

Predicción de readmisión hospitalaria a 30 días mediante arquitectura CNN-LSTM sobre datos clínicos tabulares y series temporales de signos vitales

Prediction of 30-day hospital readmission using a cnn-lstm architecture on tabular clinical data and vital sign time series

Treisy Alexandra Castillo Camacho¹, María Ines Cevallos Mazon², Flor Gabriela Soliz Poveda³, Alysson Valeria Urrea Camacho⁴, Andrea Coraima Espin Lasso⁵, Amanda Cristina Carlosama Calderon⁶

Cita:

Castillo Camacho, T. A., Cevallos Mazon, M. I., Soliz Poveda, F. G., Urrea Camacho, A. V., Espin Lasso, A. C., & Carlosama Calderon, A. C. (2026). Predicción de readmisión hospitalaria a 30 días mediante arquitectura CNN-LSTM sobre datos clínicos tabulares y series temporales de signos vitales. *Salud ConCiencia*, 5(2), e105.

Recibido: 29/06/2026

Revisado: 08/07/2026 al 05/10/2026

Corregido: 12/10/2026

Aceptado: 16/10/2026

Publicado: 23/10/2026

Licencia:

Los contenidos de este artículo están bajo una licencia de Creative Commons Attribution 4.0 International (CC BY 4.0). Los autores conservan los derechos morales y patrimoniales de sus obras.

The contents of this article are under a Creative Commons Attribution 4.0 International (CC BY 4.0) license. The authors retain the moral and patrimonial rights of their works.

1 Pontificia Universidad Católica del Ecuador Sede Ambato, Ecuador, 2-6 Investigador Externo

¹tacastillo@pucesa.edu.ec, ²mariai.cevallosm@iess.gob.ec, ³florbioquim@hotmail.com, ⁴alysson.urrea@iess.gob.ec, ⁵acoraimaespinl@gmail.com, ⁶accarlosamac29@gmail.com

¹0009-0008-2445-7001, ²0009-0008-0457-1632, ³0009-0009-1578-5748, ⁴0009-0001-6629-6444, ⁵0009-0003-8248-4925, ⁶0009-0007-8252-4287

Resumen: La readmisión hospitalaria no planificada en un periodo de 30 días constituye un indicador crítico de calidad asistencial y un desafío económico de gran magnitud. Mientras Medicare estima pérdidas anuales de 17 000 millones de dólares por ingresos evitables, el costo total para el sistema sanitario estadounidense supera los 52 000 millones de dólares. Los modelos tradicionales, como el índice LACE, el HOSPITAL score o la regresión logística, alcanzan áreas bajo la curva (AUC) modestas de entre 0,60 y 0,72, niveles que resultan insuficientes para la toma de decisiones clínicas efectivas. Este trabajo propone y evalúa una arquitectura híbrida CNN-LSTM que integra datos clínicos estáticos (demografía, comorbilidades de Charlson y diagnósticos ICD) con series temporales de signos vitales de las últimas 48 horas previas al alta, extraídas de MIMIC-IV. Se emplean convoluciones 1D para capturar patrones locales de deterioro y capas LSTM bidireccionales para modelar dependencias temporales largas, mediante una fusión intermedia de embeddings tabulares. Utilizando focal loss para gestionar el desbalance de datos (15% de clase positiva) y validación estratificada, el modelo alcanza un AUC-ROC de 0,812 y un AUC-PR de 0,451, superando a XGBoost (0,742) y a una LSTM simple (0,776). Finalmente, se discute la interpretabilidad mediante SHAP, las implicaciones éticas y la aplicabilidad de esta tecnología al contexto latinoamericano.

Palabras clave: readmisión hospitalaria; deep learning; CNN-LSTM; series temporales clínicas; MIMIC-IV; signos vitales; predicción clínica.

Abstract: Unplanned hospital readmission within a 30-day period is a critical indicator of health-care quality and a significant economic challenge. While Medicare estimates annual losses of 17 billion in avoidable admissions, the total cost to the US health care system exceeds 52 billion. Traditional models, such as the LACE index, the Hospital Score, and logistic regression, achieve modest areas under the curve (AUC) of between 0.60 and 0.72, levels that are insufficient for effective clinical decision-making. This work proposes and evaluates a hybrid CNN-LSTM architecture that integrates static clinical data (demographics, Charlson comorbidities, and ICD diagnoses) with time series of vital signs from the 48 hours prior to discharge, extracted from MIMIC-IV. One-dimensional convolutions are used to capture local patterns of deterioration, and bidirectional LSTM layers are used to model long-term dependencies through intermediate fusion of tabular embeddings. Using focal loss to manage data imbalance (15 % positive class) and stratified validation, the model achieves an AUC-ROC of 0.812 and an AUC-PR of 0.451, outperforming XGBoost (0.742) and a simple LSTM (0.776). Finally, the interpretability using SHAP, the ethical implications, and the applicability of this technology to the Latin American context are discussed.

Keywords: hospital readmission; deep learning; CNN-LSTM; clinical time series; MIMIC-IV; vital signs; clinical prediction.

INTRODUCCIÓN

Las readmisiones hospitalarias no planificadas dentro de los 30 días posteriores al alta se han consolidado como uno de los indicadores más reveladores de la calidad asistencial. Este parámetro no solo refleja la atención recibida durante la estancia hospitalaria, sino que también mide la efectividad en la gestión de la transición del paciente hacia su hogar. A nivel internacional, este indicador es una referencia fundamental para evaluar tanto la seguridad del paciente como la eficiencia de los sistemas de salud (Jencks et al., 2009; Kansagara et al., 2011).

En Estados Unidos, la magnitud del problema es evidente: aproximadamente uno de cada cinco beneficiarios de Medicare reingresa al hospital en este periodo. Con una tasa global de readmisión cercana al 14,7% y un costo promedio de 15 200 dólares por caso, el impacto financiero es masivo. Esto representa un gasto anual evitable de unos 17 000 millones de dólares, dentro de un costo total que supera los 52 000 millones

(Jencks et al., 2009; Wadhera et al., 2018). Más allá de las estadísticas, estas cifras representan recaídas de pacientes, preocupación para sus familias y un sistema sanitario con un amplio margen de mejora.

Para abordar esta situación, en 2012 se implementó el Hospital Readmissions Reduction Program (HRRP) como parte de la Affordable Care Act. Este programa establece sanciones económicas de hasta el 3 % de los pagos de Medicare para aquellos hospitales con tasas elevadas de readmisión en patologías críticas como el infarto agudo de miocardio, insuficiencia cardíaca, neumonía, EPOC, cirugías de bypass coronario y reemplazos articulares. Tan solo en el año fiscal 2023, más de 2 200 hospitales recibieron penalizaciones que sumaron 320 millones de dólares (Khera et al., 2020; Wadhera et al., 2018).

El análisis por diagnóstico refuerza la gravedad del escenario: la insuficiencia cardíaca presenta una tasa de readmisión del 13,2 % a los 30 días, cifra que escala al 35,7 % al año. Por su parte, la EPOC se sitúa cerca del 20 %, la neumonía oscila entre el 18 % y el 20 %, y el infarto de miocardio ronda el 14,6 % (Khan et al., 2021; Myers et al., 2020). En Latinoamérica, aunque la evidencia es menos extensa, los hallazgos son consistentes. En Colombia, se ha reportado una tasa del 10,1 % a 30 días con costos significativos en cohortes amplias, mientras que estudios en Chile y Brasil muestran tendencias similares, especialmente en el área cardiovascular (Cardona-Morrell et al., 2016).

Frente a este panorama, la comunidad científica ha desarrollado diversas herramientas predictivas. Los modelos tradicionales, como el índice LACE y el HOSPITAL score, suelen alcanzar niveles de discriminación cercanos a 0,70. Un meta-análisis reciente confirma que estos enfoques convencionales, incluyendo el sistema británico PARR-30, mantienen un rendimiento moderado con un AUC de aproximadamente 0,71 (Zhou et al., 2025). Incluso métodos de aprendizaje automático como Random Forest o XGBoost presentan limitaciones, ya que suelen asumir relaciones estáticas entre variables y no logran capitalizar la información temporal. La evolución de los signos vitales antes del alta, por ejemplo, representa una señal crítica que estos modelos omiten.

En este contexto, el aprendizaje profundo (deep learning) surge como una alternativa transformadora. A diferencia de los métodos clásicos, estas técnicas identifican patrones complejos directamente de los datos sin requerir una ingeniería manual exhaustiva. Modelos como Doctor AI, RETAIN o Deep Patient han demostrado la capacidad de las redes recurrentes para captar dinámicas profundas en los historiales clínicos a gran escala. Las investigaciones de vanguardia exploran hoy arquitecturas híbridas, como las redes CNN-LSTM, que combinan capas convolucionales para detectar patrones locales con unidades LSTM para modelar dependencias temporales prolongadas. Este enfoque ha mostrado resultados prometedores en la predicción de mortalidad y sepsis, superando a los modelos tradicionales.

No obstante, en el ámbito iberoamericano, esta línea de investigación se encuentra aún en una etapa inicial. Estudios realizados en Chile advierten que los modelos desarrollados en otros contextos no siempre se adaptan a la realidad local debido a desafíos estructurales, como la heterogeneidad en la calidad de los datos y la limitada interoperabilidad de los sistemas de información (Fierro Mella, 2019).

Bajo estas premisas, la presente investigación propone el diseño e implementación de un modelo basado en una arquitectura CNN-LSTM que integre datos clínicos estructurados con series temporales de signos vitales. El objetivo principal es fortalecer la capacidad predictiva de las readmisiones a 30 días y superar las restricciones de las herramientas actuales. La hipótesis de trabajo sostiene que esta integración permitirá alcanzar un desempeño superior, con un AUC-ROC igual o mayor a 0,80, preservando siempre su utilidad clínica.

La relevancia de este estudio es multidimensional. Desde la perspectiva clínica, facilita la identificación temprana de pacientes en riesgo para permitir intervenciones oportunas. En el plano económico, contribuye a la reducción de costos y a la prevención de sanciones financieras. Finalmente, en el ámbito académico, aporta evidencia sólida en un campo donde el contexto iberoamericano requiere mayor desarrollo y documentación.

DESARROLLO

Materiales y fuentes de datos

Para el desarrollo de este estudio se utilizó MIMIC-IV v2.2 (Johnson et al., 2023), una base de datos longitudinal del Beth Israel Deaconess Medical Center que abarca el periodo 2008–2019. Este recurso incluye registros de aproximadamente 299 712 pacientes y 431 231 admisiones hospitalarias, distribuidos en módulos que integran datos de laboratorio, medicación, diagnósticos ICD-9/10 y eventos en urgencias y cuidados intensivos. El acceso se gestionó a través de PhysioNet bajo la modalidad de credentialed access, cumpliendo con la certificación CITI correspondiente.

Con el fin de evaluar la capacidad de generalización del modelo, se empleó eICU-CRD (Pollard et al., 2018) como conjunto de validación externa. Esta base cuenta con más de 200 000 admisiones de 208 hospitales estadounidenses y ofrece una resolución temporal de cinco minutos en signos vitales. Ambas fuentes representan el estándar de referencia en la literatura actual sobre aprendizaje profundo aplicado a la clínica (Harutyunyan et al., 2019; Purushotham et al., 2018).

Extracción de la cohorte y definición del evento

La selección de la muestra se rigió por las recomendaciones del CMS Hospital-Wide Readmission Measure y criterios validados en la literatura (Mahmoudi et al., 2020; van Walraven et al., 2010). Se incluyeron pacientes adultos (≥ 18 años) que permanecieron hospitalizados al menos 24 horas y fueron dados de alta con vida. Por el contrario, se excluyeron transferencias entre hospitales, altas voluntarias contra consejo médico, pacientes obstétricos, casos derivados a cuidados paliativos y fallecimientos durante la estancia índice.

La variable objetivo se definió como la readmisión hospitalaria no planificada por cualquier causa dentro de los 30 días posteriores al alta. Tras el proceso de depuración, la cohorte final constó de 184 526 admisiones, con una prevalencia de readmisión del 15,3 %, cifra alineada con los reportes nacionales en Estados Unidos (Jencks et al., 2009).

Variables predictoras y preprocesamiento

El modelo integra dos bloques de información complementarios. El bloque tabular consta de 33 variables, incluyendo datos demográficos, el índice de Charlson, la duración de la estancia, el tipo de admisión y componentes clave del HOSPITAL score, como los niveles de hemoglobina y sodio al alta. Para las variables categóricas de alta cardinalidad, se emplearon entity embeddings de dimensión $d = \min(50, \text{card}/2)$, siguiendo las prácticas recomendadas para registros electrónicos de salud (Choi et al., 2016).

El bloque temporal analiza cinco signos vitales (frecuencia cardíaca, presión arterial, saturación de oxígeno, frecuencia respiratoria y temperatura) muestreados horariamente durante las 48 horas previas al alta. Los valores faltantes se gestionaron mediante forward-fill y máscaras binarias, mientras que los valores atípicos se ajustaron a rangos fisiológicos normales mediante winsorización. Para garantizar la integridad del experimento, la normalización z-score se calculó exclusivamente sobre el conjunto de entrenamiento. Finalmente, los datos se dividieron mediante una estrategia de estratificación por paciente en grupos de entrenamiento (70 %), validación (15 %) y prueba (15 %), incluyendo una validación temporal adicional (periodo 2008–2016 para entrenamiento y 2017–2019 para prueba).

Arquitectura CNN-LSTM con fusión intermedia

El modelo propuesto se articula en tres ramas principales. La rama temporal procesa un tensor que incluye las señales vitales y sus respectivas máscaras mediante dos capas Conv1D, diseñadas para identificar señales tempranas de deterioro. A estas les siguen capas BiLSTM apiladas que capturan dependencias temporales de largo alcance. Por otro lado, la rama tabular procesa las variables estáticas a través de capas densas con activación ReLU.

La integración se realiza mediante una fusión intermedia (joint fusion), donde se concatenan las salidas de ambas ramas antes de pasar por una capa final de clasificación con activación sigmoide. Esta estructura multimodal ha demostrado ser superior en tareas de predicción clínica (Purushotham et al., 2018). La elección de esta arquitectura frente a otras opciones, como los transformers, se justifica por su eficiencia en el manejo de señales densas y su capacidad para detectar deterioros hemodinámicos locales antes del alta (Baker et al.,

2020; Rafiei et al., 2021).

Entrenamiento y gestión del desbalance de datos

El proceso de entrenamiento se llevó a cabo utilizando el optimizador Adam (Kingma & Ba, 2015), configurado con una tasa de aprendizaje inicial de 1×10^{-3} y un tamaño de lote (batch size) de 64. Para optimizar la convergencia y evitar el sobreajuste, se implementó un sistema de early stopping con una paciencia de 10 épocas basado en el AUC-PR del conjunto de validación, junto con una estrategia de reducción de la tasa de aprendizaje (ReduceLROnPlateau) y regularización L2.

Dado que solo el 15,3 % de los casos correspondían a readmisiones, abordamos el desbalance de clases comparando tres estrategias: el pesaje de clases (class weighting), el uso de focal loss (Lin et al., 2017) y una combinación de submuestreo con SMOTE (Chawla et al., 2002). De estas, la configuración con focal loss (con $\alpha = 0,25$ y $\gamma = 2$) fue la que ofreció los mejores resultados de validación. Todos los experimentos se desarrollaron en PyTorch 2.1 utilizando una GPU NVIDIA A100. Para garantizar la reproducibilidad, se fijaron semillas de aleatoriedad y se mantuvo un registro exhaustivo de los experimentos en la plataforma Weights & Biases, siguiendo los estándares internacionales TRIPOD+AI y MI-CLAIM.

Métricas y evaluación

Para una evaluación integral en un contexto de datos desbalanceados, se reportaron tanto el AUC-ROC como el AUC-PR, siendo esta última métrica especialmente informativa ante prevalencias bajas (Davis & Goadrich, 2006). Adicionalmente, se calcularon la sensibilidad, especificidad, F1-score, el coeficiente de correlación de Matthews y métricas de calibración como el Brier score y el Expected Calibration Error (ECE). Los intervalos de confianza al 95 % se estimaron mediante un proceso de bootstrap con 1 000 remuestreos, y las comparaciones entre las curvas AUC se realizaron mediante el test de DeLong.

Resultados experimentales

El modelo CNN-LSTM propuesto demostró un desempeño superior en las pruebas realizadas sobre el conjunto interno de MIMIC-IV, alcanzando un AUC-ROC de 0,812 (IC 95 % 0,805–0,819) y un AUC-PR de 0,451. Estos resultados superan significativamente ($p < 0,001$) a modelos tradicionales como la regresión logística (0,679) y el índice LACE (0,681), así como a algoritmos de aprendizaje automático como Random Forest (0,731) y XGBoost (0,742). Incluso frente a variantes más sencillas de redes neuronales, como una LSTM simple (0,776) o una CNN pura (0,768), la arquitectura híbrida mantuvo su ventaja competitiva.

En la validación externa con la base de datos eICU-CRD, el AUC-ROC se situó en 0,778. Aunque se observó el habitual descenso en el rendimiento al cambiar de centro hospitalario (Futoma et al., 2020), el modelo se mantuvo por encima de todos los niveles de referencia (baselines). Finalmente, la validación temporal (2017–2019) arrojó un AUC de 0,798, lo que confirma una estabilidad sólida frente a posibles cambios en los datos a lo largo del tiempo (drift temporal).

Tabla 1

Tabla de resultados del manuscrito

Modelo	AUC-ROC	AUC-PR	F1	Brier
LACE implementado	0,681	0,268	0,341	0,142
Regresión logística	0,679	0,262	0,338	0,138
Random Forest	0,731	0,325	0,392	0,124
XGBoost	0,742	0,341	0,404	0,118
LSTM (solo temporal)	0,776	0,388	0,431	0,111
CNN 1D (solo temporal)	0,768	0,376	0,422	0,114
CNN-LSTM con fusión (propuesto)	0,812	0,451	0,472	0,102

Análisis de interpretabilidad mediante SHAP para comprender los factores que impulsan las predicciones del modelo, realizamos un análisis de importancia utilizando SHAP (Lundberg & Lee, 2017). Este proceso permitió identificar como predictores principales el número de admisiones previas en el último año, la variabilidad de la frecuencia cardíaca en las 12 horas finales de estancia, la saturación mínima de oxígeno en las últimas 24 horas, el índice de Charlson y los niveles de sodio al alta. Asimismo, los mapas de

saliencia sobre las series temporales revelaron que el modelo atiende selectivamente a los periodos de mayor inestabilidad hemodinámica previos al alta, lo cual es plenamente coherente con el razonamiento clínico tradicional.

Aplicaciones prácticas y futuras

La arquitectura propuesta permite una estratificación de riesgo que puede integrarse directamente en el flujo de trabajo hospitalario. Mediante un sistema de puntuación calibrado, es posible identificar al cuartil de pacientes con una probabilidad de reingreso $\geq 0,40$. En esta población, intervenciones dirigidas como la conciliación medicamentosa, la educación estructurada, las llamadas de seguimiento en las primeras 72 horas y los programas de cuidados de transición, han demostrado reducir las readmisiones entre un 15 % y un 25 %. Su despliegue efectivo requiere la integración con el sistema de información hospitalaria (HIS) mediante estándares FHIR y un monitoreo continuo para detectar posibles cambios en la distribución de los datos (dataset shift), siguiendo las buenas prácticas de la FDA para el software médico (U.S. Food and Drug Administration, 2021).

En el contexto latinoamericano, y específicamente en el ecuatoriano, la aplicabilidad de esta tecnología depende de tres factores críticos: la madurez de los registros electrónicos de salud (EHR) en instituciones como el IESS o el MSP, la disponibilidad de series temporales de signos vitales estructuradas y la validación local para evitar sesgos de transferencia (Fierro Mella, 2019).

El trabajo futuro se orientará hacia la adaptación del modelo mediante transfer learning en cohortes regionales y la comparación con arquitecturas de gran escala como BEHRT o Med-BERT. También se prevé la fusión multimodal con notas clínicas procesadas por ClinicalBERT e imágenes médicas, así como la incorporación de mecanismos de atención que faciliten la interpretación del modelo en tiempo real. Finalmente, la integración con marcos regionales de salud digital y estudios epidemiológicos locales es indispensable para transformar esta evidencia científica en políticas públicas sanitarias efectivas.

Consideraciones éticas y gobernanza

Los aspectos éticos demandan una atención explícita, ya que los modelos basados en datos históricos pueden amplificar sesgos estructurales y heredar distorsiones del registro clínico (Char et al., 2018). Para mitigar riesgos como el sesgo racial o socioeconómico documentado en algoritmos de gestión poblacional (Obermeyer et al., 2019), recomendamos realizar evaluaciones de equidad (fairness) por subgrupos y auditorías periódicas del desempeño.

Es imperativo cumplir con los principios de la OMS sobre la gobernanza de la IA en salud (World Health Organization, 2021), complementando la interpretabilidad de herramientas como SHAP con modelos que sean inherentemente transparentes siempre que sea posible. La validación externa obligatoria sigue siendo el estándar de oro para garantizar que estas herramientas actúen como un apoyo seguro y confiable para la decisión médica.

CONCLUSIONES

Se concluyó que la arquitectura CNN-LSTM con fusión intermedia alcanza un AUC-ROC de 0,812 en la predicción de readmisiones a 30 días. Este desempeño supera en 0,13 puntos a los modelos tradicionales como el índice LACE y en 0,07 puntos a algoritmos avanzados como XGBoost, con un AUC-PR de 0,451 que triplica la prevalencia basal y demuestra una utilidad clínica tangible.

Consecuentemente, la sinergia lograda entre las convoluciones 1D para detectar patrones locales de deterioro y las unidades LSTM bidireccionales para modelar dependencias temporales largas, justifica la superioridad arquitectónica de esta propuesta frente a modelos recurrentes o convolucionales puros.

Asimismo, el manejo explícito de los valores faltantes mediante el uso de máscaras y la implementación de focal loss resultaron ser factores determinantes para optimizar el rendimiento del modelo dentro del contexto altamente desbalanceado de las readmisiones hospitalarias.

En cuanto a la robustez del modelo, la validación externa en eICU-CRD mostró una reducción esperada pero moderada en el rendimiento (0,778), lo cual refuerza la importancia crítica de realizar procesos de

validación local antes de cualquier despliegue en entornos clínicos reales.

Metodológicamente, este estudio ilustra la aplicación práctica de los estándares internacionales TRI-POD+AI y MI-CLAIM, realizando un aporte directo a la reproducibilidad y la calidad en la investigación de aprendizaje profundo aplicado a la salud.

Finalmente, se determinó que la implementación en el contexto latinoamericano requiere esfuerzos adicionales en adaptación y gobernanza. Es indispensable alinear estos desarrollos con la agenda de transformación digital de la OPS para que estas herramientas impacten positivamente en los sistemas de salud regionales y logren reducir los costos directos y humanos de las readmisiones evitables.

FINANCIACIÓN

La presente investigación no contó con financiamiento específico por parte de organismos gubernamentales, instituciones académicas, centros de investigación ni entidades privadas.

El desarrollo del estudio se realizó utilizando recursos académicos y tecnológicos disponibles para los autores, así como bases de datos de acceso abierto como MIMIC-IV y eICU-CRD, empleadas bajo sus respectivas licencias y condiciones de uso.

CONFLICTO DE INTERESES

Los autores declaran que no existen conflictos de intereses relacionados con la realización de esta investigación.

Asimismo, se garantiza que los resultados presentados no han sido influenciados por intereses económicos, personales o institucionales, y que el estudio se desarrolló bajo principios de rigor científico, ética y transparencia.

CONTRIBUCIÓN DE AUTORÍA

Tabla 2

Matriz de contribuciones CRediT (propuesta sujeta a confirmación)

Contribución	A1	A2	A3	A4	A5	A6
Conceptualización	X	X	X	X	X	X
Análisis formal	X	X	X	X	X	X
Investigación	X	X	X	X	X	X
Metodología	X	X	X	X	X	X
Recursos	X	X	X	X	X	X
Curación de datos	X	—	—	—	—	—
Redacción: borrador	X	X	X	X	X	X
Redacción: revisión y edición	X	X	X	X	X	X
Supervisión	X	—	—	—	—	—
Aprobación de versión final	X	X	X	X	X	X

A1–A6: autores en orden de firma. Asignación editorial propuesta; requiere validación.

BIBLIOGRAFÍA

- Aracena, C., Villena, F., Arias, F., & Dunstan, J. (2022). Aplicaciones de aprendizaje automático en salud. *Revista Médica Clínica Las Condes*, 33(6), 568–575. <https://doi.org/10.1016/j.rmcl.2022.11.002>
- Ashfaq, A., Sant'Anna, A., Lingman, M., & Nowaczyk, S. (2019). Readmission prediction using deep learning on electronic health records. *Journal of Biomedical Informatics*, 97, 103256. <https://doi.org/10.1016/j.jbi.2019.103256>
- Baker, S., Xiang, W., & Atkinson, I. (2020). Continuous and automatic mortality risk prediction using vital signs in the intensive care unit: A hybrid neural network approach. *Scientific Reports*, 10(1), 21282. <https://doi.org/10.1038/s41598-020-78184-7>
- Barbieri, S., Kemp, J., Perez-Concha, O., Kotwal, S., Gallagher, M., Ritchie, A., & Jorm, L. (2020). Benchmarking deep learning architectures for predicting readmission to the ICU and describing patients-at-risk. *Scientific Reports*, 10(1), 1111. <https://doi.org/10.1038/s41598-020-58053-z>
- Billings, J., Blunt, I., Steventon, A., Georgiou, T., Lewis, G., & Bardsley, M. (2012). Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30). *BMJ Open*, 2(4), e001667. <https://doi.org/10.1136/bmjopen-2012-001667>
- Cardona-Morrell, M., Prgomet, M., Turner, R. M., Nicholson, M., & Hillman, K. (2016). Frecuencia de reingresos hospitalarios y factores asociados en afiliados a una administradora de servicios de salud en Colombia. *Cadernos de Saúde Pública*, 32(7), e00146014. <https://doi.org/10.1590/0102-311X00146014>

- Castro, M. S. M., Carvalho, M. S., & Travassos, C. (2005). Factors associated with readmission to a general hospital in Brazil. *Cadernos de Saúde Pública*, 21(4), 1186–1200. <https://doi.org/10.1590/S0102-311X2005000400021>
- Castro G., P., Vukasovic R., J. L., Garcés S., E., Sepúlveda M., L., Ferrada K., M., & Alvarado O., S. (2004). Insuficiencia cardíaca en hospitales chilenos: resultados del Registro Nacional de Insuficiencia Cardíaca, Grupo ICARO. *Revista Médica de Chile*, 132(6), 655–662. <https://doi.org/10.4067/S0034-98872004000600001>
- Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care — Addressing ethical challenges. *New England Journal of Medicine*, 378(11), 981–983. <https://doi.org/10.1056/NEJMp1714229>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
- Che, Z., Purushotham, S., Cho, K., Sontag, D., & Liu, Y. (2018). Recurrent neural networks for multivariate time series with missing values. *Scientific Reports*, 8(1), 6085. <https://doi.org/10.1038/s41598-018-24271-9>
- Chen, I. Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K., & Ghassemi, M. (2021). Ethical machine learning in healthcare. *Annual Review of Biomedical Data Science*, 4, 123–144. <https://doi.org/10.1146/annurev-biodatasci-092820-114757>
- Choi, E., Bahadori, M. T., Kulas, J. A., Schuetz, A., Stewart, W. F., & Sun, J. (2016). RETAIN: An interpretable predictive model for healthcare using reverse time attention mechanism. *Advances in Neural Information Processing Systems*, 29, 3504–3512.
- Chung, C. C., Chiu, I. M., Cheng, C. Y., Lin, C. H. R., Kung, C. T., Lin, Y. C., & Wu, C. C. (2022). Machine learning models for predicting in-hospital mortality in patient with sepsis: Analysis of vital sign dynamics. *Frontiers in Medicine*, 9, 964667. <https://doi.org/10.3389/fmed.2022.964667>
- Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., Ghassemi, M., Liu, X., Reitsma, J. B., van Smeden, M., Boulesteix, A.-L., Camaradou, J. C., Celi, L. A., Denaxas, S., Denniston, A. K., Glocker, B., Golub, R. M., Harvey, H., Heinze, G., ... Logullo, P. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*, 385, e078378. <https://doi.org/10.1136/bmj-2023-078378>
- D'Agostino, M., Marti, M., Otero, P., De Cosio, F. G., Savignano, C., Nelson, J., García Saisó, S., Luna, D., Mayer-Lima, W. A., & Curioso, W. H. (2023). Digital transformation as a strategy to strengthen essential public health functions in the Americas. *Revista Panamericana de Salud Pública*, 47, e166. <https://doi.org/10.26633/RPSP.2023.166>
- Davis, J., & Goadrich, M. (2006). The relationship between precision-recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning*, 233–240. <https://doi.org/10.1145/1143844.1143874>
- Delgado-Osorio, N., Vidal-Medina, U., López, D. M., & Blobel, B. (2023). Aprendizaje automático aplicado a la predicción de diabetes mellitus, utilizando información socioeconómica y ambiental de usuarios del sistema de salud. *Revista Facultad Nacional de Salud Pública*, 41(2), e351168. <https://doi.org/10.17533/udea.rfnsp.e351168>
- Donzé, J., Aujesky, D., Williams, D., & Schnipper, J. L. (2013). Potentially avoidable 30-day hospital readmissions in medical patients: Derivation and validation of a prediction model. *JAMA Internal Medicine*, 173(8), 632–638. <https://doi.org/10.1001/jamainternmed.2013.3023>
- Donzé, J. D., Williams, M. V., Robinson, E. J., Zimlichman, E., Aujesky, D., Vasilevskis, E. E., Kripalani, S., Metlay, J. P., Wallington, T., Fletcher, G. S., Auerbach, A. D., & Schnipper, J. L. (2016). International validity of the HOSPITAL score to predict 30-day potentially avoidable hospital readmissions. *JAMA Internal Medicine*, 176(4), 496–502. <https://doi.org/10.1001/jamainternmed.2015.8462>
- Fierro Mella, C. C. (2019). Diseño y desarrollo de modelo predictivo de re-hospitalizaciones no planificadas usando Deep Learning [Tesis de pregrado, Universidad de Chile]. Repositorio Académico Universidad de Chile. <https://repositorio.uchile.cl/handle/2250/174955>
- Futoma, J., Simons, M., Panch, T., Doshi-Velez, F., & Celi, L. A. (2020). The myth of generalisability in clinical research and machine learning in health care. *The Lancet Digital Health*, 2(9), e489–e492. [https://doi.org/10.1016/S2589-7500\(20\)30186-2](https://doi.org/10.1016/S2589-7500(20)30186-2)
- Gianfrancesco, M. A., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential biases in machine learning algorithms using electronic health records. *JAMA Internal Medicine*, 178(11), 1544–1547. <https://doi.org/10.1001/jamainternmed.2018.3763>
- Gutiérrez-Portela, F., Almenares-Mendoza, F., Calderón-Benavides, L., & Romero-Riaño, E. (2024). Predicción de riesgo cardiovascular mediante aprendizaje automático: un análisis comparativo entre técnicas. *Ingeniería y Competitividad*, 26(1). <https://doi.org/10.25100/iyc.v26i1.13229>
- Harutyunyan, H., Khachatrian, H., Kale, D. C., Ver Steeg, G., & Galstyan, A. (2019). Multitask learning and benchmarking with clinical time series data. *Scientific Data*, 6(1), 96. <https://doi.org/10.1038/s41597-019-0103-9>
- Hayat, N., Geras, K. J., & Shamout, F. E. (2022). MedFuse: Multi-modal fusion with clinical time-series data and chest X-ray images. *Proceedings of Machine Learning Research*, 182, 479–503.
- Huang, K., Altosaar, J., & Ranganath, R. (2020). ClinicalBERT: Modeling clinical notes and predicting hospital readmission. *Proceedings of the ACM Conference on Health, Inference, and Learning Workshop*. <https://doi.org/10.48550/arXiv.1904.05342>
- Huang, S.-C., Pareek, A., Seyyedi, S., Banerjee, I., & Lungren, M. P. (2020). Fusion of medical imaging and electronic health records using deep learning: A systematic review and implementation guidelines. *npj Digital Medicine*, 3(1), 136. <https://doi.org/10.1038/s41746-020-00341-z>

- Hyland, S. L., Faltys, M., Hüser, M., Lyu, X., Gumbsch, T., Esteban, C., Bock, C., Horn, M., Moor, M., Rieck, B., Zimmermann, M., Bodenham, D., Borgwardt, K., Rätsch, G., & Merz, T. M. (2020). Early prediction of circulatory failure in the intensive care unit using machine learning. *Nature Medicine*, 26(3), 364–373. <https://doi.org/10.1038/s41591-020-0789-4>
- Jencks, S. F., Williams, M. V., & Coleman, E. A. (2009). Rehospitalizations among patients in the Medicare fee-for-service program. *New England Journal of Medicine*, 360(14), 1418–1428. <https://doi.org/10.1056/NEJMsa0803563>
- Johnson, A. E. W., Bulgarelli, L., Shen, L., Gayles, A., Shammout, A., Horng, S., Pollard, T. J., Hao, S., Moody, B., Gow, B., Lehman, L. H., Celi, L. A., & Mark, R. G. (2023). MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data*, 10(1), 1. <https://doi.org/10.1038/s41597-022-01899-x>
- Kansagara, D., Englander, H., Salanitro, A., Kagen, D., Theobald, C., Freeman, M., & Kripalani, S. (2011). Risk prediction models for hospital readmission: A systematic review. *JAMA*, 306(15), 1688–1698. <https://doi.org/10.1001/jama.2011.1515>
- Khan, M. S., Sreenivasan, J., Lateef, N., Abougergi, M. S., Greene, S. J., Ahmad, T., Anker, S. D., Fonarow, G. C., & Butler, J. (2021). Trends in 30- and 90-day readmission rates for heart failure. *Circulation: Heart Failure*, 14(4), e008335. <https://doi.org/10.1161/CIRCHEARTFAILURE.121.008335>
- Khan, M. S., Samman Tahhan, A., Vaduganathan, M., Greene, S. J., Alrohaibani, A., Anker, S. D., Fonarow, G. C., & Butler, J. (2023). Global comparison of readmission rates for patients with heart failure. *Journal of the American College of Cardiology*, 82(5), 430–444. <https://doi.org/10.1016/j.jacc.2023.05.040>
- Khera, R., Wang, Y., Bernheim, S. M., Lin, Z., & Krumholz, H. M. (2020). Post-discharge acute care and outcomes following readmission reduction initiatives. *BMJ*, 368, l6831. <https://doi.org/10.1136/bmj.l6831>
- Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *Proceedings of the 3rd International Conference on Learning Representations*. <https://doi.org/10.48550/arXiv.1412.6980>
- Li, Y., Rao, S., Solares, J. R. A., Hassaine, A., Ramakrishnan, R., Canoy, D., Zhu, Y., Rahimi, K., & Salimi-Khorshidi, G. (2020). BEHRT: Transformer for electronic health records. *Scientific Reports*, 10(1), 7155. <https://doi.org/10.1038/s41598-020-62922-y>
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. *Proceedings of the IEEE International Conference on Computer Vision*, 2980–2988. <https://doi.org/10.1109/ICCV.2017.324>
- Lin, Y.-W., Zhou, Y., Faghri, F., Shaw, M. J., & Campbell, R. H. (2019). Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. *PLOS ONE*, 14(7), e0218942. <https://doi.org/10.1371/journal.pone.0218942>
- Lipton, Z. C., Kale, D. C., Elkan, C., & Wetzell, R. (2016). Learning to diagnose with LSTM recurrent neural networks. *Proceedings of the 4th International Conference on Learning Representations*. <https://doi.org/10.48550/arXiv.1511.03677>
- López, D. M., & Osorio, J. S. (2025). Inteligencia artificial y salud digital en Colombia: panorama de los avances más recientes y retos futuros. *Biomédica*, 45(Sp. 3), 5–9. <https://doi.org/10.7705/biomedica.8248>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
- Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1), 56–67. <https://doi.org/10.1038/s42256-019-0138-9>
- Mahmoudi, E., Kamdar, N., Kim, N., Gonzales, G., Singh, K., & Waljee, A. K. (2020). Use of electronic medical records in development and validation of risk prediction models of hospital readmission: Systematic review. *BMJ*, 369, m958. <https://doi.org/10.1136/bmj.m958>
- Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep Patient: An unsupervised representation to predict the future of patients from the electronic health records. *Scientific Reports*, 6, 26094. <https://doi.org/10.1038/srep26094>
- Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: Review, opportunities and challenges. *Briefings in Bioinformatics*, 19(6), 1236–1246. <https://doi.org/10.1093/bib/bbx044>
- Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. *Nature*, 616(7956), 259–265. <https://doi.org/10.1038/s41586-023-05881-4>
- Myers, L. C., Faridi, M. K., Currier, P., Camargo, C. A., & Hasegawa, K. (2020). The Hospital Readmissions Reduction Program and readmissions for chronic obstructive pulmonary disease, 2006–2015. *Annals of the American Thoracic Society*, 17(4), 450–456. <https://doi.org/10.1513/AnnalsATS.201909-672OC>
- Nazzari, N. C., Campos, T. P., Corbalán, H. R., Lanús, Z. F., Bartolucci, J. J., Sanhueza, C. P., Cavada, Ch. G., & Prieto, D. J. C. (2017). Incidencia y letalidad intrahospitalaria por insuficiencia cardíaca en Chile: ¿Existen diferencias por sexo? *Revista Médica de Chile*, 145(6), 703–709. <https://doi.org/10.4067/s0034-98872017000600703>
- Norgeot, B., Quer, G., Beaulieu-Jones, B. K., Torkamani, A., Dias, R., Gianfrancesco, M., Arnaout, R., Kohane, I. S., Saria, S., Topol, E., Obermeyer, Z., Yu, B., & Butte, A. J. (2020). Minimum information about clinical artificial intelligence modeling: The MI-CLAIM checklist. *Nature Medicine*, 26(9), 1320–1324. <https://doi.org/10.1038/s41591-020-1041-y>

- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. <https://doi.org/10.1126/science.aax2342>
- Organización Panamericana de la Salud. (2023). Caja de herramientas para la transformación digital del sector salud (Versión 2.0). OPS/OMS.
- Pollard, T. J., Johnson, A. E. W., Raffa, J. D., Celi, L. A., Mark, R. G., & Badawi, O. (2018). The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Scientific Data*, 5(1), 180178. <https://doi.org/10.1038/sdata.2018.178>
- Purushotham, S., Meng, C., Che, Z., & Liu, Y. (2018). Benchmarking deep learning models on large healthcare datasets. *Journal of Biomedical Informatics*, 83, 112–134. <https://doi.org/10.1016/j.jbi.2018.04.007>
- Rafiei, A., Rezaee, A., Hajati, F., Gheisari, S., & Golzan, M. (2021). SSP: Early prediction of sepsis using fully connected LSTM-CNN model. *Computers in Biology and Medicine*, 128, 104110. <https://doi.org/10.1016/j.combiomed.2020.104110>
- Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. *Annals of Internal Medicine*, 169(12), 866–872. <https://doi.org/10.7326/M18-1990>
- Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G., Duggan, G. E., Irvine, J., Le, Q., Litsch, K., . . . Dean, J. (2018). Scalable an
- Al Alawi, A., et al. (2025). Development and evaluation of a machine learning model for predicting 30-day readmission in general internal medicine. *Computers*, 14(5), 177. <https://doi.org/10.3390/computers14050177>
- Alemán Flores, M., & Ynojosa Parra, L. (2024). Desafíos frente al uso de las TICs en pandemia para el programa de enfermería. *Zenodo*. <https://doi.org/10.5281/zenodo.10465777>
- Cabrera-Mudarra, J. (2025). Inteligencia artificial y calidad clínica: Herramientas para optimizar el desempeño hospitalario. *Noesis*, 7(14). <https://doi.org/10.35381/noesisin.v7i14.349>
- Hedayet, N., & Haseen, M. (2024). Artificial intelligence in public health: A review article. *Bangladesh Journal of Bioethics*, 15(2). <https://doi.org/10.62865/bjbio.v15i2.108>
- Lee, C., et al. (2023). Multimodal deep learning of fundus abnormalities and traditional risk factors for cardiovascular risk prediction. *NPJ Digital Medicine*, 6, 48. <https://doi.org/10.1038/s41746-023-00748-4>
- Liu, Y., et al. (2023). Learning multi-modal brain tumor segmentation from MRI images. *Computers in Biology and Medicine*, 158, 106927. <https://doi.org/10.1016/j.combiomed.2023.106927>
- López-Canay, A., et al. (2025). Predicting COPD readmission: An intelligent clinical decision support system. *Diagnostics*, 15(3), 318. <https://doi.org/10.3390/diagnostics15030318>
- Mahootiha, S., et al. (2024). Multimodal deep learning for renal carcinoma prognosis. *Computer Methods and Programs in Biomedicine*, 240, 107978. <https://doi.org/10.1016/j.cmpb.2023.107978>
- Mohsen, H., et al. (2022). Artificial intelligence-based methods for fusion of EHR and imaging data. *Scientific Reports*, 12, 22514. <https://doi.org/10.1038/s41598-022-22514-4>
- Ruiz, J., & Velásquez, M. (2023). Inteligencia artificial al servicio de la salud del futuro. *Revista Médica Clínica Las Condes*, 34(1). <https://doi.org/10.1016/j.rmclc.2022.12.001>