Publicación Científica de la Universidad de Ciencias Médicas de Ciego de Ávila
REVISTA MÉDICA ELECTRÓNICA DE CIEGO DE ÁVILA
Editorial Ciencias Médicas
2026;32: e4345
ISSN: 1029-3035
RNPS: 1821

Artículo original

Propuesta de plataforma en línea para generar conjuntos de datos anotados de señales biomédicas

Proposal for an online platform to generate annotated datasets of biomedical signals

Proposta de plataforma online para gerar conjuntos de dados anotados de sinais biomédicos

Julio Companioni-Martínez[1] https://orcid.org/0000-0001-7412-7758
Denis Hernández-Pacheco[2*] https://orcid.org/0000-0002-7208-7817
Carlos Lázaro Jiménez-Puerto[3] https://orcid.org/0000-0001-8967-2935
José Carlos Pernas-Farray[4] https://orcid.org/0009-0002-1827-1629
Alberto Taboada-Crispi[5] https://orcid.org/0000-0002-7797-1441

[1]Mastér en Informática. Ingeniero en Ciencias Informáticas. Profesor Asistente del Departamento de Informática-FABLAB, Facultad de Ciencias Técnicas y Económicas (FCTyE). Universidad de Sancti Spíritus "José Martí Pérez". Sancti Spíritus, Cuba.
[2]Mastér en Ciencias de la Computación. Ingeniero en Telecomunicaciones y Electrónica. Profesor Instructor del Departamento de Informática-FABLAB, Facultad de Ciencias Técnicas y Económicas (FCTyE). Universidad de Sancti Spíritus "José Martí Pérez". Sancti Spíritus, Cuba.
[3]Doctor en Ciencias de la Educación. Ingeniero en Ciencias Informáticas. Vicedecano de Investigación y Posgrado. Profesor titular del Departamento de Informática-FABLAB, Facultad de Ciencias Técnicas y Económicas (FCTyE). Universidad de Sancti Spíritus "José Martí Pérez". Sancti Spíritus, Cuba.
[4]Ingeniero en Telecomunicaciones y Electrónica. Profesor del Departamento de Informática-FABLAB, Facultad de Ciencias Técnicas y Económicas (FCTyE). Universidad de Sancti Spíritus "José Martí Pérez". Sancti Spíritus, Cuba.
[5]Doctor en Ingeniería Biomédica, Ingeniero Electrónico. Profesor Titular. Investigador Titular. Centro de Investigaciones de la Informática, Facultad de Matemática, Física y Computación. Universidad

Esta obra está bajo una licencia CreativeCommons Attribution4.0/International/Deed

Central de Las Villas "Marta Abreu". Santa Clara, Cuba.

*Autor para la correspondencia. Correo electrónico: hernandezpachecodenis@gmail.com

RESUMEN

Introducción: para que las herramientas diagnósticas y predictivas basadas en inteligencia artificial sean útiles y efectivas en el ámbito médico cubano, se requieren conjuntos de datos anotados de señales biomédicas con las características propias de la población local.

Objetivo: presentar una plataforma web orientada a la creación de conjuntos de datos de bioseñales con anotación diagnóstica, adaptada a las características de la población cubana, para soportar sistemas de apoyo a las decisiones médicas.

Métodos: investigación de desarrollo tecnológico, en cuatro fases: identificación de necesidades clínicas mediante talleres colaborativos con 12 médicos especialistas; diseño de la base de datos como almacén de las señales y sus anotaciones; definición de la tecnología informática de soporte, y protocolo piloto para recolectar electrocardiogramas de pacientes cubanos, con garantía de anonimato y seguridad de los datos.

Resultados: la plataforma incluyó 19 requisitos funcionales, 22 no funcionales, un modelo de datos validado por un colectivo de expertos. Su arquitectura de tres capas se basó en tecnologías de código abierto. Los prototipos con 100 señales de electrocardiogramas resultaron funcionales.

Conclusiones: el desarrollo y uso de una plataforma web para gestionar y anotar señales biomédicas de pacientes cubanos, basada en un modelo relacional normalizado y tecnologías de código abierto, es viable. Reduce la dependencia de datos externos, y mejora la equidad y precisión de los sistemas de apoyo clínico; todo bajo estándares éticos y de seguridad. El estudio en sí mismo es un aporte crucial porque pone a disposición de los especialistas datos biomédicos propios, anotados y contextualizados.

Palabras clave: conjunto de datos; desarrollo tecnológico; informática; inteligencia artificial; seguridad computacional; pacientes.

ABSTRACT

Introduction: for diagnostic and predictive tools based on artificial intelligence to be useful and effective within the Cuban medical context, annotated datasets of biomedical signals—reflecting the specific

Esta obra está bajo una licencia CreativeCommons Attribution4.0/International/Deed

characteristics of the local population—are required.

Objective: to present a web-based platform designed for the creation of biosignal datasets with diagnostic annotations, tailored to the characteristics of the Cuban population, to support medical decision-support systems.

Methods: this technological development research was conducted in four phases: identifying clinical needs through collaborative workshops with 12 medical specialists; designing a database to store signals and their corresponding annotations; defining the supporting information technology; and establishing a pilot protocol for collecting electrocardiograms from Cuban patients, ensuring data anonymity and security.

Results: the platform incorporates 19 functional and 22 non-functional requirements, with a data model validated by a group of experts. Its three-tier architecture is based on open-source technologies. Prototypes containing 100 electrocardiogram signals proved to be functional.

Conclusions: the development and use of a web-based platform for managing and annotating biomedical signals from Cuban patients—built on a normalized relational model and open-source technologies—is feasible. It reduces reliance on external data and enhances both the equity and accuracy of clinical decision-support systems, all while adhering to ethical and security standards. The study itself represents a crucial contribution by providing specialists with their own, annotated, and context-specific biomedical data.

Keywords: artificial intelligence; computer security; dataset; informatics; patients; technological development.

RESUMO

Introdução: para que as ferramentas diagnósticas e preditivas baseadas em inteligência artificial sejam úteis e eficazes no contexto médico cubano, são necessários conjuntos de dados anotados de sinais biomédicos que apresentem as características próprias da população local.

Objetivo: apresentar uma plataforma web destinada à criação de conjuntos de dados de biosinais com anotação diagnóstica, adaptada às características da população cubana, para dar suporte a sistemas de apoio à decisão médica.

Esta obra está bajo una licencia CreativeCommons Attribution4.0/International/Deed

2026;32: e4345
ISSN: 1029-3035
RNPS: 1821

Métodos: pesquisa de desenvolvimento tecnológico, estruturada em quatro fases: identificação de necessidades clínicas por meio de oficinas colaborativas com 12 médicos especialistas; concepção do banco de dados para o armazenamento dos sinais e suas respectivas anotações; definição da tecnologia da informação de suporte; e estabelecimento de um protocolo-piloto para a coleta de eletrocardiogramas de pacientes cubanos, garantindo o anonimato e a segurança dos dados.

Resultados: a plataforma contemplou 19 requisitos funcionais e 22 não funcionais, além de um modelo de dados validado por um grupo de especialistas. Sua arquitetura de três camadas baseou-se em tecnologias de código aberto. Os protótipos, contendo 100 sinais de eletrocardiogramas, mostraram-se funcionais.

Conclusões: o desenvolvimento e a utilização de uma plataforma web para gerenciar e anotar sinais biomédicos de pacientes cubanos , baseada em um modelo relacional normalizado e em tecnologias de código aberto são viáveis. Essa abordagem reduz a dependência de dados externos e aprimora a equidade e a precisão dos sistemas de apoio à decisão clínica, em conformidade com padrões éticos e de segurança. O estudo, por si só, constitui uma contribuição fundamental, ao disponibilizar aos especialistas dados biomédicos próprios, devidamente anotados e contextualizados.

Palavras chave: conjunto de dados; desenvolvimento tecnológico; informática; inteligência artificial; pacientes; seguridad computacional.

Recibido: 12/08/2026
Aprobado: 14/10/2026
Publicado: 10/11/2026

INTRODUCCIÓN

El empleo de sistemas expertos para tomar decisiones ha aumentado en diversos ámbitos de la vida cotidiana, debido al desarrollo acelerado de la inteligencia artificial (IA) en las últimas décadas.[1] Existen aplicaciones con niveles elevados de pericia en dominios complejos que, incluso, superan a los especialistas más experimentados en sus áreas de estudios respectivas, tanto en los tiempos de ejecución

Esta obra está bajo una licencia CreativeCommons Attribution4.0/International/Deed

de diversas tareas como en la precisión de sus criterios.

El desarrollo de estas aplicaciones puede estar condicionado al cumplimiento de criterios estrictos o restricciones importantes, en dependencia de los dominios donde se utilizan. Por ejemplo, si las decisiones a tomar son sensibles, con incidencia en grupos poblacionales grandes, procesos industriales complejos, actuaciones médicas, o eventos que requieran recursos considerables, entre otros.

Uno de los dominios complejos, con tolerancia baja a los errores, en los cuales los sistemas expertos se aplican más en la actualidad, es el biomédico. En este entorno, tomar decisiones afecta directamente la salud humana.[2] Por tanto, un requisito fundamental durante el desarrollo de estos sistemas es la posibilidad de entrenarlos con datos cuyas características coincidan con aquellas que se emplearán durante su operación. Así, se garantiza que los sistemas expertos estén plenamente adaptados a las particularidades de la población objetivo de su uso.

Por otro lado, los datos son recursos estratégicos de valor económico alto; generan un mercado global muy concentrado en potencias tecnológicas y corporaciones multinacionales.[3] Cuando esa información se emplea para influir en áreas claves del desarrollo (política, economía, seguridad nacional), adquiere importancia crítica.

El contexto cubano refleja también esta situación; ante la demanda creciente de datos para desarrollar sistemas inteligentes –especialmente en el ámbito de la salud–, se deben sentar las bases para contar, en un futuro inmediato, con conjuntos de datos propios;[4] los cuales posibilitarían crear sistemas asistidos. Así, Cuba se sumaría al grupo selecto de países capaces de desarrollar este tipo de sistemas a medida.

En Internet existen conjuntos de datos diversos, vinculados al ámbito biomédico; la mayoría desarrollada por países líderes en la creación de aplicaciones de IA.[5] Los cuales mantienen, desde hace décadas, una política de generación y actualización constante de sus conjuntos de datos. Si bien muchos de estos recursos se ofrecen de forma gratuita en la red, aquellos que contienen información más sensible permanecen restringidos debido a su relevancia para los desarrollos tecnológicos actuales.[6] Entre los repositorios de datos biomédicos más conocidos que se distribuyen libremente, pueden citarse los siguientes:

Base de datos de arritmia del Instituto de Tecnología de Massachusetts y el Hospital Beth Israel (*MIT-BIH*)[7]; base de datos de la Comisión de Seguridad de Productos para el Consumidor de EE.UU. (*CPSC*, del inglés *Consumer Product Safety Commission*);[8] base de datos INCART (Instituto de Técnicas Cardiológicas);[9] base de datos *PTB-XL* (Instituto Federal de Física y Técnica, versión XL);[10] base de

Esta obra está bajo una licencia CreativeCommons Attribution4.0/International/Deed

datos *SPH*;[11] entre otras que son conjuntos de datos de señales electrocardiográficas (ECG, del inglés *electrocardiogram* y el español electrocardiograma).

Algunas son conjuntos de datos de señales electroencefalográficas (EEG): *Siena Scalp EEG Database*;[12] *EEG During Mental Arithmetic Tasks*;[13] *EEG Signals from an RSVP Task*;[14] *CHB-MIT Scalp EEG Database*;[15]

Otras, son conjuntos de datos de señales (bidimensionales) de radiografías de tórax (Rx_T): *Chest X-ray Dataset with Lung Segmentation*;[16] *BRAX, a Brazilian labeled chest X-raydataset*;[17] *Radiology Report Generation Models Evaluation Dataset for Chest X-rays* (*RadEvalX*);[18] *Chest X-ray segmentation images based on MIMIC-CXR*;[19] entre otros.

Y, otras tantas son conjuntos de datos de señales respiratorias: *A Temporal Dataset for Respiratory Support in Critically Ill Patients*;[20] *Respiratory dataset from PEEP study with expiratory occlusion*;[21] *Simulated Obstructive Disease Respiratory Pressure and Flow*;[22] *Preterm Infant Cardio-Respiratory Signals Database*.[23] Además de estos, en Internet existen muchos más conjuntos de datos de señales: temperatura, presión arterial, fonocardiografía, electrohisterografía, oximetría.

Es importante destacar que, en Cuba, el desarrollo de sistemas expertos orientados al área práctica de la medicina no ha alcanzado niveles de desarrollo en correspondencia con la experiencia vasta acumulada en el país. Téngase en cuenta que existe una red amplia de centros de estudios dedicados al tema, tanto en universidades como en empresas. En muchos de estos centros –que han desarrollado algoritmos en el área del procesamiento de señales, algunos incluso con impactos relevantes– se han utilizado señales biomédicas de poblaciones de otros países. Esto puede verse como una vulnerabilidad importante, que podría acentuarse aún más cuando los desarrollos se basan en algoritmos de IA.

Para que las herramientas diagnósticas y predictivas basadas en IA sean útiles y efectivas en el ámbito médico cubano, se requieren conjuntos de datos anotados de señales biomédicas con las características propias de la población local. La personalización de estos conjuntos mediante criterios de selección sobre sus atributos, posibilita optimizar los recursos computacionales y los tiempos de procesamiento, al tiempo que mejora la generalización y confiabilidad de los modelos. En este contexto, el presente trabajo tiene como objetivo presentar una plataforma web orientada a la creación de conjuntos de datos de bioseñales con anotación diagnóstica, adaptada a las características de la población cubana, para soportar sistemas de apoyo a las decisiones médicas.

Esta obra está bajo una licencia CreativeCommons Attribution4.0/International/Deed

MÉTODOS

La plataforma se concibió con un enfoque de investigación aplicada de desarrollo tecnológico, estructurada en cuatro fases iterativas e integradoras: ingeniería de requisitos, modelado de datos, definición de la arquitectura tecnológica, y recolección y gestión inicial de datos. Cada fase incluyó mecanismos de validación con actores clave de los dominios biomédico e informático.

Se utilizó una adaptación del Proceso Racional Unificado,[24] seleccionado por su capacidad para operar con requisitos cambiantes y su enfoque en la documentación de artefactos. Y se utilizaron varias técnicas de elicitación: talleres colaborativos, evaluación comparativa, análisis de escenarios, y creación de prototipos de baja fidelidad.

Se efectuaron cinco talleres colaborativos, con 12 especialistas (seis médicos cardiólogos, dos neurólogos, dos radiólogos, y dos ingenieros biomédicos) del Hospital General Provincial "Camilo Cienfuegos", de Sancti Spíritus, y del Centro de Investigaciones de la Informática de la Universidad Central de Las Villas “Marta Abreu”. En estos talleres se identificaron flujos de trabajo clínicos, criterios de anotación, y necesidades de personalización de conjuntos de datos.

En la evaluación comparativa se analizaron 15 repositorios internacionales de señales biomédicas (*MIT-BIH*,[7] *PTB-XL*,[10] *CHB-MIT*,[15] entre otros), y se identificaron buenas prácticas en anotación, formatos de almacenamiento, y políticas de acceso. Para el análisis de escenarios, se construyeron ocho escenarios de uso que abarcan desde la importación de señales hasta la exportación de conjuntos de datos personalizados para entrenamiento de modelos de IA. En la creación de prototipos de baja fidelidad, se diseñaron maquetas de interfaces, validadas por los médicos en dos sesiones de retroalimentación.

Los requisitos resultantes se clasificaron en funcionales (RF1 a RF19), no funcionales (RNF1 a RNF22), y éticos (consentimiento informado, anonimato). Se utilizó la herramienta Jira para el trazado de requisitos y la gestión de cambios. Se diseñó un modelo relacional normalizado en tercera forma normal (3FN) que utiliza el gestor de bases de datos *PostgreSQL 14+*.

En el proceso de modelado se identificaron 14 entidades principales, a partir del glosario generado en los talleres; entre ellas: usuario, rol, señal, tipo de señal, anotación, estándar de anotación, variable clínica, y conjunto de datos personalizado (Fig. 1).

Esta obra está bajo una licencia CreativeCommons Attribution4.0/International/Deed

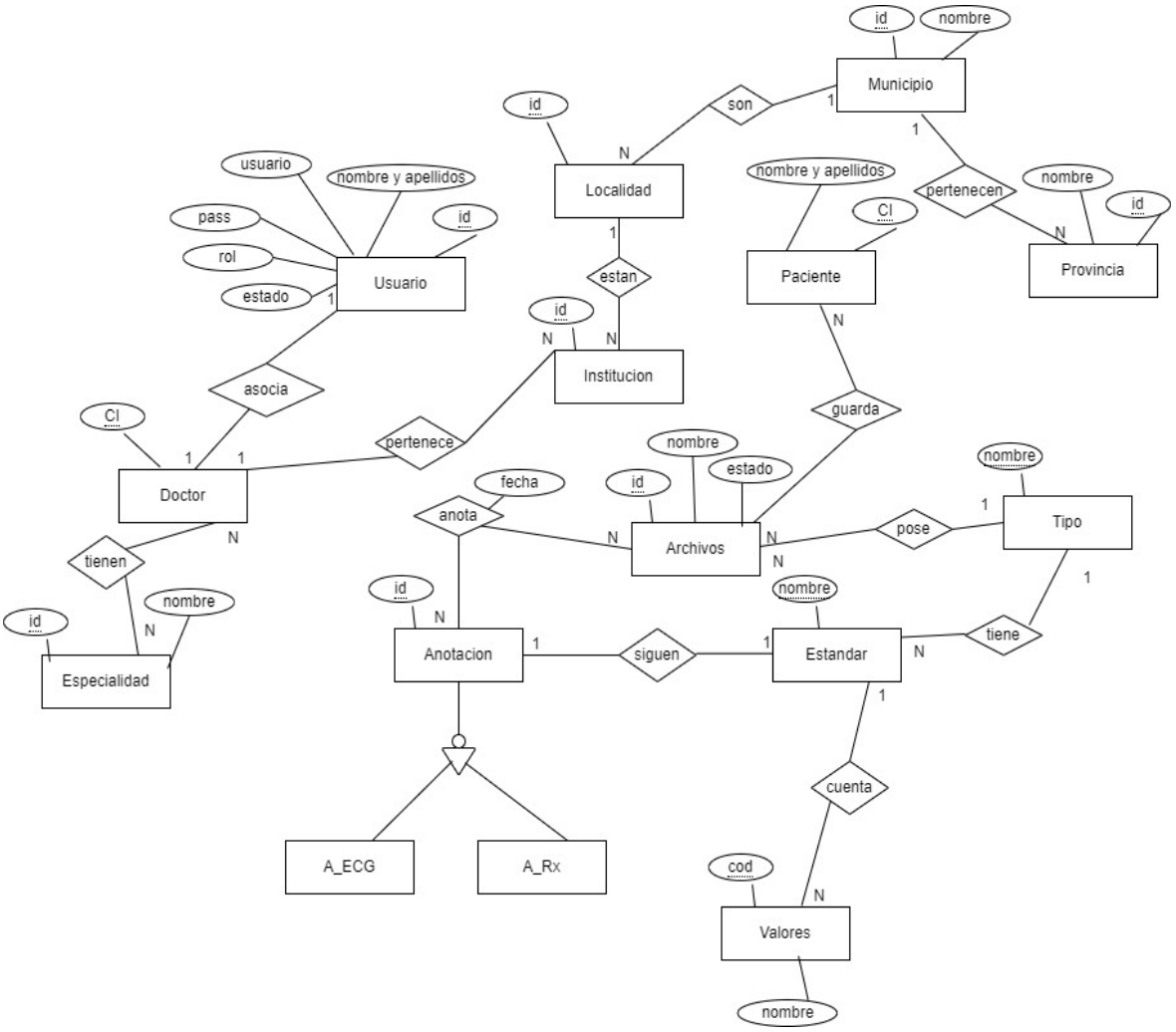


Fig.1 - Estructura de la base de datos relacional de la plataforma para la creación de conjuntos de datos de señales biomédicas.

Fuente: elaboración propia (2026).

Se definieron relaciones de herencia para los roles (médico anotador, analista, docente, y estudiante), y jerárquicas para la organización territorial (provincia, municipio, localidad, e institución). Médicos especialistas efectuaron la validación semántica, para asegurar que los atributos clínicos reflejaran los estándares *AHA/ACC/HRS*,[25] y la implementación del diccionario de datos tuviese los 86 atributos documentados con tipo, rango, nulidad, y restricciones de integridad referencial.

El modelo se implementó en un entorno de pruebas con datos sintéticos que simuló 5000 registros, para validar el rendimiento de las consultas y la escalabilidad. Se optó por una arquitectura de tres capas (presentación, lógica de negocio, y datos), con separación entre la interfaz de usuario y el servidor de

Esta obra está bajo una licencia CreativeCommons Attribution4.0/International/Deed

aplicaciones, que utilizó exclusivamente tecnologías de código abierto para garantizar la soberanía tecnológica y la sostenibilidad del proyecto.

Interfaz: *React* 18[26] con *hooks* y *context API* para la gestión de estado; *Bootstrap* 5 para diseño responsivo adaptable a dispositivos móviles; *Plotly.js* para visualización interactiva de señales ECG; *Axios* para comunicación HTTP con el *backend*.

Servidor: *Fast API* 0.104+[27] como estructura o marco de trabajo principal; se eligió por su rendimiento alto, su documentación automática (*Open API*), y su validación de tipos basada en *Pydantic*.

Autenticación: *OAuth2* (protocolo de autorización), con flujo de código de autorización y *tokens JWT* (*token web JSON*) de corta duración (30 minutos) más *tokens* de actualización.

Cifrado: los datos de los pacientes se almacenaron cifrados con el estándar AES-256 (utilizado por los bancos), y se transmiten mediante *TLS* 1.3; de modo que ninguna persona sin autorización pueda leerlos, aunque acceda al servidor. Esto forma parte de los requisitos de confidencialidad de la historia clínica.

Procesamiento de señales: librería *wfdb* (*Python*) para lectura de formatos fisión, y *h5py* para escritura de archivos *HDF*5.[28]

Para garantizar la extensibilidad a otros tipos de señales (EEG, radiografías, señales respiratorias), se diseñó un sistema de procesadores basado en el patrón *Strategy*. Cada tipo de señal implementa una interfaz común *SignalProcessor* con métodos para carga, visualización, y obtención de estándares de anotación. Esto permite añadir nuevos formatos sin modificar el núcleo de la plataforma.

Base de datos: *PostgreSQL 14+* con extensión *pgcrypto* para funciones de cifrado.

Se definieron índices compuestos sobre tipo de señal, fecha de registro, médico anotador, y estado de la anotación, para acelerar las consultas más frecuentes.

Se garantizó el respaldo automatizado mediante *scripts* diarios con cifrado asimétrico por *GPG*. Se utilizó *Docker Compose*[29] para la orquestación de contenedores (*frontend*, *backend*, base de datos, *redis* para caché); y *NGINX* como *proxy* reverso y balanceador de carga. Se concibieron las opciones de despliegue en servidores institucionales o en la nube (*AWS EC2*, con instancias aisladas).

Para soportar la carga esperada de más de 50 000 señales nuevas anuales, se diseñó una arquitectura horizontalmente escalable. Las tareas de larga duración (importación masiva, conversión de formatos, anonimato por lotes) se encolan en *RabbitMQ* y se procesan por trabajadores *Celery*. La tabla señal se particiona por rango de fechas (trimestralmente), y por tipo de señal. Además, se emplea un clúster de *Redis* para almacenar en caché las consultas personalizadas más frecuentes. El despliegue se realiza en

Esta obra está bajo una licencia CreativeCommons Attribution4.0/International/Deed

contenedores *Docker* orquestados con *Docker Swarm*, lo que permite añadir réplicas del *backend* y los trabajadores bajo demanda.

Como el objetivo es la creación de conjuntos de datos cubanos anotados, se estableció un protocolo piloto de recolección, aprobado por el Consejo Científico de la Facultad de Ciencias Técnicas y Económicas. En el cual, los criterios de inclusión fueron: pacientes mayores de 18 años, con diagnósticos confirmados según los estándares clínicos, que hubiesen firmado el consentimiento informado digital.

Se adquirieron electrocardiógrafos *Cardioline ar1200* (ECG de 12 derivaciones) y *Holter Mortara H12+*. Las señales se exportan en formato XML (propietario) y se convierten automáticamente a HDF5 mediante un módulo de ingesta desarrollado en *Python*.

El sistema elimina automáticamente todo identificador directo (nombre, carné de identidad) y reemplaza el identificador interno por un *UUID*, con lo cual se garantiza el anonimato de los pacientes. Las fechas se redondean al mes. Además, se implementó un módulo para el anonimato robusto, que aplica generalización jerárquica sobre atributos cuasidentificadores (edades por rangos quinquenales, localidades agregadas por áreas de salud), y suprime los grupos con menos de cinco individuos para evitar números únicos. Para las consultas agregadas sobre conjuntos de datos exportados, ofrece un modo de privacidad diferencial con ruido laplaciano (ε=0,1 configurable), lo que garantiza protección frente a ataques de enlace.

Las señales crudas y anotadas se almacenan en particiones cifradas con *LUKS*, y las copias de seguridad se guardan en un servidor fuera de línea, ubicado en una institución académica distinta.

El estudio se rigió por los principios contenidos en la *Declaración de Helsinki*[30] y la Resolución 141/2023 del MINSAP (Cuba)[31] sobre el uso de datos clínicos para investigación. Cada paciente o su tutor firmaron un documento de consentimiento informado, electrónico, que especifica el uso exclusivo de los datos para investigación en salud, y la imposibilidad de reidentificación. El proyecto fue aprobado por los Consejos Científicos de la Universidad de Sancti Spíritus "José Martí Pérez" y de la Universidad Central de Las Villas "Marta Abreu".

Durante la preparación de este artículo, se utilizaron las herramientas *Claude*, en su versión *Claude Sonnet* 5.5 (disponible en modalidad de acceso libre con limitaciones, versión de pago *Pro* y licencia institucional *Enterprise*), y *Consensus* 2.0.0 (disponible bajo un modelo *premiun* libre, con plan *Pro* y acceso institucional a través de bibliotecas universitarias asociadas). Su uso se limitó a la corrección de estilo, traducción, y elaboración del resumen de ideas.

Esta obra está bajo una licencia CreativeCommons Attribution4.0/International/Deed

Los autores revisaron todo el contenido generado por la herramienta, y son los únicos responsables de la exactitud e integridad del manuscrito. Todas las afirmaciones, referencias, y contenidos posteriormente los verificaron confrontándolos con las fuentes científicas originales. La IA no se considera una fuente suficiente para sustentar datos científicos, por lo que todas las afirmaciones se apoyan en artículos, guías, o documentos académicos verificables.

RESULTADOS

Se especificaron los requisitos mediante la aplicación del Proceso Racional Unificado; se capturaron y documentaron 19 requisitos funcionales (RF) y 22 no funcionales (RNF).

Los requisitos funcionales se distribuyeron en siete módulos:

Módulo 1- Administración. Incluye gestión de usuarios con roles específicos (administrador, médico anotador, docente, estudiante, analista), y autenticación segura (*OAuth* 2.0); gestión de médicos, instituciones, localidades, municipios y provincias (RF1 a RF6).

Módulo 2- Gestión de señales y datos clínicos. Importación de señales biomédicas (ECG, EEG) a formato estándar (*HDF*), clasificación por tipo y registro de variables clínicas (RF7 a RF9).

Módulo 3- Anotación diagnóstica. Interfaz de anotación basada en estándares (*AHA/ACC/HRS*), mecanismo de validación por coincidencias, asignación de señales a especialistas, y cálculo de concordancia (RF10 a RF13).

Módulo 4- Notificaciones y exportación. Alertas automáticas de anotaciones pendientes y exportación a *CSV* con filtros personalizados (RF14 a RF15).

Módulo 5- Nomencladores. Gestión de especialidades médicas y establecimiento de estándares de anotación (RF16 a RF17).

Módulo 6- Entrenamiento para residentes. Creación de pruebas con imágenes anotadas por expertos, registro de residentes vinculados a profesores, banco de preguntas, y calificación automática por porcentaje de coincidencia (RF18, RF18.1 a RF18.3).

Módulo 7- Comunicación interna. *Chat* en tiempo real con notificaciones *push*, soporte para adjuntos e historial por paciente (RF19, RF19.1 a -RF19.3).

Los requisitos no funcionales abarcaron aspectos críticos: usabilidad (interfaz médica con curva de aprendizaje <30 min, *WCAG 2.1 AA*); rendimiento (respuesta <2 s para 10 000 registros, 100 usuarios

Esta obra está bajo una licencia CreativeCommons Attribution4.0/International/Deed

concurrentes); seguridad (cifrado *AES-256*, autenticación multifactor); ética (consentimiento informado digital, anonimato automático); compatibilidad multiplataforma y navegadores; capacidad de mantenimiento (código modular, documentación *Swagger/OpenAPI*); escalabilidad (arquitectura de microservicios para más de 50000 señales al año); portabilidad (compatibilidad con sistemas operativos, despliegue *cloud/on-premise*, exportación *HDF/CSV*); y parámetros de *software* (*Frontend React.js 18+*, *Backend Fast API Python 3.9+*, bases de datos *PostgreSQL 14+/MongoDB 6+*, licencias *GPLv3*, orquestación de contenedores *Docker*).

Se diseñó un modelo relacional normalizado en tercera forma normal (3FN), que integra las entidades de gestión de señales, usuarios, y anotaciones. La estructura separa de forma explícita los datos de pacientes, señales, y anotaciones, que permiten consultas complejas para la generación de conjuntos de datos personalizados (véase la Fig. 1). La base de datos separa explícitamente tres tipos de información: los datos del paciente (anónimos), las señales registradas, y las anotaciones diagnósticas realizadas por los médicos. Esta separación garantiza que las anotaciones puedan auditarse, y que los datos personales no se mezclen con los clínicos. De este modo, se facilita el cumplimiento de la normativa de protección de datos.

La plataforma se estructuró sobre un conjunto tecnológico compuesto por *React.js* y *Fast API*, con orquestación de contenedores *Docker*. Se implementó un adaptador nativo hacia *HL7 FHIR R4*, con puntos de acceso para exportar anotaciones y metadatos. Las señales en formato *HDF5* se referenciaron mediante *Document Reference*, y se reservó el mapeo a *DICOM SR* para imágenes radiológicas futuras.

Para validar la plataforma (concordancia interanotador) se evaluaron 100 señales ECG estratificadas por complejidad (40 normales, 30 con arritmias claras, 30 con hallazgos ambiguos). Cuatro cardiólogos efectuaron anotaciones en dos rondas (intervalo de 15 días). El coeficiente *Kappa* de *Fleissglobal* fue de 0,87 (IC 95%: 0,82-0,91). Un *Kappa* de 0,87 indica una concordancia casi perfecta entre los cardiólogos: en la práctica; esto significa que dos especialistas que anotan de forma independiente la misma señal, llegan al mismo diagnóstico en la gran mayoría de los casos. Lo cual respalda la fiabilidad de las anotaciones incluidas en el conjunto de datos.

La concordancia intraanotador (prueba y reprueba) resultó en un *Kappa* medio de 0,89 (rango 0,84-0,93). En el subgrupo de casos ambiguos, el *Kappa* descendió a 0,74, lo que dictaminó la inclusión de un flujo de consenso obligatorio mediante una segunda anotación experta.

Esta obra está bajo una licencia CreativeCommons Attribution4.0/International/Deed

DISCUSIÓN

El desarrollo de la plataforma propuesta es una contribución al ecosistema de salud digital en Cuba. Solventa una brecha crítica en la disponibilidad de conjuntos de datos de bioseñales anotados y representativos de la población local. La aplicación del Proceso Racional Unificado posibilitó capturar y documentar requisitos de manera sistemática. Este enfoque es efectivo en entornos de salud donde la interacción temprana con los médicos es crucial.

En este sentido, la inclusión del módulo de entrenamiento es una innovación respecto a plataformas similares como *PALMS* o *EASELAN*. Integra funcionalidades docentes que permiten a los residentes practicar la anotación bajo supervisión. Mientras que *PALMS* se centra en la anotación manual asistida por algoritmos para señales de tiempo continuo, esta propuesta incorpora módulos específicos para anotación diagnóstica estandarizada basada en criterios clínicos establecidos.

La selección de *Python* y *React* para la interfaz del cliente, complementada con *Fast API* en el servidor de aplicaciones, es una elección alineada con las tendencias actuales en plataformas de datos biomédicos. *Fast API* ofrece un rendimiento asincrónico comparable a *Node.js* o *Go*, y su tipado estático mediante *Pydantic* reduce errores en parámetros clínicos sensibles. A diferencia de las herramientas mencionadas, esta solución incorpora un adaptador nativo hacia el estándar *HL7 FHIR R4*, que evita la creación de silos de datos, y posibilita su despliegue como un módulo interoperable dentro de los ecosistemas hospitalarios existentes. Esta es una flexibilidad necesaria, por las limitaciones de infraestructura en contextos de recursos limitados.

Por otra parte, el modelo relacional normalizado en tercera forma normal (3FN) diseñado, separa de forma explícita los datos de pacientes, señales, y anotaciones, de modo que se facilitan las consultas complejas para la generación de conjuntos de datos personalizados. Esta estructura se alinea con los principios FAIR (localizables, accesibles, interoperables y reutilizables, por sus siglas en inglés); los cuales son recomendaciones actuales, fundamentales para la gestión de datos de investigación biomédica.[29] El uso de *HDF5* (formato de datos jerárquico versión 5, por sus siglas en inglés) como formato de almacenamiento, garantiza la preservación de las señales originales junto con sus metadatos en un contenedor único, y facilita la replicación e intercambio de conjuntos de datos a largo plazo.

La estrategia de anotación, basada en expertos múltiples con un algoritmo de coincidencias, respondió a la necesidad –documentada ampliamente en la literatura– de etiquetar los datos biomédicos. Aquí la

Esta obra está bajo una licencia CreativeCommons Attribution4.0/International/Deed

variabilidad interobservador es una fuente principal de ruido.[32] Las estrategias de consenso entre anotadores múltiples en tareas de clasificación de señales biomédicas, mejoran significativamente (entre 12,60% y 26,00%) la calidad de las anotaciones en el área bajo la curva de precisión y recuperación, en comparación con las anotaciones individuales.[33] La implementación de un sistema de anotación colaborativa, similar al enfoque adoptado por *CrowdEEG* para datos de electroencefalografía, posibilita aprovechar la experiencia de especialistas múltiples, y gestionar de manera transparente los casos de divergencias diagnósticas.

La evaluación rigurosa de la concordancia (que incluye casos difíciles, y medición de la fiabilidad intraanotador), supera lo reportado en la mayoría de las plataformas similares. Estudios clásicos de variabilidad en señales electrocardiográficas, demuestran que la concordancia entre cardiólogos ante casos complejos frecuentemente cae por debajo de un coeficiente *Kappa* de 0,60. Y muchas plataformas omiten el reporte de la fiabilidad intraanotador, o carecen de mecanismos de resolución para los casos ambiguos.[34,35]

En la validación de la plataforma propuesta, se registró un *Kappa* moderado (0,74) para los casos ambiguos. Este resultado valida la decisión de implementar un flujo de consenso obligatorio mediante una segunda anotación experta. Según la clasificación de Landis y Koch[36], un *Kappa* global de 0,87 refleja una concordancia casi perfecta, que protege la calidad del conjunto de datos frente a los sesgos de un observador único.

Disponer de esta plataforma es un hecho que trasciende el ámbito técnico; posibilitará desarrollar sistemas de inteligencia artificial para diagnóstico asistido con las características de la población cubana. Ello incidirá positivamente en la equidad de acceso y la precisión de las herramientas clínicas, y fomentará la colaboración interinstitucional mediante un repositorio nacional estandarizado.

Desde una perspectiva metodológica, esta investigación demuestra las vastas posibilidades de aplicar la ingeniería de requisitos y el modelado de datos en contextos de recursos limitados; lo cual la convierte en una referencia valiosa para proyectos similares en otros países en desarrollo. Entre las experiencias de este estudio, se reconoce que fue fundamental la participación activa de los médicos para validar la relevancia clínica de los estándares analizados. También, que la alfabetización digital heterogénea del personal de la salud condiciona los diseños de las interfaces de usuarios, y justifican los requisitos estrictos de usabilidad impuestos.

No obstante, lo positivo de la experiencia, se reconocen limitaciones que deberán abordarse en trabajos

Esta obra está bajo una licencia CreativeCommons Attribution4.0/International/Deed

futuros. La validación inicial se ha centrado exclusivamente en señales ecocardigráficas; la arquitectura basada en procesadores de extensión permitirá incorporar progresivamente otras señales (electroencefalográficas, y radiográficas de tórax). La validación médica completa requerirá la incorporación de señales reales de pacientes cubanos, sujeta a la aprobación de los comités de ética institucionales.

En cuanto a la escalabilidad del sistema en el país, la arquitectura actual incluye particionamiento y colas de mensajes. Pero su despliegue simultáneo en más de 10 hospitales, requerirá evaluar el rendimiento del particionamiento cruzado y la latencia de replicación. Por ello, se implementará un sistema de federación de repositorios, donde cada provincia mantenga su nodo local y los catálogos globales se sincronicen mediante un bus de eventos. Si bien el sistema actual incorpora el anonimato por generalización, se avanzará hacia esquemas de privacidad diferencial en tiempo real para las consultas que generan conjuntos de datos personalizados.

CONCLUSIONES

Se demostró la viabilidad técnica de una plataforma web para gestionar y anotar señales biomédicas de pacientes cubanos, basada en un modelo relacional normalizado en tercera forma normal y tecnologías de código abierto. La solución integró flujos colaborativos de anotación, validación por consenso, y personalización de conjuntos para inteligencia artificial; lo cual la hace un avance estratégico para la salud digital en Cuba. Junto a los repositorios representativos de la población local, incluye funcionalidades docentes y de control diagnóstico. Ello reduce la dependencia de datos externos, y mejora la equidad y precisión de los sistemas de apoyo clínico; todo bajo estándares éticos y de seguridad. La validación completa requerirá estudios piloto multicéntricos, con señales reales (inicialmente ecocardiográficas), y la adición progresiva de otros datos biomédicos. El trabajo futuro se orienta a la implementación completa de los módulos, la evaluación de la usabilidad en entornos médicos, y el despliegue nacional de la plataforma. El estudio en sí mismo es un aporte crucial porque pone a disposición de los especialistas datos biomédicos propios, anotados y contextualizados. Esta es una condición indispensable para el desarrollo soberano de la inteligencia artificial en el sector sanitario cubano.

Esta obra está bajo una licencia CreativeCommons Attribution4.0/International/Deed

REFERENCIAS BIBLIOGRÁFICAS

1. Dwivedi YK, Hughes L, Ismagilova E, Aarts G, Coombs C, Crick T, et al. Artificial intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. Int J Inf Manage [Internet]. 2021 [citado 18 Ene 2026];57:101994. Disponible en: https://www.sciencedirect.com/science/article/pii/S026840121930917X

2. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Future Health C J [Internet]. 2021 [citado 11 Ene 2026];8(2):e188-94. Disponible en: https://www.sciencedirect.com/science/article/pii/S2514664524005277

3. Hummel P, Braun M, Tretter M, Dabrock P. Data sovereignty: a review. Big Data Soc [Internet]. 2021 [citado 18 Ene 2026];8(1):2053951720982012. Disponible en: https://journals.sagepub.com/doi/full/10.1177/2053951720982012

4. Vega-Abascal JB, Piriz-Assa AR, Nápoles-Riaño DA, Guimará-Mosqueda MR. Predicción del riesgo cardiovascular con tablas de la Organización Mundial de la Salud y modelo predictivo diseñado con inteligencia artificial. Rev Cubana Med [Internet]. 2025 [citado 18 Ene 2026];64:e5019. Disponible en: https://revmedicina.sld.cu/index.php/med/article/download/5019/2627/15414

5. Celi LA, Cellini J, Charpignon ML, Dee EC, Dernoncourt F, Eber R, et al. Sources of bias in artificial intelligence that perpetuate health care disparities: a global review. PLOS Digit Health [Internet]. 2022 [citado 18 Ene 2026];1(3):e0000022. Disponible en: https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000022

6. Scheibner J, Raisaro JL, Troncoso-Pastoriza JR, Ienca M, Fellay J, Vayena E, et al. Revolutionizing medical data sharing using advanced privacy-enhancing technologies: technical, legal, and ethical synthesis. J Med Internet Res [Internet]. 2021 [citado 18 Ene 2026];23(2):e25120. Disponible en: https://www.jmir.org/2021/2/e25120

7. Merdjanovska E, Rashkovska A. A framework for comparative study of databases and computational methods for arrhythmia detection from single-lead ECG. Sci Rep [Internet]. 2023 [citado 18 Ene 2026];13:11682. Disponible en: https://pmc.ncbi.nlm.nih.gov/articles/PMC10356811/

8. Perez-Alday EA, Gu A, Shah AJ, Robichaux C, Wong AKI, Liu C, et al. Classification of 12-lead ECGs: the Physio Net/Computing in Cardiology Challenge 2020. Physiol Meas [Internet]. Ene 2021 [citado 18 Ene 2026];41(12):124003. Disponible en: https://physionet.org/content/challenge-2020/1.0.2/

Esta obra está bajo una licencia CreativeCommons Attribution4.0/International/Deed

9. Reyna MA, Sadr N, Perez-Alday EA, Gu A, Shah AJ, Robichaux C, et al. Will two do? Varying dimensions in electrocardiography: the PhysioNet/Computing in Cardiology Challenge 2021. Physiol Meas [Internet]. 2022 [citado 7 Feb 2026];43(8):084001. Disponible en: https://physionet.org/content/challenge-2021/1.0.3/

10. Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, el al. PTB-XL, a large publicly available ECG dataset. Sci Data [Internet]. 2020 [citado 4 Mar 2026];7:154. Disponible en: https://physionet.org/content/ptb-xl/1.0.3/

11. Liu H, Chen D, Chen D, Zhang X, Li H, Bian L, et al. A large-scale multi-label 12-lead electrocardiogram database with standardized diagnostic statements. Sci Data [Internet]. 2022 [citado 6 Mar 2026];9:272. Disponible en: https://www.nature.com/articles/s41597-022-01403-5

12. Zhang X, Zhang X, Huang Q, Chen F. A review of epilepsy detection and prediction methods base don EEG signal processing and deep learning. Front Neurosci [Internet]. 2024 [citado 12 Mar 2026];18:1468967. Disponible en: https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2024.1468967/full

13. Oftadeh-Balani S, Al-Hussainy A, Shalal A, Ubaid M, Aluquaily Z, Alamoori J, et al. Exploring brain activity in different mental cognitive workloads. Iran J Psychiatry [Internet]. 2024 [citado 11 Feb 2026];19(4):356-66. Disponible en: https://ijps.tums.ac.ir/index.php/ijps/article/view/3939

14. Matran-Fernández A, Poli R. Towards thea utomated localization of targets in rapid image-sifting by collaborative brain-computer interfaces. PLoS One [Internet]. 2017 [citado 18 Feb 2026];12(5):e0178498. Disponible en: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0178498

15. Rasheed K, Qayyum A, Qadir J, Sivathamboo S, Kwan P, Kuhlmann L, et al. Machine learning for predicting epileptic seizuresusing EEG signals: a review. IEEE Rev Biomed Eng [Internet]. 2021 [citado 11 Ene 2026];14:139-55. Disponible en: https://arxiv.org/abs/2002.01925

16. Indeewara W, Hennayake M, Rathnayake K, Ambegoda T, Meedeniya D. Chest X-ray dataset with lung segmentation. PhysioNet [Internet]. 2023 [citado 16 Feb 2026];SCR_007345. Disponible en: https://www.physionet.org/content/chest-x-ray-segmentation/1.0.0/

17. Reis EP, Paiva JPQ, Bueno-Da Silva MC, Sousa-Ribeiro GA, Fornasiero-Paiva V, Bulgarelli L, et al. BRAX, Brazilian labeled chest X-ray dataset. Sci Data [Internet]. 2022 [citado 1 Mar 2026];9:487. Disponible en: https://www.nature.com/articles/s41597-022-01608-8

Esta obra está bajo una licencia CreativeCommons Attribution4.0/International/Deed

18. Calamida A, Nooralahzadeh F, Rohanian M, Nishio M, Fujimoto K, Krauthammer M. Radiology report generation models evaluation dataset for chest X-rays (Rad Eval X) (versión 1.0.0). PhysioNet; [Internet]. 2023 [citado 11 Feb 2026]. Disponible en: https://physionet.org/content/rad-eval-x/1.0.0/

19. Chen LC, Kuo PC, Wang R, Gichoya J, Celi LA. Chest X-ray segmentation images based on MIMIC-CXR (versión 1.0.0). PhysioNet [Internet]. 2022 [citado 23 Ene 2026]. Disponible en: https://physionet.org/content/lung-segment-mimic-cxr/1.0.0/

20. Moukheiber M, Moukheiber L, Moukheiber D, Hao S, Celi LA, Lee HC. A temporal dataset for respiratory support in critically ill patients (versión 1.1.0). PhysioNet [Internet]. 2023 [citado 21 Ene 2026]. Disponible en: https://physionet.org/content/temporal-respiratory-support/1.1.0/

21. Guy EFS, Clifton JA, Caljé-Van der Klei T, Chen R, Knopp J, Moeller K, et al. Respiratory dataset from PEEP study with expiratory occlusion (versión 1.0.0). PhysioNet [Internet]. 2023 [citado 16 Feb 2026]. Disponible en: https://physionet.org/content/respiratory-dataset/1.0.0/

22. Clifton JA, Guy EFS, Caljé-Van der Klei T, Knopp J, Chase JG. Simulated obstructive disease respiratory pressure and flow (versión 1.0.0). PhysioNet [Internet]. 2023 [citado 1 Mar 2026]. Disponible en: https://physionet.org/content/simulated-obstructive-disease/1.0.0/

23. Porta-García MA, Quiroz-Salazar A, Abarca-Castro EA, Reyes-Lagos JJ. Bradycardia may decrease cardio respiratory coupling in preterm infants. Entropy [Internet]. 2023 [citado 25 Feb 2026];25(12):1616. Disponible en: https://www.mdpi.com/1099-4300/25/12/1616

24. Jacobson I, Booch G, Rumbaugh J. The unified software development process. IEEE Software [Internet]. 1999. [citado 25 Feb 2026].96-102 Disponible en: https://people.eecs.ku.edu/~hossein/810/Readings/rup-unified-process.pdf

25. Joglar JA, Chung MK, Armbruster AL, Benjamin EJ, Chyou JY, Cronin EM, et al. 2023 ACC/AHA/ACCP/HRS guideline for the diagnosis and management of atrial fibrillation. Circulation [Internet]. 2024 [citado 21 Feb 2026];149(1):e1-156. Disponible en: https://www.ahajournals.org/doi/10.1161/CIR.0000000000001193

26. Meta Open Source. React: reference of Hooks [Internet]. Menlo Park (CA): Meta Platforms; 2026 [citado 1 Mar 2026]. Disponible en: https://react.dev/reference/react/hooks

27. FastAPI [Internet]. Features. 2026 [citado 21 Ene 2026]. Disponible en: https://fastapi.tiangolo.com/features/

28. h5py [Internet]. HDF5 for Python. 2026 [citado 21 Ene 2026]. Disponible en:

Esta obra está bajo una licencia CreativeCommons Attribution4.0/International/Deed

https://docs.h5py.org/en/stable/

29. Inau ET, Sack J, Waltemath D, Zeleke AA. Initiatives, concepts, and implementation practices of the findable, accessible, interoperable, and reusable data principles in health data stewardship: scopingreview. J Med Internet Res [Internet]. 2023 [citado 26 Mar 2026];25:e45013. Disponible en: https://www.jmir.org/2023/1/e45013

30. Asociación Médica Mundial. Declaración de Helsinki de la AMM. Principios éticos para las investigaciones médicas en seres humanos. Ratificada en la 75th WMA General Assembly, Helsinki, Finland, october 2024 [Internet]. Helsinki: 18ª Asamblea Mundial; 1964 [citado 15 Ene 2026]. Disponible en: https://www.wma.net/policies-post/wma-declaration-of-helsinki/

31. Ministerio de Justicia (Cuba). Protección de Datos Personales Ley No. 149 [Internet]. Gaceta Oficial de la República de Cuba, Edición Ordinaria No. 90; (25-08-2022). [citado 23 Mar 2026]. Disponible en: https://www.gacetaoficial.gob.cu/sites/default/files/goc-2022-o90_0.pdf

32. Shama DM, Venkataraman A. Bayesian uncertainty-aware deep learning with noisy labels: tackling annotation ambiguity in EEG seizure detection. PLoS One [Internet]. 2026 [citado 27 Mar 2026];21:e0352191. Disponible en: https://pmc.ncbi.nlm.nih.gov/articles/PMC13289946/

33. Ji W, Yu S, Wu J, Ma K, Bian C, Bi Q, et al. Learning calibrated medical image segmentation via multi-rateragreement modeling. En: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) [Internet]. 2021 [citado 26 Mar 2026]. p. 12341-51. Disponible en: https://openaccess.thecvf.com/content/CVPR2021/papers/Ji_Learning_Calibrated_Medical_Image_Segmentation_via_Multi-Rater_Agreement_Modeling_CVPR_2021_paper.pdf

34. Engstrom N, Dobson GP, Ng K, Lander K, Win K, Gupta A, et al. Validation of Cal ECG software for primary prevention heart failure patients: reducing inter-observer measurement variability. J Electrocardiol [Internet]. 2022 [citado 25 Mar 2026];74:128-33. Disponible en: https://researchonline.jcu.edu.au/76198/

35. Somani S, Russak AJ, Richter F, Zhao S, Vaid A, Chaudhry F, et al. Deep learning and the electrocardiogram: review of the current state-of-the-art. Europace [Internet]. 2021 [citado 23 Mar 2026;23(8):1179-91. Disponible en: https://doi.org/10.1093/europace/euaa377https://academic.oup.com/europace/article/23/8/1179/6132071?login=false

36. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics

Esta obra está bajo una licencia CreativeCommons Attribution4.0/International/Deed

[Internet]. 1977 [citado 27 Mar 2026];33(1):159-74. Disponible en: https://www.jstor.org/stable/2529310?origin=crossref

Conflictos de intereses

Los autores declaran que no existen conflictos de intereses.

Contribuciones de los autores

Julio Companioni-Martínez: conceptualización, curación de datos, investigación, metodología, validación, visualización, y redacción del borrador original.

Denis Hernández-Pacheco: conceptualización, curación de datos, metodología, administración del proyecto, validación, redacción, revisión, y edición.

Carlos Jiménez Puerto: investigación, análisis formal, supervisión, redacción, revisión, y edición

José Carlos Pernas-Farray: *software*, visualización, y redacción del borrador original.

Alberto Taboada-Crispi: conceptualización, investigación, adquisición de fondos, redacción, revisión, y edición.

Financiación

Universidad Central de Las Villas “Marta Abreu”. Santa Clara, Cuba.

Universidad de Sancti Spíritus “José Martí Pérez”. Sancti Spíritus, Cuba.

Esta obra está bajo una licencia CreativeCommons Attribution4.0/International/Deed