Ciencia y Educación
(L-ISSN: 2790-8402 E-ISSN: 2707-3378)
Vol. 7 No. 1.1
Edición Especial I 2026

PREDICCIÓN DE RESULTADOS ADVERSOS EN PACIENTES CON COVID-19 MEDIANTE MODELOS DE CLASIFICACIÓN BASADOS EN REGISTROS HOSPITALARIOS

PREDICTION OF ADVERSE OUTCOMES IN COVID-19 PATIENTS USING CLASSIFICATION MODELS BASED ON HOSPITAL RECORDS

Autores: [1]Romel Elian Haro Asipuela y [2]José Renato Cumbal Simba.

[1]ORCID ID: https://orcid.org/0009-0000-1230-5029
[2]ORCID ID: https://orcid.org/0000-0001-8182-5343
[1]E-mail de contacto: rharo1@est.ups.edu.ec
[2]E-mail de contacto: rcumbal@ups.edu.ec
Afiliación: [1*2*]Universidad Politécnica Salesiana (Ecuador).
Artículo recibido: 15 de Enero del 2026
Artículo revisado: 28 de Enero del 2026
Artículo aprobado: 06 de Febrero del 2026
[1]Estudiante de Ingeniería Biomédica en la Universidad Politécnica Salesiana, (Ecuador).
[2]Ingeniero en Electrónica y Telecomunicaciones. Magíster en Gerencia de Sistemas de Información egresado de la Universidad Politécnica Salesiana, (Ecuador). Profesor de la Universidad Politécnica Salesiana, Quito, Ecuador, y miembro del Grupo de Investigación en Telecomunicaciones, (GIETEC). Doctorante en Ingeniería, Universidad Pontificia Bolivariana, (Colombia).

Resumen

La necesidad de herramientas para el diagnóstico temprano de pacientes con alto riesgo de mortalidad se hizo notoria durante la pandemia del COVID-19. Este estudio utilizó una base de datos clínica de acceso abierto y elaboró modelos de clasificación para la predicción de mortalidad hospitalaria. La base de datos se sometió a limpieza, imputación mediante KNN, estandarización y manejo del desbalance de clases mediante ponderación. Se utilizaron Random Forest, SVM, Gradient Boosting y una red neuronal multicapa (MLP), cuyas predicciones se compararon mediante métricas de desempeño y el área bajo la curva ROC. Los modelos basados en ensamble presentaron el mejor desempeño. Random Forest se clasificó como el mejor en discriminación con un AUC de 0.835 y sensibilidad de 0.733, mientras que Gradient Boosting presentó el mejor accuracy con 0.801 y especificidad de 0.866. La presión arterial media, la edad, la saturación de oxígeno, los biomarcadores inflamatorios y renales se establecieron como los predictores más significativos. Estos resultados establecen una propuesta metodológica inicial para validaciones con datos hospitalarios en el contexto ecuatoriano.

Palabras clave: Aprendizaje automático, COVID-19, Biomarcadores clínicos, Modelos de clasificación, Mortalidad hospitalaria, Predicción clínica.

Abstract

The need for tools for early diagnosis of patients at high risk of mortality became evident during the COVID-19 pandemic. This study used an open-access clinical database and developed classification models for predicting hospital mortality. The database underwent cleaning, KNN imputation, standardization, and class imbalance management through weighting. Random Forest, SVM, Gradient Boosting, and a multilayer neural network (MLP) were used, with predictions compared using performance metrics and area under the ROC curve. Ensemble-based models showed the best performance. Random Forest ranked highest in discrimination with an AUC of 0.835 and sensitivity of 0.733, while Gradient Boosting showed the best accuracy with 0.801 and specificity of 0.866. Mean blood pressure, age, oxygen saturation, and inflammatory and renal biomarkers emerged as the most significant predictors. These results establish an initial methodological proposal for validations with hospital data in the Ecuadorian context.

Keywords: Machine learning, COVID-19, Clinical biomarkers, Classification models, Hospital mortality, Clinical prediction.

Sumário

A necessidade de ferramentas para o diagnóstico precoce de pacientes com alto risco de mortalidade tornou-se evidente durante a pandemia de COVID-19. Este estudo utilizou uma base de dados clínica de acesso aberto e desenvolveu modelos de classificação para prever mortalidade hospitalar. A base de dados foi submetida a limpeza, imputação mediante KNN, padronização e tratamento do desequilíbrio de classes mediante ponderação. Foram utilizados Random Forest, SVM, Gradient Boosting e uma rede neural multicamada (MLP), cujas previsões foram comparadas mediante métricas de desempenho e área sob a curva ROC. Os modelos baseados em ensemble apresentaram o melhor desempenho. Random Forest classificou-se como o melhor em discriminação com AUC de 0.835 e sensibilidade de 0.733, enquanto Gradient Boosting apresentou o melhor accuracy com 0.801 e especificidade de 0.866. A pressão arterial média, a idade, a saturação de oxigênio, os biomarcadores inflamatórios e renais estabeleceram-se como os preditores mais significativos. Estes resultados estabelecem uma proposta metodológica inicial para validações com dados hospitalares no contexto equatoriano.

Palavras-chave: Aprendizagem automática, COVID-19, Biomarcadores clínicos, Modelos de classificação, Mortalidade hospitalar, previsão clínica.

Introducción

La extensión de la crisis generada por COVID-19 refleja una de las crisis sistémicas más profundas del siglo XXI. La Organización Mundial de la Salud (OMS) declaró oficialmente el brote como pandemia en marzo de 2020, marcando el inicio de una etapa crítica que exigió respuestas coordinadas por parte de los servicios hospitalarios y la comunidad científica (World Health Organization, 2020). Desde las primeras etapas, numerosos estudios clínicos se enfocaron en caracterizar el curso del COVID-19 e identificar factores asociados a desenlaces adversos. Investigaciones tempranas permitieron describir el perfil clínico de pacientes ingresados y la frecuencia de complicaciones graves. La edad avanzada y comorbilidades preexistentes se asociaban consistentemente con mayor riesgo de muerte (Zhou et al., 2020; Richardson et al., 2020). Complementariamente, estudios identificaron biomarcadores con valor pronóstico: linfopenia, elevación de D-dímero, proteína C-reactiva, ferritina y empeoramiento de la función renal se correlacionaban con mayor probabilidad de ingreso a cuidados intensivos y mortalidad (Yan et al., 2020; Ikram y Pillay, 2022; Ramlall et al., 2020).

A partir de esta evidencia, surgieron intentos por desarrollar herramientas predictivas capaces de estimar el riesgo de desenlaces adversos. Algunos estudios propusieron scores clínicos basados en modelos estadísticos tradicionales, integrando variables demográficas, signos vitales y resultados de laboratorio. Entre estos, destaca el score de severidad propuesto por Altschul et al., que demostró utilidad para estratificación temprana del riesgo (Altschul et al., 2020). Posteriormente, la creciente disponibilidad de grandes bases de datos impulsó la adopción de enfoques basados en aprendizaje automático. Diversos autores comenzaron a aplicar algoritmos de clasificación para modelar relaciones complejas y no lineales entre múltiples variables clínicas. Estudios pioneros demostraron que modelos como Random Forest, Support Vector Machines y Gradient Boosting podían alcanzar desempeños superiores a métodos estadísticos convencionales en la predicción de mortalidad (Yan et al., 2020; Pourhomayoun y Shakibi, 2021; Souza et al., 2021). Investigaciones más avanzadas incorporaron técnicas de selección de características, imputación de datos faltantes

y enfoques interpretables. Casiraghi et al. propusieron un modelo de aprendizaje automático explicable para evaluación temprana del riesgo (Casiraghi et al., 2020), mientras que Hasan et al. compararon múltiples modelos evidenciando que enfoques basados en ensambles ofrecían mejor equilibrio entre desempeño predictivo y generalización (Hasan et al., 2022).

En el último tiempo, la literatura ha sumado esquemas complejos, incluyendo modelos de aprendizaje profundo y técnicas de interpretación como SHAP. Sin embargo, varios análisis críticos advierten que una cantidad importante de modelos sufren de limitaciones metodológicas: sesgos estructurales, muestras no representativas y deficiencias en validación (Wynants et al., 2020). Estas limitaciones se acentúan en contextos con menores niveles de digitalización clínica. En América Latina, la fragmentación de sistemas de salud, la escasa interoperabilidad entre instituciones y la limitada disponibilidad de bases de datos anonimizadas han restringido el desarrollo de modelos predictivos robustos adaptados a la realidad regional (Castaño, 2025). En Ecuador, la evidencia científica sobre predicción de mortalidad por COVID-19 mediante aprendizaje automático es aún escasa. Los estudios disponibles presentan restricciones metodológicas significativas: tamaños muestrales reducidos, selección limitada de variables clínicas y ausencia de procesos de validación cruzada (Carrión y García, 2022). Estas limitaciones se ven agravadas por la insuficiente infraestructura tecnológica hospitalaria y la implementación incipiente de marcos regulatorios en protección de datos personales (Registro Público del Ecuador, 2024).

Bajo este esquema, resulta fundamental desarrollar estudios que evalúen comparativamente múltiples modelos de aprendizaje automático bajo un marco metodológico homogéneo y reproducible, empleando bases de datos clínicas amplias y validadas. El objetivo del presente estudio es construir, comparar y evaluar distintos enfoques de clasificación mediante técnicas de aprendizaje automático para la predicción de mortalidad hospitalaria de pacientes con COVID-19, determinando cuantitativamente cuál de los modelos obtiene el mejor desempeño predictivo. Se evaluaron modelos de Gradient Boosting, redes neuronales, Random Forest y Support Vector Machines, midiendo su desempeño con indicadores estándar: precisión, sensibilidad, especificidad y área bajo la curva ROC. El propósito de estos resultados es ofrecer una base comparativa sólida que sirva como hoja de ruta metodológica para futuros desarrollos contextualizados a la realidad sanitaria del Ecuador.

Materiales y Métodos

El presente estudio se desarrolló bajo un enfoque cuantitativo y aplicado, con diseño no experimental, transversal y retrospectivo. La unidad de análisis correspondió a registros clínicos de pacientes hospitalizados con diagnóstico confirmado de COVID-19 provenientes de una base de datos de acceso abierto y anonimizada. La variable objetivo fue la mortalidad hospitalaria, formulada como problema de clasificación binaria. Las variables independientes incluyeron características demográficas, signos vitales y biomarcadores de laboratorio. La hoja de ruta metodológica consistió en implementación, contraste y evaluación de diversos modelos de clasificación supervisada. Se mantuvo un esquema estandarizado: todos compartieron el mismo set de variables, idéntico preprocesamiento y la

misma partición de datos. Se garantizó la reproducibilidad mediante control estricto de semillas aleatorias en cada fase del entrenamiento.

Se utilizó un conjunto de datos de acceso abierto disponible en Figshare, correspondiente al estudio Mortality incidence, sociodemographic and clinical data in COVID-19 patients (COVID-19 anonymized clinical dataset, 2020). Este conjunto fue recopilado originalmente para analizar incidencia de mortalidad en pacientes hospitalizados e incluye información sociodemográfica, signos vitales, parámetros de laboratorio y desenlaces clínicos. El dataset ha sido empleado como referencia en literatura científica previa de predicción clínica y aprendizaje automático, debido a la disponibilidad de biomarcadores relevantes y su carácter anonimizado, garantizando cumplimiento de principios éticos (Altschul et al., 2020). Se desarrolló un proceso propio de construcción de un conjunto de datos procesado, incluyendo depuración exhaustiva de registros inconsistentes, validación de rangos fisiológicos plausibles y selección de subconjunto de variables clínicamente relevantes.

El conjunto de datos original contiene más de 80 variables clínicas, sociodemográficas y de laboratorio, correspondientes a 4,711 pacientes hospitalizados con diagnóstico confirmado de COVID-19 (COVID-19 anonymized clinical dataset, 2020). Se desarrolló un proceso de depuración y preprocesamiento que permitió construir un dataset procesado compuesto por 32 variables clínicamente relevantes con disponibilidad consistente de datos. Se realizó un proceso adicional de selección de características, guiado por criterios clínicos y metodológicos, para encontrar un subconjunto limitado de variables que posibiliten la predicción de mortalidad hospitalaria. Esta estrategia permitió incrementar la optimización del modelo, eludir el ajuste excesivo y favorecer mejor interpretación de resultados en el ámbito clínico.

La selección se basó en tres criterios principales: (i) evidencia clínica reportada como factores asociados a mortalidad por COVID-19, (ii) disponibilidad consistente dentro de la base procesada y (iii) aplicabilidad clínica en escenarios hospitalarios reales. Se priorizaron variables con respaldo fisiopatológico y uso clínico frecuente, descartando aquellas con alta proporción de datos faltantes, baja relevancia clínica o redundancia informativa. La variable dependiente del estudio es mortalidad hospitalaria (Death), codificada binariamente (0: superviviente; 1: fallecido), permitiendo tratar el problema como tarea de clasificación binaria (Yan et al., 2020; Wynants et al., 2020). Las variables predictoras seleccionadas incluyen parámetros demográficos, signos vitales y biomarcadores de laboratorio ampliamente reconocidos como predictores de severidad y mortalidad en COVID-19. La edad (Age) ha sido consistentemente identificada como uno de los factores de riesgo más relevantes, asociándose directamente con aumento en mortalidad hospitalaria (Zhou et al., 2020; Richardson et al., 2020). La saturación de oxígeno (OsSats) y presión arterial media (MAP) reflejan estado respiratorio y hemodinámico, siendo indicadores clave de deterioro clínico temprano y falla orgánica (Yan et al., 2020; Ikram y Pillay, 2022). La temperatura corporal (Temp) se considera marcador de respuesta inflamatoria y gravedad de infección.

En cuanto a biomarcadores de laboratorio, el dímero-D (Ddimer) fue seleccionado por su

asociación con estados de hipercoagulabilidad y mayor riesgo de eventos trombóticos, ampliamente reportados en pacientes con desenlaces adversos (Zhou et al., 2020; Ramlall et al., 2020). La ferritina (Ferritin) y el nitrógeno ureico en sangre (BUN) se incorporaron como indicadores de respuesta inflamatoria sistémica y compromiso metabólico-renal, mecanismos fisiopatológicos vinculados a progresión hacia enfermedad grave (Yan et al., 2020; Wynants et al., 2020). La proteína C reactiva (CrctProtein) se incluyó como marcador inflamatorio de uso clínico habitual, asociado a severidad y mortalidad hospitalaria (Richardson et al., 2020). Finalmente, la troponina (Troponin) y la creatinina (Creatinine) fueron seleccionadas por su relación con lesión miocárdica y deterioro de función renal, respectivamente, ambos reportados como factores pronósticos relevantes en pacientes hospitalizados con COVID-19 (Zhou et al., 2020; Ikram y Pillay, 2022). En conjunto, este subconjunto de 10 variables capturó dimensiones clínicas clave del estado del paciente (respiratoria, inflamatoria, cardiovascular y renal), proporcionando base sólida y clínicamente interpretable para entrenamiento y comparación de modelos.

Antes de entrenar los modelos, el conjunto de datos se sometió a preprocesamiento exhaustivo destinado a garantizar calidad, consistencia y plausibilidad clínica de la información. Esta etapa es especialmente relevante en análisis de datos clínicos reales, donde es frecuente la presencia de errores de registro, valores atípicos, inconsistencias en formatos y datos faltantes que pueden afectar significativamente desempeño, estabilidad y capacidad de generalización de modelos predictivos. El proceso inició con selección de variables clínicas de interés e identificación de valores no fisiológicos. En aquellas variables donde el valor cero no es clínicamente válido (signos vitales y biomarcadores de laboratorio), dichos valores fueron recodificados como datos faltantes. Valores no realistas de edad fueron considerados inválidos. Posteriormente, se realizó conversión forzada de todas las variables predictoras a formato numérico, corrigiendo inconsistencias derivadas del uso de separadores decimales y eliminando caracteres no válidos. Se aplicó filtrado basado en rangos clínicos plausibles definidos para cada variable, según referencias ampliamente aceptadas en literatura médica. Los valores fuera de estos rangos fueron tratados como datos faltantes, preservando coherencia fisiológica sin descartar registros completos de pacientes. Previo a imputación, se cuantificó la proporción de valores faltantes por variable, observándose mayor presencia en OsSats (187 registros), MAP (306), Temp (180) y BUN (13). La gestión de valores faltantes se resolvió mediante Imputación por Vecinos más Cercanos (KNN), configurada con cinco vecinos y ponderación por distancia:

$$\hat{x}_{ij} = \left(\sum_{l \in N_5(i)} 1/(d(i,l))\, x_{lj}\right) / \left(\sum_{l \in N_5(i)} 1/(d(i,l))\right)$$

La imputación mediante KNN se aplicó sobre el conjunto completo antes de partición en entrenamiento y validación, preservando coherencia clínica global y evitando exclusión de registros incompletos. Este método estima omisiones basándose en proximidad clínica entre pacientes, preservando estructura original de información y reduciendo sesgo asociado a pérdida de datos críticos. Posteriormente, la base se segmentó en subconjuntos de entrenamiento y validación mediante partición estratificada (70/30), conservando proporción de variable objetivo Death. Se optó por única partición con fin de mantener evaluación homogénea y facilitar comparación directa

entre modelos. Finalmente, las variables predictoras fueron normalizadas mediante escalamiento tipo z-score, calculando parámetros exclusivamente sobre conjunto de entrenamiento para evitar fuga de información:

$$x_{ij}^* = (x_{ij} - \mu_j)/\sigma_j$$

Se evaluaron cuatro modelos de clasificación que contrastan enfoques lineales y no lineales: Regresión Logística, Random Forest, SVM y Red Neuronal Multicapa (MLP). Random Forest: Modelo ensamble basado en 300 árboles de decisión. Cada árbol produce predicción y la salida se determina por votación mayoritaria. La probabilidad de mortalidad hospitalaria se obtiene como promedio de probabilidades estimadas por los árboles:

$$P(y_i=1|x_i^*) = 1/T \sum_{t=1}^{T} h_t(x_i^*)$$

Support Vector Machines (SVM): Se implementó kernel radial de base gaussiana (RBF) con C = 1.0 y ponderación balanceada. El modelo identifica relaciones no lineales mediante:

$$K(x_i^*, x_j^*) = \exp(-\gamma \|x_i^* - x_j^*\|^2)$$

Las salidas se calibraron mediante Platt scaling para obtener probabilidades interpretables. Gradient Boosting: Modelo secuencial de 300 árboles con tasa de aprendizaje 0.05 y profundidad máxima 4. Cada árbol se ajusta para corregir errores del modelo previo mediante gradiente de función de pérdida:

$$F_t(x) = F_{t-1}(x) + \eta h_t(x)$$

Red Neuronal Multicapa (MLP): Arquitectura con dos capas ocultas (64 y 32 neuronas) con activación ReLU. La propagación hacia adelante calcula representación no lineal:

$$h^{(l)} = \phi(W^{(l)} h^{(l-1)} + b^{(l)})$$

El entrenamiento se realizó con optimizador Adam, regularización L2 y parada temprana. Para transformar probabilidades en predicciones binarias, se estableció umbral de decisión $\tau = 0.3$ para todos los clasificadores, priorizando sensibilidad del modelo frente a especificidad. Este criterio clínico busca incrementar detección de clase positiva a costa de aumento controlado en tasa de falsos positivos.

El desbalance de clases se abordó exclusivamente sobre conjunto de entrenamiento mediante ponderación balanceada de clases (class_weight), evitando contaminación del conjunto de prueba. El rendimiento se evaluó mediante métricas complementarias: matriz de confusión, exactitud (Accuracy), sensibilidad (Recall), especificidad, precisión, F1-score y área bajo la curva ROC (AUC). Todas las métricas se calcularon sobre conjunto de validación independiente bajo umbral de decisión unificado, garantizando comparativa objetiva y libre de sesgos externos. Se evaluó también la importancia de variables en modelos basados en árboles (Random Forest y Gradient Boosting), estimando contribución relativa de cada característica a partir de reducción de impureza durante entrenamiento. Este análisis permitió identificar biomarcadores con mayor peso predictivo en estimación del riesgo de mortalidad hospitalaria.

Resultados y Discusión

Análisis de correlación de variables clínicas

La matriz de correlación de Pearson mostró correlaciones bajas a moderadas entre la mayoría de los predictores, sugiriendo baja multicolinealidad y adecuada diversidad de información clínica para entrenamiento de modelos supervisados. No se identificaron correlaciones extremadamente altas ($r > 0.8$),

por lo que no fue necesario excluir variables por redundancia. Se destacó correlación positiva entre marcadores de función renal, con asociación moderada entre BUN y creatinina (r = 0.59), consistente con compromiso renal en pacientes con cuadros graves. Asimismo, se observaron correlaciones entre biomarcadores inflamatorios y hemostáticos: proteína C reactiva y dímero-D (r = 0.33), ferritina y proteína C reactiva (r = 0.21), reflejando respuesta inflamatoria sistémica asociada a progresión clínica. La saturación de oxígeno (OsSats) presentó correlación negativa con proteína C reactiva (r = -0.33), sugiriendo que mayores niveles de inflamación se asocian con peor oxigenación.

Desempeño comparativo del modelo

Tabla 1. *Desempeño comparativo de los modelos de clasificación*

Modelo	Acc.	Prec.	Sens.	Esp.	F1	AUC
Random Forest	0.746	0.487	0.739	0.748	0.587	0.835
SVM	0.790	0.559	0.661	0.832	0.606	0.819
Gradient Boosting	0.801	0.591	0.600	0.866	0.596	0.828
MLP	0.777	0.536	0.626	0.825	0.578	0.818

Fuente: Elaboración propia

Los modelos superaron el rendimiento esperado de un clasificador aleatorio, evidenciando capacidad para identificar patrones clínicos asociados a mortalidad hospitalaria por COVID-19. Gradient Boosting alcanzó el mayor valor de accuracy (0.801) y especificidad (0.866), indicando mayor capacidad para clasificar correctamente pacientes sobrevivientes. Random Forest obtuvo la mayor sensibilidad (0.739), el valor más alto de AUC (0.835) y F1-score competitivo, reflejando mayor capacidad discriminativa global y detección más efectiva de pacientes fallecidos.

Los modelos SVM y MLP presentaron desempeño intermedio, con valores de accuracy, AUC y F1-score comparables entre sí, aunque inferiores a modelos basados en ensambles. Los resultados evidencian diferencias claras en comportamiento según métrica considerada.

Curvas ROC y matrices de confusión

Las curvas ROC mostraron que todos los modelos se sitúan por encima de la diagonal correspondiente a clasificador aleatorio, evidenciando capacidad discriminativa superior al azar. Random Forest presentó la curva más cercana a la esquina superior izquierda, alcanzando el mayor valor de AUC (0.835). Gradient Boosting mostró desempeño competitivo, aunque con curva ligeramente más cercana a la diagonal. Las matrices de confusión, utilizando umbral $\tau = 0.3$, permitieron examinar distribución de verdaderos positivos (TP), falsos negativos (FN), verdaderos negativos (TN) y falsos positivos (FP). Random Forest presentó la mayor cantidad de verdaderos positivos y menor cantidad de falsos negativos, reafirmando su sensibilidad y capacidad de clasificar correctamente pacientes de alto riesgo de morir. Este atributo es muy importante en práctica clínica, ya que falsos negativos pueden poner a pacientes en situaciones de riesgo al retrasar intervenciones necesarias.

Los modelos SVM y Gradient Boosting presentaron mayor número de verdaderos negativos, traduciéndose en mayor especificidad y mejor tasa de detección de sobrevivientes. Sin embargo, esto vino acompañado de mayor número de falsos negativos. El desempeño de MLP se situó en punto medio, logrando equilibrio aceptable entre sensibilidad y especificidad. La selección de umbral reducido ($\tau = 0.3$) favoreció la

detección de clase positiva (fallecimiento), incrementando sensibilidad a expensas de proporción más elevada de falsos positivos. Este enfoque es coherente con objetivo clínico del estudio, enfocado en privilegiar detección temprana de pacientes con elevado riesgo de mortalidad hospitalaria.

Variables clínicas determinantes

El análisis de importancia de variables mostró que la presión arterial media (MAP) presenta la mayor importancia relativa en predicción de mortalidad hospitalaria en mayoría de modelos evaluados. La edad se mantiene como una de las variables con mayor contribución en todos los clasificadores. Biomarcadores asociados a inflamación sistémica y función renal (proteína C reactiva, creatinina, BUN y dímero-D) muestran contribución relevante, especialmente en modelos de ensamble. En contraste, variables como temperatura corporal y troponina exhiben menor importancia relativa cuando se consideran conjuntamente con resto de predictores clínicos. A pesar de diferencias en valores absolutos de importancia entre modelos y métodos de estimación, la jerarquía de variables más influyentes se mantiene consistente.

Los resultados demuestran que los modelos de aprendizaje automático son herramientas eficaces para predicción de mortalidad hospitalaria en pacientes con COVID-19, siendo los métodos de ensamble los que alcanzaron mejor desempeño. Random Forest y Gradient Boosting mostraron mayor capacidad predictiva en comparación con SVM y MLP, tanto en discriminación global como en equilibrio entre sensibilidad y especificidad. Al contrastar estos hallazgos con literatura existente, se observa que el mejor desempeño de métodos de ensamble coincide con lo reportado en estudios previos, donde se destaca su capacidad para modelar relaciones no lineales, manejar interacciones complejas entre variables clínicas y mantener estabilidad frente al ruido presente en datos hospitalarios reales (Altschul et al., 2020; Pourhomayoun y Shakibi, 2021; Hasan et al., 2022).

Los valores de AUC alcanzados en este estudio (0.835 para Random Forest y 0.828 para Gradient Boosting) se sitúan dentro del rango superior reportado en literatura, donde se describen valores entre 0.78 y 0.85 para modelos entrenados con variables clínicas similares (Yan et al., 2020; Casiraghi et al., 2020; Barough et al., 2023). Estos resultados indican que el enfoque metodológico aplicado (depuración clínica rigurosa, imputación controlada de valores faltantes y normalización adecuada) permitió obtener desempeño competitivo y comparable con estudios multicéntricos de mayor escala. Un aspecto relevante identificado es la priorización de sensibilidad clínica mediante uso de umbral de decisión reducido. Esta estrategia permitió mejorar detección de pacientes con alto riesgo de fallecimiento, aun a costa de incrementar falsos positivos. Desde punto de vista clínico, este enfoque es coherente con literatura, que enfatiza importancia de minimizar falsos negativos en escenarios críticos como pandemia por COVID-19 (Wynants et al., 2020).

El análisis de importancia de variables identificó consistentemente edad, presión arterial media, saturación de oxígeno y biomarcadores inflamatorios y renales como predictores más relevantes de mortalidad hospitalaria. Estos hallazgos concuerdan con fisiopatología descrita del COVID-19 severo, donde disfunción hemodinámica, compromiso renal y respuesta inflamatoria sistémica exacerbada han sido ampliamente asociados con desenlaces adversos (Zhou et al., 2020;

Richardson et al., 2020; Yan et al., 2020). Si bien el uso de repositorio de datos internacional constituye limitación en términos de representatividad local, los resultados obtenidos permiten evaluar comportamiento de modelos en contexto clínico heterogéneo y comparable con investigaciones previas. Además, la metodología propuesta es reproducible y escalable, facilitando su futura validación con datos provenientes de hospitales ecuatorianos o latinoamericanos.

Conclusiones

Se desarrolló y evaluó un conjunto de modelos de clasificación supervisada orientados a predicción de mortalidad intrahospitalaria en pacientes con COVID-19, empleando variables clínicas pre procesadas bajo protocolo metodológico riguroso. Los resultados demuestran que aprendizaje automático permite identificar patrones clínicos asociados a desenlaces desfavorables, superando ampliamente desempeño esperado de clasificadores aleatorios. La comparación entre Random Forest, Support Vector Machines, Gradient Boosting y Perceptrón Multicapa evidenció que métodos de ensamble presentan mejor desempeño global en términos de capacidad discriminativa y estabilidad. Random Forest destacó por alcanzar mayor sensibilidad y mayor valor de AUC, posicionándose como enfoque más adecuado cuando se prioriza detección temprana de pacientes con alto riesgo de fallecimiento, aspecto crítico en práctica clínica.

El análisis de correlación y de importancia de variables permitió identificar predictores clínicamente coherentes con fisiopatología del COVID-19 severo, entre los que se destacan edad, presión arterial media, saturación de oxígeno y biomarcadores inflamatorios y renales. La consistencia de estos hallazgos entre distintos modelos refuerza interpretabilidad y validez clínica de resultados obtenidos. La adopción de umbral de decisión desplazado permitió privilegiar sensibilidad de modelos, optimizando detección de pacientes en riesgo a costa de incremento controlado de falsos positivos. Esta estrategia se alinea con necesidades reales del entorno hospitalario, donde identificación temprana de casos críticos resulta determinante para toma de decisiones terapéuticas y gestión eficiente de recursos. Si bien el estudio se basó en base de datos internacional, lo que constituye limitación en términos de representatividad local, el enfoque metodológico propuesto es fácilmente adaptable y escalable. Esto abre posibilidad de su validación futura con datos de hospitales ecuatorianos o latinoamericanos, así como su integración en sistemas de apoyo a decisión clínica.

Agradecimientos

Se agradece a la Universidad Politécnica Salesiana y al Grupo de Investigación en Telecomunicaciones (GIETEC) por el apoyo brindado en el desarrollo de esta investigación

Referencias Bibliográficas

Altschul, D., Unda, S., Benton, J., Garza, R., Cezayirli, P., Mehler, M., y Eskandar, E. (2020). A novel severity score to predict inpatient mortality in COVID-19 patients. *Scientific Reports*, 10.

Barough, S., Safavi, S., Siavoshi, F., Tamimi, A., Ilkhani, S., Akbari, S., Ezzati, S., Hatamabadi, H., y Pourhoseingholi, M. (2023). Generalizable machine learning approach for COVID-19 mortality risk prediction using on-admission clinical and laboratory features. *Scientific Reports*, 13.

Carrión, J., y García, M. (2022). Aplicación de modelos predictivos para COVID-19 en Ecuador. Repositorio Institucional ESPOL.

Casiraghi, E., Malchiodi, D., Trucco, G., Frasca, M., Cappelletti, L., Fontana, T.,

Esposito, A., Avola, E., Jachetti, A., Reese, J., Rizzi, A., Robinson, P., y Valentini, G. (2020). Explainable machine learning for early assessment of COVID-19 risk prediction in emergency departments. *IEEE Access*, 8, 196299–196325.

Castaño, S. (2025). Artificial intelligence in public health: Opportunities, ethical challenges and future perspectives.

COVID-19 anonymized clinical dataset. (2020). Figshare.

Hasan, M., Bath, P., Marincowitz, C., Sutton, L., Pilbery, R., Hopfgartner, F., Mazumdar, S., Campbell, R., Stone, T., Thomas, B., Bell, F., Turner, J., Biggs, K., Petrie, J., y Goodacre, S. (2022). Pre-hospital prediction of adverse outcomes in patients with suspected COVID-19: Development, application and comparison of machine learning and deep learning methods. *Computers in Biology and Medicine*, 151.

Ikram, A., y Pillay, S. (2022). Admission vital signs as predictors of COVID-19 mortality: A retrospective cross-sectional study. *BMC Emergency Medicine*, 22.

Pourhomayoun, M., y Shakibi, M. (2021). Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. *Smart Health*, 20.

Ramlall, V., Thangaraj, P., Meydan, C., Foox, J., Butler, D., Kim, J., May, B., Freitas, J., Glicksberg, B., Mason, C., Tatonetti, N., y Shapira, S. (2020). Immune complement and coagulation dysfunction in adverse outcomes of SARS-CoV-2 infection. *Nature Medicine*, 26, 1609–1615.

Registro Público del Ecuador. (2024). Proyecto de ley de protección de datos personales.

Richardson, S., Hirsch, J., Narasimhan, M., Crawford, J., McGinn, T., y Davidson, K. (2020). Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. *JAMA*, 323(20), 2052–2059.

Souza, F., Hojo-Souza, N., Santos, E., Silva, C., y Guidoni, D. (2021). Predicting the disease outcome in COVID-19 positive patients through machine learning: A retrospective cohort study with Brazilian data. *Frontiers in Artificial Intelligence*, 4.

World Health Organization. (2020). WHO characterizes COVID-19 as a pandemic.

Wynants, L., Van Calster, B., Collins, G., Riley, R., Heinze, G., Schuit, E., Albu, E., Arshi, B., Bellou, V., Bonten, M., Dahly, D., Damen, J., Debray, T., Jong, V., Vos, M., Dhiman, P., Ensor, J., Gao, S., Haller, M., y Smeden, M. (2020). Prediction models for diagnosis and prognosis of COVID-19: Systematic review and critical appraisal. *BMJ*, 369.

Yan, L., Zhang, H., Goncalves, J., Xiao, Y., Wang, M., Guo, Y., Sun, C., Tang, X., Jing, L., Zhang, M., Huang, X., Cao, H., Chen, Y., Ren, T., Wang, F., Tan, X., y Yuan, Y. (2020). An interpretable mortality prediction model for COVID-19 patients. *Nature Machine Intelligence*, 2, 283–288.

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., Guan, L., Wei, Y., Li, H., Wu, X., Xu, J., Tu, S., Zhang, Y., Chen, H., y Cao, B. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. *The Lancet*, 395, 1054–1062.