostrados de Variables Seleccionadas (7) son calculados utilizando un umbral del 50 % para la clasificación, pero si movemos este a 35 % el desempeño del *recall* o sensibilidad queda en 0.97, con una especificidad de 0.76, lo cual es aún mejor que el resultado recién discutido del estado del arte.

Como se puede comprobar con el análisis, el modelo Variables Seleccionadas (7) es preferible sobre el Variables Seleccionadas (9), tanto en desempeño como en facilidad de uso en el escenario clínico. Los resultados de éste primer modelo son comparables con el estado del arte, y dependiendo del umbral que se elija, pueden alcanzar el desempeño más elevado dentro de lo que se observa en la literatura. Sobre esto cabe destacar que las métricas utilizadas para hacer la comparación están limitadas a problemas de clasificación (a excepción de ROC Área), y que en la práctica no se elegiría un umbral de predicción, si no que el modelo entregaría directamente la probabilidad de ocurrencia del evento VMI. Finalmente, se aprecia que las variables seleccionadas en el modelo de predicción diseñado coinciden en su mayor parte por las utilizadas en la literatura. Una diferencia a notar es la variable  $paO_2$ , que es elegida tanto por XGBoost como por el análisis SHAP como una de las que tiene más importancia en la predicción, y no es utilizado por ninguno de los estudios mencionados, exceptuando por Youssef, A. *et al.* [22] que muestra el mejor desempeño en ROC Área.

A modo de conclusión, se muestra que el modelo de predicción de ventilación mecánica invasiva diseñado Variables Seleccionadas (7) logra discernir entre pacientes que requieren VMI en el futuro cercano y los que no, entregando a su salida la probabilidad de que ocurra el evento. Para ello 7 variables claves son utilizadas, las cuales alcanzan y mejoran el rendimiento del modelo que utiliza las 57 variables clínicas puestas a estudio. Debido a su rendimiento superior, requerimiento de menos variables y que las variables que utiliza son más simples de obtener en la práctica clínica, se elige el modelo de Variables Seleccionadas (7) sobre el de 9 variables. La propuesta alcanza un desempeño a la par de lo que se ve en la literatura, teniendo énfasis en la correcta predicción de verdaderos positivos demostrado en su *recall* o sensibilidad de 0.93, que puede alcanzar hasta un 0.97 según el umbral de clasificación que se elija.

## 4.2. Predicción de Mortalidad

Para el caso de mortalidad, se referirá a la propuesta de modelo de predicción como modelo de “Variables Seleccionadas”, cuyas características de entrada son extraídas de:

**Variables Seleccionadas:** pCO<sub>2</sub>, paO<sub>2</sub>, Frecuencia Cardiaca, SatO<sub>2</sub>, Nitrógeno Ureico, PCR, FiO<sub>2</sub>.

En la Figura 4.10 (a) se ve la matriz de confusión representativa de un *accuracy* de 0.82, simétrica para ambas clases. En la Tabla 4.5 se muestran las métricas de desempeño del modelo, en donde se destaca el *recall* de 82% de aciertos para verdaderos positivos. La métrica *precision* nos indica por otro lado que del total de “Fallece” predichos, solo un 27% de ellos realmente fallecen en los próximos cinco días. Al igual que en el caso de VMI, esto es debido al desbalance en las clases, donde el 18% de pacientes falsos positivos son 133 datos, y el 82% de verdaderos positivos son 28.

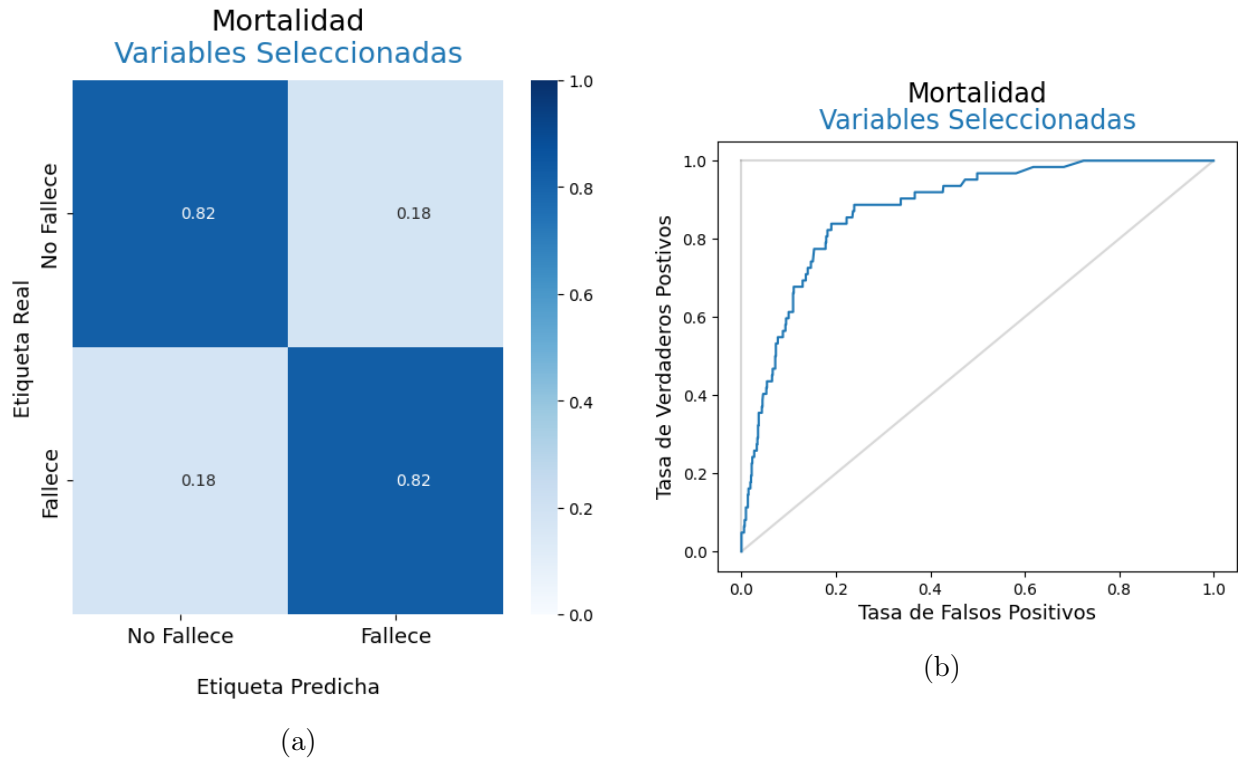


Figura 4.10: Resultados del modelo de predicción de Mortalidad asociado a las Variables Seleccionadas. (a) Matriz de confusión. (b) Curva ROC.

Tabla 4.5: Métricas de desempeño para el modelo de predicción de Mortalidad Variables Seleccionadas.

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	Especificidad	ROC Área
0.82	0.27	0.82	0.82	0.82

En la Figura 4.10 (b), se aprecia que la curva ROC comienza a converger cerca de un 45%

de la tasa de falsos positivos, con un desempeño de más del 95 % en la predicción de la clase “Fallece”. Es decir, se puede elegir un umbral en el modelo que exagere la clase 1 “Fallece”, clasificando al rededor de un 45 % de personas que no fallecerán de manera equivocada, con la compensación de tener un acierto del 95 % de las personas que si fallecerán. Si se desea ser un poco menos conservador, se puede tener un *recall* de 0.9 sacrificando falsos positivos a una tasa del 25 %. En su esencia, el modelo tiene el potencial de predecir con gran cantidad de aciertos sobre la clase “Fallece” pero hay que ser consciente del aumento en los falsos positivos que ésto produce, tal como lo muestra la curva ROC.

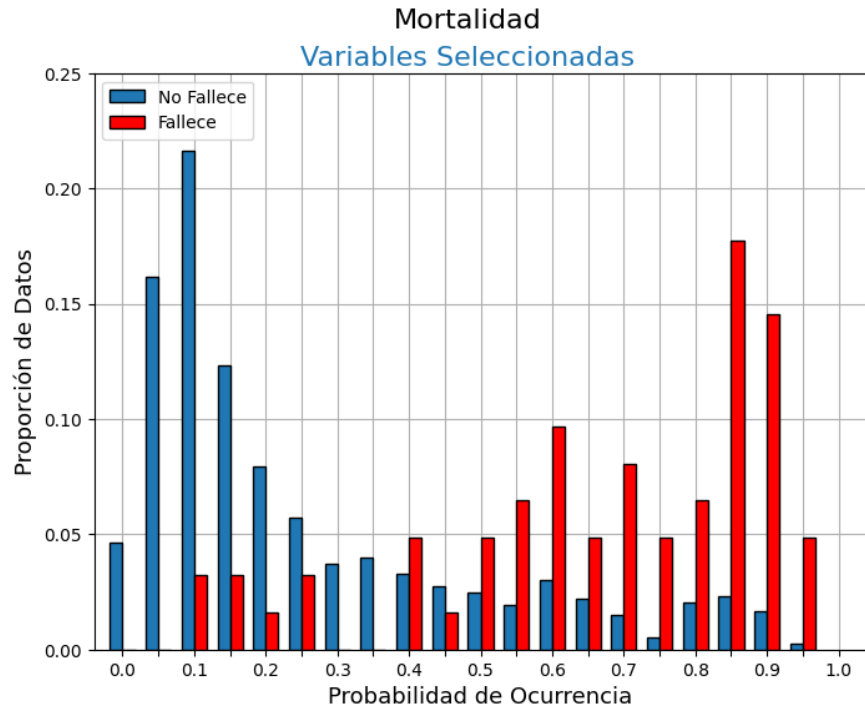


Figura 4.11: Histograma de probabilidad predicha para el modelo de predicción de Mortalidad asociado a las Variables Seleccionadas. En azul los pacientes de clase 0 “No Fallece” y en rojo los de clase 1 “Fallece”.

Al analizar el histograma de probabilidades de ocurrencia en la Figura 4.11, este da a parecer que la clase “No Fallece” tiene más certeza en la predicción, teniendo más datos cercanos a la probabilidad 0, que datos cercanos a la probabilidad 1 para la clase “Fallece”, cuya distribución es más ancha. Como contraste, la matriz de confusión en la Figura 4.10 (a) nos muestra un comportamiento simétrico, aludiendo a un desempeño similar sobre ambas clases. Ésto se debe a que la matriz de confusión asume un umbral de clasificación de 0.5, mientras que el histograma muestra distribuciones de probabilidad, que representan de manera diferente el escenario de predicción de probabilidad de fallecer que una métrica de clasificación que fuerza a una u otra clase. La intersección de las distribuciones de ambas clases no es menor, a diferencia de lo visto en los modelos de predicción de VMI, lo cual refleja la dificultad que tiene el modelo de desempeñar sobre este problema.

En la Figura 4.12 se muestra el análisis SHAP del modelo en cuestión. Las características correspondientes a desviación estándar y pendiente son las que menos impacto tienen en el modelo, mostrándose a lo más dos veces y en variables que ya se encuentran en el top 15.

La variable Frecuencia Cardíaca está asociada a la clase “Fallece” a mayor sea su valor, con un comportamiento característico. Su aporte a la clase “No Fallece” es relativamente bajo en comparación a los demás, generando un gran cúmulo de puntos cerca del valor SHAP 0.0. Al tomar un valor alto, ésta salta a ser una de las que más impacto tiene sobre la clase “Fallece”, con los datos muy esparcidos en el eje horizontal, dejando un hueco en los valores intermedios.

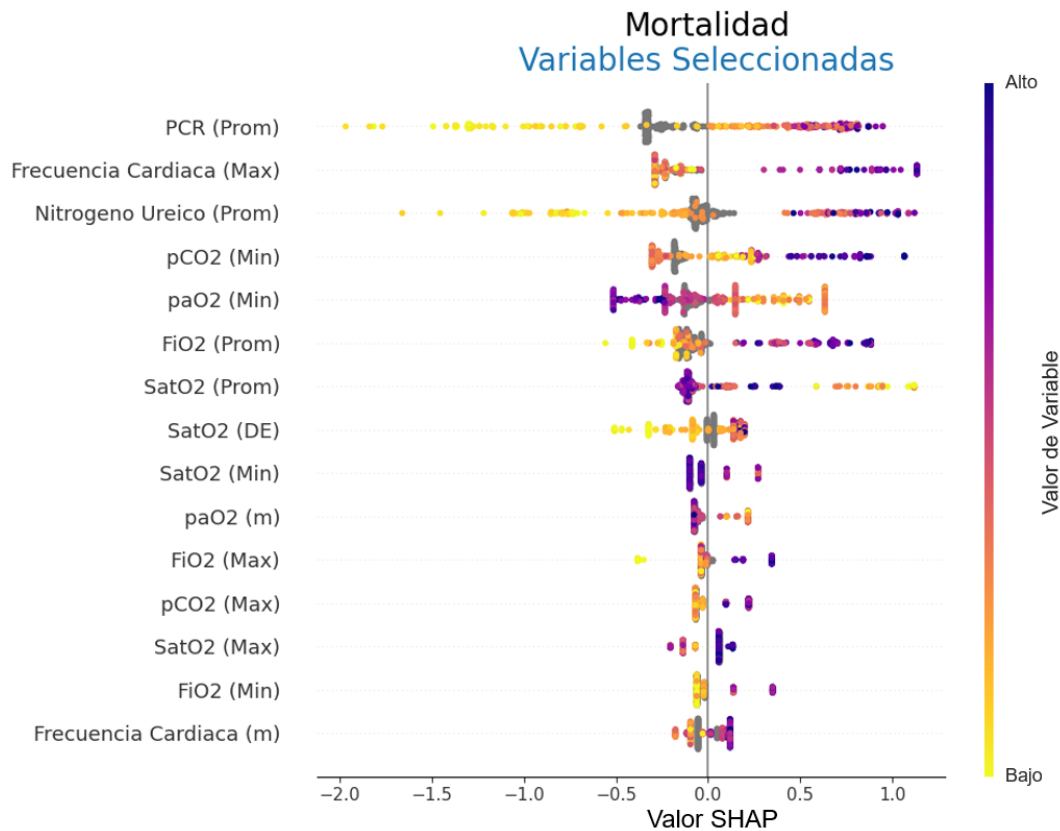


Figura 4.12: Análisis SHAP para el modelo de predicción de Mortalidad asociado a las Variables Seleccionadas. Un valor SHAP positivo está relacionado con la clase “Fallece”, y uno negativo con “No Fallece”. Un color más oscuro representa un valor mayor relativo a la variable y gris un valor faltante.

(Prom: Promedio, Min: Mínimo, Max: Máximo, DE: Desviación Estándar, m: Pendiente).

La variable Proteína C Reactiva (PCR) vuelve a aparecer en predicción de Mortalidad, en donde un valor relativamente alto empuja hacia la clase “Fallece”. Se aprecia un cúmulo de valores faltantes con valor SHAP negativo, lo que implica que los pacientes a los que no se les toma éste examen suelen no estar en estado crítico. Esta vez ésta variable muestra ser la que más impacta en el modelo, según su propagación de los valores SHAP.

Continuando con Nitrógeno Ureico, es de interés como los valores faltantes en esta variable se encuentran centrados en el valor SHAP 0.0 a diferencia del resto de los exámenes. Esto se puede interpretar como que éste examen se toma sin mayor prejuicio del estado crítico del paciente. Un valor relativo alto de Nitrógeno Ureico esta directamente relacionado con la probabilidad de Fallecer del paciente según el análisis SHAP.

Para las demás variables, el comportamiento se condice con lo visto en los modelos de

predicción de VMI, lo cual tiene sentido teniendo en cuenta que ambos escenarios muestran un estado crítico del paciente. En particular se ve el par  $paO_2$ ,  $FiO_2$  con una muy similar influencia sobre la decisión, mismo análisis que se abordó en el caso anterior.

Analizando el desempeño del modelo de Variables Seleccionadas con uno que utilice todas las Variables Basales, se pueden comparar en la Figura 4.13 las curvas ROC de ambos. La curva del modelo de Variables Basales tiende a estar sobre la del modelo propuesto en algunos tramos, aunque la diferencia no es particularmente sustancial. El comportamiento de ambas curvas tienen la misma tendencia en forma y velocidad con la que escalan, y con simetría para ambas clases. Ésto muestra que el modelo de Variables Seleccionadas se logra ajustar a lo que los datos del estudio permiten obtener, y utilizando tan solo 7 variables.

En cuanto a las métricas de desempeño, se entrena un modelo basado en *Random Forest* con las mismas condiciones que el propuesto, con tal de comparar el rendimiento de XGBoost frente a otros algoritmos de inteligencia artificial. Éste es superado en todas las métricas como se ve en la Tabla 4.6, estando al menos 0.05 por debajo del modelo Variables Seleccionadas. Por otro lado, se observa como el modelo de Variables Basales obtiene 0.01 más de rendimiento en *accuracy* o exactitud, especificidad y ROC Área sobre el de Variables Seleccionadas. Ya que el *recall* o sensibilidad es el mismo en ambos, se concluye que esto es debido a que el de Variables Basales predice un tanto mejor a los pacientes No Fallece en comparación, lo cual se ve reflejado también en la curva ROC. A pesar de esto, el modelo de Variables Seleccionadas presenta la ventaja de que en la práctica clínica el personal de salud requerirá ingresar solo 7 variables para que el modelo funcione en su máximo desempeño, y manteniendo un desempeño similar a utilizar 57.

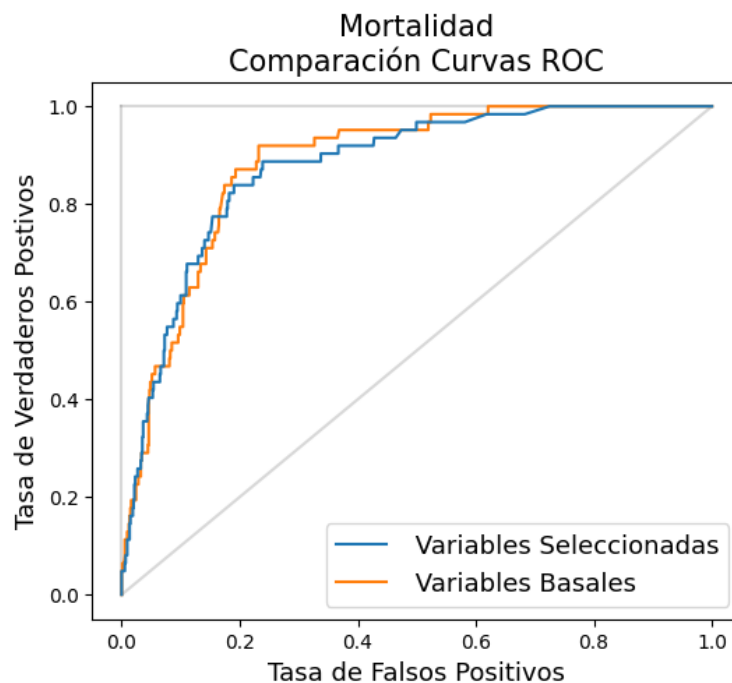


Figura 4.13: Curvas ROC del modelo de predicción de Mortalidad de Variables Seleccionadas y de Variables Basales.

Al igual que en el caso anterior, se desea examinar el comportamiento del modelo a más

Tabla 4.6: Comparación de métricas de desempeño para los modelos de predicción de Mortalidad Variables Seleccionadas, Variables Basales y *Random Forest*. “VS” representa Variables Seleccionadas, “VB” Variables Basales y “RF” *Random Forest*.

Modelo	<i>Accuracy</i> o		<i>Recall</i> o		
	Exactitud	<i>Precision</i>	Sensitividad	Especificidad	ROC Área
RF	0.77	0.20	0.69	0.77	0.73
<b>VS</b>	0.82	0.27	<b>0.82</b>	0.82	0.82
<b>VB</b>	<b>0.83</b>	<b>0.28</b>	<b>0.82</b>	<b>0.83</b>	<b>0.83</b>

de solo 5 días en su predicción. En la Figura 4.14 se muestran los 13 pacientes que Fallecen del conjunto de prueba y la probabilidad dada por su predicción de mortalidad durante su hospitalización. Manteniendo la misma idea de representatividad que en el caso de VMI, y debido a que existen pacientes con más de 40 días de hospitalización dejando el gráfico difícil de leer, se acota la información a 15 días antes del evento. Cada paciente es representado por una línea de color, su primer dato corresponde al día de ingreso al servicio de urgencias (con pacientes que ingresan antes de lo que muestra la Figura) y su último dato al día antes de fallecer.

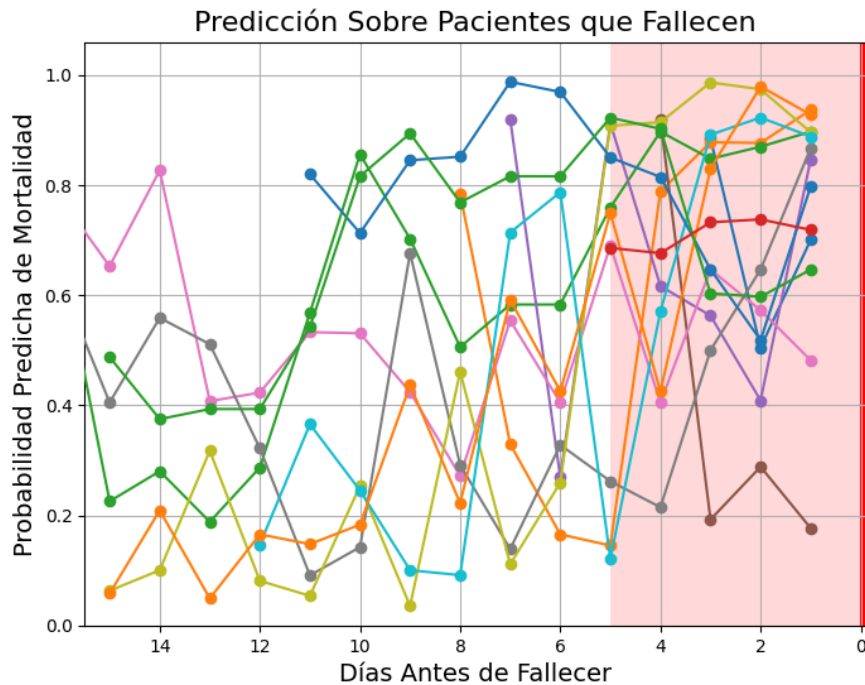


Figura 4.14: Probabilidad de fallecer predicha por el modelo Variables Seleccionadas de pacientes que fallecen, calculada por día. En el eje X, Días Antes de Fallecer, el día 0 demarcado con una línea roja a la derecha es el día de defunción. Cada punto corresponde a una predicción, y las predicciones de un mismo paciente se encuentran unidas con una línea. La zona destacada en rojo representa la ventana de predicción de 5 días horas elegida.

El comportamiento del modelo sobre las predicciones se ve más ruidoso que en el caso de VMI a pensar de tener cinco pacientes menos. Una razón de ésto es que, a diferencia del caso anterior en el que los pacientes sufren del evento clínico a los pocos días del ingreso de manera consistente, para el caso de mortalidad los paciente pueden estar hospitalizados



durante bastante más tiempo antes de fallecer, y tiene sentido pues es un suceso, no una decisión clínica como VMI.

Se puede notar como el día antes del evento, casi todos los pacientes son predichos con más de 50% de probabilidad de fallecer exceptuando por uno, y para los siguientes días hasta el quinto, los pacientes se encuentran en mayor parte sobre el 40%. En el sexto día el escenario se comporta más difuso, y de ahí a más días el modelo tiende a entregar menos probabilidad de fallecer en general, aunque con una gran dispersión.

En la Figura 4.15 se presentan los diagramas de caja hasta el día 15 antes de fallecer, los cuales se condicen con lo mencionado. En general, las cajas poseen una gran dispersión si se comparan con las revisadas en Predicción de VMI, dando a parecer que predecir mortalidad es una tarea más difícil de afrontar. La gran dispersión permite que los *boxplot* identifique un único *outlier*, tomando los demás valores extremos como parte de los cuartiles.

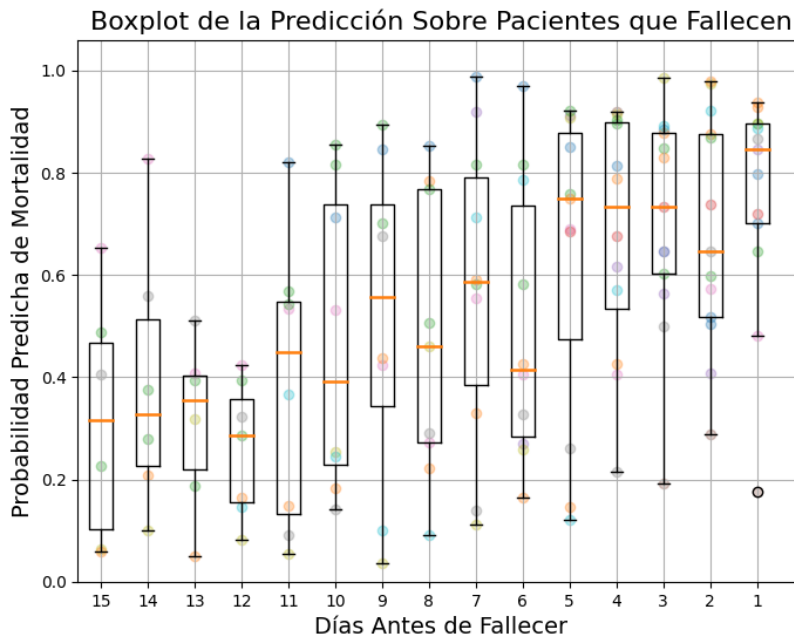


Figura 4.15: Diagramas de caja generados sobre la predicción de Mortalidad por día. La línea naranja dentro de las cajas corresponde a la mediana, y los puntos externos a los *outliers*. Cada punto de color corresponde a una predicción, las cuales son las mismas a las previamente mostradas en la Figura 4.14.

Como es de esperarse, las cajas tienen tendencia a bajar en el eje y a medida que se alejan de la fecha de defunción, lo que quiere decir que el modelo entrega una mayor probabilidad de fallecer en general cuando el paciente está más cerca del evento. Es importante notar también que durante los cinco días antes de fallecer, la mediana de la probabilidad predicha es mayor a un 60%, y excluyendo el día 2, mayor a un 70%.

El hecho de que el modelo tenga mejores resultados exactamente cinco días antes de fallecer, puede deberse a que fue entrenado justamente para ese propósito, y de haber sido entrenado con una ventana de predicción más amplia, es posible que los resultados entregues probabilidades más altas para días más lejanos.

Entrando en la discusión, considerando otros trabajos realizados en el área, existe una extensa cantidad de modelos de predicción de mortalidad propuestos en la literatura, en comparación al escenario anterior de VMI. Una recopilación del estado del arte se muestra en la Tabla 4.7 ordenada según el desempeño de los modelos sobre ROC Área, métrica compartida por todos.

Tabla 4.7: Comparación de modelos de predicción de Mortalidad en el estado del arte y en el presente trabajo.

Trabajo	Modelo Base	# Pacientes (Fallece)	<i>Acc</i>	<i>Prec</i>	<i>Rec</i>	<i>Esp</i>	ROC Área
Halasz G. <i>et al.</i> [42]	Naive Bayes	852 (293)	0.67		0.88	0.55	0.80
<b>Variables Seleccionadas</b>	<b>XGBoost</b>	<b>395 (58)</b>	<b>0.82</b>	<b>0.27</b>	<b>0.82</b>	<b>0.82</b>	<b>0.82</b>
Emami H. <i>et al.</i> [16]	GBT	4,120 (~25%)	0.64		0.81	0.61	0.85
Estiri H. <i>et al.</i> [17]	GBT	16,709 (830)					0.90
Yan L. <i>et al.</i> [12]	XGBoost	375 (174)	0.94				0.95
Karthikeyan A. <i>et al.</i> [43]	LR	370 (169)	0.97				0.99

**Acc:** Accuracy o Exactitud. **Prec:** Precision. **Rec:** Recall o Sensitividad. **Esp:** Especificidad. **GBT:** Gradient Boosting Tree. **LR:** Logistic Regression.

Halasz G. *et al.* [42] presenta un modelo de predicción basado en Naive Bayes, obteniendo un ROC Área de 0.80 y un *recall* o sensibilidad de 0.88, a costa de un *accuracy* relativamente bajo de 0.67 y una especificidad de 0.55. 852 pacientes fueron enrolados en el estudio, de los cuales el 34% falleció en el Hospital. Se realiza selección de variables según importancia de estas en el modelo, eligiendo: Edad, Hemoglobina,  $paO_2/FiO_2$ , Temperatura, Previo Infarto Cerebrovascular y Género. De estas seis,  $paO_2$  y  $FiO_2$  son variables en común con el modelo de Variables Seleccionadas.

Emami H. *et al.* [16] y Estiri H. *et al.* [17] desarrollan modelos de predicción basados en *Gradient Boosting Trees* con datos de más de mil pacientes, lo que es una cantidad relativamente alta para el presente escenario. En Emami H. *et al.* [16] utilizan 38 variables de registro de salud electrónico sin detallar cuales, obteniendo un ROC Área de 0.85 y un *recall* de 0.81, otra vez a costa de un menor desempeño sobre la clase “No Fallece” obteniendo 0.61 de especificidad. En Estiri H. *et al.* [17] utilizan 51 variables clínicas elegidas bajo un meticuloso proceso de selección. Con esto obtienen un ROC Área de 0.90, un resultado muy positivo en comparación a los demás modelos descritos, sin reportar más métricas.

En este último trabajo también se da consideración a la importancia de la variable Edad. A diferencia de la aproximación tomada en el modelo Variables Seleccionadas, Estiri H. *et al.* [17] deciden separar el problema en tres escenarios distintos: 45-65, 65-85 y 85+ años, tomando ventaja de la gran cantidad de datos disponibles. Separar los datos por grupos de edades les entrega un comportamiento negativo en la predicción de mortalidad en comparación a su resultado global, con ROC Áreas de 0.8, 0.75 y 0.68 respectivamente.

En el trabajo de Yan L. *et al.* [12] se utiliza un *dataset* de Wuhan, China de pacientes registrados en una etapa temprana de la pandemia entre el 10 de Enero y el 18 de Febrero. Luego de un proceso de exclusión se consideran 375 pacientes COVID-19 positivos, de los cuales 174 fallecen en su periodo de hospitalización. Éste estudio ampliamente citado en la literatura presenta un desempeño de 0.94 de *accuracy* o exactitud y ROC Área de 0.95,

utilizando un modelo de predicción basado en XGBoost. Para obtener estos resultados se utilizan tres variables claves de entrada: LDH, PCR y Linfocitos, las cuales pertenecen a exámenes de laboratorio. De éstas se tiene en común con el modelo Variables Seleccionadas la variable PCR, que según el análisis SHAP presentado en la Figura 4.12, tiene un impacto considerable en el modelo. En la publicación se menciona como el modelo tiene un *accuracy* mayor a 90 % para predecir mortalidad con incluso más de 10 días de antelación, siendo uno de los pocos estudios que presenta la variable temporal de la predicción.

Karthikeyan A. *et al.* [43] presenta diversos modelos de inteligencia artificial, en donde el que alcanza el mejor desempeño es uno basado en regresión logística con un ROC Área de 0.99 y *accuracy* o exactitud de 0.97. Es de interés recalcar que el estudio es llevado a cabo utilizando la misma base de datos que Yan L. *et al.* [12], con la diferencia de que con sus criterios de exclusión se consideran 370 pacientes de los cuales 169 fallecen. La publicación presenta los mejores resultados en cuanto a predicción de mortalidad para el escenario planteado, reportando cinco variables claves: Neutrófilos, PCR, Edad, Linfocitos y LDH, las cuales se condicen con la publicación anterior, y agrega un examen de laboratorio más y la edad del paciente.

Estos últimos dos trabajos muestran desempeños particularmente buenos para el escenario planteado, y han sido ampliamente citados en la literatura por ello. La cantidad de pacientes que utilizan en el estudio es cercana a la cantidad de pacientes en el presente documento, pero tienen la diferencia de tener los pacientes fallecidos durante la hospitalización balanceados con los que no fallecen. Ambos estudios utilizan el mismo *dataset* de Wuhan, China, lo que es relevante pues es posible que dicha base de datos sea particularmente consistente para predecir mortalidad, haciéndolos sobresalir en el estado del arte sin lograr ser alcanzados en desempeño por otros estudios.

Fuera de éstos, los resultados obtenidos por el modelo Variables Seleccionadas son comparables con los del estado del arte, obteniendo un desempeño similar, aunque con una menor ROC Área en general. Las bases de datos utilizadas por los esfuerzos realizados en el escenario planteado son en su mayor parte abundantes, con miles de datos, lo que permite compensar el gran desbalance entre pacientes que fallecen y los que no. El modelo de Variables Seleccionadas obtiene valor al ser entrenado con una base de datos local, con una acotada cantidad de pacientes y gran desbalance entre ellos, y lograr desempeñar la tarea propuesta a un nivel similar al de otros trabajos en la literatura. Como se muestra en la Tabla 4.7, el modelo con menor cantidad de datos posee 825 pacientes, de los cuales fallecen 293, cantidad muy superior a la utilizadas por el modelo Variables Seleccionadas, con aproximadamente cinco veces menos pacientes de la clase Fallece.

Así, en conclusión, se muestra que el modelo de predicción de mortalidad diseñado, Variables Seleccionadas, logra discernir entre pacientes que fallecerán en el futuro cercano y los que no, entregando a su salida la probabilidad de que ocurra el evento. Para ello, al igual que en el caso de VMI, 7 variables claves son utilizadas, las cuales alcanzan el rendimiento del modelo que utiliza las 57 Variables Basales puestas a estudio, aunque en éste caso ellas se componen en gran medida de exámenes de laboratorio, las cuales son por su naturaleza más complejas de obtener. La propuesta de modelo alcanza un desempeño comparable con lo que se ve en la literatura tal como fue discutido, mostrando una tendencia simétrica tanto

para el evento Fallece como No Fallece, reflejado en su *accuracy* o exactitud de 0.82 aun con las clases altamente desbalanceadas.

# Capítulo 5

## Conclusiones

Un sistema de predicción de eventos clínicos adversos para pacientes COVID-19 ha sido diseñado basado en árboles de decisión XGBoost, en base a datos recolectados en el Hospital Clínico de la Universidad de Chile (HCUCH) en los primeros 100 días de la pandemia. Éste consta de dos propuestas de modelo, un modelo de predicción para indicación de Ventilación Mecánica Invasiva (VMI) en los próximos 2 días, y uno de predicción de mortalidad en los próximos 5 días. Una exhaustiva revisión y curación de datos es llevada a cabo sobre la base de datos, con ayuda de personal de salud cooperadores del proyecto, logrando sobreponerse a las imperfecciones de calidad de datos presentes, producidas por factor humano al momento de ser registrados. Este proceso es fundamental para la creación de modelos exitosos, en especial en el escenario de medicina, en donde se trabaja con información sensible como lo son los datos clínicos.

En cuanto a la tarea de predicción, para el caso de VMI, el modelo propuesto es entrenado con 7 variables clínicas que demostraron tener más impacto sobre el escenario: Presión Arterial de Oxígeno, Fracción Inspirada de Oxígeno, Frecuencia Respiratoria, Tiempo Hospitalizado, Temperatura Axilar, Frecuencia Cardíaca y Saturación de Oxígeno. Exceptuando por la primera que es un examen de laboratorio, estas variables corresponden a signos vitales, las que requieren de un proceso de baja complejidad para su obtención. El desempeño del modelo presenta un *recall* o sensibilidad de 0.93 (considerando el umbral de 50%) y área bajo la curva ROC de 0.88. Un segundo modelo es entrenado con 9 variables clínicas: las mismas 7 anteriores más Proteína C-reactiva y Lactato Deshidrogenasa. Ésta propuesta es descartada en el análisis de resultados, pues su desempeño es similar (y en casos peor) que el del modelo de 7 variables. En particular, sus distribuciones de probabilidad muestran una menor certeza de predicción, y no menos importante, la inclusión de estos dos exámenes de laboratorio hacen de este modelo más complejo de utilizar a su máxima capacidad en la práctica clínica, por la dificultad y tiempo que toma obtenerlos.

Para el caso de predicción de mortalidad, el modelo propuesto posee como variables más importantes la Presión Arterial de Dióxido de Carbono, Presión Arterial de Oxígeno, Frecuencia Cardíaca, Saturación de Oxígeno, Nitrógeno Ureico, Proteína C-reactiva y Frecuencia Inspirada de Oxígeno. A diferencia del caso anterior, cinco de las siete variables de mayor importancia para el modelo corresponden a exámenes de laboratorio, lo que añade a la pre-

dicción de mortalidad complejidad relativa de uso en la práctica en comparación al escenario de VMI. El desempeño del modelo presenta un *recall* o sensibilidad de 0.82 (considerando el umbral de 50 %) y área bajo la curva ROC de 0.82.

La selección de variables en los modelos es llevada a cabo en primera instancia con apoyo de profesionales de la salud, los que eligen por conocimiento experto variables a considerar dada su relevancia, logrando disminuir la cantidad de éstas de más de 300 a 57. Desde este punto, un análisis de sensibilidad es realizado tomando ventaja de XGBoost para analizar la importancia de cada variable sobre la predicción. En términos del modelo, la selección de variables permite hacer la predicción con un menor costo computacional, manteniendo el desempeño en el caso de mortalidad, y mejorándolo en el caso de VMI. Para la práctica clínica, poder utilizar el modelo con solo 7 variables tiene un impacto aún mayor. Cada vez que un dato clínico del paciente es actualizado, este debe ser ingresado de manera manual en el computador, lo que hace que tener 57 variables sea una solución inviable, dando al personal de salud una labor cansadora y que consume tiempo, o sencillamente quitarles deseo de usar el modelo. La idea de este trabajo tiene como núcleo generar una herramienta de apoyo al personal, lo que convierte a la selección de variables en una tarea esencial.

En el diseño de los modelos se realiza extracción de características considerando un día de datos de cada variable, con tal de que el modelo tenga noción de la evolución temporal del paciente. Las características utilizadas son el promedio, la desviación estándar, el máximo, el mínimo y la pendiente, esta última calculada a partir de una regresión lineal proyectada. A pesar de lo esperado, los modelos tienden a utilizar los valores promedios, máximos y mínimos por sobre otros con más información evolutiva, como por ejemplo la pendiente.

Los resultados obtenidos en los modelos de predicción son comparables con lo mostrado en la literatura, y dependiendo del umbral que se elija, estos alcanzan el máximo desempeño en torno al *recall* o sensibilidad del estado del arte, o lo superan en el caso del modelo de predicción de VMI. El sistema propuesto muestra una aproximación poco explorada, realizando extracción de características y con ello considerando el factor temporal de los datos, aunque como mencionó, ésta no parece tener particular relevancia en la predicción, pues hay tendencia a utilizar características como el promedio en vez de la pendiente. Se puede mencionar también que las variables seleccionadas se asemejan en gran medida a las utilizadas en otros trabajos para ambos eventos adversos. Por último, los modelos destacan por lograr desempeñarse a pesar de la acotada base de datos disponible, inferiores en cantidad a las disponibles en la literatura. Queda en discusión el caso de mortalidad, el que muestra un desempeño relativamente menor en torno a ROC Área, y que una posibilidad de esto es que el escenario requiere de una mayor cantidad de datos para alcanzar la representatividad.

Los modelos son validados gracias a los resultados obtenidos sobre conjuntos de prueba, y dándoles sentido clínico a partir del análisis SHAP por profesionales de la salud. Se considera de todas maneras este hito como parcialmente logrado, puesto que no se logra poner el sistema a prueba en la práctica clínica para su validación final. Esto es debido a que el sistema fue entrenado y validado con datos pertenecientes al primer escenario de la pandemia, con pacientes COVID-19 positivos no vacunados. Se solicitaron datos de pacientes vacunados para probar el desempeño del modelo, pero tal como se mencionó durante el presente, el escenario crítico de los centros de salud hizo de esto una tarea larga. La segunda vacuna fue

administrada antes de recibir los datos, dificultando aún más el proceso, puesto que ahora se necesitaban datos de pacientes con dos vacunas. A esto se suma que debido a la disposición de datos en el hospital, no es directo obtener los datos clínicos de los pacientes de manera automatizada, como tampoco lo es instalar el sistema de predicción, lo cual escapa de los alcances de la tesis.

Gracias al trabajo interdisciplinario del proyecto en el que se sitúa esta tesis, se logra diseñar y validar los modelos de predicción a partir de datos chilenos, demostrando su utilidad para uso local. Destacar de esto la disposición e importancia de un equipo de trabajo con conocimiento experto en las áreas a trabajar, ya que el personal de salud que apoyó a este trabajo cumplió un rol fundamental en la recolección de datos, preprocesamiento y curación de éstos, selección y exclusión de variables según sentido clínico, y validación de resultados en torno al comportamiento de las variables sobre el modelo. Algo tan mínimo como el lenguaje utilizado en el área y sus diferencias, puede ser una gran limitante al momento de realizar un proyecto.

Limitaciones: El sistema ha sido entrenado y validado para predecir el evento indicación de VMI en las próximas 48hr, y mortalidad en los próximos cinco días, para pacientes COVID-19 no vacunados, y su uso en el estado actual debe considerar este escenario para su correcto funcionamiento.

## 5.1. Trabajo Futuro

El presente trabajo deja abierta la posibilidad de ahondar más en la solución desde distintas aristas. Se requiere realizar una validación de los modelos con datos de pacientes del escenario actual, y en caso de ser necesario, reentrenarlos para conseguir el desempeño requerido. Introducir el sistema de predicción en la práctica clínica, lo que requiere tener acceso a los datos del hospital, levantar un servidor con el sistema, y todo lo que conlleve. Probar el sistema sobre otras enfermedades, respiratorias o no respiratorias, rediseñando los modelos de manera oportuna. Es posible entrenar los modelos con distintas ventanas de predicción, y con ello probar si mejora el rendimiento sobre predicciones más lejanas en días al evento. Y finalmente, incluir datos imagenológicos de los pacientes en los modelos, con lo que se espera que el desempeño mejore, tarea que no fue posible en el trabajo debido a la acotada cantidad de usuarios con exámenes de Rayos-X o tomografías computadas dentro del estudio.

# Bibliografía

- [1] I. Siches, J. Vega, M. Chomalí, B. Yarza, R. Estay, M. Goyenechea, P. Jiménez, A. Aliaga, F. Águila, J. Troncoso, and J. Pacheco. El impacto de covid19 en el sistema de salud y propuestas para la reactivación. <https://www.colegiomedico.cl/wp-content/uploads/2020/08/reactivacion-sanitaria.pdf>, Agosto, 2020.
- [2] W. Navarrete, D. Miranda, G. Briceño, L. Gramusset, M. Allende, N. Céspedes, N. Pizarro, F. Bravo, and S. González. Características y resultados de pacientes covid-19 tratados con ventilación mecánica invasiva en hospital san juan de dios. marzo-junio 2020. *SciELO*, 149, 2021.
- [3] A. Lindberg. Developing theory through integrating human and machine pattern recognition. *Journal of the Association for Information Systems*, 21(1), 2020.
- [4] T Chen and C Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, 2016.
- [5] W. Zhang, R. Zhang, C. Wu, A. Goh, C. Teck, S Lacasse, Z. Liu, and H. Liu. State-of-the-art review of soft computing applications in underground excavations. *Geoscience Frontiers*, 11(4):1095–1106, 2020.
- [6] F. Junliang, W. Xiukang, W Lifeng, Z. Hanmi, Z. Fucang, Y. Xiang, L. Xianghui, and X. Youzhen. Comparison of support vector machine and extreme gradient boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in china. *Energy Conversion and Management*, 164(15):102–111, 2018.
- [7] R. Sheridan, W. Wang, A. Liaw, J. Ma, and E. Gifford. Extreme gradient boosting as a method for quantitative structure-activity relationships. *Journal of chemical information and modeling*, 56(12):2353–2360, 2016.
- [8] I. Babajide and F. Saeed. Bioactive molecule prediction using extreme gradient boosting. *Molecules*, 21(8):983, 2016.
- [9] M. Hasan, M. Islam, M. Sadeq, W. Fung, and J. Uddin. Review on the evaluation and development of artificial intelligence for covid-19 containment. *Sensors*, 23(1):527, 2023.



- [10] A. Eleni, M. Konstantinos, and N. Konstantina. Artificial intelligence in clinical care amidst covid-19 pandemic: A systematic review. *Computational and Structural Biotechnology Journal*, 19:2833–2850, 2021.
- [11] Y. Xin, H. Li, Q. Zhou, Y. Yang, W. Mu, H. Xiao, Z. Zhuo, H. Liu, H. Wang, X. Qu, C. Wang, H. Liu, and K. Yu. The accuracy of artificial intelligence in predicting covid-19 patient mortality: a systematic review and meta-analysis. *BMC medical informatics and decision making*, 23(1), 155., 23(1):155, 2023.
- [12] L. Yan, H. Zhang, J. Goncalves, Y. Xiao, M. Wang, Y. Guo, C. Sun, X. Tang, L. Jing, M. Zhang, X. Huang, Y. Xiao, H. Cao, Y. Chen, T. Ren, F. Wang, Y. Xiao, S. Huang, X. Tan, N. Huang, B. Jiao, C. Cheng, Y. Zhang, A. Luo, L. Mombaerts, J. Jin, Z. Cao, S. Li, H. Xu, and Y. Yuan. An interpretable mortality prediction model for covid-19 patients. *Nature Machine Intelligence*, 2(5):283–288, 2020.
- [13] W. Galanter, J Rodríguez, K. Chow, S. Harford, K. Kochendorfer, M. Pishgar, J. Theis, J. Zulueta, and H. Darabi. Predicting clinical outcomes among hospitalized covid-19 patients using both local and published models. *BMC Medical Informatics and Decision Making*, 21(1):224, 2021.
- [14] T Kuno, Y. Sahashi, S. Kawahito, M. Takahashi, M. Iwagami, and N. Egorova. Prediction of in-hospital mortality with machine learning for covid-19 patients treated with steroid and remdesivir. *Journal of medical virology*, 94(3):958–964, 2022.
- [15] J. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232, 2001.
- [16] H. Emami, R. Rabiei, S. Sohrabei, and A. Atashi. Predicting the mortality of patients with covid-19: A machine learning approach. *Health science reports.*, 6(4):e1162, 2023.
- [17] H. Estiri, Z. Strasser, J. Klann, P. Naseri, K. Waghlikar, and S. Murphy. Predicting covid-19 mortality with electronic medical records. *npj Digital Medicine.*, 4(1):15, 2021.
- [18] J. Bae, S. Kapse, G. Singh, R. Gattu, S. Ali, N. Shah, C. Marshall, J. Pierce, T. Phatak, A. Gupta, J. Green, N. Madan, and P. Prasanna. Predicting mechanical ventilation and mortality in covid-19 using radiomics and deep learning on chest radiographs: A multi-institutional study. *Diagnostics (Basel)*, 11(10):1812, 2021.
- [19] M. Marcos, M. Belhassen, A. Sánchez, J. Sampedro, R. Azibeiro, P. Dorado, E. Marcano, C. García, M. Moreira, N. Cubino, M. Pérez, B. Rodríguez, D. Encinas, S. Peña, E. Sobejano, S. Inés, C. Carbonell, M. López, F. Andrade, A. López, C. Lorenzo, A. Carpio, D. Polo, M. Sánchez, R. Borrás, V. Sagredo, P. Sanchez, A. Soriano, and J. Martín. Development of a severity of disease score and classification model by machine learning for hospitalized covid-19 patients. *PloS One*, 16(4):e0240200, 2021.
- [20] H. Burdick, C. Lam, S. Mataraso, A. Siefkas, G. Braden, R. Dellinger, A. McCoy, J. Vincent, A. Green, G. Barnes, J. Hoffman, J. Calvert, E. Pellegrini, and R. Das. Prediction of respiratory decompensation in covid-19 patients using machine learning: The ready trial. *Computers in Biology and Medicine*, 124:103949, 2020.

- [21] S. Wollenstein, C. Cassandras, and I. Paschalidis. Personalized predictive models for symptomatic covid-19 patients using basic preconditions: Hospitalizations, mortality, and the need for an icu or ventilator. *International journal of medical informatics*, 142:104258, 2020.
- [22] A. Youssef, S. Kouchaki, F. Shamout, J. Armstrong, R. El-Bouri, T. Taylor, D. Birrenkott, B. Vasey, A. Soltan, T. Zhu, D. Clifton, and D. Eyre. Development and validation of early warning score systems for covid-19 patients. *Healthcare technology letters.*, 8(5):105–117, 2021.
- [23] L. McInnes, J. Healy, and J. Melville. Umap: Uniform manifold approximation and projection for dimension reduction, arxiv. <https://arxiv.org/abs/1802.03426>, 2020.
- [24] D. Sylvestre and J. Hulin. *Book Title: Riemannian Geometry*. Springer Berlin, Heidelberg, 2005.
- [25] How umap works - documentation. [https://umap-learn.readthedocs.io/en/latest/how\\_umap\\_works.html](https://umap-learn.readthedocs.io/en/latest/how_umap_works.html), Octobre, 2023.
- [26] B. Ghojogh, A. Ghodsi, F. Karray, and M. Crowley. Uniform manifold approximation and projection (umap) and its variants: Tutorial and survey, arxiv. <https://arxiv.org/abs/2109.02508>, 2021.
- [27] Y. Izza, A. Ignatiev, and J. Silva. On explaining decision trees, arxiv. <https://arxiv.org/abs/2010.11034>, 2021.
- [28] L. Rokach and O. Maimon. *Book Title: Data Mining and Knowledge Discovery Handbook - Decision Trees*. Springer US, 2005.
- [29] J. Friedman. Stochastic gradient boosting. *Computational Statistics Data Analysis*, 38(4):367–378, 2002.
- [30] A. Natekin and A. Knoll. Gradient boosting machines, a tutorial. *Frontiers in Neuro-robotics*, 7:21, 2013.
- [31] R. Johnson and T. Zhang. Learning nonlinear functions using regularized greedy forest. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(5):942–954, 2014.
- [32] Xgboost documentation - xgboost parameters. <https://xgboost.readthedocs.io/en/stable/parameter.html>, Octobre, 2023.
- [33] Xgboost documentation - python api reference. [https://xgboost.readthedocs.io/en/stable/python/python\\_api.html](https://xgboost.readthedocs.io/en/stable/python/python_api.html), Octobre, 2023.
- [34] S. Lundberg and S. Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Curran Associates Inc, 2017.
- [35] S. Hart. *Book Title: Game Theory - Shapley Value*. Palgrave Macmillan UK, 1989.

- [36] A. Oppenheim and A. Willsky. *Book Title: Signals and Systems*. Prentice-Hall, Inc., 1996.
- [37] T. Hastie, R. Tibshirani, and J. Friedman. *Book Title: The Elements of Statistical Learning: Data Mining, Inference, and Prediction - Relative Importance of Predictor Variables*. Springer New York, 2009.
- [38] P. Ades and P. Savage. The obesity paradox: perception vs knowledge. *Mayo Clinic proceedings.*, 85(2):112–114, 2010.
- [39] H. Lennon, M. Sperrin, E. Badrick, and A. Renehan. The obesity paradox in cancer: a review. *Current Oncology Reports.*, 18(9):56, 2016.
- [40] R. Smith, Vijayaharan T., V. Linehan, Z. Sun, J. Hai Ein Yong, S. Harris, H. Mariathas, and Bhatia. Efficacy of risk prediction models and thresholds to select patients for lung cancer screening. *Canadian Association of Radiologists Journal*, 73(4):672–679, 2022.
- [41] S. Shashikumar, G. Wardi, P. Paul, , Carlile., L. Brenner, K. Hibbert, C. North, S. Mukerji, G. Robbins, Y. Shao, B. Westover, S. Nemati, and A. Malhotra. Development and prospective validation of a deep learning algorithm for predicting need for mechanical ventilation. *CHEST.*, 159(6):2264 – 2273, 2021.
- [42] G. Halasz, M. Sperti, M. Villani, U. Michelucci, P. Agostoni, A. Biagi, L. Rossi, A. Botti, C. Mari, M. Maccarini, F. Pura, L. Roveda, A. Nardecchia, E. Mottola, M. Nolli, E. Salvioni, M. Mapelli, M. A. Deriu, D. Piga, and M. Piepoli. A machine learning approach for mortality prediction in covid-19 pneumonia: Development and evaluation of the piacenza score. *Journal of medical Internet research.*, 23(5):e29058, 2021.
- [43] A. Karthikeyan, A. Garg, P. Vinod, and U. Priyakumar. Machine learning based clinical decision support system for early covid-19 mortality prediction. *Frontiers in public health.*, 9:626697, 2021.

# Anexo A

## Acrónimos

### General

**HCUCH:** Hospital Clínico de la Universidad de Chile

**IA:** Inteligencia Artificial

**ML:** *Machine Learning*

**TI:** Tecnología e Información

**UPC:** Unidad de Pacientes Críticos

**VMI:** Ventilación Mecánica Invasiva

### Variables Clínicas

**CK:** Creatina Quinasa

**CREA:** Creatina Plasmática

**EPOC:** Enfermedad Pulmonar Obstructiva Crónica

**FiO<sub>2</sub>:** Fracción Inspirada de Oxígeno

**GOT:** Aspartato Aminotransferasa

**GPT:** Alanina Aminotransferasa

**HGT:** Hemoglucotest

**INR:** *International Normalized Ratio*

**LDH:** Lactato Deshidrogenasa

**paO<sub>2</sub>:** Presión Arterial de Oxígeno

**PAS:** Presión Arterial Sistólica

**PAD:** Presión Arterial Diastólica

**PAM:** Presión Arterial Media

**PCR:** Proteína C Reactiva

**pCO<sub>2</sub>:** Presión Arterial de Dióxido de Carbono

**PCT:** Procalcitonina

**SatO<sub>2</sub>:** Saturación de Oxígeno

**TTPa:** Tiempo de Tromboplastina Parcial activado

**VHS:** Virus del Herpes Simple

# Anexo B

## Variables Datos Estáticos

Tabla B.1: Variables incluidas en el estudio a pacientes COVID-19 ingresados al servicio de urgencia durante los primeros 100 días, datos estáticos.

Agrupación	Variables
Identificación (3)	↪ Nombre ↪ Run ↪ Fecha de Nacimiento
Demográfica (3)	→ Edad → Género → Ingreso por causa respiratoria
Comorbilidades (23)	↪ Tabaquismo ↪ Hipertensión Arterial (HTA) ↪ Diabetes tipo 2 (Diabetes mellitus) ↪ Diabetes Mellitus con daño de órgano blanco ↪ Asma ↪ Obesidad ↪ Enfermedad Coronaria ↪ Insuficiencia Cardíaca ↪ Enfermedad pulmonar obstructiva crónica (EPOC) ↪ Malignidad ↪ Enfermedad Renal Crónica (ERC) ↪ Accidente Cerebrovascular (ACV) ↪ Hemiplejía ↪ Demencia ↪ Enfermedad Arterial Oclusiva (EAO) ↪ Úlcera Péptica ↪ Daño Hepático Crónico (DHC) ↪ Inmunosupresión Celular (HIV) ↪ Enfermedad del tejido conectivo ↪ Taco (uso anticoagulación / tromboprofilaxis) ↪ Severidad de Daño Hepático Crónico ↪ Barthel

	↪ Número de Comorbilidades
Síntomas (15)	<ul style="list-style-type: none"> <li>→ Fiebre</li> <li>→ Tos</li> <li>→ Anorexia</li> <li>→ Mialgias, Compromiso de Estado General o Fatiga</li> <li>→ Disnea</li> <li>→ Expectoración</li> <li>→ Odinofagia</li> <li>→ Diarrea</li> <li>→ Vómitos</li> <li>→ Rinorrea (congestión nasal)</li> <li>→ Cefalea</li> <li>→ Hemoptisis</li> <li>→ Artralgias</li> <li>→ Hiposmía</li> <li>→ Disgeusia (Hipo)</li> </ul>
Signos Vitales y Examen Físico (9)	<ul style="list-style-type: none"> <li>↪ Saturación de Oxígeno</li> <li>↪ Frecuencia Respiratoria</li> <li>↪ Frecuencia Cardíaca</li> <li>↪ Presión Arterial Sistólica (PAS)</li> <li>↪ Presión Arterial Media (PAM)</li> <li>↪ Temperatura</li> <li>↪ Compromiso de Conciencia</li> <li>↪ GCS (<i>Glasgow Coma Scale</i>)</li> <li>↪ Fracción inspirada de Oxígeno (FiO<sub>2</sub>)</li> </ul>
Laboratorio (24)	<ul style="list-style-type: none"> <li>→ Hemoglobina</li> <li>→ Glóbulos Blancos (linfocitos)</li> <li>→ Neutrófilos</li> <li>→ Plaquetas</li> <li>→ Albúmina</li> <li>→ Aspartato Aminotransferasa (GOT)</li> <li>→ Alanina Aminotransferasa (GPT)</li> <li>→ Creatinina Plasmática</li> <li>→ Nitrógeno Ureico Sanguíneo (BUN)</li> <li>→ Lactato Deshidrogenasa (LDH)</li> <li>→ Creatina Quinasa (CK)</li> <li>→ Lactato</li> <li>→ Troponinas</li> <li>→ INR (<i>International Normalized Ratio</i>)</li> <li>→ Procalcitonina (PCT)</li> <li>→ Proteína C-reactiva (PCR)</li> <li>→ Bilirrubina Total</li> <li>→ Presión Arterial de Oxígeno (PaO<sub>2</sub>)</li> <li>→ Presión Arterial de Dióxido de Carbono (PCO<sub>2</sub>)</li> <li>→ Velocidad de Sedimentación (VHS)</li> <li>→ Dímero D</li> <li>→ Tiempo de Tromboplastina Parcial Activado (TTPa)</li> </ul>

	→ Fibrinógeno
	→ Porcentaje de Linfocitos
Complicaciones (9)	↳ Trombosis Venosa Profunda (TVP) ↳ Tromboembolismo Pulmonar (TEP) ↳ Enfermedad Renal Aguda (AKI) ↳ Injuria Miocárdica Aguda ↳ Shock Séptico ↳ Infección Secundaria Demostrada ↳ Coagulación Intravascular Disemina (CID) ↳ Rabdomiólisis ↳ Arritmia
Evaluación Clínica (12)	→ PaFi (PaO <sub>2</sub> /FiO <sub>2</sub> ) → Puntaje SOFA PaFi → Puntaje SOFA Creatina → Puntaje SOFA Bilirrubina → Puntaje SOFA Sistema Nervioso Central → Puntaje SOFA Plaquetas → Puntaje SOFA PAM → Puntaje SOFA total → Puntaje <i>Quick</i> SOFA → Puntaje CURB65 → Puntaje de severidad según ATS → Escala de severidad de WHO
Terapias Farmacológicas (21)	↳ Indicación Ceftriaxona ↳ Días en Tratamiento Ceftriaxona ↳ Indicación Azitromicina ↳ Días en Tratamiento Azitromicina ↳ Indicación de otros Antibióticos ↳ Indicación de Antivirales ↳ Indicación de Hidroxicloroquina ↳ Indicación de Corticoides ↳ Tipo de Corticoides Indicado ↳ Dosis total Corticoides primer día hospitalización ↳ Fecha inicio tratamiento Corticoides ↳ Fecha fin tratamiento Corticoides ↳ Dosis equivalente de Prednisolona ↳ Indicación de Antifúngicos ↳ Fecha inicio tratamiento Antifúngicos ↳ Fecha fin tratamiento Antifúngicos ↳ Indicación de Inmunoglobulina Intravenosa ↳ Tratamiento con Heparina ↳ Tipo de Heparina ↳ Dosis de Heparina ↳ N° de veces administración Heparina al día
Intervenciones (24)	→ Uso de Cánula Nasal de Alto Flujo (CNAF) → Fecha de inicio CNAF → Fecha de término CNAF

- Uso de Ventilación Mecánica no invasiva (VMNI)
- Fecha de inicio VMNI
- Fecha de término VMNI
- Uso de Prono Vigil
- Fecha de inicio Prono Vigil
- Fecha de término Prono Vigil
- Uso de Ventilación Mecánica Invasiva (VMI)
- Fecha de inicio VMI
- Fecha de término VMI
- Indicación Prono con Intubación Orotraqueal (IOT)
- Fecha de inicio Prono con IOT
- Oxigenación por Membrana Extracorpórea (ECMO)
- Fecha de inicio ECMO
- Fecha de término ECMO
- Indicación de Terapia de Reemplazo Renal (TRR)
- Fecha de inicio TRR
- Fecha de término TRR
- Indicación de Drogas Vasoactivas (DVA)
- Dosis más alta de DVA
- Evaluación por Cuidados Paliativos
- Fecha evaluación por cuidados paliativos

---

Eventos (33)

- ⇒ Muerte
- ⇒ Alta
- ⇒ Lugar de Muerte
- ⇒ Traslado a Unidad de Pacientes Críticos (UPC)
- ⇒ Tiempo entre consulta y traslado a UPC
- ⇒ Fecha de traslado UPC
- ⇒ Unidad a la que Ingresa a Hospitalización
- ⇒ Fecha de primer traslado durante hospitalización
- ⇒ Unidad a la que se realiza primer traslado
- ⇒ Fecha de segundo traslado durante hospitalización
- ⇒ Unidad a la que se realiza segundo traslado
- ⇒ Fecha de tercer traslado durante hospitalización
- ⇒ Unidad a la que se realiza tercer traslado
- ⇒ Fecha de cuarto traslado durante hospitalización
- ⇒ Unidad a la que se realiza cuarto traslado
- ⇒ Fecha de quinto traslado durante hospitalización
- ⇒ Unidad a la que se realiza quinto traslado
- ⇒ Fecha de sexto traslado durante hospitalización
- ⇒ Unidad a la que se realiza sexto traslado
- ⇒ Días de hospitalización
- ⇒ Fecha inicio de síntomas (Primer síntoma manda)
- ⇒ Fecha de muerte, alta o censura
- ⇒ Días que dura seguimiento desde fecha inicio síntomas
- ⇒ Días entre inicio de síntomas e ingreso
- ⇒ Diagnóstico Síndrome de Respiratorio Agudo (SDRA)
- ⇒ Días desde ingreso a diagnóstico de SDRA



- ↪ Fecha de diagnóstico de SDRA
- ↪ Días entre inicio de síntomas e inicio de VMI
- ↪ Fecha de inicio de VMI o Fecha de muerte
- ↪ Días entre inicio de síntomas e ingreso a UCI
- ↪ Días entre inicio de síntomas e inicio de Corticoides
- ↪ Días entre inicio de síntomas e indicación CNAF
- ↪ Días entre inicio de síntomas y diagnóstico de TEP

---

Otras (102)

- Edad Charlson
- Índice de Charlson
- Charlson sin edad
- Enfisema
- Enfisema severo
- Enfermedad Respiratoria Crónica Avanzada
- Charlson covid
- Charlson covid CAT (*Critical Appraisal Topics*)
- Lactato 2
- PaFi 300
- CAT Escenario
- Puntaje Chavez
- Puntaje Metabólico
- Fecha Muerte 2
- Fecha Alta 2
- Fecha Cierre 2
- Fecha Consulta 2
- Fecha inicio sintomas 2
- Uso de Antibióticos (ATB)
- Saturación 95
- Anciano
- Lactato CAT 2
- Lactato CAT miss
- Lactato CAT 3
- Tropo CAT 03
- LDH 250
- Muerte 2
- Días consulta evento 28
- Charlson CAT 3
- Charlson reCAT
- DD2000
- PCR 300
- PCT 6
- Periodo
- paO2 imp
- imp PaFi
- PaFi análisis
- Charlson seba
- Perf alterada
- Fecha 56

- Fecha 28
- Sexo
- Fecha muerte ok
- Fecha alta ok
- Fecha cierre ok
- Fecha muerte okk
- Fecha alta okk
- Fecha cierre okk
- Fecha cierre okk2
- Fecha cierre ok3
- Fecha cierre ok4
- Fecha evento
- Días consulta evento
- Días síntomas consulta
- Linfocitos cont
- Fecha inicio VMI
- Fecha fin VMI
- Fecha inicio VMI ok
- Fecha fin VMI ok
- Tiempo con CNAF
- Días CNAF
- Tiempo con VMI
- Días VMI
- Dimero D 1000
- PCR 10
- PCT 05
- Pafi SOFA análisis
- SOFA noresp
- PaFi Escenarios
- PaFi Escenarios 2
- Escenarios 2
- Escenarios
- Escenarios CATE
- Escenarios CATE 2
- CURB 65 CATE
- TEP TVP
- Muerte Escenario
- ROC Escenarios CATE
- FPR Escenarios CATE
- Fiebre ADM
- SOFA análisis
- SOFA noresp análisis
- qSOFA CAT
- Agudo
- PaFi Análisis 2
- Escenarios 3
- Escenarios otro

- Escenarios otro 2
  - Agudo final
  - Crónico final
  - Escenarios final
  - PaFi análisis 100
  - Barthel CAT
  - Escenarios ag PaFi 2
  - Escenarios final 2
  - Escenarios ag PaFi
  - Uso Heparina
  - \_st
  - \_d
  - \_origin
  - \_t
  - \_t0
-