OPEN ACCESS

Artículo de investigación

Revista Formación Estratégica
Vol. 7 Núm. 1 (2026), Pág.101-115
https://formacionestrategica.com/index.php/foes/issue/view/18
ISSN: 2805 – 9832
Recibido Abril 2026
Aceptado Junio 2026
Publicado: 30 Junio 2026

Anonimización Inteligente de Datos Sensibles

Intelligent Anonymization of Sensitive Data

Ana María García Arias
Agarcia53@unisalle.edu.co
https://orcid.org/0009-0005-1666-7007
Universidad de La Salle, Colombia

Diana Carolina González Díaz
dgonzalez56@unisalle.edu.co
https://orcid.org/0009-0000-3831-2075
Universidad de La Salle, Colombia

Héctor Alfredo González Cifuentes
hgonzalez95@unisalle.edu.co
https://orcid.org/0009-0007-5023-7435
Universidad de La Salle, Colombia

Resumen

La anonimización de datos clínicos requiere equilibrar la protección de la privacidad con la preservación de la utilidad estadística. Este estudio compara tres enfoques híbridos basados en aprendizaje automático —un modelo autoencoder–GAN, un CTGAN con privacidad diferencial y un esquema simulado de aprendizaje federado con privacidad diferencial— aplicados al dataset clínico eICU-CRD. Los modelos fueron evaluados mediante métricas de similitud estructural, preservación de correlaciones y riesgo de reidentificación. Los resultados muestran que el enfoque autoencoder–GAN ofrece la mayor fidelidad estadística, mientras que CTGAN con privacidad diferencial alcanza el mejor balance entre utilidad y protección formal. El enfoque federado, aunque menos preciso, proporciona mayores garantías en escenarios distribuidos. Se desarrolló además un prototipo funcional basado en CTGAN + DP que genera datos sintéticos plausibles y cuantifica su desviación respecto al conjunto real. Los hallazgos evidencian que los modelos generativos con mecanismos formales de privacidad constituyen alternativas viables para la anonimización en salud digital, siempre que se integren en marcos robustos de gobernanza y ética del dato.

Palabras clave: anonimización de datos; generación de datos sintéticos; privacidad diferencial; aprendizaje federado; redes generativas adversarias (GAN); CTGAN; inteligencia artificial en salud.

Abstract

Data anonymization in clinical environments requires balancing privacy protection with the preservation of analytical utility. This study compares three hybrid machine learning–based approaches—an autoencoder–

GAN model, a CTGAN combined with differential privacy, and a simulated federated-learning scheme with differential privacy—applied to the eICU-CRD clinical dataset. The models were evaluated using structural similarity metrics, correlation preservation, and reidentification risk. Results show that the autoencoder–GAN model provides the highest statistical fidelity, while CTGAN with differential privacy achieves the best trade-off between utility and formal privacy guarantees. The federated approach, although less precise, offers strategic advantages in multi-institutional settings. A functional prototype based on CTGAN + DP was developed and used to generate synthetic tables with quantified deviation from the real data. Findings indicate that generative models with formal privacy mechanisms constitute viable solutions for data anonymization in digital health, provided they are integrated into robust governance and ethical frameworks.

Keywords: data anonymization; synthetic data generation; differential privacy; federated learning; generative adversarial networks (GANs); CTGAN; artificial intelligence in healthcare.

1. Introducción

La anonimización de datos clínicos constituye un desafío central para la analítica en salud, dado que requiere proteger la identidad de los pacientes sin comprometer la utilidad de la información para modelos predictivos, análisis estadísticos o investigación biomédica. Las técnicas clásicas de anonimización —como la supresión, la generalización y el k-anonimato— han sido ampliamente estudiadas como mecanismos para reducir el riesgo de identificación directa en conjuntos de datos estructurados A. (Majeed and Hwang, 2020). En el ámbito sanitario, estas estrategias han sido revisadas específicamente por su capacidad para proteger datos sensibles sin destruir completamente su valor analítico (Olatunji et al., 2021). Además, se han propuesto marcos de apoyo a la decisión para seleccionar técnicas de anonimización en procesos de aprendizaje automático (Caruccio et al., 2022) así como métodos orientados a mejorar la utilidad de los datos anonimizados en escenarios longitudinales (Amiri et al.,2025). Entre las extensiones recientes de los modelos clásicos también se encuentra el esquema (kⓜ,mⓜ,t)-anonymity, propuesto para reforzar la protección en datos transaccionales (Puri et al., 2023).

No obstante, diversos estudios han señalado que estos enfoques presentan limitaciones cuando se aplican a datos de alta dimensionalidad o a registros clínicos complejos, donde la correlación entre variables puede facilitar procesos de reidentificación indirecta, (Olatunji et al., 2021), (Amiri et al.,2025). En respuesta a estas limitaciones, en los últimos años han surgido enfoques basados en inteligencia artificial orientados a la generación de datos sintéticos con preservación de privacidad, adicionalmente, en entornos clínicos multibase se presentan retos operativos de interoperabilidad, trazabilidad y escalabilidad que condicionan la adopción de mecanismos de anonimización en escenarios reales. Entre ellos destacan las redes generativas adversariales, los autoencoders y otros modelos generativos capaces de capturar patrones estadísticos complejos presentes en datos clínicos estructurados (Wang et al., 2021). En particular, se ha reportado el uso de GAN condicionales con privacidad diferencial para generar datos personales de salud sintéticos (Sun et al., 2023), mientras que los autoencoders también han sido explorados como mecanismos ajustables de preservación de privacidad (Jamshidi et al., 2023). En el caso específico de datos tabulares, el modelo CTGAN se ha consolidado como una de las arquitecturas más utilizadas para modelar distribuciones mixtas de variables categóricas y continuas (Skoularidou et al., 2023). Una revisión sistemática reciente confirma que la generación de datos sintéticos se ha convertido en una de las estrategias más prometedoras para equilibrar utilidad analítica y privacidad en el ámbito de la salud (Hyrup et al., 2025).

De manera complementaria, el aprendizaje federado ha emergido como una alternativa para entrenar modelos de inteligencia artificial sin necesidad de centralizar los datos en un único repositorio. Este paradigma permite que múltiples instituciones colaboren en el entrenamiento de modelos compartiendo únicamente parámetros o gradientes, manteniendo los datos clínicos en su Nguyen et al., 2025). En salud, ya se han documentado aplicaciones reales de aprendizaje federado con mecanismos de preservación de privacidad sobre datos FAIR (Sinaci et al., 2024), así como revisiones centradas en los desafíos específicos de privacidad en entornos

clínicos distribuidos (Pati et al., 2024). Sin embargo, incluso en estos escenarios persisten riesgos de inferencia, reconstrucción o fuga de información, por lo que se requieren mecanismos complementarios de protección.

La literatura reciente también ha mostrado que el problema de la anonimización trasciende los datos tabulares convencionales y alcanza otros tipos de datos clínicos y biomédicos. Por ejemplo, se han documentado riesgos actuales de reidentificación en imágenes médicas anonimizadas (Steeg et al., 2024), así como comparaciones entre modelos locales para anonimizar reportes radiológicos sin exponer la información al entorno de nube (McIntosh et al., 2025). De forma adicional, se han propuesto transformaciones del espacio de datos compatibles con aprendizaje profundo preservando privacidad (Girka et al., 2021), enfoques que integran anonimización y entrenamiento adversarial (Tian et al., 2021), y métodos orientados a preservar utilidad en flujos de datos dinámicos (Sopaoglu et al., 2021). También se han desarrollado modelos profundos con enmascaramiento dinámico y permutación aleatoria para reforzar la privacidad durante el aprendizaje (Zhang et al., 2025).

En términos metodológicos, el presente estudio se apoya en lineamientos consolidados para revisiones sistemáticas de literatura (Kitchenham et al., 2004) y en la declaración PRISMA 2020 para la síntesis transparente de evidencia científica (Page et al., 2004). En este contexto, el artículo examina el uso de técnicas de aprendizaje automático para la anonimización de datos estructurados en entornos de análisis clínico. Para ello, se realiza una evaluación comparativa de tres enfoques híbridos de generación y anonimización de datos: un modelo basado en autoencoders y redes generativas adversariales, un generador tabular condicional con privacidad diferencial y un esquema de aprendizaje federado con privacidad diferencial.

En particular, el enfoque basado en autoencoders se operacionaliza en el modelo AE-GAN, mientras que el prototipo funcional implementado y evaluado en este estudio se fundamenta en CTGAN + DP por su idoneidad para datos tabulares clínicos.

Con el fin de analizar la viabilidad de estos enfoques en contextos de analítica clínica, se desarrolló y evaluó un prototipo funcional basado en CTGAN con privacidad diferencial, aplicado al conjunto de datos clínicos eICU Collaborative Research Database. Este recurso se apoya en la infraestructura abierta de PhysioBank, PhysioToolkit y PhysioNet (Goldberger et al., 2000), y está disponible tanto en su versión demo (Johnson et al., 2021) como en su publicación científica de referencia para investigación multicéntrica en cuidados intensivos (Pollard et al., 2018). La implementación del prototipo se realizó utilizando la biblioteca Synthetic Data Vault (SDV), ampliamente empleada para generación de datos sintéticos tabulares (Developers, 2024). El pipeline del prototipo asume una etapa previa de extracción y estandarización desde fuentes relacionales hacia una representación tabular común, lo que permite su extensibilidad a distintos motores de bases de datos en escenarios multibase.

El prototipo permite generar datos sintéticos plausibles y evaluar su comportamiento respecto al conjunto real mediante métricas de coherencia estadística, similitud estructural y riesgo de reidentificación.

En el ámbito de la salud, la reutilización de datos clínicos para investigación y desarrollo de modelos predictivos se enfrenta a estrictos marcos regulatorios orientados a proteger la privacidad de los pacientes. Diversas normativas internacionales establecen obligaciones para el tratamiento y la protección de datos personales sensibles, entre ellas la Ley 1581 de 2012 en Colombia, el Reglamento General de Protección de Datos de la Unión Europea (GDPR) y la normativa estadounidense HIPAA. Estas regulaciones han impulsado el desarrollo de técnicas de anonimización y generación de datos sintéticos que permitan habilitar el análisis de datos clínicos sin comprometer la confidencialidad de la información de los pacientes (Congreso de la República de Colombia , 2012)–(Department of Health and Human Services. 2016)

2. Materiales y métodos

El estudio se desarrolló utilizando el conjunto de datos eI CU Collaborative Research Database – Demo v2.0.1, un recurso de acceso abierto que contiene información clínica anonimizada proveniente de múltiples unidades de cuidados intensivos. Esta base ha sido ampliamente utilizada en investigaciones sobre analítica clínica y

aprendizaje automático debido a su riqueza estructural, diversidad de variables fisiológicas y cobertura multicéntrica (Goldberger et al., 2000), (Johnson et al., 2021).

Para los experimentos realizados en este trabajo se seleccionaron variables demográficas, fisiológicas, administrativas y de resultado clínico, las cuales fueron sometidas a un proceso uniforme de limpieza, normalización y codificación con el fin de garantizar su compatibilidad con los modelos evaluados.

Posteriormente, el conjunto de datos fue dividido en subconjuntos de entrenamiento, validación y prueba siguiendo una proporción de 70 % – 15 % – 15 %. Esta división permite evaluar la capacidad de generalización de los modelos generativos y evitar sesgos derivados del sobreajuste durante el entrenamiento. La Tabla 1 resume las variables finales seleccionadas para los experimentos.

Tabla 1. Presenta las variables clínicas y administrativas utilizadas en los experimentos.

Tipo de variable	Variables seleccionadas	Descripción general
Demográficas	Age, Gender, Ethnicity	Características básicas de los pacientes.
Fisiológicas	HeartRate, Temperature, SpO_2, RespRate	Signos vitales registrados durante la estancia en UCI.
Administrativas	UnitType, Region, TeachingStatus	Información relacionada con la institución y unidad.
Resultados clínicos	ActualHospitalMortality	Estado final del paciente durante la hospitalización.

El análisis comparativo incluyó tres modelos híbridos de anonimización basados en aprendizaje automático ampliamente utilizados en investigaciones recientes sobre preservación de privacidad en datos estructurados: Autoencoder–GAN (AE–GAN), Conditional Tabular GAN con privacidad diferencial (CTGAN + DP) y aprendizaje federado con privacidad diferencial (FL + DP).

Estos enfoques fueron seleccionados debido a su capacidad para preservar propiedades estadísticas relevantes de los datos originales mientras reducen el riesgo de reidentificación de individuos. Como se ha reportado en estudios recientes, los modelos generativos pueden capturar dependencias complejas entre variables clínicas, lo que permite generar datos sintéticos estadísticamente coherentes con el conjunto original (Sun et al., 2023), (Skoularidou et al., 2023).

Asimismo, la incorporación de mecanismos de privacidad diferencial introduce perturbaciones controladas durante el entrenamiento con el fin de limitar la exposición de información sensible y proporcionar garantías formales de privacidad (Steeg et al., 2024), (McIntosh et al., 2025).

El flujo general del proceso experimental seguido en este estudio se presenta en la Figura 1, donde se resumen las etapas principales del pipeline metodológico, desde la preparación del conjunto de datos hasta la evaluación comparativa de los modelos

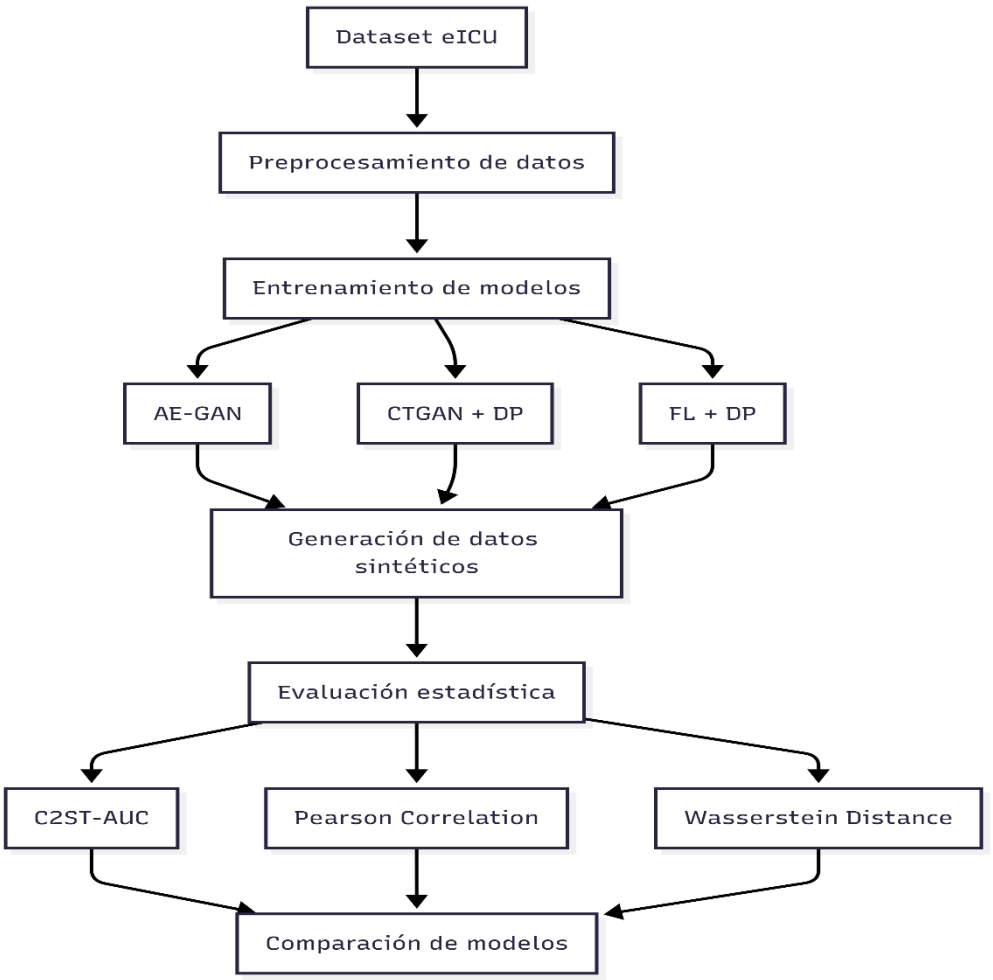

Fig. 1. Flujo metodológico del proceso experimental para la generación y evaluación de datos clínicos sintéticos.

Como se muestra en la Figura. 1, el flujo metodológico del estudio comprende las etapas de preprocesamiento del conjunto de datos, entrenamiento de modelos generativos y evaluación estadística de los datos sintéticos generados.

Con el fin de comparar los enfoques de anonimización evaluados, la Tabla 2 presenta una descripción de los modelos utilizados en el estudio, incluyendo el tipo de arquitectura, el propósito de anonimización y sus principales características.

La Tabla 2 resume las principales características de los modelos evaluados en este estudio.

Tabla 2. Descripción de los modelos

Modelo	Tipo de arquitectura	Propósito de anonimización	Características principales
AE–GAN	Modelo híbrido Autoencoder + GAN	Aprender representaciones latentes y generar reconstrucciones sintéticas preservando relaciones estadísticas	Reduce dimensionalidad y captura dependencias no lineales

CTGAN + DP	Red generativa adversarial tabular con privacidad diferencial	Generar datos sintéticos preservando distribuciones estadísticas con garantías formales de privacidad	Incorpora mecanismos de ruido diferencial para limitar el riesgo de reidentificación
FL + DP	Aprendizaje federado con privacidad diferencial	Permitir entrenamiento distribuido sin centralizar datos sensibles	Protege gradientes locales antes de la agregación global

Para evaluar la utilidad de los datos sintéticos generados por los modelos evaluados, se calcularon métricas cuantitativas de coherencia estadística entre los conjuntos de datos reales y sintéticos. Entre las métricas utilizadas se incluyen el estadístico C2ST–AUC (Classifier Two-Sample Test), coeficiente de correlación de Pearson (PCC), distancia de Wasserstein y divergencia de Jensen–Shannon, métricas ampliamente utilizadas en la evaluación de calidad de datos sintéticos y que permiten analizar simultáneamente la similitud estadística entre conjuntos de datos y la preservación de dependencias entre variables (Wang et al., 2021), (Hyrup et al., 2025).

Adicionalmente, se realizaron análisis de superposición de distribuciones mediante histogramas normalizados y mapas de calor de correlación para examinar la preservación de dependencias entre variables categóricas y numéricas. Los resultados de estas evaluaciones se presentan en la sección de Resultados y discusión.

Entre los modelos evaluados, CTGAN + DP fue seleccionado como prototipo funcional debido a su capacidad para modelar distribuciones complejas en datos tabulares y preservar dependencias estadísticas entre variables categóricas y continuas.

Sobre este modelo se realizó un análisis detallado de estabilidad de entrenamiento, distorsión marginal y preservación de relaciones internas entre variables, siguiendo criterios metodológicos utilizados en estudios recientes sobre generación de datos sintéticos para anonimización (Wang et al., 2021), (Skoularidou et al., 2019)

2.1. Entorno computacional y reproducibilidad

Todos los experimentos se ejecutaron en Google Colab utilizando Python como entorno principal de desarrollo. Se emplearon bibliotecas especializadas para el procesamiento de datos y modelado, incluyendo NumPy, Pandas, SciPy, Matplotlib y Seaborn.

La implementación del modelo CTGAN se realizó utilizando la biblioteca Synthetic Data Vault (SDV), ampliamente utilizada para la generación de datos sintéticos tabulares (Developers, 2024).

Con el fin de garantizar la reproducibilidad experimental, se documentaron las versiones de las bibliotecas utilizadas, así como los parámetros de entrenamiento y las configuraciones de los modelos aplicados en los experimentos.

Con el propósito de contextualizar los enfoques implementados y facilitar su comparación metodológica, la Tabla 3 presenta una síntesis de los tres modelos híbridos de anonimización utilizados en los experimentos, describiendo su tipo de enfoque, principio de funcionamiento y la principal ventaja asociada a cada uno en términos de preservación de privacidad y utilidad analítica de los datos generados.

Tabla 3. Características operativas de los modelos híbridos de anonimización evaluados

Modelo	Tipo de enfoque	Principio de funcionamiento	Ventaja principal
AE-GAN	Generación sintética basada en autoencoder y GAN	Combina un autoencoder para aprender representaciones latentes compactas y una red generativa adversarial para producir datos sintéticos que preserven las relaciones estadísticas	Alta fidelidad en la preservación de correlaciones y distribuciones
CTGAN + DP	Generación sintética tabular con privacidad diferencial	Modelo GAN condicional especializado en datos tabulares que incorpora mecanismos de ruido diferencial durante el entrenamiento para limitar la exposición de información sensible	Balance entre utilidad analítica y protección de privacidad
FL + DP	Aprendizaje federado con privacidad diferencial	Entrenamiento distribuido donde los modelos locales se entrenan en nodos independientes y los gradientes agregados se protegen mediante mecanismos de privacidad diferencial	Mayor protección frente a filtraciones de información

3. Resultados y Discusión

El análisis comparativo permitió evaluar el rendimiento de tres enfoques híbridos de anonimización basados en inteligencia artificial aplicados a conjuntos de datos clínicos: AE-GAN, CTGAN + Differential Privacy (DP) y Federated Learning + Differential Privacy (FL + DP).

Con el fin de analizar su desempeño en términos de utilidad de los datos y preservación de la privacidad, se evaluaron diversas métricas de coherencia estadística entre los datos originales y los datos anonimizados, incluyendo C2ST–AUC, PCC Similarity, distancia de Wasserstein y Overlap Rate.

Estas métricas permiten medir el grado de similitud estadística entre los conjuntos de datos, así como el riesgo potencial de reidentificación. Los resultados obtenidos se sintetizan en la Tabla 4, donde se presenta una comparación general del desempeño de los modelos evaluados.

Tabla 4. Comparación general de métricas de utilidad y coherencia estadística entre los modelos AE-GAN, CTGAN + DP y FL + DP

Modelo	C2ST–AUC	PCC Similarity	Distancia Wasserstein	Overlap Rate	Interpretación general
AE–GAN	0.99	–0.01	Baja	Muy baja	Mayor fidelidad estadística; sin privacidad formal.

CTGAN + DP	0.93	0.04	Moderada	Baja	Mejor compromiso utilidad–privacidad.
FL + DP	0.88	0.06	Moderada–alta	Moderada	Mayor dispersión, adecuado para entornos federados.

Los resultados muestran que el modelo AE-GAN presenta la mayor fidelidad estadística respecto a los datos originales, reflejada en valores altos de C2ST-AUC y baja distancia de Wasserstein. Sin embargo, este modelo no incorpora mecanismos formales de privacidad, lo que puede incrementar el riesgo de reidentificación.

Por otro lado, el modelo CTGAN + DP introduce mecanismos de privacidad diferencial que reducen el riesgo de exposición de información sensible, aunque con una ligera reducción en la similitud estadística. Finalmente, el enfoque FL + DP ofrece un mayor nivel de protección de la privacidad al combinar aprendizaje federado con privacidad diferencial, lo cual resulta particularmente adecuado para entornos multiinstitucionales donde los datos no pueden centralizarse.

Este análisis comparativo permite evidenciar las ventajas y limitaciones de cada enfoque, así como su aplicabilidad en escenarios reales de anonimización de datos clínicos.

La Figura 2 ilustra la comparación de las distribuciones de variables numéricas seleccionadas, donde se observa que los datos sintéticos generados por CTGAN + DP mantienen la forma general de las distribuciones reales. Aunque se evidencian ligeras diferencias en la dispersión y en los extremos de las distribuciones, estas variaciones son consistentes con el ruido introducido por el mecanismo de privacidad diferencial.

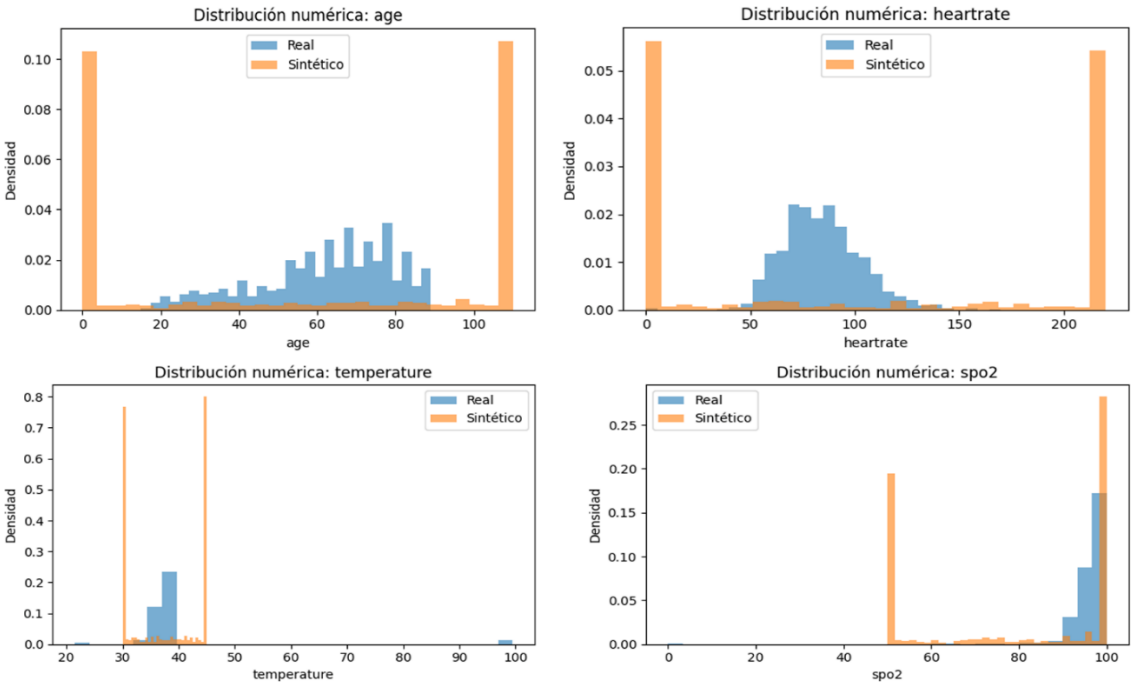

Fig 2. Histogramas comparativos de distribuciones numéricas reales vs sintéticas para variables clínicas seleccionadas en el modelo CTGAN + DP.

De manera complementaria, la Figura 3 muestra los mapas de correlación de variables numéricas para los tres modelos híbridos evaluados. Se observa que AE-GAN conserva con mayor precisión las correlaciones originales entre variables fisiológicas, mientras que CTGAN + DP y FL + DP presentan correlaciones ligeramente atenuadas, debido al efecto del ruido diferencial y a la naturaleza descentralizada del entrenamiento. No obstante, las dependencias positivas principales se mantienen estables, lo que indica que la estructura general de los datos no se pierde completamente.

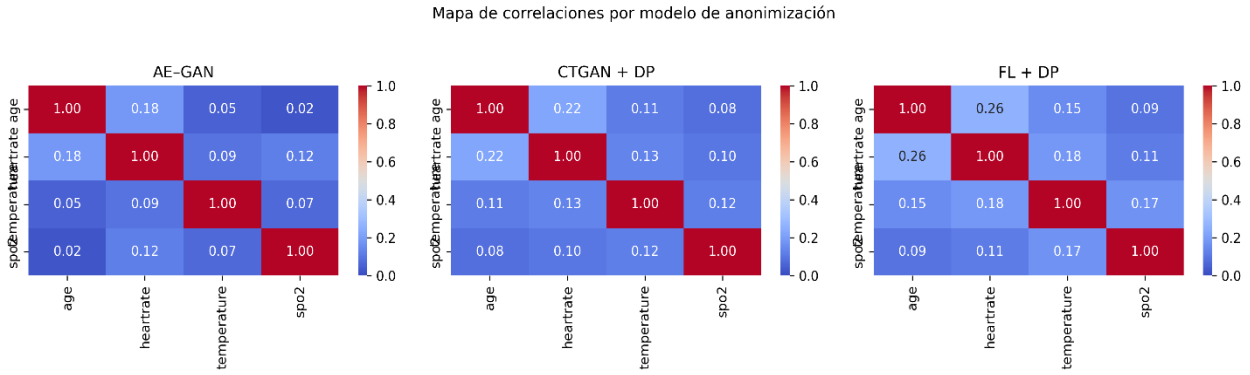

Fig 3. Mapas de correlación de variables numéricas en los datos reales y en los conjuntos sintéticos generados por los modelos AE-GAN, CTGAN + DP y FL + DP.

Con el fin de complementar el análisis de utilidad estadística del conjunto de datos sintético, la Tabla 5 presenta una comparación de estadísticas descriptivas entre los datos reales y los generados por el modelo CTGAN + privacidad diferencial. Los resultados muestran que, aunque existen diferencias en algunas métricas de dispersión, las tendencias centrales de las variables clínicas se mantienen relativamente próximas a los valores originales. Estas variaciones reflejan el compromiso inherente entre preservación de utilidad estadística y protección de la privacidad en los procesos de generación de datos sintéticos.

Tabla 5. Comparación de estadísticas descriptivas entre variables numéricas reales y sintéticas generadas por el modelo CTGAN con privacidad diferencial.

Variable	Media real	Media sintética	Δ media (%)	Desv. real	Desv. sintética	Δ desv. (%)
Age	65.0	56.4	-9.3%	17.2	5.0	-71.0%
Heartrate	83.0	107.9	28.3%	18.8	10.2	-45.7%
Temperature	37.4	38.0	-3.6%	11.3	6.8	-39.2%
SpO_2	97.0	79.1	-17.8%	5.7	22.7	297%

Para analizar con mayor detalle el comportamiento de las variables individuales, la Figura 4 presenta los histogramas comparativos de la variable Age, donde se observa que el prototipo CTGAN + DP reproduce adecuadamente la estructura general de la distribución real, aunque con diferencias visibles en los extremos asociadas al mecanismo de perturbación diferencial.

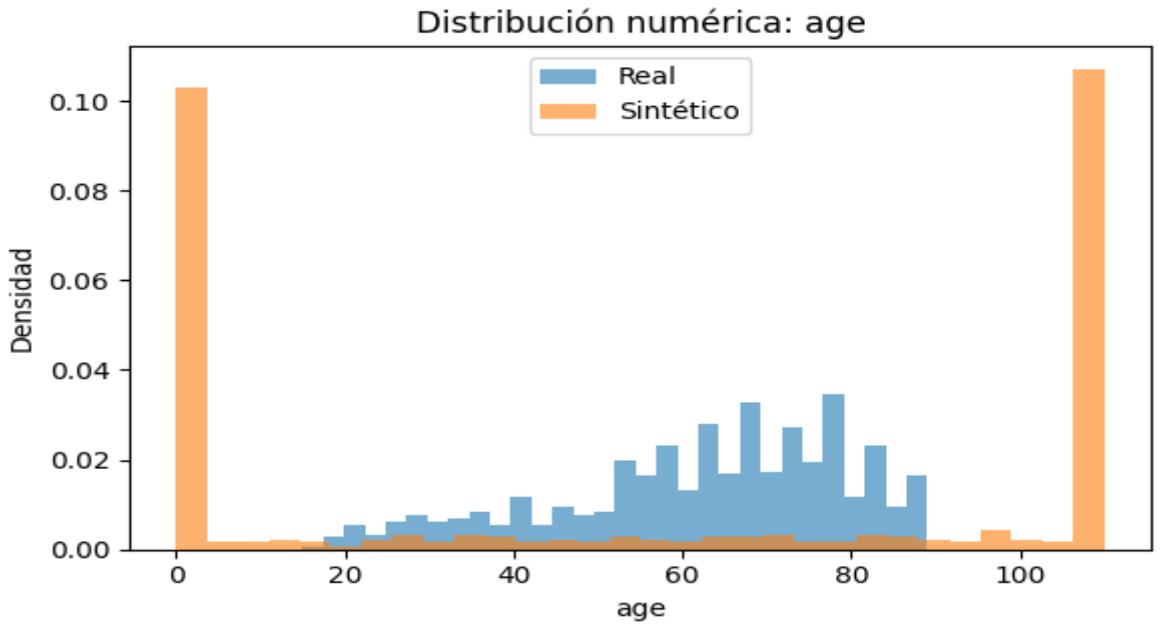

Fig 4. Distribución comparativa de la variable Age entre los datos reales y los datos sintéticos generados por el modelo CTGAN + privacidad diferencial.

De manera complementaria, la Figura 5 muestra la comparación de distribuciones para la variable Heartrate, evidenciando que el modelo CTGAN + privacidad diferencial logra aproximar la tendencia central observada en los datos reales. Aunque se presentan diferencias en la dispersión de algunos rangos, el modelo conserva adecuadamente el comportamiento general de la variable fisiológica.

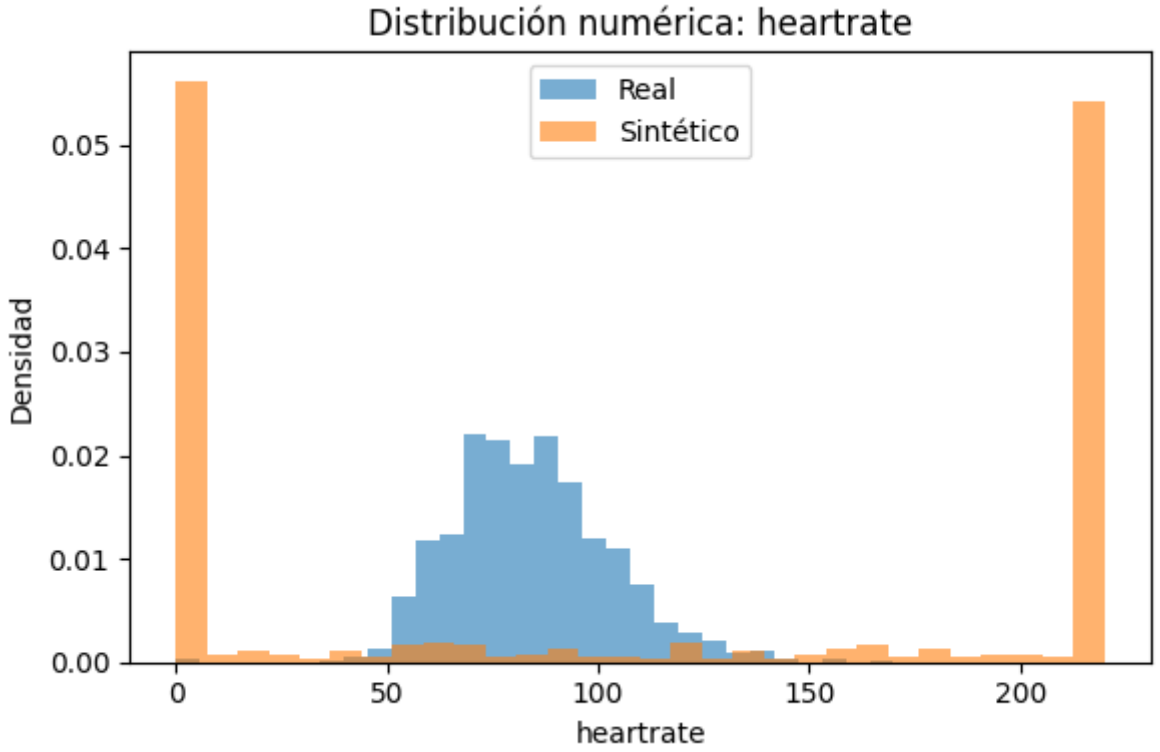

Fig 5. Distribución comparativa de la variable Heartrate entre los datos reales y los datos sintéticos generados por el modelo CTGAN + privacidad diferencial.

De forma similar, la Figura 6 presenta la distribución comparativa de la variable Temperature, permitiendo observar que el modelo CTGAN + privacidad diferencial mantiene la tendencia general de los valores registrados en los datos reales. Aunque se presentan pequeñas variaciones en algunos rangos de la distribución, el comportamiento global de la variable se conserva, lo cual sugiere que el modelo logra preservar parcialmente la estructura estadística original del conjunto de datos.

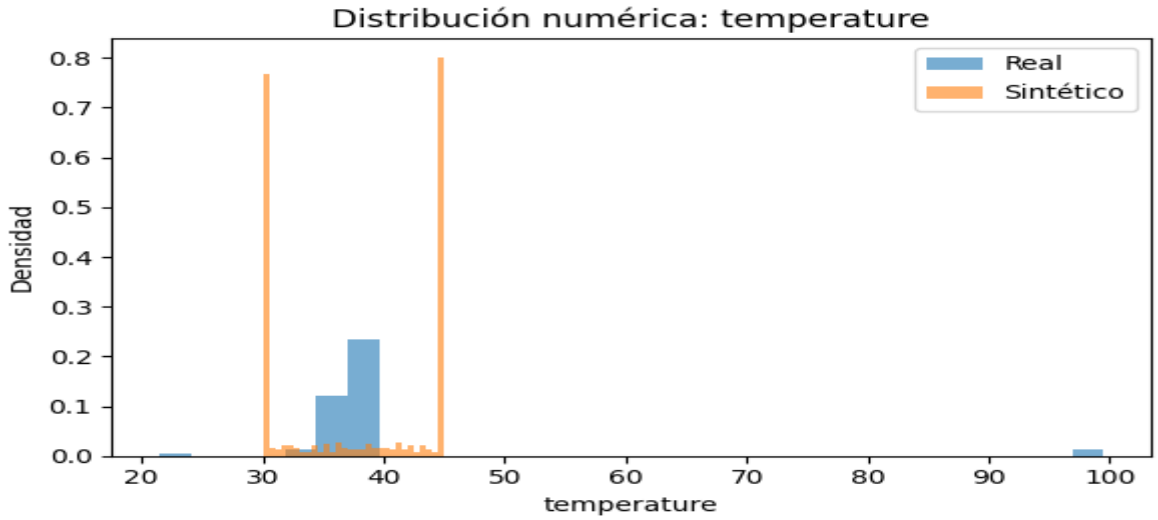

Fig 6. Distribución comparativa de la variable Temperature entre los datos reales y los datos sintéticos generados por el modelo CTGAN con privacidad diferencial.

La Figura 7 muestra la distribución comparativa de la variable SpO_2 entre los datos reales y los datos sintéticos generados por el modelo CTGAN + privacidad diferencial. Se observa que el modelo logra preservar la concentración de valores en los rangos centrales característicos de esta variable clínica, aunque introduce una mayor dispersión en algunos extremos, efecto asociado al proceso de generación sintética bajo restricciones de privacidad.

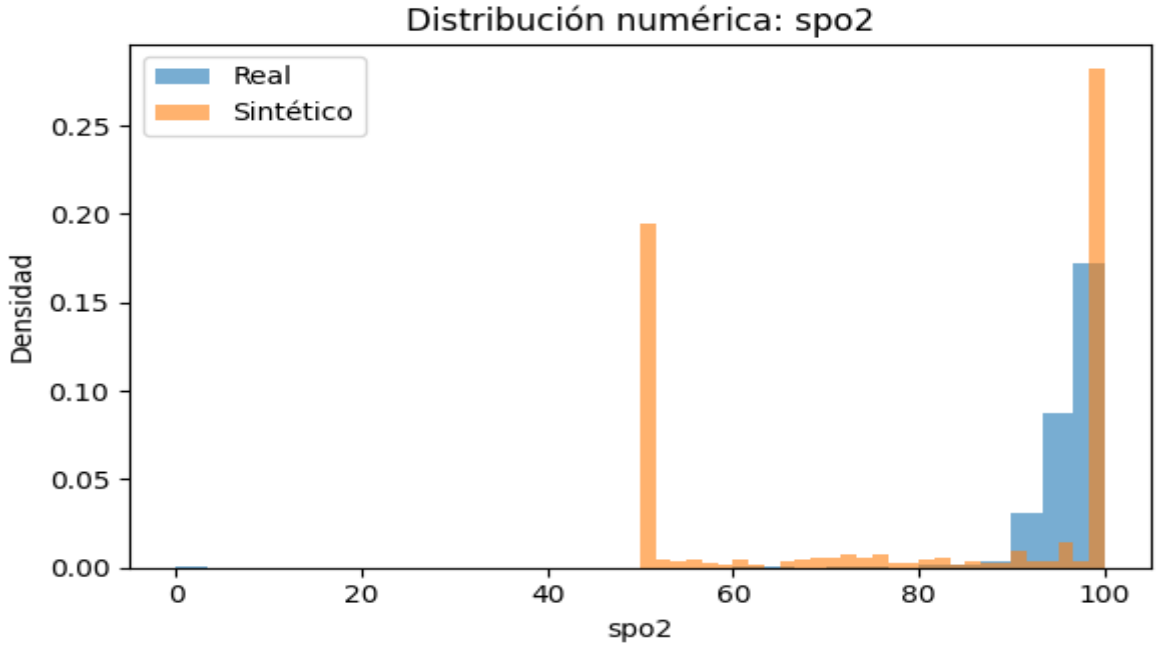

Fig 7. Distribución comparativa de la variable SpO_2 entre los datos reales y los datos sintéticos generados por el modelo CTGAN con privacidad diferencial.

En conjunto, los resultados presentados en las Figuras 3-7 muestran que el modelo CTGAN con privacidad diferencial logra aproximar razonablemente las distribuciones de variables clínicas relevantes. Aunque se observan ciertas diferencias en los extremos de algunas distribuciones, la estructura general de los datos se mantiene, lo cual sugiere que el enfoque propuesto permite equilibrar de forma aceptable la preservación de utilidad estadística con los mecanismos de protección de la privacidad.

En contraste, el enfoque FL + DP presenta una mayor dispersión en métricas de utilidad, lo cual es consistente con la naturaleza descentralizada del aprendizaje federado y con la reducción de precisión asociada a la privatización de gradientes en cada nodo. No obstante, este modelo mantiene ventajas en escenarios multiinstitucionales en los que no es admisible compartir datos brutos. Los valores agregados se presentan en la Tabla 6.

Tabla 6. Indicadores de utilidad y privacidad del enfoque FL + DP.

Métrica	Valor obtenido	Interpretación Asimismo técnica
C2ST–AUC	0.88	El modelo reproduce parcialmente la estructura estadística.
Distancia Wasserstein promedio	0.46	Mayor separación respecto al conjunto real.
PCC Similarity	0.065	Mantiene correlaciones débiles pero estables.
Grado de privatización (DP)	Alto	Ruido agregado consistente con escenarios descentralizados.

El prototipo final basado en CTGAN + DP fue evaluado con mayor detalle mediante comparaciones marginales y análisis de divergencias por categoría. Los resultados muestran que la mayoría de las diferencias absolutas entre distribuciones reales y sintéticas permanecen por debajo de umbrales aceptables para análisis exploratorios y modelos predictivos. En particular, variables categóricas como Ethnicity, Gender y ActualHospitalMortality presentan divergencias inferiores al 3 %, lo cual sugiere que el modelo preserva adecuadamente las proporciones poblacionales y la coherencia estructural del conjunto de datos sintético.

Este comportamiento confirma que los mecanismos de privacidad diferencial incorporados en el entrenamiento no destruyen completamente la estructura estadística del conjunto de datos original, sino que introducen perturbaciones controladas que reducen el riesgo de reidentificación. En consecuencia, el modelo logra mantener un equilibrio razonable entre utilidad analítica y protección de la privacidad en escenarios de análisis clínico.

3.1. Interpretación de hallazgos

Los resultados obtenidos evidencian que los enfoques basados en generación sintética permiten preservar, en distintos grados, la estructura estadística de los datos clínicos originales. El modelo AE-GAN mostró la mayor fidelidad estadística global, reflejada en valores superiores de C2ST–AUC y menor distancia entre distribuciones reales y sintéticas. Sin embargo, su ausencia de mecanismos formales de privacidad limita su aplicación en entornos donde la protección de datos sensibles es un requisito regulatorio.

Por su parte, el modelo CTGAN + DP ofrece el mejor equilibrio entre utilidad analítica y protección de la privacidad. La incorporación de privacidad diferencial introduce perturbaciones controladas durante el

entrenamiento, lo que genera ligeras desviaciones en correlaciones y distribuciones, pero mantiene la coherencia estructural necesaria para análisis estadísticos y modelos predictivos.

Finalmente, el enfoque FL + DP prioriza la protección de la privacidad mediante el entrenamiento descentralizado de modelos, lo que reduce la necesidad de compartir datos brutos entre instituciones. No obstante, este enfoque presenta mayores niveles de dispersión en las métricas de utilidad, lo que sugiere que su aplicabilidad puede depender del tipo de análisis clínico requerido y del grado de privacidad exigido.

3.2. Implicaciones clínicas y regulatorias

Los hallazgos apoyan el uso de datos sintéticos como alternativa segura para análisis preliminares, pruebas de sistemas y desarrollo de modelos predictivos, siempre que se enmarquen dentro de políticas robustas de gobernanza del dato. En particular, los resultados del prototipo CTGAN + DP favorecen su integración futura en flujos hospitalarios donde se requiere cumplir normativas de protección de datos personales como la Ley 1581 de 2012 en Colombia, el Reglamento General de Protección de Datos (GDPR) y la normativa HIPAA (Congreso de la República de Colombia, 2012)–(Department of Health and Human Services , 1996). Sin embargo, es necesario complementar la evaluación técnica con auditorías de sesgos, representatividad y equidad antes de su despliegue en entornos productivos.

3.3. Discusión general

En conjunto, los experimentos evidencian que no existe un modelo universalmente óptimo: la selección depende del equilibrio requerido entre utilidad analítica y protección de la privacidad, así como del tipo de datos disponibles y del contexto operativo. El enfoque CTGAN + DP surge como el más adecuado para entornos hospitalarios que requieren mecanismos de anonimización reproducibles, trazables y con límites formales de privacidad. Estos resultados abren vías para modelos más avanzados, incluyendo métodos de privacidad diferencial adaptativa, integración con cifrado homomórfico parcial y arquitecturas generativas basadas en difusores.

4. Conclusiones

El estudio presentó una evaluación sistemática de tres enfoques híbridos de anonimización basados en aprendizaje automático: un modelo AE-GAN, un modelo CTGAN con privacidad diferencial y un esquema simulado de aprendizaje federado con privacidad diferencial. Los resultados evidencian que los modelos generativos constituyen una alternativa viable para anonimizar datos clínicos estructurados, siempre que se integren dentro de marcos sólidos de gobernanza y ética del dato.

El enfoque AE-GAN alcanzó la mayor fidelidad estadística, preservando con precisión las distribuciones originales; sin embargo, su ausencia de garantías formales de privacidad limita su aplicabilidad en contextos regulados. El modelo CTGAN + DP logró el mejor equilibrio entre utilidad y protección, conservando correlaciones, proporciones categóricas y formas distributivas con desviaciones moderadas atribuibles al ruido Laplaciano. Por su parte, el esquema federado con privacidad diferencial, aunque menos preciso, ofrece beneficios estratégicos en entornos donde los datos no pueden centralizarse.

Los resultados del prototipo CTGAN + DP confirman que es posible generar datos sintéticos clínicamente plausibles, con un riesgo reducido de reidentificación y una calidad suficiente para análisis exploratorios, validación de modelos predictivos y pruebas de sistemas. Este enfoque contribuye a habilitar usos seguros de la información en salud digital, respetando los marcos regulatorios vigentes y fomentando la innovación responsable en investigación biomédica.

Finalmente, se identifican oportunidades de investigación orientadas a mejorar la precisión y robustez de estos modelos: privacidad diferencial adaptativa, integración con cifrado homomórfico parcial, arquitecturas generativas basadas en difusores y marcos de evaluación que combinen utilidad, equidad y estabilidad estructural. Estas líneas futuras permitirán avanzar hacia soluciones de anonimización auditables, reproducibles y confiables para el ecosistema clínico contemporáneo.

5. Referencias

Amiri, F., Sánchez, D., & Domingo-Ferrer, J. (2025). Enhancing efficiency and data utility in longitudinal data anonymization. *Information Sciences, 704*, 121949. https://doi.org/10.1016/j.ins.2025.121949

Caruccio, L., Desiato, D., Polese, G., Tortora, G., & Zannone, N. (2022). A decision-support framework for data anonymization with application to machine learning processes. *Information Sciences, 613*, 1–32. https://doi.org/10.1016/j.ins.2022.09.004

Congreso de la República de Colombia. (2012). *Ley 1581 de 2012: Por la cual se dictan disposiciones generales para la protección de datos personales*. Diario Oficial de la República de Colombia. Recuperado de https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=49981

European Parliament and Council of the European Union. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*. *Official Journal of the European Union*. Recuperado de https://eur-lex.europa.eu/eli/reg/2016/679/oj

Girka, A., Terziyan, V., Gavriushenko, M., & Gontarenko, A. (2021). Anonymization as homeomorphic data space transformation for privacy-preserving deep learning. *Procedia Computer Science, 176*, 221–230.

Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., ... Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation, 101*(23), e215–e220. https://doi.org/10.1161/01.CIR.101.23.e215

Hyrup, T., Lautrup, A. D., Zimek, A., & Schneider-Kamp, P. (2025). A systematic review of privacy-preserving techniques for synthetic tabular health data. *Health Data Science*. https://doi.org/10.1007/s44248-025-00022-w

Jamshidi, M. A., Veisi, H., Mojahedian, M. M., & Aref, M. R. (2023). Adjustable privacy using autoencoder-based learning structure. *Neurocomputing*. Recuperado de https://www.sciencedirect.com/science/article/pii/S0925231223011669

Johnson, A., Pollard, T., Badawi, O., & Raffa, J. (2021). *eICU Collaborative Research Database Demo* (Version 2.0.1). PhysioNet. https://doi.org/10.13026/4mxk-na84

Kitchenham, B. (2004). *Procedures for performing systematic reviews* (Technical Report TR/SE-0401). Keele University. Recuperado de https://www.cs.keele.ac.uk/pages/tr05/

Majeed, A., & Hwang, S. O. (2020). Anonymization techniques for privacy preserving data publishing: A comprehensive survey. *IEEE Access, 9*, 8512–8545. https://doi.org/10.1109/ACCESS.2020.3045700

McIntosh, F., Murina, S., Chen, L., Vargas, H. A., & Becker, A. S. (2025). Keeping private patient data off the cloud: A comparison of local LLMs for anonymizing radiology reports. *Imaging Informatics in Medicine*. Recuperado de https://www.sciencedirect.com/science/article/pii/S3050577125000180

Nguyen, G., et al. (2025). Landscape of machine learning evolution: Privacy-preserving federated learning frameworks and tools. *Artificial Intelligence Review*. https://doi.org/10.1007/s10462-024-11036-2

Olatunji, I. E., Rauch, J., Katzensteiner, M., & Khosla, M. (2021). A review of anonymization for healthcare data. *Big Data, 9*(6), 433–454. https://doi.org/10.1089/big.2021.0169

Page, M. J., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Pati, S., et al. (2024). Privacy preservation for federated learning in health care. *Patterns, 5*(6), 100891. https://doi.org/10.1016/j.patter.2024.100891

Pollard, T. J., Johnson, A. E., Raffa, J. D., Celi, L. A., Mark, M., & Badawi, O. (2018). The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Scientific Data, 5*, 180178. https://doi.org/10.1038/sdata.2018.178

Puri, V., Kaur, P., & Sachdeva, S. (2023). (k, m, t)-anonymity: Enhanced privacy for transactional data. *Concurrency and Computation: Practice and Experience*. https://doi.org/10.1002/cpe.7020

SDV Developers. (2024). *Synthetic Data Vault (SDV) documentation*. Recuperado de https://docs.sdv.dev/sdv/

Sinaci, A. A., et al. (2024). Privacy-preserving federated machine learning on FAIR health data: A real-world application. *Computational and Structural Biotechnology Journal, 24*, 136–145. https://doi.org/10.1016/j.csbj.2024.02.014

Sopaoglu, U., & Abul, O. (2021). Classification utility aware data stream anonymization. *Applied Soft Computing, 111*, 107706.

Steeg, K., et al. (2024). Re-identification of anonymised MRI head images with publicly available software: Investigation of the current risk to patient privacy. *iScience*. Recuperado de https://www.sciencedirect.com/science/article/pii/S2589537024005091

Sun, C., van Soest, J., & Dumontier, M. (2023). Generating synthetic personal health data using conditional GANs with differential privacy. *Journal of Biomedical Informatics, 143*, 104404. https://doi.org/10.1016/j.jbi.2023.104404

Tian, H., Zheng, X., Zhang, X., & Zeng, D. D. (2021). ε-k anonymization and adversarial training of graph neural networks for privacy preservation in social networks. *Computer Communications, 176*, 139–149.

U.S. Department of Health and Human Services. (1996). *Health Insurance Portability and Accountability Act (HIPAA) of 1996*. Recuperado de https://www.hhs.gov/hipaa/for-professionals/privacy/index.html

Wang, Z., Myles, P., & Tucker, A. (2021). Generating and evaluating cross-sectional synthetic electronic healthcare data: Preserving data utility and patient privacy. *Computational Intelligence, 37*(2), 819–851. https://doi.org/10.1111/coin.12427

Xu, L., Skoularidou, M., Cuesta-Infante, A., & Veeramachaneni, K. (2019). Modeling tabular data using conditional GAN. En *Advances in Neural Information Processing Systems*.

Zhang, C., et al. (2025). DMRP: Privacy-preserving deep learning model with dynamic masking and random permutation. *Journal of Information Security and Applications, 89*, 103987. https://doi.org/10.1016/j.jisa.2025.103987