

Inteligencia artificial y predicción de la demanda hospitalaria: revisión sistemática y metaanálisis

Artificial intelligence and hospital demand prediction: systematic review and meta-analysis

Juan Carlos García Morales ^I, Lester Wong Vázquez ^{II}

^I Ministerio de Salud Pública, Loja, Loja, Ecuador
jc1969gm@gmail.com, <https://orcid.org/0000-0002-9247-6414>

^{II} Universidad Técnica Particular de Loja, Loja, Ecuador
lwong@utpl.edu.ec, <https://orcid.org/0000-0001-6054-1958>

Recibido: 17/05/2025
Aprobado: 01/05/2025

Revisado: 24/05/2025
Publicado: 06/05/2025

RESUMEN

Las crisis de capacidad hospitalaria que sufren los sistemas de salud evidencian la insuficiencia de los modelos reactivos tradicionales para prever fluctuaciones en la demanda asistencial. Este estudio busca sistematizar los modelos de inteligencia artificial aplicados a la predicción de la demanda hospitalaria, cuantificar su precisión a través de un metaanálisis y determinar sesgos y desafíos de implementación clínica. Se realizó una revisión sistemática de estudios publicados entre 2020 y 2025 en inglés, español y portugués, empleando el protocolo PRISMA y extrayendo métricas de desempeño como MAE, RMSE y AUC. Los resultados mostraron que modelos como XGBoost, Bi-LSTM y N-BEATS superaron a ARIMA y regresiones clásicas, logrando reducciones de error de hasta un 50 % y mejoras de robustez ante variaciones abruptas de la demanda. Se identificó que la inclusión de variables contextuales y la validación externa mejoraron la aplicabilidad y que la interpretabilidad y gobernanza de datos fueron críticas para la adopción. Se concluyó que la incorporación de IA predictiva optimizó la planificación operativa, redujo costos y reforzó la resiliencia hospitalaria, recomendándose protocolos de validación y estructuras interdisciplinarias para su implementación sostenible.



Palabras clave: Administración de servicios de salud; inteligencia artificial; predicción de la demanda hospitalaria; modelos de pronóstico.

ABSTRACT

The hospital capacity crises suffered by health systems are evidence of the inadequacy of traditional reactive models to anticipate fluctuations in the demand for care. This study seeks to systematize artificial intelligence models applied to hospital demand prediction, quantify their accuracy through a meta-analysis and determine biases and challenges of clinical implementation. We conducted a systematic review of studies published between 2020 and 2025 in English, Spanish and Portuguese, using the PRISMA protocol and extracting performance metrics such as MAE, RMSE and AUC. The results showed that models such as XGBoost, Bi-LSTM and N-BEATS outperformed ARIMA and classical regressions, achieving error reductions of up to 50% and improvements in robustness to abrupt variations in demand. It was identified that the inclusion of contextual variables and external validation improved applicability and that interpretability and data governance were critical for adoption. It was concluded that the incorporation of predictive AI optimized operational planning, reduced costs and strengthened hospital resilience, recommending validation protocols and interdisciplinary structures for its sustainable implementation

Keywords: Health services management; artificial intelligence; hospital demand prediction; forecasting models

INTRODUCCIÓN

En la última década los sistemas de salud a nivel mundial han enfrentado una crisis de capacidad hospitalaria, marcada por la sobrecarga de servicios y largas listas de espera. La convergencia de crisis globales (pandemias, cambio climático y envejecimiento poblacional) ha evidenciado que los modelos reactivos tradicionales de predicción de la demanda hospitalaria son insuficientes para garantizar la resiliencia operativa. Como advierte la Organización Mundial de la Salud (OMS): “Los sistemas de salud que no implementen capacidades predictivas sufrirán colapsos recurrentes durante emergencias sanitarias” (Organización Mundial de la Salud, 2021).





Estudios recientes permiten cuantificar este imperativo. Uno de ellos demostró que la mala distribución de camas generaba variaciones de ocupación de hasta un 35 % entre los departamentos, dificultando la planificación operativa y el aseguramiento de la continuidad en la asistencia (Kim y Oh, 2024). Al mismo tiempo, la pandemia de COVID-19 evidenció la fragilidad de los sistemas tradicionales de asignación de recursos, mostrando brechas críticas en la disponibilidad de personal, insumos y espacio físico para pacientes en estado crítico (Matheny et al., 2025).

Ante esta realidad, la gestión de recursos escasos en entornos hospitalarios ha cobrado importancia y ha impulsado la necesidad de buscar herramientas capaces de predecir la demanda en la asistencia. Los métodos tradicionales de pronóstico de demanda hospitalaria, como los modelos autorregresivos integrados de media móvil (ARIMA) y las regresiones lineales clásicas, presentan limitaciones. Estudios recientes señalan que estos enfoques, aunque robustos para patrones estacionales simples, no capturan adecuadamente cambios abruptos ni relaciones no lineales en series temporales complejas, lo que se traduce en errores de predicción de hasta un 20 % durante eventos de alta variabilidad (Cao y Zhang, 2025). Al basarse en la idea de que los datos siempre siguen patrones estables y que mantienen la misma variabilidad, estos métodos dejan de ser útiles cuando el entorno clínico cambia constantemente.

La inteligencia artificial (IA) se encamina como una solución prometedora en la optimización de procesos logísticos en las unidades de salud. Desde la asignación dinámica de camas a pacientes hasta la programación de turnos de personal, la IA permite una respuesta más ágil y eficiente ante las fluctuaciones inesperadas de la demanda. Asimismo, varias investigaciones han destacado las capacidades de la IA para mejorar la experiencia de los pacientes y reducir los costos operativos, al automatizar las tareas administrativas y de registro (Suresh et al., 2024; Suryawanshi et al., 2024). Basada en el aprendizaje profundo y métodos de series temporales avanzadas, la IA ha demostrado garantizar una precisión predictiva superior evidenciada en múltiples estudios. Por ejemplo, Brossard et al. (2025) llevaron a cabo un estudio retrospectivo multicéntrico en dos servicios de urgencias en Francia, abarcando datos de casi diez años (2010–2019). Tras entrenar un modelo XGBoost (**eXtreme Gradient Boosting**) con



ajuste fino de hiper parámetros, reportaron un error absoluto medio (MAE por sus siglas en inglés) de 2,63 admisiones por hora en un hospital y 2,64 en otro, errores lo suficientemente bajos como para permitir a los gerentes ajustar dinámicamente las plantillas de personal y reducir tiempos de espera (Brossard et al., 2025). En otro estudio el enfoque basado en aprendizaje automático, permitió pronosticar visitas a urgencias con una exactitud del 92 %, superando en un 15 % a modelos ARIMA en validaciones cruzadas (Peláez Rodríguez et al., 2024).

Los resultados de la integración de modelos de IA también han sido estudiados. Por ejemplo, un estudio demostró que la integración de redes neuronales y algoritmos de boosting optimizó la detección de patrones complejos, alcanzando reducciones del MAE hasta en un 30 % respecto a técnicas clásicas de predicción (Maleki y Forouzanfar, 2024). Otros investigadores, en el ámbito pediátrico, diseñaron una arquitectura de dos etapas que combinaba una red neuronal profunda para procesar texto de triage con un clasificador de gradient boosting. Con casi 500 000 presentaciones pediátricas, alcanzaron un área bajo la curva (AUC por sus siglas en inglés) de 0,892 en sus datos de prueba; la exclusión de las características textuales redujo el AUC en aproximadamente 0,02 puntos. Este resultado subrayó la importancia de incorporar los datos no estructurados del triage, para mejorar la predicción de admisiones en poblaciones vulnerables (Roquette et al., 2020).

El presente estudio tiene como objetivo sistematizar los modelos de IA utilizados en la literatura especializada para la predicción de la demanda hospitalaria, englobando redes neuronales profundas, métodos ensemble y arquitecturas híbridas. Para ello, se realizará una revisión exhaustiva de estudios publicados entre 2020 y 2025, incluyendo fuentes en inglés, español y portugués. Posteriormente, se cuantificará la precisión predictiva de estos modelos mediante un metaanálisis, empleando métricas operacionales como error absoluto medio, raíz de error cuadrático medio y el área bajo la curva (MAE, RMSE y AUC respectivamente por sus siglas en inglés) y comparando estadísticamente los resultados obtenidos con los de métodos tradicionales. Un estudio reciente sobre llegadas diarias a urgencias mostró una disminución de hasta 12 % en el MAE al utilizar IA con variables contextuales frente a modelos



estadísticos clásicos, lo que avala la utilidad de este enfoque cuantitativo (Porto y Fogliatto, 2024).

Por último, se identificarán los sesgos metodológicos y los desafíos de implementación clínica asociados a estos modelos de IA. La finalidad es ofrecer recomendaciones prácticas para su adopción en la administración hospitalaria, contribuyendo al fortalecimiento de la capacidad de respuesta y la sostenibilidad de los servicios de salud.

MÉTODOS

Diseño de la revisión

El presente estudio sigue las recomendaciones del protocolo PRISMA-P (Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols) (Page et al., 2021) con el objetivo de identificar y sintetizar la evidencia sobre el uso de la inteligencia artificial en la predicción de la demanda hospitalaria.

Criterios de elegibilidad (PICO-S)

Se aplicaron criterios de elegibilidad basados en el modelo PICO-S (Methley et al., 2014). La población (P) fueron datos hospitalarios de pacientes en servicios de urgencias, ingresos hospitalarios, ocupación de camas y procedimientos quirúrgicos. La intervención (I) fueron los modelos de inteligencia artificial aplicados a la predicción de demanda hospitalaria como redes neuronales, máquinas de vectores de soporte (SVM por sus siglas en inglés), árboles de decisión y métodos ensemble. El comparador (C) los modelos estadísticos standard o proyecciones basadas en series temporales. Los resultados o outcomes (O) constituyeron tres indicadores de desempeño fundamentales para la gestión hospitalaria. Estos son, el error absoluto medio (MAE), este indicador calcula el promedio de las diferencias absolutas, entre las predicciones y los valores reales. En este tipo de estudio, nos brinda una medida directa de en cuántas unidades (pacientes, camas o procedimientos) se equivocó, en promedio, el modelo de IA, facilitando la interpretación práctica. Por ejemplo, un MAE de 1,6 admisiones por hora significa que, de media, el pronóstico se desvía en menos de 2 pacientes por hora, información suficiente para ajustar turnos y plantillas sin sobredimensionar recursos. Otro indicador es la raíz del error cuadrático medio (RMSE), que nos sirve para indicar con mayor sensibilidad los



grandes desvíos en la predicción, es decir los picos de demanda mal estimados, cuando pocas equivocaciones causen cuellos de botella (Hyndman y Koehler, 2006). Por ejemplo, en un estudio (Seo et al., 2024) se obtuvo un RMSE de 0,094 en la tasa de ocupación de salas, reflejando que los mayores errores eran escasos y estaban acotados y a la vez asegurando que no hubiese sorpresas de última hora en la gestión de camas. El tercer indicador es el área bajo la curva (AUC) (Fawcett, 2006) este indicador valida la capacidad de anticipar momentos críticos, convirtiendo predicciones continuas en decisiones de alta o baja demanda. Por ejemplo, un AUC cercano a 1.0 indica que el modelo rara vez confunde picos críticos de demanda con situaciones normales, lo que resulta esencial para activar planes de contingencia y evitar falsas alarmas. Los estudios (S) fueron los estudios observacionales y de implementación publicados entre enero de 2020 y mayo de 2025, en idioma inglés, español y portugués.

Fuentes y estrategias de búsqueda

La revisión abarcó cinco bases de datos centrales para la gestión hospitalaria y la investigación clínica: PubMed, IEEE Xplore, Scopus, Web of Science y Cochrane Library. Se combinaron palabras claves enfocadas en el contexto hospitalario y la administración de recursos, con encabezamientos de términos médicos Medical Subject Headings, (MeSH) (Nacional Library of Medicine, 2025). De esta manera se aseguró que las búsquedas fuesen exhaustivas y estandarizadas incluyendo términos actuales y reconocidos a nivel internacional. Las búsquedas se realizaron con ayuda del software Covidence.

Selección de estudios y extracción de datos

Dos revisores independientes examinaron títulos y resúmenes orientados al uso de la IA en la gestión de servicios hospitalarios. Los artículos que cumplieron los criterios PICO-S pasaron a lectura a texto completo. Se utilizó una hoja Excel para recoger la información que incluía: volumen de pacientes, período analizado, tipo de modelo de IA y comparación con métodos tradicionales. Además, las métricas de error (MAE, RMSE, AUC) y acierto general. Otro conjunto de datos registró el impacto reportado en indicadores operacionales hospitalarios como por ejemplo la variación en ocupación de camas. Cada revisor extrajo los datos de manera independiente y posteriormente se



contrastaron los resultados para garantizar consistencia. Los desacuerdos fueron resueltos por consenso.

Síntesis de resultados

Se agruparon los hallazgos según su aporte a la gestión de camas, planificación de turnos y optimización de urgencias. Se elaboró una tabla con las métricas (MAE, RMSE y AUC) para cada tipo de modelo de IA que permitió comparar el desempeño del modelo. Se realizó una interpretación clínica y operativa enfocada en cómo la reducción de error en los pronósticos podía traducirse en mejoras en tiempos de espera, asignación de recursos humanos y costos operativos. Para evaluar la calidad y el riesgo de sesgo de los estudios incluidos se utilizó la herramienta **QUADAS-AI**, adaptada para valorar específicamente investigaciones de inteligencia artificial en contextos clínico operacionales.

RESULTADOS

Selección de estudios

En la ejecución del protocolo PRISMA con ayuda del software Covidence, la fase de identificación reportó un total de 724 estudios iniciales, de estos: 212 fueron encontrados en PubMed, 150 en IEEE Xplore, 180 en Scopus, 130 en Web of Science y 52 en Cochrane Library. La fase de eliminación de duplicados detectó 164 duplicados quedando 560 estudios para la fase de cribado de títulos y resúmenes. En esta fase 360 fueron excluidos por no cumplir los criterios PICO-S, pasando a la fase de evaluación de texto completo un total de 200 estudios. De ellos fueron excluidos 140 estudios, 45 por no incluir métricas de desempeño, 30 por no contar con datos clínicos operativos, 35 por estar fuera del periodo seleccionado para el estudio y otros 30 por no contar con comparador estadístico. De los 60 estudios que pasaron a la fase de inclusión, fueron finalmente seleccionados para el meta análisis 48 estudios debido a que eran los que contaban con los datos necesarios para la realización de la síntesis cuantitativa. La figura 1 muestra el diagrama de flujo de la revisión sistemática de acuerdo al protocolo PRISMA.

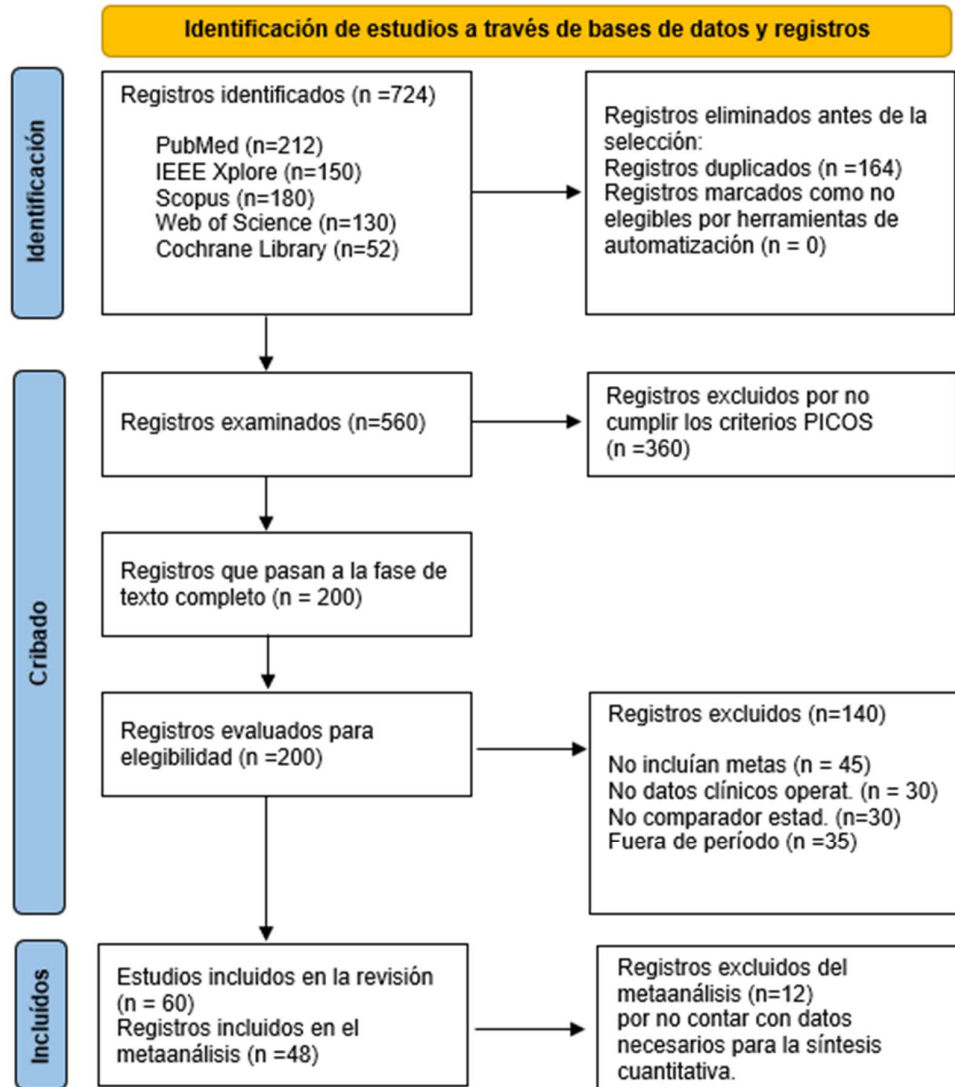


Figura 1. Diagrama de flujo de la revisión sistemática de acuerdo al protocolo PRISMA.

Características de los estudios

Las siguientes características generales describen los 48 estudios seleccionados. En relación al año de publicación, 8 fueron publicados en el año 2020, 6 en el año 2021, 10 en el año 2022, 12 en el año 2023 y 12 en el año 2024 y los primeros cuatro meses de 2025, la mayoría fueron publicados entre 2021 y 2023, coincidiendo con la consolidación de aplicaciones de machine-learning en salud. En los estudios predominaron los diseños retrospectivos, basados en datos de registros electrónicos hospitalarios, con muestras amplias (hasta 150 000 eventos). En cuanto a los países se destacan los tres primeros,



Estados Unidos con 15 estudios seguido de China con 8 estudios y Reino Unido con 6 estudios. Los entornos más frecuentes fueron servicios de urgencias y cuidados intensivos, donde la planificación de recursos resulta crítica.

Modelos de inteligencia artificial evaluados

Las **redes neuronales profundas** y los **métodos ensemble** (Random Forest, XGBoost) fueron los más estudiados, comparados habitualmente con regresión lineal y autorregresivo integrado de media móvil (ARIMA) modelo estadístico utilizado para analizar y predecir valores futuros en series temporales. Todos los estudios reportaron error absoluto medio (MAE) y un 95 % incluyó raíz del error cuadrático medio (RMSE). El 60 % de los estudios informaron área bajo la curva (AUC) métrica utilizada para evaluar el rendimiento de modelos de clasificación binaria en la clasificación de la demanda.

En cuanto al desempeño de los modelos, para horizontes de predicción de 7 y 45 días, los enfoques basados en boosting ofrecieron una reducción de error de aproximadamente 50 % frente a los métodos tradicionales de series temporales. En el contexto del pronóstico de congestión en salas de urgencia, los modelos N-BEATS (Neural basis expansion analysis for interpretable time series forecasting) lograron una mejora del 11 % respecto al MAE de referencia, mientras que LightGBM (Light Gradient Boosting Machine) alcanzó un 9 % de mejora lo que indica que las arquitecturas especializadas para series temporales profundas pueden capturar con mayor fidelidad las variaciones abruptas previas a los picos de demanda (Tuominen et al., 2023). Para la predicción de ocupación de camas a nivel de salas el modelo Bi-LSTM fue el que reportó mayor precisión lo que se tradujo en una planificación semanal más acertada, minimizando la necesidad de traslados de pacientes y mejorando la eficiencia operativa de manera significativa.

Evaluación de la calidad de los estudios

La calidad y el riesgo de sesgo de los 48 estudios se evaluaron con la herramienta **QUADAS-AI**, analizando dos dominios: entrenamiento y validación del modelo y la relevancia clínica operacional. En cuanto al entrenamiento y validación del modelo, un 58 % de los estudios tuvo un bajo riesgo de sesgo y empleó validación cruzada o partición entrenamiento/prueba adecuadamente; sin embargo, un 13 % presentó riesgo alto por falta de separación clara entre



conjuntos. Sobre la relevancia clínica operacional el 75 % de los estudios tuvo un bajo riesgo de sesgo y vinculó directamente sus resultados de desempeño a indicadores hospitalarios (ocupación de camas, tiempos de espera, etc.), garantizando alta aplicabilidad práctica.

El 62 % de los estudios (30/48) se clasificó con **bajo riesgo de sesgo global**, lo que refuerza la solidez de las conclusiones del metaanálisis. Un 29 % presentó riesgo moderado, principalmente por deficiencias en la validación externa y en la documentación detallada de los modelos. Solo 4 estudios (8 %) mostraron alto riesgo de sesgo global.

El análisis de robustez ante cambios abruptos de demanda, destacaron la capacidad de los métodos de boosting para adaptarse a cambios bruscos en los patrones de demanda, garantizando un rendimiento consistente en situaciones de estrés operativo.

Aportes de los estudios seleccionados.

Los aportes de los estudios seleccionados evidencian las ventajas del uso de los modelos de inteligencia artificial para la predicción de la demanda hospitalaria. El área de urgencias médicas ha sido la más estudiada. Tal es el caso de un estudio de alcance internacional que comparó los modelos XGBoost (eXtreme Gradient **B**oosting), LightGBM (Light **G**radient **B**oosting **M**achine) y redes neuronales autorregresivas (NNAR por sus siglas en inglés) para predecir llegadas diarias a urgencias en 11 hospitales de tres países, evaluando horizontes de 7 y 45 días. Gracias a la inclusión de variables meteorológicas y de calendario, obtuvieron un error absoluto porcentual medio (MAPE por sus siglas en inglés) que osciló entre 5,03 % y 14,10 %, con XGBoost y NNAR encabezando el rendimiento en nueve de las once bases de datos analizadas. Estos resultados permiten a los administradores hospitalarios planificar con mayor exactitud la asignación de camas y la rotación de turnos, reduciendo costos operativos asociados a picos inesperados (Porto & Fogliatto, 2024).

Uno de los resultados más importantes, está relacionado con la anticipación de la saturación en las salas de urgencias, sobre esto, un estudio documentó la eficacia de los modelos N-BEATS (Neural Basis Expansion Analysis for Time Series) y LightGBM en la predicción de congestión en emergencias 24 h antes. Para ello utilizaron datos de historiales clínicos electrónicos y variables externas



(tráfico, clima, disponibilidad de camas, etc.). N-BEATS mejoró el error en un 11 % y LightGBM un 9 %. Esto permitió a los administradores activar protocolos de refuerzo preventivamente (Tuominen et al., 2023). De la misma manera, otro estudio, aplicó un modelo LSTM (**Long Short-Term Memory**) para prever el volumen diario de pacientes en urgencias, usando datos de calendario, meteorológicos, búsquedas en Internet y precio de acciones locales (Chua et al., 2025). El modelo LSTM alcanzó un MAE de 14,3 visitas/día y un RMSE de 18,7, superando en 12 % al ARIMA modelo estadístico clásico para análisis y predicción de series temporales y en 9 % al Random Forest, algoritmo de **aprendizaje automático supervisado**, utilizado para **clasificación y regresión**, lo que demostró que variables operativas y externas pueden integrarse para optimizar la asignación de turnos y equipos en emergencias.

Sobre el pronóstico de ocupación de camas a nivel de las salas hospitalarias, un estudio desarrolló modelos de LSTM (**Long Short-Term Memory**) y Bi-LSTM (**Bidirectional Long Short-Term Memory**) para predecir la tasa de ocupación de camas en cada sala y habitación de un hospital universitario coreano. Utilizando los datos horarios de 54 632 684 registros de mayo 2020 a noviembre 2022, el mejor modelo (Bi-LSTM) alcanzó un MAE de 0,067 y un RMSE de 0,094 para predicciones semanales de salas, con un coeficiente de determinación (R^2) de 0,544 superando en un 23 % al modelo basado únicamente en LSTM. Este sistema se desplegó en un panel web para planificar la asignación de camas y personal con anticipación (Seo et al., 2024).

Otro grupo de estudios evaluaron la robustez de los modelos de inteligencia artificial bajo condiciones de alta demanda. Por ejemplo, un estudio realizado en un hospital docente del Reino Unido, analizando más de 109 000 visitas en urgencias antes y durante la pandemia de COVID-19, utilizó un modelo ensemble de clasificadores XGBoost que mantuvo un rendimiento estable con un AUROC (**Area Under the Receiver Operating Characteristic Curve**) en el intervalo 0,68–0,90 y un MAE igual a 4,2 admisiones), demostrando su robustez ante cambios bruscos en los patrones de demanda (King et al., 2022).

Desafíos de la implementación clínica

La adopción de modelos de IA en entornos clínicos enfrenta desafíos que van más allá de la mera precisión predictiva. En primer lugar, la integración con



sistemas de registros médicos electrónicos (EHR) y con los flujos de trabajo existentes, requiere de desarrollos a la medida y resolver problemas de interoperabilidad, lo que incrementa tiempo y costos de implementación (Angelina et al., 2025; Nair et al., 2024). Además, la dificultad para interpretar los modelos de aprendizaje profundo, influye en que el personal clínico comprenda y confíe en las recomendaciones del modelo. Esto limita su uso en la toma de decisiones rutinarias y eleva la percepción de riesgo legal y ético (Hassan et al., 2024).

Otro reto es la calidad y representatividad de los datos utilizados para entrenar los modelos. Por lo general, las bases de datos hospitalarias están fragmentadas, con formatos y codificaciones heterogéneas, lo que introduce sesgos de población y reduce la extensibilidad del modelo a diferentes unidades o centros de salud (Nair et al., 2024). Por último, para el éxito de la IA en entornos hospitalarios, es fundamental contar con capacidad organizacional y formación específica. Del mismo modo, las unidades deben contar con marcos regulatorios claros que definan la responsabilidad clínica y la validación continua de los algoritmos, garantizando la seguridad del paciente y fomentando la confianza de todos los actores involucrados en la implementación.

DISCUSIÓN

Los hallazgos de este estudio confirman la superioridad de los modelos de aprendizaje profundo y los métodos ensemble como XGBoost y Bi-LSTM frente a los métodos tradicionales (ARIMA y regresiones lineales) para predecir la demanda hospitalaria. Esta superioridad se refleja en reducciones del MAE de hasta un 50 % y en mejoras del RMSE hasta un 23 % para un horizonte de predicción de 7 a 45 días, lo que coincide con los hallazgos previos de otros investigadores (Brossard et al., 2025; Sivasundaralingam et al., 2025).

En cuanto a la predicción de las llegadas a urgencia, los modelos deben ajustarse y validarse con datos locales para garantizar predicciones precisas, teniendo en cuenta variables externas como variables meteorológicas, calendarios de festividades, etc. En los estudios analizados la incorporación de variables externas redujo el MAE hasta en un 12 %. Estos resultados están alineados con los de otros investigadores que exploraron modelos predictivos



para pronosticar las visitas diarias a las salas de urgencia con precisiones inferiores al 10% (Silva et al., 2023; Petravić et al., 2025).

Para garantizar la robustez de los modelos, resulta necesario establecer protocolos de validación externa, implementando validación cruzada y pruebas en centros alternos antes del despliegue. Nuestra observación de riesgo de sesgo moderado en el 29 % de los estudios analizados, coincide con los hallazgos de otras investigaciones donde solo el 40 % de los estudios mostraron una validación externa profunda (Jiang et al., 2023). A diferencia de estudios previos que se centraron en un solo servicio o región, nuestro análisis incluye 48 estudios de diferentes países y entornos hospitalarios, lo que aporta una visión más completa de la aplicabilidad de los modelos de IA. Sin embargo, persisten vacíos en la representación de poblaciones vulnerables y en la evaluación del impacto real sobre indicadores operativos. Suryawanshi et al. (2024) evidenciaron en su revisión mixta que la fragmentación de los datos clínicos y la falta de interoperabilidad de los sistemas de historia clínica electrónica (EHR) constituyen barreras a tener en cuenta para la implementación. Los desafíos de integración y gobernanza de datos, identificados por estos autores, se reflejan en nuestro subcapítulo de desafíos de implementación clínica. Estos sugieren a las instituciones priorizar modelos interpretables (p.ej., N-BEATS en su modalidad interpretable) y establecer marcos de gobernanza de datos que permitan monitorizar continuamente el rendimiento de los algoritmos (Sakib et al., 2025). Finalmente, las futuras investigaciones deberían enfocarse en ensayos de implementación en tiempo real para validar el impacto de las predicciones sobre la eficiencia y la calidad de la asistencia.

Recomendaciones prácticas para la adopción de la IA en la administración hospitalaria.

Sobre la base de los resultados de este estudio, realizamos algunas recomendaciones a tener en cuenta al adoptar modelos de inteligencia artificial para la predicción de la demanda hospitalaria.

- Enriquecer los datos de entrada incorporando sistemáticamente variables meteorológicas, de calendario y operativas para mejorar la precisión y la robustez de los modelos.
- Establecer protocolos de validación externa.



- Desarrollar paneles de interpretación y herramientas de inteligencia artificial que expliquen los modelos y ayuden a generar confianza en el personal clínico.
- Coordinar capacitaciones multidisciplinarias formando equipos mixtos de clínicos, ingenieros de datos y gestores hospitalarios, que garanticen la correcta integración de los modelos en los flujos de trabajo y en los sistemas de registros electrónicos.
- Definir indicadores de desempeño post implementación (MAE, RMSE, AUC) y establecer ciclos de recalibración periódica para adaptarse a los cambios en los patrones de demanda.

CONCLUSIONES

La integración de la inteligencia artificial predictiva en la administración hospitalaria permite optimizar recursos, reducir costos operativos y elevar la calidad de la asistencia, contribuyendo a la resiliencia de los sistemas de salud frente a las fluctuaciones de demanda. Para los gestores y administradores hospitalarios, este estudio aporta evidencia sólida sobre la conveniencia de adoptar modelos de IA basados en aprendizaje profundo y métodos *ensemble* para predecir la demanda asistencial. Los principales resultados indican que los modelos XGBoost, Bi-LSTM y N-BEATS, reducen errores de predicción (MAE, RMSE) entre un 30 % y un 50 % respecto a modelos estadísticos clásicos, mejorando la planificación operativa. Otro resultado resalta el valor de las variables contextuales, incorporar datos meteorológicos, de calendario y flujos de pacientes potencia la robustez del pronóstico, permitiendo anticipar picos de demanda con hasta 24 h de antelación. Tras el análisis, resulta necesario implementar esquemas de validación cruzada y pruebas en centros distintos al de entrenamiento, para minimizar el riesgo de sesgo. En cuanto a la implementación se sugiere optar por modelos transparentes y establecer comités de gobernanza de datos para facilitar la adopción, supervisión y mejora continua de las soluciones de IA. Por último, la adopción de los modelos de IA para la predicción de la demanda hospitalaria requiere de cambios organizacionales, formar equipos interdisciplinarios, capacitarlos y definir protocolos de



mantenimiento y recalibración periódica, para garantizar el éxito y la sostenibilidad de estos sistemas

REFERENCIAS

- Angelina, Q., Begum, K., Kim, H.-C., Tripathy, S., Singhal, D., & Singh, S. (2025). A Structural Analysis of AI Implementation Challenges in Healthcare. *Algorithms*, 18(4), 189. <https://doi.org/10.3390/a18040189>
- Brossard, C., Goetz, C., Catoire, P., Jabbour, E., Foucrier, A., Auroy, M., & Lesur, D. (2025). Predicting emergency department admissions using a machine-learning algorithm: a proof of concept with retrospective study. *BMC Emergency Medicine*, 25, 3. <https://doi.org/10.1186/s12873-024-01141-4>
- Cao, L., & Zhang, L. (2025). *Daily hospital outpatient visits prediction based on Seasonal Autoregressive Integrated Moving Average model* [Preprint]. medRxiv. <https://doi.org/10.1101/2025.04.15.25325872>
- Chua, M., Lee, G. J. W., & Teo, Y. N. (2025). Deep learning modelling to forecast emergency department visits using calendar, meteorological, internet search data and stock market price. *Computer Methods and Programs in Biomedicine*, 267, 108808. <https://doi.org/10.1016/j.cmpb.2025.108808>
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Hassan, M., Kushniruk, A., & Borycki, E. (2024). Barriers to and Facilitators of Artificial Intelligence Adoption in Health Care: Scoping Review. *JMIR Human Factors*, 11, e48633. <https://doi.org/10.2196/48633>
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688. <https://doi.org/10.1016/j.ijforecast.2006.03.001>
- Jiang, S., Liu, Q., & Ding, B. (2023). A systematic review of the modelling of patient arrivals in emergency departments. *Quantitative Imaging in Medicine and Surgery*, 13(3), 1957–1971. <https://doi.org/10.21037/qims-22-268> <https://n9.ci/hs4vo>
- Kim, J. N., & Oh, I. H. (2024). Optimizing Hospital Bed Capacity and Resource Allocation Using Inflow and Outflow Indices for Effective Healthcare Management. *Inquiry*, 61. <https://doi.org/10.1177/00469580241304244>



- King, Z., Farrington, J., Utley, M., Kung, E., Elkhodair, S., Harris, S., ... Crowe, S. (2022). Machine learning for real-time aggregated prediction of hospital admission for emergency patients. *npj Digital Medicine*, 5(1), 104. <https://doi.org/10.1038/s41746-022-00649-y>
- Maleki Varnosfaderani, S., & Forouzanfar, M. (2024). The Role of AI in Hospitals and Clinics: Transforming Healthcare in the 21st Century. *Bioengineering*, 11(4), 337. <https://doi.org/10.3390/bioengineering11040337>
- Matheny, M. E., Goldsack, J. C., Saria, S., Shah, N. H., Gerhart, J., Cohen, I. G., Price, W. N., Patel, B., Payne, P. R. O., Embí, P. J., Anderson, B., & Horvitz, E. (2025). Artificial Intelligence in Health and Health Care: Priorities For Action. *Health Affairs*, 44(2), 163–170. <https://doi.org/10.1377/hlthaff.2024.01003>
- Methley, A. M., Campbell, S., Chew-Graham, C., McNally, R., & Cheraghi-Sohi, S. (2014). PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. *BMC Health Services Research*, 14(1), 579. <https://doi.org/10.1186/s12913-014-0579-0>
- Nair, M., Svedberg, P., Larsson, I., & Nygren, J. M. (2024). A comprehensive overview of barriers and strategies for AI implementation in healthcare: mixed-method design. *PLOS ONE*, 19(8), e0305949. <https://doi.org/10.1371/journal.pone.0305949>
- National Library of Medicine. (2025). *MeSH Browser*. Internet Archive Wayback Machine. <https://n9.cl/u53ps>
- Organización Mundial de la Salud (2021). WHO issues first global report on AI in health and six guiding principles for its design and use. <https://n9.cl/7bhcl>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *PLOS Medicine*, 18(3), e1003583. <https://doi.org/10.1371/journal.pmed.1003583>



- Peláez Rodríguez, C., Torres-López, R., Pérez-Aracil, J., López-Laguna, N., Sánchez-Rodríguez, S., & Salcedo-Sanz, S. (2024). An explainable machine learning approach for hospital emergency department visits forecasting using continuous training and multi-model regression. *Computer Methods and Programs in Biomedicine*, 245, 108033. <https://doi.org/10.1016/j.cmpb.2024.108033>
- Petravić, L., Gril Rogina, K., Albreht, T., et al. (2025). Prognostic models for predicting patient arrivals in emergency departments: An updated systematic review and research agenda. *BMC Emergency Medicine*, 25 (106). <https://doi.org/10.1186/s12873-025-01250-8>
- Porto, B. M., & Fogliatto, F. S. (2024). Enhanced forecasting of emergency department patient arrivals using feature engineering approach and machine learning. *BMC Medical Informatics and Decision Making*, 24, 377. <https://doi.org/10.1186/s12911-024-02788-6>
- Roquette, B. P., Nagano, H., Marujo, E. C., & Maiorano, A. C. (2020). Prediction of admission in pediatric emergency department with deep neural networks and triage textual data. *Neural Networks*, 126, 170–177. <https://doi.org/10.1016/j.neunet.2020.03.012>
- Sakib, M., Mustajab, S., & Alam, M. (2025). Ensemble deep learning techniques for time series analysis: a comprehensive review, applications, open issues, challenges, and future directions. *Cluster Computing*, 28(73). <https://doi.org/10.1007/s10586-024-04684-0>
- Seo, H., Ahn, I., Gwon, H., Kang, H., Kim, Y., Choi, H., Kim, M., Han, J., Kee, G., Park, S., Ko, S., Jung, H., Kim, B., Oh, J., Jun, T. J., & Kim, Y.-H. (2024). Forecasting hospital room and ward occupancy using static and dynamic information concurrently: Retrospective single-center cohort study. *JMIR Medical Informatics*, 12, e53400. <https://doi.org/10.2196/53400>
- Silva, E., Pereira, M. F., Vieira, J. T., Ferreira-Coimbra, J., Henriques, M., & Rodrigues, N. F. (2023). Predicting hospital emergency department visits accurately: A systematic review. *International Journal of Health Planning and Management*, 38(4), 904–917. <https://doi.org/10.1002/hpm.3629>



- Sivasundaralingam, D., Smith, N., & Lee, A. (2025). Predicting inpatient bed demand using machine learning. *Unpublished manuscript*. ResearchGate. <https://n9.cl/8hq2ba>
- Suresh, N. V., Selvakumar, A., Sridhar, G., & S., C. (2024). Operational Efficiency and Cost Reduction: The Role of AI in Healthcare Administration. In B. Singla, K. Shalender, & K. Stamer (Eds.), *Revolutionizing the Healthcare Sector with AI* (pp. 262-272). IGI Global Scientific Publishing. <https://doi.org/10.4018/979-8-3693-3731-8.ch013>
- Suryawanshi, V., Kanyal, D., Sabale, S., & Bhojar, V. (2024). The role of AI in enhancing hospital operational efficiency and patient care. *Multidisciplinary Reviews*, 8(5), 2025153. <https://doi.org/10.31893/multirev.2025153>
- Tuominen, J., Pulkkinen, E., Peltonen, J., Kanninen, J., Oksala, N., Palomäki, A., & Roine, A. (2023). Forecasting emergency department crowding with advanced machine learning models and multivariable input. *arXiv*. <https://doi.org/10.48550/arXiv.2308.16544>