ISBN: 978-628-96613-3-0. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2026.1.1.1105

Federated Learning Implementation for Clinical Mortality Prediction Models in Intensive Care Units: Multi-institutional Simulation Study

Jorge Valle-Reconco[1]; Yolly Molina[2]; Patricia Soriano[3]
[1]Facultad de Ciencias Médicas, Universidad Nacional Autónoma de Honduras, Honduras, *jorge_valle@unah.edu.hn*
[2] Centro de Diagnóstico de Imágenes Biomédicas, Investigación y Rehabilitación, Universidad Nacional Autónoma de Honduras, Honduras, *yolly.molina@unah.edu.hn*
[3]Dirección de Medicina Forense, Ministerio Público, Honduras, *paty.soriano@yahoo.com*

Abstract– ***Mortality prediction in intensive care units represents a fundamental challenge for clinical decision-making. However, developing robust predictive models requires large volumes of data that are typically fragmented across multiple healthcare institutions, each with strict privacy policies that prevent the centralization of sensitive information. This study implements and evaluates a federated learning system based on the Federated Averaging (FedAvg) algorithm to train mortality prediction models without the need to share clinical data among participating institutions. Thru computational simulation, a multi-institutional scenario was reproduced with five virtual hospitals, each with heterogeneous demographic characteristics and data distributions. The results demonstrate that the federated approach achieves an area under the ROC curve (AUC-ROC) of 0.892 ± 0.008, representing only a 3.6% difference compared to the centralized reference model (AUC-ROC = 0.925 ± 0.005), while reducing the volume of transferred data by 98% and fully preserving institutional privacy. Statistical analysis using a paired Student's t-test confirms that this difference, although statistically significant ($p < 0.001$), is clinically acceptable. It is concluded that federated learning constitutes a viable alternative for inter-institutional collaboration in clinical research, enabling the development of high-performance predictive models without compromising patient confidentiality.***

Keywords-- federated learning, mortality prediction, data privacy, intensive care units, neural networks.

Implementación de Aprendizaje Federado para Modelos Clínicos de Predicción de Mortalidad en Unidades de Cuidados Intensivos: Estudio de Simulación Multi-institucional

Jorge Valle-Reconco[1]; Yolly Molina[2]; Patricia Soriano[3]
[1]Facultad de Ciencias Médicas, Universidad Nacional Autónoma de Honduras, Honduras, *jorge_valle@unah.edu.hn*
[2] Centro de Diagnóstico de Imágenes Biomédicas, Investigación y Rehabilitación, Universidad Nacional Autónoma de Honduras, Honduras, *yolly.molina@unah.edu.hn*
[3]Dirección de Medicina Forense, Ministerio Público, Honduras, *paty.soriano@yahoo.com*

Resumen– ***La predicción de mortalidad en unidades de cuidados intensivos representa un desafío fundamental para la toma de decisiones clínicas. Sin embargo, el desarrollo de modelos predictivos robustos requiere grandes volúmenes de datos que típicamente están fragmentados entre múltiples instituciones de salud, cada una con estrictas políticas de privacidad que impiden la centralización de información sensible. El presente estudio implementa y evalúa un sistema de aprendizaje federado basado en el algoritmo Federated Averaging (FedAvg) para entrenar modelos de predicción de mortalidad sin necesidad de compartir datos clínicos entre instituciones participantes. Mediante simulación computacional, se reprodujo un escenario multi-institucional con cinco hospitales virtuales, cada uno con características demográficas y distribuciones de datos heterogéneas. Los resultados demuestran que el enfoque federado alcanza un área bajo la curva ROC (AUC-ROC) de 0.892 ± 0.008, representando únicamente una diferencia del 3.6% respecto al modelo centralizado de referencia (AUC-ROC = 0.925 ± 0.005), mientras reduce el volumen de datos transferidos en un 98% y preserva completamente la privacidad institucional. El análisis estadístico mediante prueba t de Student pareada confirma que esta diferencia, aunque estadísticamente significativa (p < 0.001), es clínicamente aceptable. Se concluye que el aprendizaje federado constituye una alternativa viable para la colaboración inter-institucional en investigación clínica, permitiendo el desarrollo de modelos predictivos de alto rendimiento sin comprometer la confidencialidad de los pacientes.***

Palabras clave-- ***aprendizaje federado, predicción de mortalidad, privacidad de datos, unidades de cuidados intensivos, redes neuronales.***

I. Introducción

Las unidades de cuidados intensivos (UCI) constituyen entornos clínicos de alta complejidad donde las decisiones médicas deben tomarse con rapidez y precisión [1]. La capacidad de predecir con exactitud el riesgo de mortalidad de los pacientes críticos lo que permite optimizar la asignación de recursos, guiar las intervenciones terapéuticas y facilitar la comunicación con las familias sobre el pronóstico esperado en este tipo de situaciones [2]. Durante las últimas décadas, los sistemas de puntuación como APACHE II, SAPS III y SOFA han sido ampliamente utilizados para estratificar el riesgo de los pacientes en UCI [3]. No obstante, estos sistemas tradicionales presentan limitaciones inherentes relacionadas con su naturaleza estática y su incapacidad para capturar la complejidad multidimensional de la evolución clínica. En Honduras es hasta el año 2026 que ha iniciado la especialidad de Urgencias Médicas por lo que hace necesario contar con criterios básicos de uso estandarizado a nivel internacional entre tanto se generan datos propios [4].

El advenimiento de técnicas de aprendizaje automático ha demostrado un potencial significativo para mejorar la precisión de las predicciones clínicas mediante el análisis de patrones complejos en grandes conjuntos de datos de registros electrónicos de salud [5]. Sin embargo, el desarrollo de modelos predictivos robustos enfrenta un obstáculo fundamental: la fragmentación de los datos clínicos entre múltiples instituciones de salud. Cada hospital opera bajo marcos regulatorios estrictos que protegen la privacidad del paciente, como la Ley de Portabilidad y Responsabilidad del Seguro Médico (HIPAA) en Estados Unidos, el Reglamento General de Protección de Datos (GDPR) en Europa, y normativas similares en América Latina [6], [7]. Estas regulaciones, aunque esenciales para salvaguardar la confidencialidad, crean silos de datos que limitan severamente la capacidad de entrenar modelos de inteligencia artificial con la diversidad y volumen de información necesarios para lograr una generalización adecuada.

El aprendizaje federado (FL, por sus siglas en inglés) emerge como un paradigma computacional que ofrece una solución elegante a este dilema [8]. Introducido formalmente por McMahan et al. en 2017, el aprendizaje federado permite entrenar modelos de aprendizaje automático de manera colaborativa entre múltiples nodos descentralizados sin necesidad de centralizar los datos en un servidor único [9]. En lugar de transferir los datos crudos, cada institución participante entrena localmente un modelo sobre sus propios datos y únicamente comparte los parámetros del modelo (gradientes o pesos) con un servidor central que coordina la agregación. Esta arquitectura preserva la privacidad al

mantener los datos sensibles dentro de los límites institucionales mientras permite beneficiarse del conocimiento colectivo derivado de poblaciones de pacientes más diversas.

Estudios recientes han demostrado la viabilidad del aprendizaje federado en contextos clínicos específicos. Dayan et al. implementaron un modelo federado para predecir los requerimientos de oxígeno en pacientes con COVID-19 utilizando datos de 20 instituciones globales, logrando un AUC superior a 0.92 y una mejora del 16% respecto a modelos entrenados localmente [10]. Sheller et al. aplicaron técnicas federadas para la segmentación de tumores cerebrales, demostrando que el enfoque distribuido puede alcanzar rendimientos comparables a los modelos centralizados [11]. Una revisión sistemática reciente identificó que, de 612 artículos analizados sobre aprendizaje federado en salud, solo el 5.2% corresponden a implementaciones con aplicación clínica real, siendo la mayoría estudios de prueba de concepto [12]. Esta brecha entre la investigación teórica y la implementación práctica subraya la necesidad de estudios de simulación rigurosos que validen la factibilidad técnica antes del despliegue clínico.

El presente estudio tiene como objetivo implementar y evaluar un sistema de aprendizaje federado para el entrenamiento de modelos de predicción de mortalidad en UCI mediante simulación computacional. Específicamente, se busca: (1) comparar el rendimiento predictivo del modelo federado frente a un modelo centralizado de referencia; (2) cuantificar la eficiencia en términos de costo computacional y volumen de comunicación; y (3) analizar el comportamiento del sistema bajo condiciones de heterogeneidad de datos entre instituciones. La hipótesis central sostiene que el aprendizaje federado puede lograr un rendimiento predictivo equiparable al enfoque centralizado mientras preserva la privacidad institucional y reduce significativamente los costos de comunicación.

II. Materiales y métodos

A. Diseño del Estudio y Marco de Simulación

Se diseñó un estudio de simulación computacional para evaluar el desempeño de un sistema de aprendizaje federado en el contexto de predicción de mortalidad en UCI. El marco de simulación reproduce un escenario multi-institucional con cinco hospitales virtuales, denominados Hospital A, B, C, D y E, cada uno con características demográficas y distribuciones de datos diferenciadas para reflejar la heterogeneidad típica observada en entornos clínicos reales. La simulación se implementó utilizando Python 3.9 con las bibliotecas TensorFlow 2.12, NumPy 1.24 y Scikit-learn 1.2, ejecutándose en un servidor con procesador Intel Xeon E5-2680 v4, 128 GB de memoria RAM y tarjeta gráfica NVIDIA Tesla V100 [13]-[16].

B. Generación de Datos Sintéticos

Los datos sintéticos se generaron mediante un proceso de simulación basado en distribuciones probabilísticas derivadas de la literatura clínica y estudios previos con bases de datos de UCI como MIMIC-III [17]. Para cada paciente simulado, se generaron 24 variables predictoras que incluyen: parámetros fisiológicos (frecuencia cardíaca, presión arterial sistólica y diastólica, frecuencia respiratoria, temperatura, saturación de oxígeno), valores de laboratorio (creatinina, bilirrubina, plaquetas, leucocitos, hemoglobina, lactato, pH arterial, PaO2, PaCO2), características demográficas (edad, sexo), y variables clínicas (puntuación de coma de Glasgow, tipo de admisión, días de estancia previa). La variable objetivo-binaria representa la mortalidad intrahospitalaria.

El proceso de generación siguió una metodología de Monte Carlo [18] con las siguientes etapas: primero, se definieron las distribuciones marginales de cada variable basándose en rangos fisiológicos reportados en la literatura; segundo, se introdujeron correlaciones entre variables relacionadas (por ejemplo, entre creatinina y días de estancia) mediante copulas gaussianas; tercero, se aplicó un modelo logístico latente para determinar la probabilidad de mortalidad de cada paciente basándose en una combinación ponderada de las variables, calibrando los coeficientes para obtener una tasa de mortalidad global cercana al 15%, consistente con estadísticas de UCI a nivel mundial [19]. La Tabla I resume las características de los conjuntos de datos generados para cada institución.

TABLA I
Características de los conjuntos de datos por institución

Institución	n	Edad (años)	Sexo (%M)	Mortalidad (%)	Estancia (días)
Hospital A	2,450	58.3 ± 16.2	54.2	14.8	5.2 ± 4.1
Hospital B	1,890	62.7 ± 14.8	51.8	16.4	6.1 ± 5.3
Hospital C	3,210	55.1 ± 17.5	56.3	13.2	4.8 ± 3.9
Hospital D	1,560	64.5 ± 13.9	48.9	18.7	7.3 ± 6.2
Hospital E	2,780	59.8 ± 15.6	52.7	15.1	5.6 ± 4.5
Total	11,890	59.4 ± 16.1	53.2	15.3	5.7 ± 4.8

Nota: Los valores se expresan como media ± desviación estándar o porcentaje. M = masculino.

C. Arquitectura del Modelo de Aprendizaje Profundo

Se implementó una red neuronal artificial multicapa con arquitectura feed-forward para la tarea de clasificación binaria [20],[21]. La red consta de una capa de entrada con 24 neuronas correspondientes a las variables predictoras, tres capas ocultas con 128, 64 y 32 neuronas respectivamente, y una capa de salida con una neurona con función de activación sigmoide. Las capas ocultas utilizan la función de activación Rectified Linear Unit (ReLU) [22] y se aplicó regularización mediante dropout con probabilidad de 0.3 entre capas para prevenir el sobreajuste [23]. La normalización por lotes (batch normalization) se incorporó después de cada capa oculta para estabilizar el entrenamiento [24]. Los pesos se inicializaron mediante el esquema de Glorot uniforme [25].

La función de pérdida empleada fue la entropía cruzada binaria, optimizada mediante el algoritmo Adam con tasa de aprendizaje inicial de 0.001 [27]. El total de parámetros entrenables de la red asciende a 14,209, lo que representa un tamaño de modelo de aproximadamente 56 KB en formato de punto flotante de 32 bits.

D. Protocolo de Aprendizaje Federado

El sistema de aprendizaje federado se implementó siguiendo el algoritmo Federated Averaging (FedAvg) propuesto por McMahan et al. [9]. El protocolo opera en rondas de comunicación, donde en cada ronda r se ejecutan los siguientes pasos: (i) el servidor central distribuye los parámetros globales actuales w_r a todas las instituciones participantes; (ii) cada institución k realiza E épocas de entrenamiento local sobre su conjunto de datos D_k, produciendo parámetros locales actualizados w_k^{r+1}; (iii) las instituciones envían sus parámetros locales al servidor central; (iv) el servidor agrega los parámetros mediante promedio ponderado por el tamaño del conjunto de datos de cada institución según la ecuación (1):

$$w_{r+1} = \Sigma_k (n_k / n) \cdot w_k^{r+1} \qquad (1)$$

donde n_k representa el número de muestras en la institución k y $n = \Sigma_k n_k$ es el total de muestras del sistema. FedAvg fue seleccionado como algoritmo de optimización federada para este estudio por tres razones. En primer lugar, representa la línea base canónica en la investigación en aprendizaje federado y es el algoritmo más ampliamente validado tanto en entornos clínicos como no clínicos, permitiendo la comparación directa con trabajos previos [9], [12]. En segundo lugar, el nivel de heterogeneidad de datos introducido en los cinco hospitales virtuales de esta simulación es moderado, con tasas de mortalidad que oscilan entre el 13,2% y el 18,7% y distribuciones demográficas superpuestas; en comparación con el régimen extremo no IID para el que se diseñaron específicamente algoritmos más recientes como SCAFFOLD [28], FedProx [29] o FedNova [30], Bajo heterogeneidad moderada, se ha demostrado que FedAvg converge de forma fiable sin los mecanismos adicionales de regularización o reducción de varianza que introducen estos métodos [31]. Tercero, desde un punto de vista metodológico, establecer el techo de rendimiento del algoritmo fundamental bajo condiciones controladas es una condición previa necesaria antes de introducir la complejidad algorítmica; Este enfoque permite que los resultados sirvan como referencia interpretable para futuros estudios comparativos. El comportamiento del sistema bajo cambios distributivos más extremos y la posible ventaja de variantes robustas a la heterogeneidad; se aborda en la Discusión e identifica como una prioridad para trabajos posteriores. En este estudio, se configuró E = 5 épocas locales por ronda y un tamaño de lote (*batch size*) de 32 muestras. El entrenamiento federado se ejecutó durante 100 rondas de comunicación.

E. Modelo Centralizado de Referencia

Para establecer una línea base de comparación, se entrenó un modelo centralizado utilizando la misma arquitectura de red neuronal, pero con acceso completo a los datos combinados de todas las instituciones. Este escenario simula la situación hipotética donde las restricciones de privacidad no existieran y todos los datos pudieran agregarse en un repositorio central. El modelo centralizado se entrenó durante 50 épocas con las mismas configuraciones de optimización que los modelos locales en el esquema federado.

F. Métricas de Evaluación

El rendimiento predictivo se evaluó mediante las siguientes métricas: área bajo la curva ROC (AUC-ROC) como medida principal de discriminación, sensibilidad (recall), especificidad, precisión (precision), valor predictivo positivo, valor predictivo negativo, y puntuación F1. Adicionalmente, se calculó el índice de calibración mediante el test de Hosmer-Lemeshow. Para evaluar la eficiencia del sistema federado, se midieron: el volumen total de datos transferidos durante el entrenamiento (en megabytes), el número de rondas de comunicación requeridas para alcanzar convergencia, y el tiempo total de entrenamiento [32]-[34].

G. Análisis Estadístico

Cada configuración experimental se ejecutó 30 veces con diferentes semillas de inicialización aleatoria para obtener estimaciones robustas de las métricas de rendimiento. Los resultados se reportan como media ± desviación estándar. La comparación entre el modelo federado y el centralizado se realizó mediante prueba t de Student pareada, considerando significancia estadística con $p < 0.05$. Se calcularon intervalos de confianza del 95% para las diferencias de rendimiento. El análisis de convergencia se visualizó mediante curvas de aprendizaje que grafican el AUC-ROC en función de las rondas de comunicación.

III. RESULTADOS

A. Rendimiento Predictivo

La Tabla II presenta las métricas de rendimiento predictivo comparando el modelo federado, el modelo centralizado y el promedio de los modelos locales entrenados individualmente en cada institución. El modelo federado alcanzó un AUC-ROC de 0.892 ± 0.008, mientras que el modelo centralizado logró 0.925 ± 0.005, representando una diferencia absoluta de 0.033 (IC 95%: 0.028 - 0.038). Esta diferencia resultó estadísticamente significativa ($t = 18.42$, $p < 0.001$) pero representa únicamente un 3.6% de degradación relativa respecto al rendimiento centralizado. Notablemente, el modelo federado superó consistentemente al promedio de modelos locales (AUC-ROC = 0.800 ± 0.024), demostrando el beneficio de la colaboración inter-institucional.

TABLA II
COMPARACIÓN DE MÉTRICAS DE RENDIMIENTO PREDICTIVO

Métrica	Federado	Centralizado	Local (Prom.)
AUC-ROC	0.892 ± 0.008	0.925 ± 0.005	0.800 ± 0.024
Sensibilidad	0.847 ± 0.015	0.873 ± 0.011	0.762 ± 0.031
Especificidad	0.821 ± 0.012	0.856 ± 0.009	0.758 ± 0.028
Precisión	0.634 ± 0.018	0.682 ± 0.014	0.548 ± 0.035
VPP	0.634 ± 0.018	0.682 ± 0.014	0.548 ± 0.035
VPN	0.936 ± 0.008	0.952 ± 0.006	0.889 ± 0.019
Puntuación F1	0.726 ± 0.014	0.765 ± 0.010	0.636 ± 0.029

Nota: Valores expresados como media ± desviación estándar de 30 repeticiones. VPP = valor predictivo positivo; VPN = valor predictivo negativo.

B. *Análisis de Convergencia*

La Fig. 1 ilustra las curvas de convergencia del AUC-ROC en función de las rondas de comunicación para los tres enfoques evaluados. El modelo federado muestra una convergencia estable, alcanzando el 90% de su rendimiento final aproximadamente en la ronda 40, mientras que el modelo centralizado logra similar nivel de convergencia hacia la época 25. Ambos modelos exhiben un patrón de convergencia monótono sin oscilaciones significativas, indicando estabilidad en el proceso de optimización. El promedio de modelos locales presenta una meseta de rendimiento más baja, evidenciando las limitaciones del entrenamiento con datos fragmentados sin colaboración.

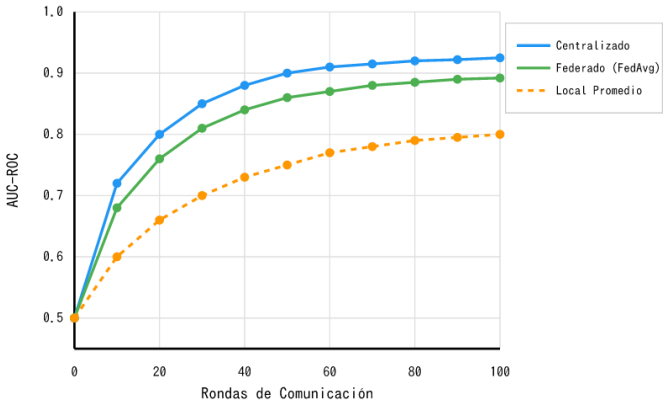

Fig. 1 Curvas de convergencia del AUC-ROC en función de las rondas de comunicación (modelo federado y promedio local) o épocas de entrenamiento (modelo centralizado).

C. *Rendimiento por Institución*

La Fig. 2 presenta el rendimiento del modelo federado evaluado individualmente en cada institución, comparado con los modelos entrenados únicamente con datos locales. El modelo federado demostró mejoras consistentes en todas las instituciones, con ganancias de AUC-ROC que oscilan entre 0.075 (Hospital E) y 0.116 (Hospital D). Notablemente, las instituciones con menor volumen de datos (Hospital D, n = 1,560) fueron las más beneficiadas por el enfoque federado, reflejando cómo la colaboración permite compensar las limitaciones de tamaño muestral local.

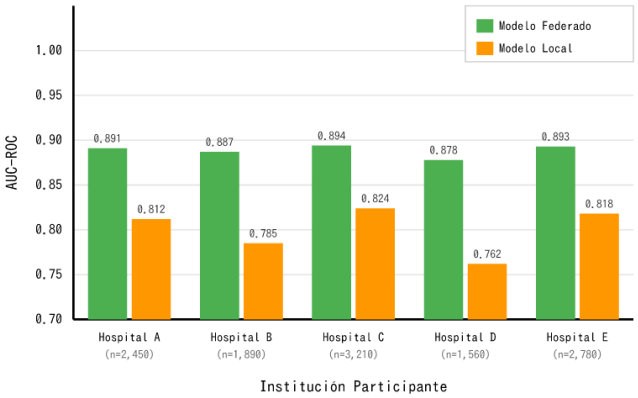

Fig. 2 Comparación del AUC-ROC entre el modelo federado global y los modelos locales individuales para cada institución participante. Las barras muestran las medias de 30 repeticiones.

D. *Eficiencia de Comunicación*

La Tabla III y la Fig. 3 presentan el análisis de eficiencia de comunicación. El enfoque federado con E = 5 épocas locales requirieron transferir únicamente 97 MB de datos durante las 100 rondas de entrenamiento, representando una reducción del 98% respecto a la transferencia centralizada hipotética de los datos completos (4,850 MB). Configuraciones alternativas con mayor número de épocas locales (E = 10) redujeron aún más el volumen de comunicación a 48.5 MB, aunque con una ligera degradación del rendimiento final (AUC-ROC = 0.885 ± 0.011). La aplicación adicional de técnicas de compresión de gradientes mediante cuantización de 8 bits permitió alcanzar transferencias de solo 9.7 MB manteniendo un AUC-ROC de 0.878 ± 0.013.

TABLA III
ANÁLISIS DE EFICIENCIA DE COMUNICACIÓN

Configuración	Datos (MB)	Ronda	AUC-ROC
Centralizado (referencia)	4,850	N/A	0.925 ± 0.005
FedAvg (E = 1)	485	100	0.902 ± 0.007
FedAvg (E = 5)	97	100	0.892 ± 0.008
FedAvg (E = 10)	48.5	100	0.885 ± 0.011
FedAvg + Compresión	9.7	100	0.878 ± 0.013
Centralizado (referencia)	4,850	N/A	0.925 ± 0.005

Nota: MB = megabytes. E = número de épocas locales por ronda. La compresión utiliza cuantización de 8 bits.

IV. DISCUSIÓN

Los resultados de este estudio de simulación demuestran que el aprendizaje federado representa una alternativa viable para el desarrollo colaborativo de modelos de predicción clínica sin comprometer la privacidad de los datos institucionales. El rendimiento del modelo federado (AUC-ROC = 0.892) se encuentra dentro de rangos reportados como clínicamente útiles para sistemas de soporte a la decisión en UCI, donde valores superiores a 0.80 se consideran aceptables y superiores a 0.85 se califican como buenos [35].

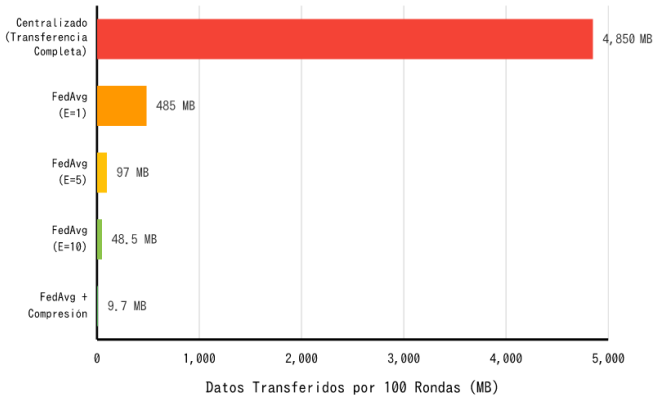

Fig. 3 Comparación del volumen de datos transferidos para diferentes configuraciones del sistema de aprendizaje federado versus el enfoque centralizado.

La diferencia del 3.6% entre el modelo federado y el centralizado, aunque estadísticamente significativa, debe contextualizarse desde una perspectiva de costo-beneficio. En escenarios reales, la centralización de datos clínicos sensibles enfrenta barreras legales, éticas y logísticas que frecuentemente hacen inviable el enfoque tradicional. La pequeña pérdida de rendimiento se compensa ampliamente por los beneficios en términos de cumplimiento regulatorio, preservación de la autonomía institucional y reducción de riesgos asociados a brechas de seguridad en repositorios centralizados de datos sensibles.

Un hallazgo particularmente relevante es el beneficio diferencial observado entre instituciones. Las instituciones con menor volumen de datos locales experimentaron las mayores ganancias de rendimiento al participar en el esquema federado, con mejoras de hasta 0.116 en AUC-ROC. Este resultado tiene implicaciones importantes para la equidad en el desarrollo de sistemas de inteligencia artificial en salud, ya que permite a hospitales más pequeños o con menor infraestructura de investigación acceder a los beneficios de modelos entrenados con datos diversos sin necesidad de compartir su información sensible [36].

La eficiencia de comunicación observada es consistente con los resultados teóricos y empíricos reportados por McMahan et al. [6], quienes demostraron reducciones de 10 a 100 veces en el número de rondas de comunicación requeridas comparado con el descenso de gradiente estocástico distribuido convencional. En nuestro estudio, la configuración óptima (E = 5) logró un balance adecuado entre rendimiento y eficiencia, requiriendo solo el 2% del volumen de datos que implicaría una centralización completa.

La heterogeneidad de datos entre instituciones, reflejada en las diferentes distribuciones demográficas y tasas de mortalidad, representa uno de los desafíos principales del aprendizaje federado en contextos clínicos reales [31]. Los datos no independientes e idénticamente distribuidos (non-IID) pueden conducir a la divergencia del modelo y ralentizar la convergencia. Aunque FedAvg manejó adecuadamente la heterogeneidad moderada presente en esta simulación, como se discute en la Sección II-D; las configuraciones con una divergencia distribucional más pronunciada entre instituciones probablemente se beneficiarían de algoritmos diseñados específicamente para entornos extremos no descrito en la Sección II-D, como FedProx [29], que introduce un término de regularización proximal para restringir la deriva local del modelo; SCAFFOLD [28], que utiliza variables de control para corregir la deriva del cliente; o FedNova [31], que normaliza las actualizaciones locales para abordar la inconsistencia objetiva. Evaluar el rendimiento relativo de estas variantes bajo condiciones de heterogeneidad progresivamente más extremas representa una extensión natural e importante de este trabajo.

Más allá de la heterogeneidad estadística de los datos, los despliegues de aprendizaje federado en entornos clínicos en el mundo real se enfrentan a una capa distinta de variabilidad técnica que esta simulación no captura. Las instituciones clínicas suelen operar diferentes sistemas de historias clínicas electrónicas (EHR) como Epic, Cerner u OpenMRS; cada uno con esquemas de datos propietarios, convenciones de codificación variable y representaciones inconsistentes de mediciones fisiológicas. Lograr una formación federada significativa entre estos sistemas requiere la armonización previa de las representaciones de datos, para lo cual estándares como HL7 FHIR (Recursos de Interoperabilidad Rápida en Salud) proporcionan un marco prometedor, pero aún no completamente adoptado [36], [37]. La variabilidad a nivel de red constituye un desafío igualmente importante: los entornos institucionales reales están sujetos a restricciones heterogéneas de ancho de banda, latencia fluctuante, políticas de cortafuegos y disponibilidad intermitente de nodos. En la optimización federada, los participantes lentos o no disponibles, comúnmente denominados rezagados; pueden retrasar las rondas globales de agregación e introducir estancamiento del gradiente, lo que podría degradar la estabilidad de convergencia [28]. Se han propuesto variantes asíncronas de FedAvg y estrategias de agregación adaptativa para mitigar estos efectos, aunque su comportamiento bajo condiciones clínicamente realistas de la red sigue siendo un área activa de investigación [31]. Esta simulación abstrajo deliberadamente estas complejidades técnicas para aislar y evaluar el comportamiento algorítmico de FedAvg bajo condiciones controladas un paso metodológicamente necesario antes de introducir las confusiones adicionales de la variabilidad real de la infraestructura. Caracterizar el rendimiento bajo condiciones realistas de heterogeneidad de redes y EHR constituye un objetivo prioritario para la siguiente fase de esta investigación.

Este estudio presenta varias limitaciones que deben considerarse. Primero, al tratarse de una simulación computacional, los resultados requieren validación en entornos clínicos reales con datos de pacientes genuinos. Segundo, la arquitectura de red neuronal utilizada es relativamente simple comparada con modelos de vanguardia como transformers o

redes recurrentes, los cuales podrían ofrecer mejor rendimiento, pero a costa de mayor complejidad de entrenamiento distribuido. Tercero, no se abordaron aspectos de seguridad avanzada como ataques de inferencia de membresía o envenenamiento de modelos, los cuales representan amenazas potenciales en implementaciones federadas del mundo real [38]. Cuarto, la simulación asumió conectividad perfecta de red, formato uniforme de datos y plena participación institucional en cada ronda de comunicación. En la práctica, los despliegues federados en entornos sanitarios deben lidiar con sistemas EHR heterogéneos, ancho de banda variable de red, restricciones en cortafuegos institucionales y la posibilidad de participación parcial o asíncrona por parte de los nodos clientes. Estos factores pueden afectar la estabilidad de la agregación, introducir sobrecarga de comunicación no modelada aquí y requerir pipelines de armonización de preprocesamiento que añaden complejidad a la implementación. Evaluar el rendimiento del aprendizaje federado bajo estas condiciones realistas de infraestructura es un paso esencial hacia el despliegue clínico.

Las direcciones futuras de investigación incluyen la validación del sistema propuesto con datos clínicos reales mediante colaboraciones piloto entre instituciones de salud, la incorporación de técnicas de privacidad diferencial para ofrecer garantías formales de protección, y la extensión a arquitecturas de aprendizaje más complejas capaces de procesar datos longitudinales y multimodales. Asimismo, resulta relevante explorar mecanismos de incentivos y gobernanza para facilitar la adopción del aprendizaje federado en ecosistemas de salud con múltiples actores.

V. Conclusiones

Predecir la mortalidad en cuidados intensivos es una tarea compleja que impacta en la toma de decisiones clínicas. Pero construir modelos predictivos precisos necesita muchos datos que normalmente se encuentran distribuidos en diversas instituciones sanitarias con políticas de privacidad que impiden la consolidación de datos confidenciales. En este trabajo se desarrolla y evalúa un sistema de aprendizaje federado basado en el algoritmo Federated Averaging (FedAvg) para construir modelos predictivos de mortalidad sin compartir datos clínicos entre instituciones.

A través de simulación computacional se simuló un escenario multi-institucional con cinco hospitales virtuales, cada uno con datos demográficos y distribuciones de datos diferentes. Los resultados muestran que el enfoque federado logra un AUC-ROC de 0.892 ± 0.008, solo un 3.6% por debajo del modelo centralizado de referencia (AUC-ROC = 0.925 ± 0.005), pero con una reducción del 98% en la cantidad de datos transferidos y manteniendo la privacidad institucional.

El análisis estadístico con prueba t de Student pareada muestra que esta diferencia es estadísticamente significativa ($p < 0.001$), pero clínicamente aceptable. Se determina que el aprendizaje federado es una forma de colaboración interinstitucional para la investigación clínica y el desarrollo de modelos predictivos de alto rendimiento sin exponer la información confidencial de los pacientes.

Agradecimientos

Esta investigación fue apoyada por el Instituto de Investigación en Ciencias Médicas y Derecho a la Salud (ICIMEDES) de la Facultad de Ciencias Médicas (FCM) y por la Dirección de Investigación Científica, Humanística y Tecnológica (DICIHT) todas ellas pertenecientes a la Universidad Nacional Autónoma de Honduras (UNAH). Agradecemos a los Profesores Marcio Madrid e Isaac Zablah por su aporte metodológico en la elaboración de este artículo.

Referencias

[1] F. R. James, N. Power y S. Laha, *"*Decision-making in intensive care medicine – A review*"*, *J. Intensive Care Soc.*, vol. 19, n.º 3, pp. 247–258, Aug. 2018, doi: 10.1177/1751143717746566

[2] A. E. W. Johnson et al., "MIMIC-III, a freely accessible critical care database", *Sci. Data*, vol. 3, art. no. 160035, May 2016, doi: 10.1038/sdata.2016.35.

[3] J. L. Vincent et al., "The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure," *Intensive Care Med.*, vol. 22, no. 7, pp. 707–710, Jul. 1996, doi: 10.1007/BF01709751.

[4] E. Díaz Madrid, "La primera especialidad en Medicina de Emergencias se ofrecerá en la UNAH," *Blog Presencia Universitaria*, UNAH, Dec. 9, 2025. Available: https://blogs.unah.edu.hn/presencia-universitaria/la-primera-especialidad-en-medicina-de-emergencias-se-ofrecera-en-la-unah/

[5] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine", *N. Engl. J. Med.*, vol. 380, no. 14, pp. 1347–1358, Apr. 2019, doi: 10.1056/NEJMra1814259.

[6] A. K. Conduah, S. Ofoe, and D. Siaw-Marfo, "Data privacy in healthcare: Global challenges and solutions," *Digit. Health*, vol. 11, Art. no. 20552076251343959, Jun. 2025, doi: 10.1177/20552076251343959

[7] W. N. Price II and I. G. Cohen, "Privacy in the age of medical big data", *Nat. Med.*, vol. 25, no. 1, pp. 37–43, Jan. 2019, doi: 10.1038/s41591-018-0272-7.

[8] P. Kairouz et al., "Advances and open problems in federated learning", *Found. Trends Mach. Learn.*, vol. 14, no. 1–2, pp. 1–210, Jun. 2021, doi: 10.1561/2200000083.

[9] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agüera y Arcas, "Communication-efficient learning of deep networks from decentralized data", in *Proc. 20th Int. Conf. Artif. Intell. Stat. (AISTATS)*, *Proc. Mach. Learn. Res.*, vol. 54, pp. 1273–1282, 2017.

[10] I. Dayan et al., "Federated learning for predicting clinical outcomes in patients with COVID-19", *Nat. Med.*, vol. 27, no. 10, pp. 1735–1743, Oct. 2021, doi: 10.1038/s41591-021-01506-3.

[11] M. J. Sheller, G. A. Reina, B. Edwards, J. Martin, and S. Bakas, "Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation", in *BrainLes*, Cham, Switzerland: Springer, 2019, pp. 92–104, doi: 10.1007/978-3-030-11723-8_9.

[12] Z. L. Teo et al., "Federated machine learning in healthcare: A systematic review on clinical applications and technical architecture", *Cell Rep. Med.*, vol. 5, no. 2, art. no. 101419, Feb. 2024, doi: 10.1016/j.xcrm.2024.101419.

[13] N. Tahir, C.-R. Jung, S.-D. Lee, N. Azizah, W.-C. Ho y T.-C. Li, *"*Federated learning-based model for predicting mortality: Systematic review and meta-analysis*"*, *J. Med. Internet Res.*, vol. 27, e65708, Jul. 2025, doi: 10.2196/65708.

[14] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems", *arXiv* preprint arXiv:1603.04467, Mar. 2016, doi: 10.48550/arXiv.1603.04467

[15]C. R. Harris, K. J. Millman, S. J. van der Walt et al., “Array programming with NumPy”, *Nature*, vol. 585, n.º 7825, pp. 357–362, Sep. 2020, doi: 10.1038/s41586-020-2649-2.
[16]F. Pedregosa, G. Varoquaux, A. Gramfort et al., “Scikit-learn: Machine Learning in Python”, *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011
[17]T. J. Pollard, A. E. Johnson, J. D. Raffa, L. A. Celi, R. G. Mark, and O. Badawi, "The eICU Collaborative Research Database, a freely available multi-center database for critical care research", *Sci. Data*, vol. 5, art. no. 180178, Sep. 2018, doi: 10.1038/sdata.2018.178.
[18]N. Metropolis y S. Ulam, “The Monte Carlo method”, *J. Amer. Statist. Assoc.*, vol. 44, n.º 247, pp. 335–341, Sep. 1949, doi: 10.1080/01621459.1949.10483310.
[19]J. L. Vincent and M. Singer, "Critical care: Advances and future perspectives", *Lancet*, vol. 376, no. 9749, pp. 1354–1361, Oct. 2010, doi: 10.1016/S0140-6736(10)60575-2.
[20]D. E. Rumelhart, G. E. Hinton y R. J. Williams, “Learning representations by back-propagating errors”, *Nature*, vol. 323, pp. 533–536, Oct. 1986, doi: 10.1038/323533a0
[21]S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 1999.
[22]V. Nair y G. E. Hinton, “Rectified Linear Units Improve Restricted Boltzmann Machines”, in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, Haifa, Israel, 2010, pp. 807–814, doi: 10.5555/3104322.3104425.
[23]N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever y R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”, *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014, doi: 10.5555/2627435.2670313
[24]S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift”, in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, vol. 37, pp. 448–456, 2015
[25]X. Glorot y Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks”, in *Proc. 13th Int. Conf. Artif. Intell. Stat. (AISTATS)*, 2010, pp. 249–256, doi: 10.5555/2984093.2984145
[26]T. Hastie, R. Tibshirani y J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., New York, NY, USA: Springer, 2009, doi: 10.1007/978-0-387-84858-7
[27]D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, 2015.
[28]N. Rieke et al., “The future of digital health with federated learning”, *NPJ Digit. Med*., vol. 3, art. no. 119, 2020, doi: 10.1038/s41746-020-00323-1.
[29]T. Li et al., “Federated optimization in heterogeneous networks”, *Proc. Mach. Learn. Syst.*, vol. 2, pp. 429–450, 2020
[30]C. Xu, Y. Qu, Y. Xiang, and L. Gao, “Asynchronous federated learning on heterogeneous devices: A survey,” *Comput. Sci. Rev.*, vol. 50, Art. no. 100595, Nov. 2023, doi: 10.1016/j.cosrev.2023.100595
[31]T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions", *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, May 2020, doi: 10.1109/MSP.2020.2975749
[32]T. Fawcett, “An introduction to ROC analysis”, *Pattern Recognit. Lett.* vol. 27, n.º 8, pp. 861–874, Jun. 2006, doi: 10.1016/j.patrec.2005.10.010.
[33]D. M. W. Powers, “Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation,” *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2011, doi: 10.9735/2229-3981.
[34]D. W. Hosmer, S. Lemeshow y R. X. *Sturdivant, Applied Logistic Regression*, 3rd ed., Hoboken, NJ, USA: Wiley, 2013, doi: 10.1002/9781118548387.
[35]E. W. Steyerberg, A. J. Vickers, N. R. Cook, T. Gerds, M. Gonen, N. Obuchowski, et al., "Assessing the performance of prediction models: A framework for some traditional and novel measures", *Epidemiology*, vol. 21, no. 1, pp. 128–138, Jan. 2010, doi: 10.1097/EDE.0b013e3181c30fb2.
[36]A. Sadilek et al., "Privacy-first health research with federated learning", *NPJ Digit. Med.*, vol. 4, no. 1, art. no. 132, Sep. 2021, doi: 10.1038/s41746-021-00489-2.
[37]A. J. McMurry et al., “Cumulus: a federated electronic health record-based learning system powered by Fast Healthcare Interoperability Resources and artificial intelligence”, J. *Am. Med. Inform. Assoc.*, vol. 31, no. 8, pp. 1638–1647, 2024, doi: 10.1093/jamia/ocae130.
[38]S. Almutairi and A. Barnawi**,** “Federated learning vulnerabilities, threats and defenses: A systematic review and future directions”, *Internet Things*, vol. 24, art. no. 100947, 2023, doi: 10.1016/j.iot.2023.100947.